AN ABSTRACT OF THE DISSERTATION OF

Erik R. Rowley for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on April 6, 2016.

Title: Genomic Resource Development for European Hazelnut (*Corylus avellana* L.).

Abstract approved:

_____

Todd C. Mockler

European hazelnut (*Corylus avellana* L.) is an important crop Oregon's Willamette Valley, producing 99% of the hazelnuts grown in North America and brings over US $60 million dollars to the region annually. Hazelnuts are rich in fiber and vitamins, as well in demand by consumers due to their popularity as the predominant flavor in a multitude of confectionary pastes and chocolate spreads. Breeding efforts are focused on developing hazelnut cultivars with enhanced agronomic traits of interest including size, blanching ability, and resistance to disease. Molecular markers have been developed for hazelnut and placed on a genetic linkage map, linking DNA marker sequences that segregate with phenotypic traits of interest. The objective of this study was to utilize high-throughput sequencing technologies to sequence the hazelnut genome, allowing for trait and marker discovery on a genome-wide scale. We have established genomic resources for hazelnut, including genomic and transcriptomic sequences to allow breeders the opportunity to exploit the wealth of genetic diversity when choosing germplasm for crosses. We chose the Eastern Filbert Blight (EFB) resistant diploid hazelnut cultivar 'Jefferson' (OSU 703.007) as the reference. The 'Jefferson' transcriptome assembly is represented by 28,255 transcript contigs, having an average length of 532bp, and an N50 of 961bp. These transcripts were characterized using both BLASTX protein homology and gene ontology (GO)

classifications, with the majority of the predicted proteins to have high conservation with the most closely related plant sequences of *Vitis*, *Populus*, and *Ricinus*. A survey of gene classes enriched among tissue types further validates the assembly and transcript models. The draft genome assembly for 'Jefferson' was assembled *de novo* using Illumina short read technology into 36,641 contigs and scaffolds, with half of the assembly contained in scaffolds and contigs greater than 21.5 Kb. We captured approximately 91% (345 Mb) of the flow-cytometry-determined genome size and identified 34,910 putative gene loci which were functionally annotated to identify candidates for future molecular validation. The majority of the annotated genes share homology with the best annotated and related genera *Vitis*, *Prunus*, *Populus*, and *Ricinus*. We also resequenced seven additional European hazelnut cultivars, detecting millions of variants between one of more of these genomes and that of 'Jefferson'. These variants were annotated based on the functional consequence each polymorphism on the affected loci. In addition, we utilized Genotyping-by-Sequence (GBS) technologies to produce a high-density genetic map within 138 individuals of an F1 hazelnut mapping population, representing a five-fold increase in marker density over the previous maps. Hazelnut genome sequencing has provided new resources to the scientific community and promises to accelerate trait discovery and enhance future breeding efforts, and serve as a tool for gene discovery and functional studies.

Genomic Resource Development for European Hazelnut (*Corylus avellana* L.)

by
Erik R. Rowley

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented April 6, 2016
Commencement June 2017

Doctor of Philosophy dissertation of Erik R. Rowley presented on April 6, 2016

APPROVED:

Major Professor, representing Molecular and Cellular Biology

Director of the Molecular and Cellular Biology Program

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Erik R. Rowley, Author

ACKNOWLEDGEMENTS

I’d like to thank everyone who has have been a part of this wild ride with me over the years. You’ve made it all worthwhile, Cheers.

CONTRIBUTION OF AUTHORS

Dr. Todd Mockler and Dr. Shawn Mehlenbacher oversaw the planning and design of this project. Dr. Shawn Mehlenbacher assisted with the background of European hazelnut and collection of the tissues. All coauthors helped to edit and approve the final versions of the chapters with which they are associated.

TABLE OF CONTENTS

Page

Chapter 1. Introduction..... 1

Chapter 2. Assembly and characterization of the European hazelnut (*Corylus avellana* L.) 'Jefferson' transcriptome..... 6

Abstract..... 7

Introduction..... 8

Materials and methods..... 10

Hazelnut samples, collection, and RNA preparation…….. 10
Random primed cDNA synthesis..... 11
SMART cDNA synthesis..... 11
Preparation of RNAseq libraries for the Illumina GAIIx..... 12
Transcriptome assembly..... 13
Functional annotation..... 14
Tissue-specific gene expression..... 14
Data availability..... 15

Results..... 14

Illumina sequencing and assembly..... 14
Functional annotation of assembled contigs..... 15
Survey of differential expression among tissues..... 15

Discussion..... 16

Literature cited..... 18

Chapter 3. A draft genome and high-density genetic map of European hazelnut (*Corylus avellana* L.) 28

Abstract..... 29

Introduction..... 29

Materials and methods..... 32

Tissue collection and Illumina HiSeq 2000 sequencing..... 32
Genome assembly and filtering..... 33
Gene prediction and functional annotation..... 34
Polymorphism discovery..... 35

TABLE OF CONTENTS (Continued)

Page

Visualization of data..... 35
Construction of the GBS-based genetic map..... 35
Data availability..... 37

Results..... 37

Assembly and characterization of the hazelnut genome..... 37
Developing a high-density map for European hazelnut..... 39
Construction of the hazelnut physical map and genome anchoring..... 62
Comparative genomics within the Rosids..... 63
Discovery and characterization of protein coding genes..... 65
Disease resistance genes..... 67
Pollen incompatibility S-loci..... 71
Paclitaxel biosynthesis precursor molecules..... 73
Resequencing and polymorphism detection in additional hazelnut cultivars..... 76
Functional consequences pf polymorphisms in protein coding loci..... 79

Discussion..... 90

Author contributions..... 93

Literature cited..... 93

Chapter 4. Conclusion..... 101

Literature cited..... 111

Appendices..... 113

LIST OF FIGURES

Figure — Page

2.1 Contig length distribution for the 'Jefferson' transcriptome assembly..... 16

2.2 Top gene ontology (GO) classifications from the functional gene associations..... 18

2.3 Organism breakdown of the BLASTX results showing transcript contigs having a match in the nonredundant protein database..... 19

2.4 Visual illustration of differential enrichment of tissue-specific transcript contigs..... 20

2.5 Multilevel pie chart showing enrichment of cellular component plant GOslimmer terms in leaf tissue..... 21

2.6 Multilevel pie chart showing enrichment of cellular component plant GOslimmer terms in whole seedlings…..... 22

3.1 K-mer coverage of the 'Jefferson' 350bp PE Illumina library..... 39

3.2 Improved genetic linkage map for European hazelnut..... 40

3.3 Synteny between the peach and hazelnut genomes based on markers from the genetic map..... 64

3.4 Percent of protein-coding sequences in the hazelnut assembly present in related genera as determined using BLASTP..... 66

3.5 Functional gene assignments derived from the Gene Ontology (GO) database..... 67

3.6 Cluster of loci encoding TIR-domain containing NBS-LRR disease resistance proteins..... 69

3.7 SNP in "putative class III acidic chitinase" in the EFB-susceptible 'Daviana' that introduces a mutation in the splice site acceptor region..... 83

3.8 'Daviana', 'Halls Giant', 'Ratoli', and 'Tonda Gentile de Langhe' SNPs in a putative pollen-expressed male determinant SI-related locus..... 89

LIST OF TABLES

Table	Page
2.1 Summary of tissues used for RNA sequencing (RNA-seq).....	13
2.2 Summary of European hazelnut transcriptome assembly.....	16
2.3 Summary of hazelnut transcript functional annotation.....	17
3.1 Summary of the 345-Mb *de novo* genome assembly of the European hazelnut cultivar 'Jefferson', the first sequenced genome in the order Fagales.....	38
3.2 Summary of hazelnut maternal genetic map.....	61
3.3 Summary of hazelnut paternal genetic map.....	61
3.4 Loci annotated participating in pollen incompatibility….....	72
3.5 Loci with homology to the NDTBT family of proteins in the 'Jefferson' genome.....	74
3.6 Seven additional European hazelnut cultivars chosen for resequencing and variant discovery.....	78
3.7A Number of SNPs by predicted effect on coding potential in resequenced cultivars compared with 'Jefferson'.....	80
3.7B Number of INDELs by predicted effect on coding potential in resequenced cultivars compared with 'Jefferson'.....	80
3.8A Unique loci containing SNPs by predicted effect on coding potential in resequenced cultivars compared with 'Jefferson'.....	81
3.9B Unique loci containing INDELs by predicted effect on coding potential in resequenced cultivars compared with 'Jefferson'.....	81
3.9 Variants in S-loci predicted to alter coding potential.....	87

LIST OF APPENDICES

Appendix — Page

Appendix Chapter 1. Rapid synthesis of a long double-stranded oligonucleotide from a single-stranded nucleotide using magnetic beads and an oligo library..... 114

Abstract..... 115

Introduction..... 116

Materials and methods..... 120

Materials..... 120
Preparation of Streptavidin-coated magnetic beads..... 122
Annealing process..... 122
Binding of Streptavidin-coated magnetic beads and biotinylated oligos..... 123
Ligation..... 123
PCR..... 123
Gel electrophoresis..... 124
Sanger DNA sequencing..... 124
One-pot dsDNA synthesis..... 124

Results and discussion..... 125

Annealing..... 125
Ligation..... 125
Sequencing of the target DNA..... 130

Conclusions..... 131

Supporting information..... 132

Acknowledgements..... 133

Author contributions..... 133

Literature cited..... 133

Appendix Chapter 2. Genome sequencing and resource development for European hazelnut..... 135

Abstract..... 136

Introduction..... 137

LIST OF APPENDICES (continued)

Appendix	Page
Materials and methods	138
Plant materials	138
Assembly of hazelnut genomes and transcriptome	138
Functional annotation and SNP discovery	141
Results and Discussion	141
Assembly and SNP discovery among hazelnut genomes	141
Transcriptome assembly	141
Literature cited	144
Appendix Chapter 3. Analysis of global gene expression in *Brachypodium distachyon* reveals extensive network plasticity in response to abiotic stress	145
Abstract	146
Introduction	147
Materials and methods	149
Experimental growth conditions and tissue sampling	149
RNA preparation, labeled cDNA synthesis, and microarray hybridization	150
Mapping of probes	150
Microarray data analysis	151
Heatmap and principal component analysis	152
GO analysis and principal component analysis	152
Network analysis	153
Promoter analysis	153
Network plasticity analysis	154
Accession number	155
Results	155
Overall differential expression analysis	155
Network analysis of stress response in *Brachypodium*	160

LIST OF APPENDICES (Continued)

Appendix	Page
Discussion	172
Stress responsive modules in *Brachypodium*. transcriptional circuitry	172
Photosynthesis	173
Plant growth	175
Cold response	176
Heat response	177
Calcium-mediated stress response	178
Translation	179
Signaling	181
Novel and uncharacterized modules	181
Network Plasticity	182
Web Resources	184
Acknowledgements	185
Funding	185
Author contributions	185
Literature cited	185
Appendix Chapter 4. Plant abiotic stress: insights from the genomics era	194
Introduction	195
Salt	198
Cold	210

LIST OF APPENDICES (Continued)

Appendix	Page
Severe desiccation and water deficit: heat and drought stress.....	226
Heat stress.....	226
Drought.....	236
Combinations of abiotic stressors: Profiles of heat and drought....	242
High light.....	245
Future of abiotic stress research: insights from the genomics revolution.....	249
Literature cited.....	252
Appendix Chapter 5. Discovery of SNP markers in expressed genes of hazelnut.....	293
Abstract.....	294
Introduction.....	295
Materials and methods.....	296
Plant materials, growth conditions, and tissue collection.	296
EST sources and database construction.....	296
SNP discovery.....	297
Results.....	297
De novo assembly of EST contigs from RNAseq data	297
SNP discovery in the hazelnut dataset.....	298
Discussion.....	300
Literature cited.....	302

LIST OF APPENDIX FIGURES

Figure	Page
Figure 1.1 Procedure for DNA synthesis.....	145
Figure 1.2 Overlapping synthesis of DNA on streptavidin-coated magnetic beads.....	146
Figure 1.3 Agarose gel electrophoresis of annealed and ligation products.....	151
Figure 1.4 Agarose gel electrophoresis of final product.....	154
Figure 1.5 DNA sequencing alignment of final synthesized product.....	154
Figure 3.1 Numbers of genes up-regulated and down-regulated shown as a function of time in hours after stress onset.....	181
Figure 3.2 Heatmap of expression differences between control and indicated stress arrays.....	158
Figure 3.3 Venn diagram showing overlap of up-regulated genes in response to the four assayed abiotic stresses.....	159
Figure 3.4 Venn diagram showing overlap of down-regulated genes in response to the four assayed abiotic stresses.....	159
Figure 3.5 Weighted gene co-expression network of *Brachypodium* stress responsive genes.....	160
Figure 3.6 Expression profiles of modules as a function of time….....	162
Figure 3.7 Scatterplot of transcription factor/target gene correlations.....	169
Figure 4.1 Depiction of the overlapping and complex cellular responses resulting from abiotic stressors.....	196
Figure 4.2 Heat shock protein network during abiotic stress response.....	231
Figure 4.3 Current model of the heat stress response pathway in tomato.....	232

LIST OF APPENDIX FIGURES (continued)

Figure Page

Figure 5.1 Alignment of RNA-seq reads to a portion of a Velvet-assembled Hazelnut EST contig and inferred SNP..... 298

Figure 5.2 Classes of single nucleotide polymorphism detected in hazelnut..... 299

LIST OF APPENDIX TABLES

Table	Page
Table 1.1 Sequence of the oligo fragments used to form the complete dsDNA	121
Table 2.1 Assembly statistics for 'Jefferson' reference genome assembly…..	137
Table 2.2 Assembly statistics for 7 additional European hazelnut accessions sequenced	140
Table 2.3 Summary of SNPs and INDELs detected between hazelnut accessions and 'Jefferson'	142
Table 3.1 Module membership and functional and regulatory enrichment…..	164
Table 3.2 Specific GO terms uniquely enriched in a selection of network modules	165
Table 3.3 Specific short DNA sequences found to be statistically enriched in the promoters of module member genes	167
Table 3.4 Putative network plasticity present between all pairwise conditional comparisons	171
Table 5.1 Description of the RNA-seq data and EST contigs used for single nucleotide polymorphism detection	297
Table 5.2 Summary statistics for Velvet output contigs	298

DEDICATION

For my Mom and my Dad for their unwavering support over the years!!

Music.

Chapter 1:

Introduction

Erik R.Rowley

The Pacific Northwest (USA) is the country's largest producer of European hazelnut (*Corylus avellana* L.), representing nearly 99% of the U.S. hazelnut market. Production is centered in the Willamette Valley of Oregon, which shares a similar climate to the epicenter of the world's hazelnut production, the Black Sea region of Turkey. Turkey accounts for approximately 70% of the global hazelnut market, producing over 1 million tons annually (Gökirmak, Mehlenbacher, and Bassil 2008). Despite accounting for 3-5% of the global market, Oregon hazelnuts bring over US $67 million dollars to the region annually (USDA NASS Oregon Field Office, 2011).

C. avellana is of worldwide importance and consumer demand due to the multitude of products made from its kernels. Hazelnuts have been popularized as the predominant flavor in a variety of popular confectionary pastes, butters, and chocolate spreads. Additionally the fiber and vitamin rich whole kernels may be either blanched or roasted, accounting for a large portion of the regional market in the Pacific Northwest. The shells of the hazelnut are commonly used for landscaping and as groundcover.

Of major concern to European hazelnut breeders in North America is a fungal disease known as eastern filbert blight (EFB). This disease is responsible for severe crop loss in the susceptible cultivars currently grown in the region, including 'Barcelona', the heirloom accession grown throughout much of the Pacific Northwest. EFB is caused by the fungus *Anisogramma anomola* (Peck) E. Müller, and manifests itself via cankers. These cankers slowly girdle the branches, leading first to a reduction in leaf growth and ultimately the death of the tree after only a few years. Several research groups are developing hazelnut cultivars with resistance to this disease, including extensive efforts using DNA markers to identify seedlings that carry the dominant allele from the accession 'Gasaway', shown to confer resistance to EFB.

In addition to developing germplasm resistant to EFB, hazelnut breeding efforts have focused on producing progeny with enhanced agronomic traits of interest including kernel oil content and size, blanching ability, and bud mite resistance. When implementing a breeding strategy aimed at crop improvement, breeders utilize a process known as complementary hybridization. Crosses are planned in order for the beneficial traits of one parent to complement the weaknesses, such as disease susceptibility, of the other. The parents of such crosses should be genetically unrelated in order to optimize success. One issue hampering the suitability of certain accessions as parents is self-incompatibility (SI).

C. avellana is monoecious; separate male and female flowers exist on the same tree. In order to prevent inbreeding and encourage genetic diversity, flowering plants have evolved self-incompatibility mechanisms to promote outcrossing. While providing self/non-self determination, it also limits the individuals that may be used in crosses. If the stigma and pollen express the same allele, SI renders the cross incompatible. This hampers hazelnut breeding efforts by limiting the parents that can successfully be used for crosses. Traditionally fluorescence microscopy is used to determine the compatibility of pollinations, and to identify S-alleles in hazelnut varieties from different geographical origins.

European hazelnut is a member of the family Betulacea, a group of six plant genera which include the birches (*Betula* L. spp.), alders *(Alnus* Mill. spp.), and hornbeans (*Carpinus L. spp.).* In addition to its agricultural importance, hazelnut possesses several traits that make it an attractive candidate for use as a model . for the family Betulacea. These include a short life cycle, producing fruit 5 years post-planting and a small stature of only ~ 5 m. Hazelnut is also a diploid with 11 chromosomes (2n =2x=22) and an empirically determined genome size of ~378 Mbp, approximately triple that of *Arabidopsis thaliana* (L.) Heynh, at 125 Mbp but is significantly smaller than other tree genomes such as *Populus trichocarpa*

at ~520 Mbp (Tuskan and Torr 2014) and loblolly pine at ~23 Gbp (Neale *et al.* 2014).

Genetic diversity exists in over 700 accessions of hazelnut collected by The United States Department of Agriculture (USDA) and Oregon State University (OSU), which are preserved in Corvallis, Oregon for use in trait improvement efforts. Among these accessions is an F1 mapping population of 138 individuals resulting from the cross of the EFB-resistant accession OSU 414.062 (paternal) and EFB-susceptible accession OSU 252.146 (maternal).

Within this population is the cultivar 'Jefferson' (OSU 703.007). ‘Jefferson’ contains the dominant EFB resistance allele ‘Gasaway’ and was released in 2009 by Dr. Shawn Mehlenbacher at Oregon State University (Mehlenbacher *et al.* 2011). Given the agricultural relevance of hazelnut, and its attractive traits for developing a model system, it is a genome that would benefit from the development of genomic tools to assist genetic improvement efforts. ‘Jefferson’ was chosen for Illumina next-generation sequencing, assembly, and annotation.

With the advent of high-throughput next-generation sequencing (NGS) technologies, specifically the Illumina short-read sequencing platform, it is possible to rapidly sequence and *de novo* assemble novel plant genome sequences. In this method, called shotgun sequencing, an unknown sequence of nucleic acid is fragmented (representing the genome or cDNA fragments in the case of RNA transcripts), the ends are repaired, and known sequences (adapters) are ligated to the now blunt-ended 5’ and 3’ ends. These adapters are attached to a glass slide, called a flowcell, in the Illumina machine along with the downstream single stranded nucleic acid sequence. The Illumina sequencing-by-synthesis technology then detects single nucleotides as they are incorporated, base-by-base, into the growing second strand DNA molecule. The resulting millions of sequencing reads, between 80 and 100 nucleotides in length, are assembled into longer pieces of contiguous sequence (contigs) based on their overlaps. There are a plethora of programs that have been developed and are available to deal with the gigabases

(Gb) of sequence data that result from an NGS run, with new and faster algorithms and methods being developed almost continuously.

Once these short reads have been assembled into contigs, other programs are used to identify regions of the sequence that may code for proteins and determine the location of genes. In order to predict the function of these genes, the amino acid sequences translated arise from these regions are functionally annotated via homology to sequences in the public domain at the curated database at the National Center for Biotechnology Information (NCBI). Once these annotations have been performed, the genic content can be characterized, novel genes can be identified, and putative candidates responsible for phenotype can be predicted for future molecular characterization and hypothesis generation.

To date, genetic and genomic resources for hazelnut have included a genetic linkage map based on random amplified polymorphic DNA (RAPD) and microsatellite marker loci (Mehlenbacher *et al.*, 2004, Mehlenbacher *et al.*, 2006), and a BAC library (Sathuvalli *et al.*, 2011). Having additional genomic tools for hazelnut, such as an annotated genome and transcriptome, will serve as a tool for gene discovery and functional studies. The genome assembly will allow for the development of DNA markers and other genomic tools for breeders, as well as allow integration of the genome sequences and the genetic and physical maps.

The main objectives of this research were to 1) develop genomic tools for *C. avellana* 'Jefferson' 2) resequencing and variant discovery among seven diverse hazelnut cultivars 3) identify additional markers in the F1 mapping population and improve the existing genetic linkage map

Chapter 2:

Assembly and characterization of the European hazelnut (*Corylus avellana* L.) 'Jefferson' transcriptome

Erik R. Rowley, Samuel E. Fox, Douglas W. Bryant, Christopher M. Sullivan, Henry D. Priest, Scott A. Givan, Shawn A. Mehlenbacher, and Todd C. Mockler

Crop Science
5585 Guilford Road
Madison, WI 53711, USA
52: 2679-2686 (2012)
DOI: 10.2135/cropsci2012.02.0065

Abstract:

European hazelnut (*Corylus avellana* L.) is of worldwide agronomic importance, with an urgent need to breed resistance to Eastern Filbert Blight (EFB), a cause of severe crop loss in much of the United States. Oregon State University recently released a heterozygous resistant cultivar, 'Jefferson' (OSU 703.007), selected for a transcriptome assembly for establishment of further genetic resources. We sequenced cDNA libraries on the Illumina platform from four unique hazelnut tissues; leaves, catkins, bark, and whole seedlings. The hazelnut transcriptome data generated was de novo assembled into 28,255 contigs (≥80bp) with an average length of 532bp, an N50 value of 961. Using the 1e-5 cutoff, 75% result in BLAST hits and 60% yield gene ontology gene product classifications. High protein product similarity to related plant transcriptomes demonstrates the validity of gene models. A survey of enriched tissue specific GO terms further validates the assembly. A searchable database is available and will be of importance to breeders in aligning future physical and genetic maps and aid in marker-assisted breeding efforts, such as the discovery and mapping of important agronomic traits.

Introduction:

European Hazelnut (*Corylus avellana* L.) is of worldwide agronomic importance due to the multitude of products made from its kernels. Butters, pastes, confectionary spreads, flours, whole kernels, and the shells are in demand by consumers worldwide. Hazelnut grows best in regions having mild winters, with 70% of the world's production coming from the Black Sea region of Turkey[1]. The tree is also of agronomic importance to growers in the Willamette Valley of Oregon (USA), introduced to the Pacific Northwest, with the moderate climate of the coastal valleys being well-suited to hazelnut production. This region provides nearly 99% of the hazelnuts grown in the United States, with an estimated annual value in Oregon of $59.6 million (http://www.nass.usda.gov/).

Hazelnuts are a clonal diploid crop, and breeders utilize the beneficial traits of one tree to complement the weaknesses of another (Gokirmak *et al.*, 2009), in a process known as complementary hybridization. The parents of the crosses should be genetically distinct, and therefore it is of great benefit to hazelnut breeders to have genetic resources in place in order to take advantage of the wealth of genetic diversity while choosing parents, where there exist hundreds of accessions worldwide.

Other factors of importance when implementing a breeding strategy include resistance to disease and pathogen susceptibility. Of major concern to European hazelnut grown in the United States is Eastern Filbert Blight (EFB), a cause of severe crop loss to the mostly non-resistant cultivars grown in the majority of the country. EFB is caused by the fungus *Anisogramma anomola* (Peck) E. Müller, which manifests itself via the growth of cankers that slowly girdle branches, limbs, and tree trunks leading to the decline of leaf growth and ultimately the death of the tree (Johnson *et al.*, 1996). There are many groups selectively breeding hazelnut cultivars with resistance to this disease (Chen *et al.*, 2005), including extensive breeding efforts using DNA markers to identify seedlings that

carry the 'Gasaway' gene, which confers dominate single gene resistance to EFB (Mehlenbacher *et al.*, 1991), prior to planting.

A genetic linkage map is also available for European hazelnut (Mehlenbacher *et al.*, 2006), with microsatellite loci being used to genetically fingerprint hazelnut cultivars, whereupon their loci are added to the linkage map. Also in existence is a BAC library (Sathuvalli *et al.*, 2011), enabling groups to focus on map-based cloning of the EFB resistance and pollen-stigma incompatibility loci.

With its worldwide importance as a crop and current available genetic resources, *C. avellana* also possesses many attributes that make it an attractive candidate for use as a model system for the family Betulaceae; a family which includes the birches and alders. *C. avellana* has a relatively short life cycle; bearing seeds at around five years, has a short stature for a tree (~5M), a small genome of ~400 Mp (approximately triple that of established monocot plant model *Arabidopsis thaliana* at 125 Mb), and is amenable to transformation. The cultivar ‘Jefferson’ (OSU 703.007) was released by Oregon State University in 2009 (Mehlenbacher *et al.*, 2011), and was selected to be the reference hazelnut accession to be sequenced.

In order to begin fully utilizing ‘Jefferson’ as a genetic resource, it is first necessary to establish a high-quality reference transcriptome assembly, a task suited to a massively parallel high-throughput RNA sequencing (RNA-seq) approach on the Illumina platform (reviewed by Fox *et al.*, 2009). The resulting short reads, 80 and 36 nucleotides in length were trimmed and *de novo* assembled based on their overlaps into contigs representing putative gene models (transcript assemblies). These transcript assemblies can then be utilized as a tool for homology searches and BLAST queries. We used this approach to create an initial transcriptome assembly for ‘Jefferson’, with various analyses offering insights into gene content, protein homology to related organisms, and gene product

enrichment levels among tissues. These analyses demonstrate the quality of the initial assembly and we offer the data to the research community for further analyses.

Materials and methods:

Hazelnut Samples, Collection, and RNA Preparation. Tissues for ribonucleic acid (RNA) isolation were collected from 1-yr-old nursery-grown trees as well as from field grown examples. To better represent the expressed genic content of the transcriptome, tissues of four different types were collected. Bark and leaves were collected from nursery-grown Jefferson trees. Catkins were collected from a field-grown example of the 'Barcelona' accession (PI 557037) while the whole seedlings, including roots, were the progeny of a cross: OSU 954.076 × OSU 976.091. Catkin collection was conducted in autumn and sampling of the other tissues was conducted in late spring when the trees were actively growing.
Total RNA was isolated using a protocol previously described (Filichkin *et al.*, 2007). Ribonucleic acid was extracted using Plant RNA Reagent (Invitrogen). Total RNA was treated for 10 min at 65°C with RNasecure reagent (Ambion). To eliminate genomic DNA contamination, all RNA preparations were treated with ribonuclease (RNase)-free Turbo DNase (deoxyribonuclease) (Ambion) for 15 min at 37°C. Total RNA was further purified using the RNeasy Mini RNA kit (Qiagen) according to the manufacturer's cleanup protocol. Isolation of messenger RNA (mRNA) essentially free of ribosomal and other nonpolyadenylated multiple RNA molecules or sequences was critical for generation of nonbiased randomly primed (RP) libraries. For the RP libraries, the poly (A) mRNA was isolated by two consecutive cycles of purification on oligo d(T) cellulose using the Micro-PolyA-Purist kit (Ambion). Concentration, integrity, and extent of contamination by ribosomal RNA were assessed using a ND-1000 spectrophotometer (Thermo Fisher Scientific) and Bioanalyzer 2100 (Agilent Technologies).

Random Primed cDNA Synthesis. The seedling complementary DNA (cDNA) libraries were prepared by a random priming method, in which first-strand cDNA was synthesized using 1 μg of poly(A) mRNA. Random hexamer primers (300 ng μg−1 RNA) and Superscript III reverse transcriptase (Invitrogen) were added to the reaction and incubated at 75°C for 5 min. Second-strand cDNA was synthesized by combining 20 μL of the first-strand reaction, 8 μL of 10x Klenow Buffer (New England Biolabs), 1 U RNase H (Invitrogen), 68.8 μL water, and 30 U DNA polymerase I, Klenow fragment (New England Biolabs). The reaction was incubated for 90 min at 15°C and cDNA was purified using a QIAquick PCR (polymerase chain reaction) Purification Kit (Qiagen).

SMART Complementary DNA Synthesis. We used a modification of the BD Clontech SMART cDNA Library Construction method that uses SMART adaptor primers (Zhu *et al.*, 2001). The bark, leaf, and catkin cDNA libraries were prepared using the SMART method. Each primer (CDS III 3′ PCR primer [Clontech] to capture the poly(A) tail and 5′ SMART IV oligonucleotide) were added to 250 to 500 ng of poly(A) RNA sample. Samples were incubated at 72°C for 5 min and then placed on ice for 2 min. We then added 2 μL of 5x First-Strand Synthesis Buffer (Clontech kit) and 1 μL of 20 mM dithiothreitol, 1 μL of 10 mM deoxyribonucleotide triphosphates (dNTPs), and 1 μL of moloney murine leukemia virus reverse transcriptase. The reaction was incubated at 42°C for 1 h followed by a PCR amplification step. Polymerase chain reactions were as follows: 80 μL sterile water, 10 μL of 10x Advantage 2 PCR buffer, 2 μL of 50x dNTPs (10 mM each), 4 μL of 5′ PCR primer II A, 2 μL of 50x Advantage 2 PCR Polymerase Mix, and 2 μL of the control first-strand cDNA (provided with BD Bio- sciences Clontech kit) amplified for 15 cycles and purified using the QIAquick PCR Purification Kit. The SMART-prepared samples were fragmented by nebulization before preparation for the Illumina sequencing. We transferred the cDNA sample (brought up to a 50 μL volume) to a nebulizer and added 750 μL of nebulization buffer (Illumina). We fragmented the DNA using compressed

N at 220 to 240 kPa for 7 min and then isolated the sheared DNA using a QIAquick PCR Purification Kit.

Preparation of RNA Sequencing Libraries for the Illumina Genome Analyzer IIx. Fragmented RP or SMART-prepared cDNA (~30 μL) was combined with 10 μL of 10 mM adenosine triphosphate in 5x T4 DNA ligase buffer (Invitrogen), 4 μL of 10 mM dNTPs mix, 2.5 μL of T4 DNA polymerase (3 U μL−1), 1 μL of Klenow DNA Pol (5 U μL−1), and 2.5 μL of T4 polynucleotide kinase (10 U μL−1) (New England Biolabs). After incubation for 30 min at 20°C the DNA was purified using QIAquick PCR Purification kit. To add deoxyadenosine to the termini, 32 μL DNA from the prior step was mixed with 5 μL of 10x Klenow buffer, 10 μL of 1 mM deoxyadenosine triphosphate, and 3 μL of Klenow exo- polymerase (3′ to 5′ exo minus, 5 U μL−1) (New England Bio- labs) and incubated for 30 min at 37°C. The DNA was purified using a QIAquick MinElute Reaction Clean-up kit (Qiagen). To ligate Illumina adapters, 10 μL of cDNA from the prior step was mixed with 5 μL of 5z T4 DNA ligase buffer, 6 μL of adaptor oligo mix, and 4 μL of T4 DNA ligase (New England Biolabs) and incubated for 15 min at 25°C. The DNA was purified using a QIAquick MinElute PCR Purification Kit. Complementary DNA was size-fractionated (with average fragment length of between 325 and 350 bp) on 3.5% (w/v) NuSieve GTG agarose. The fractionated libraries were PCR amplified using Phusion Hot Start High-Fidelity DNA polymerase (New England Bio- labs) and the following PCR protocol: 30 s at 98°C and then 10 s at 98°C, 30 s at 65°C, and 30 s at 72°C for 18 cycles followed by a 10-min extension step at 72°C and purified using a QIAquick PCR Purification kit. Illumina cluster generation was done in the Oregon State University Center for Genome Research and Biocomputing core facility using a standard Illumina protocol.

Transcriptome Assembly. Illumina reads were filtered to remove any reads having greater than 5% ambiguous basecalls (Ns; N is a position in a sequence

assembly which could not be resolved as any of the four nucleotides) as called by the Illumina CASAVA pipeline. The resulting ~32 million 36mers and ~70 million 80mers, including both paired-end reads (PEs) and single-end reads (SEs) and totaling ~6.8 Gb of sequence, were used to assemble the Jefferson transcriptome. Leaves were represented by 3.65 Gb of PE data, bark by 1.07 Gb of PE data, catkins by 0.91 Gb of PE data, and seedlings by 1.11 Gb of SE data and 0.06 Gb of PE data, respectively (**Table 2.1**).

Table 2.1. Summary of Tissues used for RNA-seq

	RNA-seq Library type	**Length (bp)**	**Gbp**
Leaf	PE	80	3.65
Bark	PE	80	1.07
Catkins	PE	80	0.91
Seedlings	PE	80	0.06
Seedlings	SE	36	1.11

The random priming method used to generate the seedling libraries may result in enrichment for chloroplast transcripts. The differences in read length, numbers of reads, and read type (single-end or paired-end) reflects changes in the Illumina platform over time during the course of this study. The reads were trimmed to compensate for the higher than average error rate associated with base calls near the 5′ and 3′ ends of reads. The 80mers were trimmed to 76mers by removing two nucleotides from each end and then further trimmed from nucleotides 2 to 52, generating 50mers, and from nucleotides 30 to 70, generating 40mers. The last two bases of the 36mers were trimmed, generating 34mers, and all the resulting reads were then pooled together for assembly. The assembly program Velvet version 1.0.13 (Zerbino and Birney, 2008) was used for de novo assembly of the Illumina reads, using a hash length of 31 and incorporating mate-distance information for the PEs. The contigs output from Velvet were used as input to the assembler MIRA version 3.0.0 (Chevreux *et al.*, 2004) along with an additional one million paired-end 76mer reads derived from seedlings that were subjected to

the same filtering and trimming process as described above in an effort to merge the Velvet contigs when possible. In its default output format, MIRA contigs contain International Union of Pure and Applied Chemistry (IUPAC) bases at locations where the overlap of bases disagree with each other. To facilitate functional annotation, SOAP (Li *et al.*, 2008) and SOAPsnp (Li *et al.*, 2009) were used to determine the most common nucleotide occurring at each specific position in each assembled contig. This consensus base replaced the IUPAC symbol at ambiguous positions, resulting in 28,255 hazelnut transcript contigs that were used for further analysis. Overall there were 2705 Ns in the final assembly of 15,037,682 nucleotides, rep- resenting 0.017% of the total sequence.

Functional Annotation. The processed contigs were used for predicting protein-coding regions using OrfPredictor, a web-based open reading frame (ORF) prediction tool (Min *et al.*, 2005), resulting in 26,375 (93.3%) predicted protein-coding ORFs of at least 33 amino acids, or 99 bases, in length. The BLASTX tool (Altschul *et al.*, 1997) was used with an E-value cut-off of 1×10^{-5} to search the National Center for Biotechnology Information (NCBI) non- redundant protein database using the translated ORFs from the assembly. The package Blast2GO (Conesa *et al.*, 2005) was used to predict gene ontology (GO) terms for the contigs by assigning functional classifications (Gene Ontology Consortium, 2000) and potential properties of gene products to the contigs.

Tissue-Specific Gene Expression. Reads per kilobase per million mapped reads (RPKM) (Mortazavi *et al.*, 2008) values were generated for each of the four tis-sue types by aligning the input Illumina reads from each tissue to the 28,255 transcript contigs of the transcriptome assembly. The RPKM values were log2 transformed and the R-statistical analysis package (R Development Core Team, 2008) was used to filter the data based on specific log2 RPKM cutoffs. The web-based tool Blast2GO (Conesa *et al.*, 2005) was used to both search for homology using BLASTX (Altschul *et al.*, 1997) and to subsequently annotate and assign

GO terms to the filtered data and generate graphs. Visualization of expression pro- files and hierarchal clustering for the generation of a heatmap corresponding to the tissue-specific log2 RPKM levels were conducted using the using the package MADE4 version 1.16.0 (Culhane *et al.*, 2005) in R.

Data Availability. All RNA-seq data obtained in this study are available in the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/ sra/ [accessed 13 June 2012]) under accession SRP013737.

Results:

Illumina Sequencing and Assembly. The aim of this project was to sequence, assemble, annotate, and broadly characterize European hazelnut transcripts. To this end, a diverse set of *C. avellana* tissues and organ types was collected: young leaves, catkins, bark, and whole young seedlings. RNA was extracted and RNA-seq libraries were prepared for sequencing on the Illumina Genome Analyzer IIx platform. The resulting dataset included both SEs and PEs, which were filtered and trimmed to account for any ambiguity or error in their 5′ and 3′ ends of each sequence. The filtered and trimmed reads were then pooled and assembled in a two-step process using the short read assembler Velvet (Zerbino and Birney, 2008) followed by MIRA (Chevreux *et al.*, 2004) to further merge the Velvet-generated contigs. The final output from MIRA contained 28,255 transcript contigs (**Table 2.2**) having an average length of 532 bp and an N50 (the minimum contig length necessary such that all contigs of equal or greater length will equal half of the bases of the assembly) of 961 bp. Our results are comparable to other recent de novo transcriptome assemblies such as chickpea (*Cicer arietinum* L.) (Garg *et al.*, 2011), which had an average contig length of 523 bp and N50 of 900 bp, the gametophyte of the brackenfern [*Pteridium aquilinum* (L.) Kuhn] (Der *et al.*, 2011) with an average contig length of 547 bp, rubber tree [*Hevea brasiliensis* (Willd. ex A. Juss.) Müll. Arg.] (Xia *et al.*, 2011) with an average contig length of 436 bp, and whitefly (*Bemisia tabaci*) (Wang *et al.*, 2011) with an average contig

size of 266 bp.

Table 2.2. Summary of Transcriptome Assembly

	Contigs	**Average (bp)**	**N50**
All Contigs	28,255	532	961
100-999 bp	22,358	246	
Contigs > 100bp	27,010	552	
Largest contig	5,490		

Of the 28,255 hazelnut transcript contigs, the majority (~79%) were 100 to 999 bp in length (**Figure 2.1**); however, 4.4% (1245) were within the range of 80 to 100 nucleotides (nt).

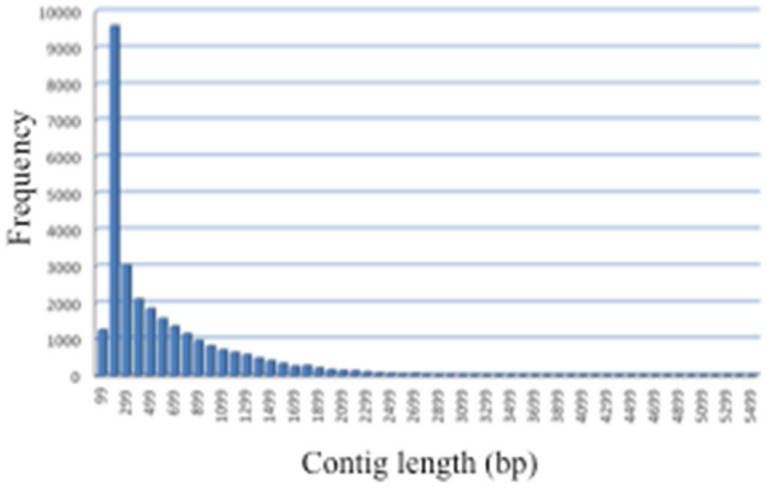


Figure. 2.1. Graph of the contig length distribution for the 'Jefferson' transcriptome assembly, with frequency on the y-axis and contig length (bp) on the x-axis. These very short contigs may represent partial transcripts or fragments of splice variants and were retained to provide broad representation of the diversity of sequences in the 'Jefferson' transcriptome.

***Functional Annotation of Assembled Contigs*.** Given the agricultural importance of hazelnut, the objective of this study was to provide a sequence resource for

breeders to use in gene and marker discovery searches as well as to create an initial transcriptome assembly for European hazelnut that will facilitate the identification and annotation of gene models in the eventual genome sequence. To achieve these goals, the initial step following transcript assembly was the descriptive annotation and assignment of putative biological functions to the transcript contigs. We used the tool OrfPredictor (Min *et al.*, 2005) to predict protein-coding regions in the assembled contigs. Of the 28,255 transcript contigs, 26,375 (93.3%) contain predicted ORFs of at least 33 amino acids (**Table 2.3**).

Table 2.3. Summary of Functional Annotation

BLAST threshold	**Contigs containing putative ORFs**	**Contigs with BLASTx hits**	**Contigs with GO terms**
10^{-5}	26,375	21,202	16,488

The BLASTX algorithm (Altschul *et al.*, 1997) was used to align the translated nucleotide sequences of the ORFs to the NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi [accessed 25 Mar. 2011]) nonredundant protein database using an E-value cutoff of 1×10^{-5} resulting in matches for 21,202 (75%) of the transcripts (Table 2). Functional classifications were performed on these 21,202 contigs using GO analysis (Gene Ontology Consortium, 2000) to survey, categorize, and define the potential properties of gene products with respect to their predicted biological contexts. We used the program Blast2GO, a tool for assigning GO terms to unknown sequences (Conesa *et al.*, 2005), to functionally annotate the 21,202 transcripts from the BLASTX output. This resulted in GO functional classifications for 16,488 (~60%) of the transcript contigs comprising 58,495 GO terms broadly grouped by GO component classifications (**Figure 2.2**).

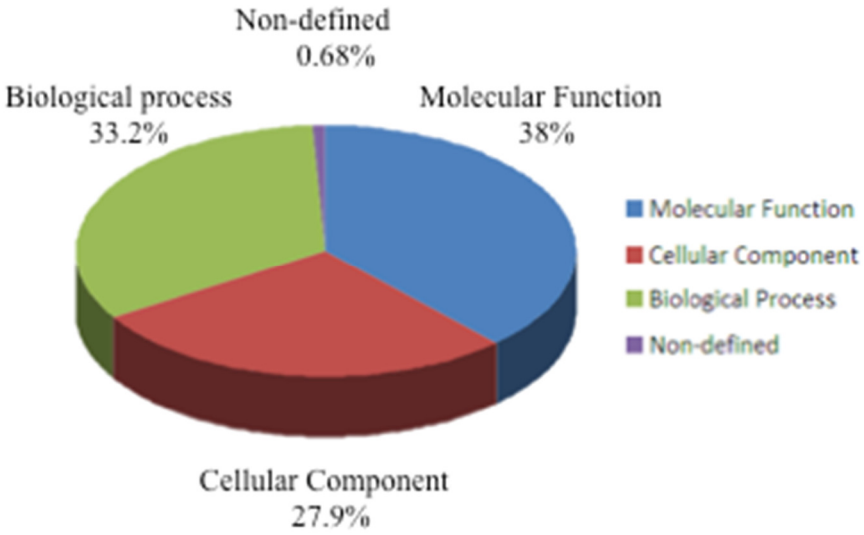


Figure 2.2. Pie chart depicting the three broad GO classifications from the functional associations. 58,495 GO terms were assigned to the 16,488 transcript models, with 27.9% assigned to Cellular Component, 33.2% assigned to Biological Process, 38% assigned to Molecular Function, and 0.68% without current assignments to the GO database at the time of writing this manuscript

Among the transcript contigs classified by Blast2GO, 27.9% were assigned to "cellular component" ontology, 33.2% were assigned to "biological process" ontology, and 38% to "molecular function" ontology, with only 0.7% lacking assignment to GO classifications.

We also surveyed the similarity of the proteins encoded by the transcript contigs to existing sequences currently in the NCBI nonredundant protein database using BLASTX (Altschul *et al.*, 1997). Matches to three organisms in particular represented a combined total of 80.8% of the *C. avellana* transcript contigs: 36.6% had best BLASTX matches in grape followed by poplar at 23.1% and castor bean with 20.9% (**Figure 2.3**).

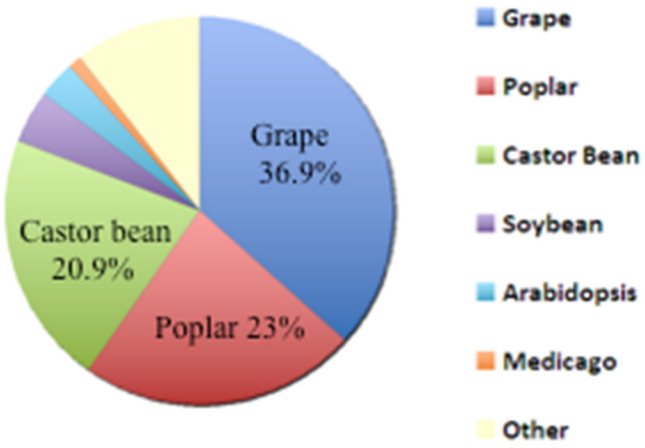


Figure 2.3. Organism breakdown of the BLASTX results showing transcript contigs having a match in the nr protein database.

The remaining 19.2% of the transcript contigs had best matches to sequences from more distantly related plants and other organisms.

Survey of Differential Expression Among Tissues. While this study was designed to identify transcripts of *C. avellana*, we took advantage of the transcriptome assembly generated from the pooled tissues and the Illumina reads for each individual tissue to identify genes that were differentially expressed in different tissues. We used the RPKM approach to normalize both for transcript length and total number of reads in each library and to facilitate comparison of transcript levels among samples (Mortazavi *et al.*, 2008). The Illumina reads from each tissue were aligned to the transcriptome assembly, and RPKM values were generated for each of the four tissue types, log2 transformed, and used to assess gene expression differences between the tissue types using standard methods (Gan *et al.*, 2010; Hebenstreit *et al.*, 2011). To assess differential expression of genes between the tissues, the log2 transformed RPKM values were filtered to identify transcript contigs displaying an absolute difference of at least 8 between pairs of tissues using R (R Development Core Team, 2008) and visualized using the Bioconductor package MADE4 (Culhane *et al.*, 2005), which performs

multivariate analysis and hierarchical clustering of gene expression data (**Figure 2.4**).

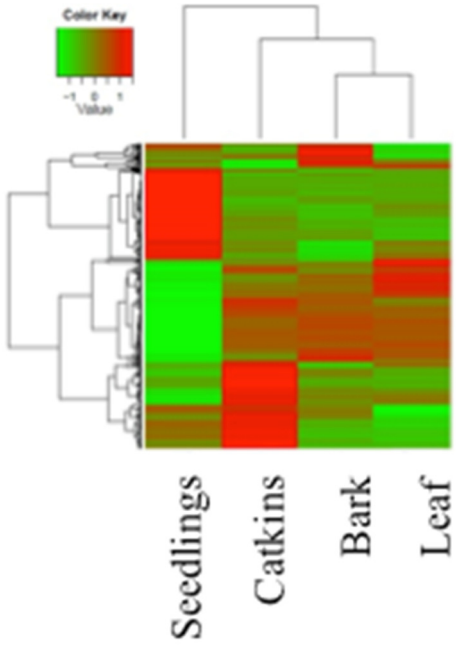


Figure 2.4. Visual illustration of differential enrichment of tissue-specific transcript contigs, filtered by those displaying absolute difference of ≥8 in log2 RPKM transformed values between one another. The hierarchically clustered heatmap was generated with the R-statistical analysis package MADE4 version 1.16.0, with values scaled by default to aid visualization.

Filtering the full set of log2 transformed RPKM values for those transcript contigs exhibiting an RPKM change of 8 or greater between tissues enabled identification of genes whose expression was highly enriched within each tissue.

This analysis identified 533 transcript contigs whose expression was enriched in bark, 542 contigs enriched in catkins, 321 contigs enriched in leaf, and 538 contigs enriched in seedlings. A precise estimate of fold changes or assignment of *p*-values or false-discovery rates was not feasible given the lack of replication in this analysis. Surveying each tissue's GO classifications using a term filter value of 5 and filtering based on the number of sequences demonstrated an enrichment of contigs classified as "chloroplast stroma" (~14%), "chloroplast envelope"

(12.5%), “apoplast” (10%), and “chloroplast thylakoid membrane” (8.7%) in leaf (**Figure 2.5**).

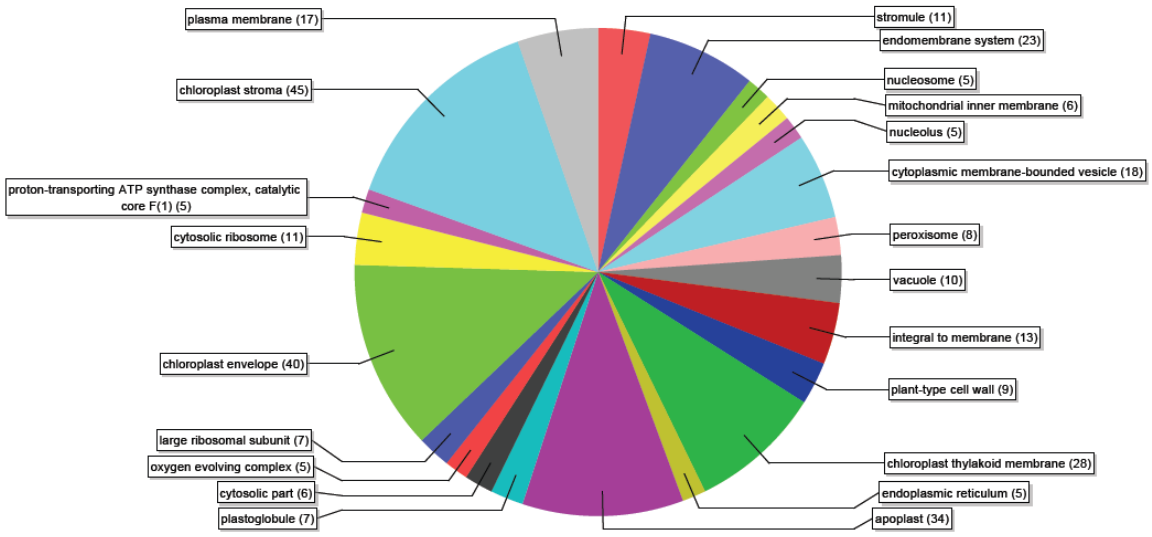


Figure 2.5. Multi-level pie chart created using BLAST2GO (BLAST2GO.org) showing enrichment of cellular component plant GOslimmer terms corresponding to transcripts in leaf having log2 RPKM values ≥8, and a GO #sequences term cutoff of 5. The three most represented GO terms, comprising 35% of the total, relate to the chloroplast and apoplast.

As another example, for seedlings (**Figure 2.6**), which contained emerging leaves, the most abundant GO terms were “integral to the membrane” (30.4%) followed by “chloroplast thylakoid membrane” (18.1%). In catkins, we observed enrichment of GO terms corresponding to the plasma membrane (16.3%) followed by the endomembrane system (12.5%) and cytoplasmic membrane-bound vesicles (12.5%) while in bark the most enriched GO terms corresponded to the plasma membrane (19.2%) followed by those for gene products predicted as integral to the membrane (13.1%).

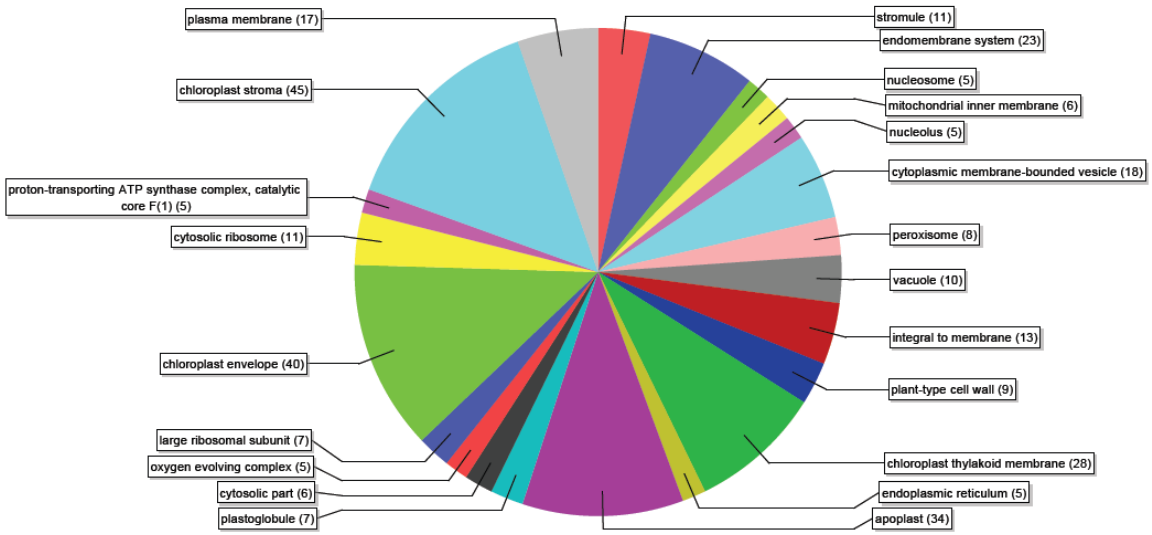


Figure 2.6. Multi-level pie chart created using BLAST2GO (BLAST2GO.org) showing enrichment of cellular component plant GOslimmer for those transcripts in seedlings (which contained emerging leaves) filtered by log2 RPKM values ≥8 and a GO # sequences term cutoff of 5. The top enriched GO terms are those integral to the membrane at 30.4%, and those corresponding to the chloroplast thylakoid membrane at 18.1%.

Discussion:

This work demonstrates how advances in high-throughput sequencing technologies coupled with bioinformatics tools and public databases enabled the rapid assembly of a large collection of transcripts in a complex nonmodel plant. We generated a hazelnut transcriptome assembly comprising 28,255 transcript contigs, the vast majority of which (93.3%) encode proteins. It is clear that the assembled *C. avellana* transcript contigs reported here are broadly representative of the hazelnut transcriptome for two reasons. First, we sampled ~6.8 Gb of sequence, which is consistent with a recent survey (Priest *et al.*, 2010) involving several plant species that demonstrated that 2 Gb of RNA-seq data was sufficient to sample ~90% of annotated genes. Second, using the basic local alignment search tool (BLAST) (Altschul *et al.*, 1990) to compare the putative proteins encoded by our assembly to the annotated proteomes of *Arabidopsis thaliana* (L.), grape, and poplar identified matches to 63, 67, and 67% of all of the annotated

proteins in these plants, respectively (data not shown). Therefore, our assembly is likely to comprehensively cover the transcribed gene content in the four tissues sampled and be representative of a large fraction of hazelnut genes. Descriptive functional annotations were made for 21,202 (~75%) of these putative encoded proteins, with ~81% displaying homology to annotated grape, poplar, and castor bean proteins.

The distributions of RNA-seq contig lengths in hazelnut are comparable to those obtained in de novo transcriptome assemblies in other species, including chickpea, brackenfern, rubber tree, and whitefly. Our assembly does contain many short contigs that may represent transcript fragments but that will nevertheless be useful for some downstream applications such as marker discovery and gene identification. Many of the contigs in the ~100 nt range contained apparent protein encoding ORFs, further supporting the hypothesis that such contigs represent either partial transcripts or fragments of unassembled alternative splice variants. Through differential expression analysis of these transcriptomic data, genes were identified as being enriched in particular tissues. For example, the leaf and seedling libraries were enriched for chloroplast-associated GO categories as expected. Such results may offer future insights into gene regulation within these tissues.

Because *C. avellana* is an important crop, the sequence resource generated here will benefit breeders in, for example, discovering the genes responsible for EFB resistance (Sathuvalli, 2011) and will thus facilitate selecting for resistance to EFB. This sequence resource will also be of use in identifying genes responsible for other traits such as kernel oil content, fatty acid composition, tocopherol content, and tree growth habit. Also, having available a hazelnut reference transcriptome sequence database facilitates annotation of a future hazelnut genome assembly and will enable comparative genomics efforts among cultivars and with other species. This initial transcriptome assembly will undoubtedly

prove useful in other studies, such as in efforts to determine the molecular basis of self-incompatibility, which hinders crosses between some cultivars. In the near future, the transcript contigs generated in this study will be used to aid annotation and validation of a draft Jefferson genome assembly (T.C. Mockler and S.A. Mehlenbacher, unpublished data, 2012). Currently, the transcriptome assembly is being used for development of expressed SSR markers (B. Peterschmidt, V. Sathuvalli, and S. Mehlenbacher, personal communication, 2012), providing another future resource to the hazelnut genetics community.

The Jefferson transcriptome is available to the community as a public web-based database (http:// hazelnut.cgrb.oregonstate.edu [accessed 8 Jan. 2010]) that includes a BLAST interface and downloadable files containing the transcript contig sequences, predicted coding sequences, predicted proteins, and descriptive annotations based on BLASTX (Altschul *et al.*, 1997) comparisons and GO assignments.

Literature Cited:

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang

W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI- BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402. doi:10.1093/nar/25.17.3389

Chen, H., S.A. Mehlenbacher, and D.C. Smith. 2007. Hazelnut accessions provide new sources of resistance to eastern filbert blight. HortScience 42:466–4695.

Chevreux, B., T. Pfisterer, B. Drescher, A.J. Driesel, W.E. Müller, T. Wetter, and S. Suhai. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 14:1147– 1159. doi:10.1101/gr.1917404

Conesa, A., S. Götz, J.M. García-Gómez, J. Terol, M. Talón, and M. Robles. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 18:3674–3676.

doi:10.1093/bioinformatics/bti610

Culhane, A.C., J. Thioulouse, G. Perrière, and D.G. Higging. 2005. MADE4: An R package for multivariate analysis of gene expression data. Bioinformatics 11:2789–2790.

Der, J.P., M.S Barker, N.J. Wickett, C.W. dePamphilis, and P.G. Wolf. 2011. De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. BMC Genomics 12:99.

Filichkin, S.A., S.P. DiFazio, A.M. Brunner, J.M. Davis, Z.K. Yang, U.C. Kalluri, R.S. Arias, E. Etherington, G.A. Tuskan, and S.H. Strauss. 2007. Efficiency of gene silencing in *Arabidopsis*: Direct inverted repeats vs. transitive RNAi vectors. Plant Biotechnol. J. 5:615–626. doi:10.1111/j.1467-7652.2007.00267.x

Fox, S.E., S.A. Filichkin, and T.C. Mockler. 2009. Applications of ultra high throughput sequencing. Methods Mol. Biol. 553:79–108. doi:10.1007/978-1-60327-563-7_5

Gan, Q., I. Chepelev, G. Wei, L. Tarayrah, K. Cui, K. Zhao, and X. Chen. 2010. Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res. 20:763–783. doi:10.1038/cr.2010.64

Garg, R., R.K. Patel, A.K. Tyagi, and M. Jain. 2011. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res. 18:53–63. doi:10.1093/dnares/dsq028

Gene Ontology Consortium. 2000. Gene ontology: Tool for the unification of biology. Nat. Genet. 1:25–29.

Gokirmak, T., S.A. Mehlenbacher, and N.V. Bassil. 2009. Characterization of European hazelnut (*Corylus avellana*) cultivars using SSR markers. Genet. Resour. Crop Evol. 56:147–172. doi:10.1007/s10722-008-9352-8

Hebenstreit, D., M. Fang, M. Gu, V. Charoensawan, A. van Oudenaarden, and S.A. Teichmann. 2011. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. Mol. Syst. Biol. 7:497. doi:10.1038/msb.2011.28

Johnson, K.B., J.N. Pinkerton, S.A. Mehlenbacher, J.K. Stone, and J.W. Pscheidt. 1996. Eastern filbert blight of European hazelnut: It's becoming a manageable disease. Plant Dis. 80:1308–1316. doi:10.1094/PD-80-1308

Li R., Y. Li, X. Fang, H. Yang, J. Wang, K. Kristiansen, and J. Wang. 2009. SNP detection for massively parallel whole- genome resequencing. Genome

Res. 19:1124–1132.

Li, R., Y. Li, K. Kristiansen, and J. Wang. 2008. SOAP: Short oligonucleotide alignment program. Bioinformatics 5:713– 714. doi:10.1093/bioinformatics/btn025

Mehlenbacher, S.A., R.N. Brown, J.W. Davis, H. Chen, N.V. Bassil, D.C. Smith, and T.L. Kubisiak. 2004. RAPD markers linked to eastern filbert blight resistance in *Corylus avellana*. Theor. Appl. Genet. 108:651–656. doi:10.1007/s00122-003-1476-9

Mehlenbacher, S.A., R.N. Brown, E.R. Nouhra, G. Tufan, N.V. Bassil, and T.L. Kubisiak. 2006. A genetic linkage map for hazelnut (*Corylus avellana* L.) based on RAPD and SSR markers. Genome 49:122–133.

Mehlenbacher, S.A., D.C. Smith, and R.L. McCluskey. 2011. "Jefferson" hazelnut. HortScience 46:662–664.

Mehlenbacher, S.A., M.M. Thompson, and R.H. Cameron. 1991. Occurrence and inheritance of resistance to eastern filbert blight in 'Gasaway' hazelnut. HortScience 26(4):410–411.

Min, X.J., G. Butler, R. Storms, and A. Sang. 2005. OrfPredictor: Predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 33:W677–W680. doi:10.1093/nar/gki394

Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5:621–628. doi:10.1038/nmeth.1226

Priest, H.D., S.E. Fox, S.A. Filichkin, and T.C. Mockler. 2010. Utility of next-generation sequencing for analysis of horticultural crop transcriptomes. Acta Hortic. 859:283–288.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org (accessed 16 Aug. 2011).

Sathuvalli, V.R. 2011. Eastern filbert blight in hazelnut (*Corylus avellana*): Identification of new resistance sources and high resolution genetic and physical mapping of a resistance gene. Ph.D. diss. Oregon State Univ., Corvallis, OR.

USDA National Agricultural Statistics Service (NASS) Oregon Field Office. 2011. 2010–2011 Oregon agriculture & fisheries statistics. USDA, Oregon Department of Agriculture, NASS, Portland, OR. http://www.oregon.gov/ODA/docs/pdf/ pubs/agripedia_stats.pdf?ga=t

(accessed 21 Feb. 2012).

Wang, X.-W., J.-B. Luan, J.-M. Li, Y.-Y. Bao, C.-X. Zhang, and S.-S. Liu. 2011. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. BMC Genomics 11:400.

Xia, Z., H. Xu, J. Zhai, D. Li, H. Luo, C. He, and X. Huang. 2011. RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. Plant Mol. Biol. 3:299–308. doi:10.1007/ s11103-011-9811.

Zerbino, D.R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Zhu Y.Y., E.M. Machleder, A. Chenchik, R. Li, and P.D. Siebert. 2001. Reverse transcriptase template switching: A SMART approach for full-length cDNA library construction. Biotechniques 4:892–897.

Chapter 3:

A Draft Genome and High-Density Genetic Map of European Hazelnut (*Corylus avellana* L.) 'Jefferson'

Erik R. Rowley, Robert VanBuren, Doug W. Bryant, Henry D. Priest[3], Shawn A. Mehlenbacher, and Todd C. Mockler

Abstract:

European hazelnut (Corylus avellana L.) is of global agricultural and economic significance, with genetic diversity existing in hundreds of accessions. Breeding efforts have focused on maximizing nut yield and quality and reducing susceptibility to diseases such as Eastern Filbert Blight (EFB). Here we present the first sequenced genome among the order Fagales, the EFB-resistant diploid hazelnut cultivar 'Jefferson' (OSU 703.007). We assembled the highly heterozygous hazelnut genome using an Illumina based approach and the final assembly has a scaffold N50 of 21.5kb. We captured approximately 91% (345 Mb) of the flow-cytometry-determined genome size and identified 34,910 putative gene loci. In addition, we identified over 2 million polymorphisms across seven diverse hazelnut cultivars and characterized their effect on coding sequences. We produced two high-density genetic maps with 3,209 markers from an F1 hazelnut population, representing a five-fold increase in marker density over the previous maps. These genomic resources will aid in the discovery of molecular markers linked to genes of interest for hazelnut breeding efforts.

Introduction:

European hazelnut (Corylus avellana, L.) is a member of the family Betulacea, a group of six plant genera that include the birches (Betula L. spp.) and alders (Alnus Mill spp.). Members of this family are either a reduced stature tree or are shrub-like with fruit produced in the form of nuts. It is these nuts which make C. avellana an agriculturally significant crop. Hazelnuts provide the predominant flavor in a variety of butters, candies, chocolate spreads, and confectionary pastes. Whole blanched and roasted kernels are in demand by consumers worldwide, and the shells are used for both landscaping and groundcover. Hazelnuts are also high in fiber, contain several essential vitamins, and have potential for use as biofuels (Moser 2012).

European hazelnut grows best in areas with mild coastal tempered winters, and the vast majority of the world's hazelnut production is centered in the Black Sea region of Turkey. This region accounts for over 70% of the hazelnuts on the global market (Gökirmak, Mehlenbacher, and Bassil 2008), producing over 1 million tons each year. The moderate climate of Oregon's (USA) Willamette Valley is also similarly suited for hazelnut production. Although Oregon provides for only 3% of the world share of hazelnuts, it produces 99% of the hazelnuts grown in North American, worth an estimated US $67.5 million dollars annually (USDA NASS Oregon Field Office, 2011).

Hazelnut is clonally propagated using traditional simple layerage, tie-off layerage (stooling), or grafting, all of which are labor-intensive and produce limited numbers of plants. Micropropagation, or in vitro propagation on a defined culture medium (Yu and Reed, 1995), allows rapid propagation of selected types, and has been particularly useful in the rapid increase of new cultivars and pollinizers. Meristems from these plantlets may be subcultured

for up to 6 years without change to their genetic structure (Nas and Read 2004), and storage at 4 °C allows maintenance of many accessions

Breeding efforts have focused on producing progeny with enhanced agronomic traits of interest such as increased oil content and size, blanching ability, bud mite resistance, and disease resistance. Specifically, development of germplasms resistant to the fungus Anisogramma anomola (Peck) E. Muller, the causal agent of Eastern filbert blight (EFB), is important. EFB is a deadly disease that initially causes cankers to form on the woody parts of the tree. These cankers act to slowly girdle the branches, leading to canopy leaf loss and death of the tree within a few years (Johnson, K.B., Pinkerton, J.N., Mehlenbacher, S.A., Stone, J.K, and Pscheidt 1996). Most of the European hazelnut cultivars grown in North America are susceptible to EFB including ‘Barcelona’, the predominant heirloom cultivar grown in Oregon and the Pacific Northwest.

The United States Department of Agriculture (USDA) and Oregon State University (OSU) have collected over 700 accessions of C. avellana L.; these cultivars are preserved in Corvallis, Oregon where they are used in breeding efforts (Gürcan, K., Mehlenbacher S. A., and Erdoğan 2010). These accessions are maintained both as plantlets in tissue culture and as adult trees grown near the Oregon State University campus.

Among the European hazelnut accessions preserved at the Smith Horticulture Research Farm is an F1 mapping population of 138 individuals resulting from the cross of the EFB-resistant accession OSU 414.062 (paternal) and EFB-susceptible accession OSU 252.146 (maternal). Among these progeny is the cultivar 'Jefferson' (OSU 703.007), which was released in 2009 by Dr. Shawn Mehlenbacher at Oregon State University (Mehlenbacher at al. 2011). ‘Jefferson’ was selected as the reference cultivar for genome assembly and

characterization due to the presence of the dominant ‘Gasaway’ allele, which confers resistance to EFB. Identifying sources of resistance and resistant cultivars will not only increase the parental germplasm available for future breeding efforts but will also the reduce fungicide requirements and EFB-associated costs to regional growers (Julian, J., C.F. Seavert 2009).

C. avellana possesses several traits that make it an attractive candidate for use as a model system for the family Betulaceae. Among these are a short life cycle; *C. avellana* seedlings produce fruit 5 years post-planting. It occupies a small habit for a tree at ~ 5 m and is amenable to Agrobacterium-mediated transformation. European hazelnut is a diploid with 11 pairs of chromosomes (2n =2x=22) and a genome of ~378 Mbp (empirically determined by flow cytometry at the USDA-ARS-NCGR in Oregon in 2012). This genome is approximately triple the size of the established model dicot Arabidopsis thaliana (L.) Heynh, at 125 Mbp but is significantly smaller than those of other sequenced tree genomes such as *Populus trichocarpa* at ~520 Mbp (Tuskan and Torr 2014) and loblolly pine at ~23 Gbp (Neale *et al.* 2014).

We sequenced the ‘Jefferson’ genome at ~93x coverage and completed a de novo genome assembly, capturing ~91% of the genome (345 Mbp) with a contig N50 of 21,540 bp. Homology-based functional annotation, restricted to whole gene models and aided by the ‘Jefferson’ transcriptome assembly (**Chapter 2**) predicted 34,910 protein coding loci. . Of these predicted loci, 22,474 have homology to an entry in the NCBI non-redundant protein database, and 82.5% of the annotated genes are presented in the best annotated and closest related genera *Vitis*, *Prunus*, *Populus*, and *Ricinus*.

Additionally we resequenced seven European hazelnut cultivars at ~ 20x coverage and discovered millions of polymorphisms between one of more of these genomes and that of ‘Jefferson’. We predicted the effects of each

polymorphism on the coding potential of affected loci and identified candidate genes for future research. Genotyping by sequence (GBS) analysis of the F1 mapping population enabled identification of 3,209 additional GBS-derived markers between the maternal (OSU 252.146) and paternal (OSU 414.062) maps. This improved high-density genetic map will be useful for marker-assisted breeding and for the identification of new, desirable traits in hazelnut.

Hazelnut genome sequencing has provided new resources to the scientific community and promises to accelerate trait discovery and enhance future breeding efforts. This resource will serve as a tool for gene discovery and functional studies, for the development of DNA markers and other genomic tools, and will allow future integration of the genome sequence with genetic and physical maps and the incorporation of new sequencing technologies..

Materials and methods:

Collection of tissues and sequencing on the Illumina HiSeq 2000. Tissues were collected from field-grown trees located in Corvallis, Oregon. Genomic DNA (gDNA) was extracted using Qiagen Plant DNeasy kits per the manufacturer's directions. The quality of the gDNA was assessed by visualization on agarose gels and using the Qubit Fluorometer (Life Technologies) prior to library construction. The construction of 250-bp and 350-bp Illumina paired-end (PE) libraries and a 4.5-Kb mate-pair (MP) library was performed at the Georgia Genomics Facility at the University of Georgia. Cluster generation on the Illumina HiSeq 2000 was performed in the Oregon State University Center for Genome Research and Biocomputing (CGRB) core facility using a standard Illumina protocol.

Genome assembly and filtering. Genomic data from two PE libraries (250-bp insert, 350-bp insert) and one MP library (4.5-Kb insert) were trimmed based on quality score (Q <30) using the FASTX toolkit version 0.0.13

(http://hannonlab.cshl.edu/fastx_toolkit/) and adaptor sequences were removed. Velvet version 1.2.01 (12) was used to generate assemblies of these quality-filtered genomic data with a k-mer hash length of 51 bp. SSPACE v2.0 (13) was then applied, using all quality-controlled data, to improve the Velvet assembly by merging and extending scaffolds where possible. Contigs shorter than 1 Kb were discarded, as were those having >25% homology and >10 unique best BLASTN hits to annotated non-plant and organelle sequences in the NCBI nucleotide database to remove any bacterial or fungal contamination.

Gene prediction and functional annotation. Low complexity and repetitive regions in filtered contigs were masked using the program RepeatMasker (Smit, AFA, Hubley, R., and Green, P. RepeatMasker Open-4.0. 2013-1015 http://www.repeatmasker.org), and putative loci were identified using the gene prediction software AUGUSTUS (Stanke *et al.* 2004). The 28,255 ESTs from the European hazelnut transcriptome assembly (**Chapter 2**) were used to guide the gene predictions, which were restricted to whole gene (open-reading frame) models only. After annotation, further filtering was implemented to remove all loci annotated as transposable elements, retroelements, and gag-polymerases. The loci removed from the annotated protein file are available for visualization on the hazelnut JBrowse portal and for download on the FTP server. In total, 22,474 protein-coding regions (64.3% of the putative loci) and 2,398 (1,178 unique) transposable elements were identified via homology to the NCBI non-redundant protein database using the BLASTP tool (Altschul *et al.* 1997) with an E-value cut-off of $1x10^{-5}$.

Polymorphism discovery. We sequenced seven unique hazelnut cultivars (‘Barcelona’, 'Tonda Gentile delle Langhe', ‘Tonda di Giffoni’,‘Ratoli’, ‘Daviana’, ‘Halls Giant’, ‘Tombul (Extra Ghiaghli)) on four lanes of an Illumina HiSeq 2000 flowcell and and BWA was used to align the resulting

reads from each of the seven cultivars to the 'Jefferson' reference genome (Li and Durbin 2009). We then applied the GATK (17) base quality score recalibration, insertion and deletion (INDEL) realignment and duplicate removal. Single-nucleotide polymorphisms (SNP) and INDEL discovery was performed across all samples simultaneously using standard hard filtering parameters according to GATK Best Practices recommendations (DePristo *et al.* 2011; Van der Auwera *et al.* 2013). SNP and INDEL predictions were further filtered using the SnpSift component of the SnpEff package (Cingolani *et al.* 2012); a minimum of 15 overlapping reads and a quality score (Q) of 30 for the SNPs and 20 for the INDELS were required. SnpEff was used to predict the functional significance of each polymorphism and effects on the coding potential of the putative loci..

Visualization of data. For visualization of the gene features and polymorphisms we uploaded the alignments, predicted AUGUSTUS gene models, and their BLASTP annotations to JBrowse (http://jbrowse.org/), a Java-based genome browser. These datasets are available for visualization on the hazelnut genome website hosted at hazelnut.mocklerlab.org/JBrowse.

Construction of the GBS-based genetic map. The high-density genetic map was constructed from a full-sib population of 144 F1 plants (138 seedlings and two parents in triplicate) from the cross of OSU 252.146 and OSU 414.062 using a two-way pseudo testcross approach (Mehlenbacher *et al.* 2006). High-quality genomic DNA from the progeny and parental plants was extracted as previously described and used for construction of GBS libraries (Elshire *et al.* 2011). GBS libraries were prepared using the restriction enzyme ApeK1 and pooled in sets of 72 uniquely barcoded individuals. Each 72 individual barcoded pool was sequenced (4 lanes total) on an Illumina HiSeq 2500 1x100 SE run. Polymorphisms were identified by implementing the UNEAK package of the TASSLE-GBS pipeline (Glaubitz *et al.* 2014)

using default settings. A minimum coverage of 5 overlapping reads was required to call each polymorphism in order to minimize false positives.

A total of eight F1 individuals were removed prior to analysis because of low coverage. Raw genotype output from TASSEL was first filtered to remove SNPs with more than 20% missing data in the population. After filtering, SNPs that were homozygous in the maternal parent (OSU 252.146) but heterozygous in the paternal parent (OSU 414.062) and heterozygous in the maternal parent (OSU 252.146) but homozygous in the paternal parent (OSU 414.062) were used for map construction., as these configurations are expected to segregate at a 1:1 ratio, and any SNPs that failed to meet this segregation pattern were discarded. The remaining 2,198 SNP markers were converted into a cross pollinator (CP) population and then mapped using JOINMAP 4.1 (Van Ooijen 2006). Markers were assigned to linkage groups (LGs) with independence LOD scores of 8.0. After classifying markers into linkage groups, the regression mapping algorithm and a maximum recombination frequency of 0.40 were employed. Genetic distances between loci were calculated with Kosambi's function. Orthologous regions between hazelnut and in peach were identified using BLAST with a minimum e-value of $1x10^{-5}$ and minimum length of 40 bp. Markers mapping with multiple high confidence hits were removed as these likely represented repetitive sequences in peach or duplicated chromosomal regions.

***Data availability*.** All data used in this study are available in the NCBI Sequence Read Archive (SRA). The current assembly for the 'Jefferson' genome is available for BLAST queries and downloads at hazelnut.mocklerlab.org; putative amino acid sequences, variant calls, and effect predictions for the seven resequenced cultivars, and the current genetic linkage map are also available at hazelnut.mocklerlab.org. Polymorphism and gene feature tracks of the seven resequenced hazelnut cultivars and 'Jefferson', along with functional annotation and variant effect predictions,

can be visualized at the European hazelnut web portal: hazelnut.mocklerlab.org/JBrowse.

Results:

***Assembly and characterization of the European hazelnut genome*.** The objective of this study was to provide hazelnut genomic resources for breeders for use in gene and marker discovery and polymorphism detection, with the ultimate goal of hazelnut improvement. We assembled the European hazelnut cultivar 'Jefferson' using two PE and one MP Illumina library (collectively representing 93x genome coverage) using the assembly programs Velvet and SSPACE. We discarded all contigs shorter than 1 Kb, and implemented a nucleotide-filtering cutoff to remove non-plant and organelle sequences from the assembly. After filtering, the draft assembly for the hazelnut genome included 36,641 contigs and scaffolds with a total sequence length of 345 Mb. This is 91% of the size of the genome determined by flow cytometry, for approximately 93x coverage. Half of the assembly is contained in scaffolds and contigs greater than 21.5 Kb, with the largest scaffold comprising 274.5 Kb (**Table 3.1**).

Table 3.1. Summary of hazelnut genome assembly and annotation

Total number of contigs	36,641
Total size of assembly	345.5 Mb
Length of largest contig	274.5 Kb
Average contig size	9.4 Kb
N50	21.5 Kb
Number of *ab initio* predicted protein-coding loci	34,910
Number of functionally annotated gene products	22,474

The fragmented nature of the assembly is likely to high within genome heterozygosity, evidenced by the bimodal K-mer distribution (**Figure 3.1**).

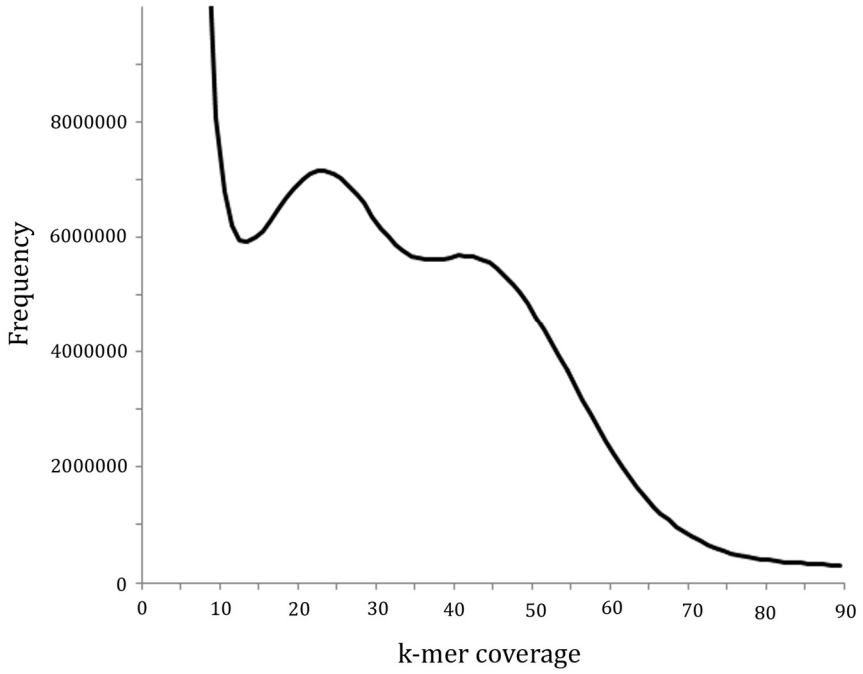
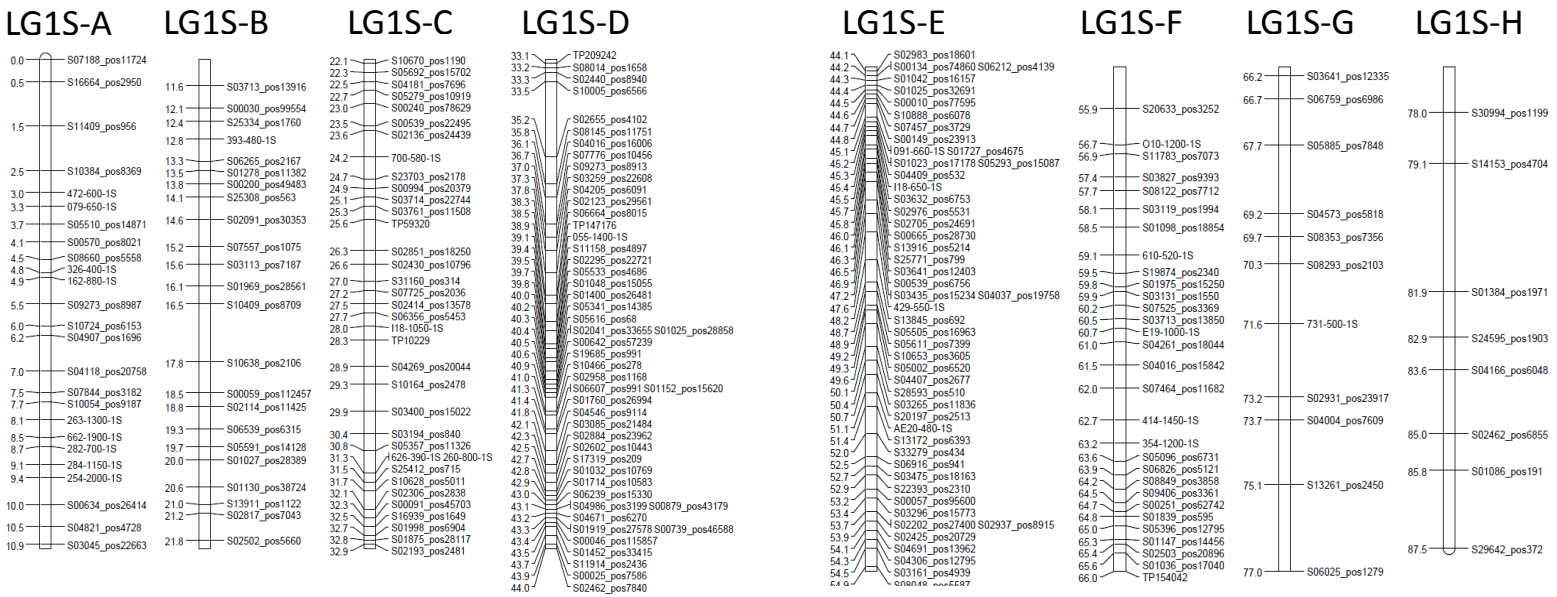


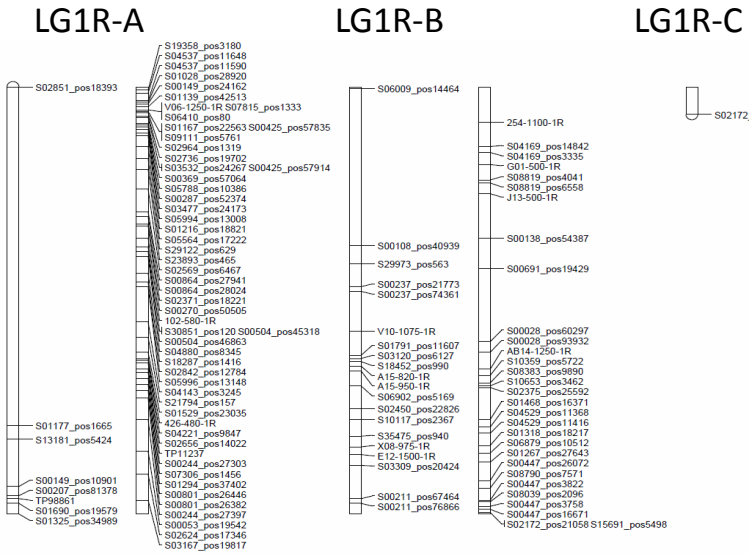
Figure 3.1. K-mer coverage of the 'Jefferson' 350bp PE Illumina library (k=21). The peak with k-mer coverage of ~22 represents heterozygous sites and the peak at coverage ~44 represents homozygous sites. The large peak at 22X coverage suggests high within-genome heterozygosity.

Developing a high-density genetic map for hazelnut. Marker development in hazelnut has been slow, but roughly 500 random amplified polymorphic DNA (RAPD) markers and ~700 simple sequence repeat (SSR) markers are currently available (Mehlenbacher *et al.* 2006). We have now constructed a high-density GBS-based map using a full-sib population of 138 F1 plants from the cross of OSU 252.146 × OSU 414.146 (**Figure 3.2**).

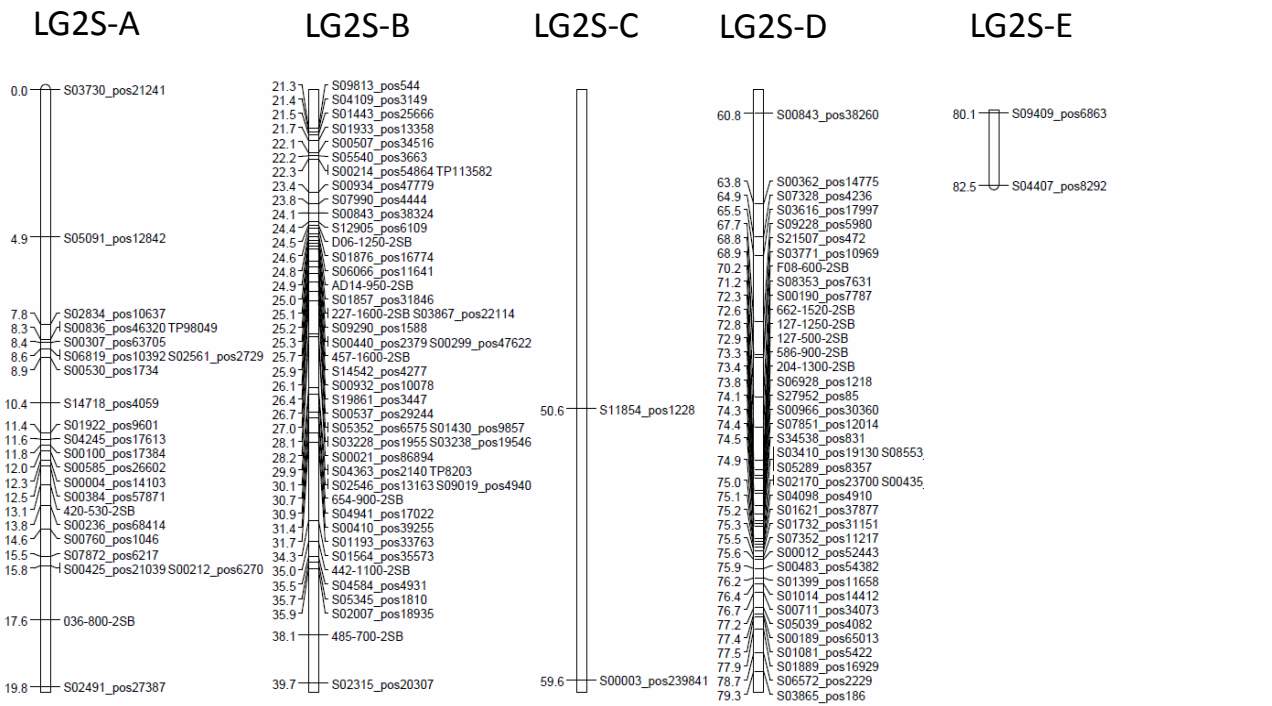
(A)


LG1S-A
0.0 S07188_pos11724
0.5 S16664_pos2950
1.5 S11409_pos956
2.5 S10384_pos8369
3.0 472-600-1S
3.3 079-650-1S
3.7 S05510_pos14871
4.1 S00570_pos8021
4.5 S08660_pos5558
4.8 326-400-1S
4.9 162-880-1S
5.5 S09273_pos8987
6.0 S10724_pos6153
6.2 S04907_pos1696
7.0 S04118_pos20758
7.5 S07844_pos3182
7.7 S10054_pos9187
8.1 263-1300-1S
8.5 662-1900-1S
8.7 282-700-1S
9.1 284-1150-1S
9.4 254-2000-1S
10.0 S00634_pos26414
10.5 S04821_pos4728
10.9 S03045_pos22663
LG1S-B
11.6 S03713_pos13916
12.1 S00030_pos99554
12.4 S25334_pos1760
12.8 393-480-1S
13.3 S06265_pos2167
13.5 S01278_pos11382
13.8 S00200_pos49483
14.1 S25308_pos563
14.6 S02091_pos30353
15.2 S07557_pos1075
15.6 S03113_pos7187
16.1 S01969_pos28561
16.5 S10409_pos8709
17.8 S10638_pos2106
18.5 S00059_pos112457
18.8 S02114_pos11425
19.3 S06539_pos6315
19.7 S05691_pos14128
20.0 S01027_pos28389
20.6 S01130_pos38724
21.0 S13911_pos1122
21.2 S02817_pos7043
21.8 S02502_pos5660
LG1S-C
22.1 S10670_pos1190
22.3 S05692_pos15702
22.5 S04181_pos7696
22.7 S05279_pos10919
23.0 S00240_pos78629
23.5 S00539_pos22495
23.6 S02136_pos24439
24.2 700-580-1S
24.7 S23703_pos2178
24.9 S00994_pos20379
25.1 S03714_pos22744
25.3 S03761_pos11508
25.6 TP59320
26.3 S02851_pos18250
26.6 S02430_pos10796
27.0 S31160_pos314
27.2 S07725_pos2036
27.5 S02414_pos13578
27.7 S06356_pos5453
28.0 I18-1050-1S
28.3 TP10229
28.9 S04269_pos20044
29.3 S10164_pos2478
29.9 S03400_pos15022
30.4 S03194_pos840
30.8 S05357_pos11326
31.3 I626-390-1S 260-800-1S
31.5 S25412_pos715
31.7 S10628_pos5011
32.1 S02306_pos2838
32.3 S00091_pos45703
32.5 S16939_pos1649
32.7 S01998_pos6904
32.8 S01875_pos28117
32.9 S02193_pos2481
LG1S-D
33.1 TP209242
33.2 S08014_pos1658
33.3 S02440_pos8940
33.5 S10005_pos6566
35.2 S02655_pos4102
35.8 S08145_pos11751
36.1 S04016_pos16006
36.7 S07776_pos10456
37.0 S09273_pos8913
37.3 S03259_pos22608
37.8 S04205_pos6091
38.3 S02123_pos29561
38.5 S06664_pos8015
38.9 TP147176
39.1 055-1400-1S
39.4 S11158_pos4897
39.5 S02295_pos22721
39.7 S05533_pos4686
39.8 S01048_pos15055
40.0 S01400_pos26481
40.2 S05341_pos14385
40.3 S05616_pos68
40.4 S02041_pos33655 S01025_pos28858
40.5 S00642_pos57239
40.6 S19685_pos991
40.9 S10466_pos278
41.0 S02958_pos1168
41.3 S06607_pos991 S01152_pos15620
41.4 S01760_pos26994
41.8 S04546_pos9114
42.1 S03085_pos21484
42.3 S02884_pos23962
42.5 S02602_pos10443
42.7 S17319_pos209
42.8 S01032_pos10769
42.9 S01714_pos10583
43.0 S06239_pos15330
43.1 S04986_pos3199 S00879_pos43179
43.2 S04671_pos6270
43.3 S01919_pos27578 S00739_pos46588
43.4 S00046_pos115857
43.5 S01452_pos33415
43.7 S11914_pos2436
43.9 S00025_pos7586
44.0 S02462_pos7840
LG1S-E
44.1 S02983_pos18601
44.2 S00134_pos74860 S06212_pos4139
44.3 S01042_pos16157
44.4 S01025_pos32691
44.5 S00010_pos77595
44.6 S10888_pos6078
44.7 S07457_pos3729
44.8 S00149_pos23913
45.1 091-660-1S S01727_pos4675
45.2 S01023_pos17178 S05293_pos15087
45.3 S04409_pos532
45.4 I18-650-1S
45.5 S03632_pos6753
45.7 S02976_pos5531
45.8 S02705_pos24691
46.0 S00665_pos28730
46.1 S13916_pos5214
46.3 S25771_pos799
46.5 S03641_pos12403
46.9 S00539_pos6756
47.2 S03435_pos15234 S04037_pos19758
47.6 429-550-1S
48.2 S13845_pos692
48.7 S05505_pos16963
48.9 S05611_pos7399
49.2 S10653_pos3605
49.3 S05002_pos6520
49.6 S04407_pos2677
50.1 S28593_pos510
50.4 S03265_pos11836
50.7 S20197_pos2513
51.1 AE20-480-1S
51.4 S13172_pos6393
52.0 S33279_pos434
52.5 S06916_pos941
52.7 S03475_pos18163
52.9 S22393_pos2310
53.2 S00057_pos95600
53.4 S03296_pos15773
53.7 S02202_pos27400 S02937_pos8915
53.9 S02425_pos20729
54.1 S04691_pos13962
54.3 S04306_pos12795
54.5 S03161_pos4939
LG1S-F
55.9 S20633_pos3252
56.7 O10-1200-1S
56.9 S11783_pos7073
57.4 S03827_pos9393
57.7 S08122_pos7712
58.1 S03119_pos1994
58.5 S01098_pos18854
59.1 610-520-1S
59.5 S19874_pos2340
59.8 S01975_pos15250
59.9 S03131_pos1550
60.2 S07525_pos3369
60.5 S03713_pos13850
60.7 E19-1000-1S
61.0 S04261_pos18044
61.5 S04016_pos15842
62.0 S07464_pos11682
62.7 414-1450-1S
63.2 354-1200-1S
63.6 S05096_pos6731
63.9 S06826_pos5121
64.2 S08849_pos3858
64.5 S09406_pos3361
64.7 S00251_pos62742
64.8 S01839_pos595
65.0 S05396_pos12795
65.3 S01147_pos14456
65.4 S02503_pos20896
65.6 S01036_pos17040
66.0 TP154042
LG1S-G
66.2 S03641_pos12335
66.7 S06759_pos6986
67.7 S05885_pos7848
69.2 S04573_pos5818
69.7 S08353_pos7356
70.3 S08293_pos2103
71.6 731-500-1S
73.2 S02931_pos23917
73.7 S04004_pos7609
75.1 S13261_pos2450
77.0 S06025_pos1279
LG1S-H
78.0 S30994_pos1199
79.1 S14153_pos4704
81.9 S01384_pos1971
82.9 S24595_pos1903
83.6 S04166_pos6048
85.0 S02462_pos6855
85.8 S01086_pos191
87.5 S29642_pos372

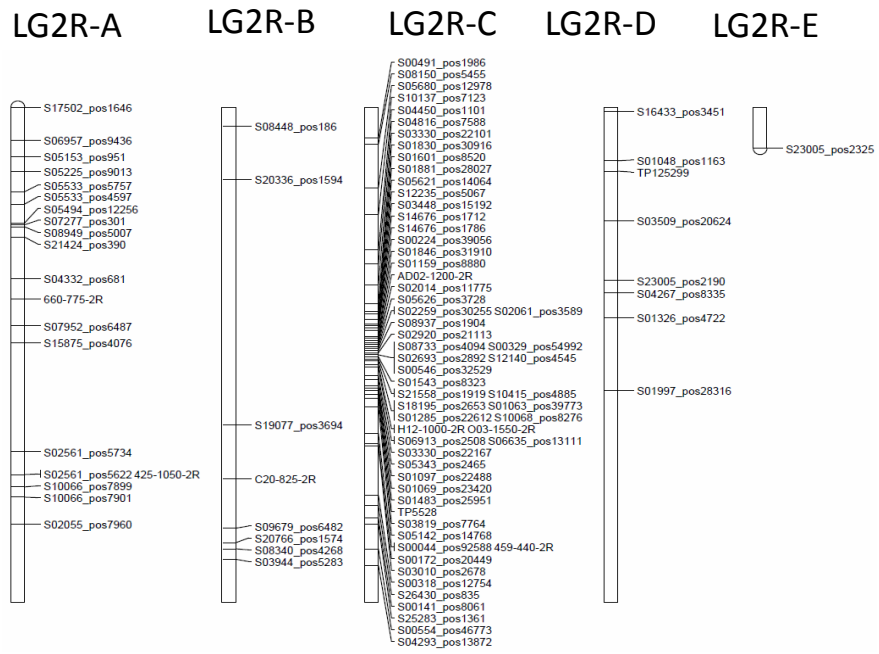
(B)


LG1R-A
LG1R-B
LG1R-C
S02851_pos18393
S01177_pos1665
S13181_pos5424
S00149_pos10901
S00207_pos81378
TP98861
S01690_pos19579
S01325_pos34989
S19358_pos3180
S04537_pos11648
S04537_pos11590
S01028_pos28920
S00149_pos24162
S01139_pos42513
V06-1250-1R S07815_pos1333
S06410_pos80
S01167_pos22563 S00425_pos57835
S09111_pos5761
S02964_pos1319
S02736_pos19702
S03532_pos24267 S00425_pos57914
S00369_pos57064
S05788_pos10386
S00287_pos52374
S03477_pos24173
S05994_pos13008
S01216_pos18821
S05564_pos17222
S29122_pos629
S23893_pos465
S02569_pos6467
S00864_pos27941
S00864_pos28024
S02371_pos18221
S00270_pos50505
102-580-1R
S30851_pos120 S00504_pos45318
S00504_pos46863
S04880_pos8345
S18287_pos1416
S02842_pos12784
S05996_pos13148
S04143_pos3245
S21794_pos157
S01529_pos23035
426-480-1R
S04221_pos9847
S02656_pos14022
TP11237
S00244_pos27303
S07306_pos1456
S01294_pos37402
S00801_pos26446
S00801_pos26382
S00244_pos27397
S00053_pos19542
S02624_pos17346
S03167_pos19817
S06009_pos14464
S00108_pos40939
S29973_pos563
S00237_pos21773
S00237_pos74361
V10-1075-1R
S01791_pos11607
S03120_pos6127
S18452_pos990
A15-820-1R
A15-950-1R
S06902_pos5169
S02460_pos22826
S10117_pos2367
S35475_pos940
X08-975-1R
E12-1500-1R
S03309_pos20424
S00211_pos67464
S00211_pos76866
254-1100-1R
S04169_pos14842
S04169_pos3335
G01-500-1R
S08819_pos4041
S08819_pos6558
J13-500-1R
S00138_pos54387
S00691_pos19429
S00028_pos60297
S00028_pos93932
AB14-1250-1R
S10359_pos5722
S08383_pos9890
S10653_pos3462
S02375_pos25592
S01468_pos16371
S04529_pos11368
S04529_pos11416
S01318_pos18217
S06879_pos10512
S01267_pos27643
S00447_pos26072
S08790_pos7571
S00447_pos3822
S08039_pos2096
S00447_pos3758
S00447_pos16671
S02172_pos21058 S15691_pos5498
S02172

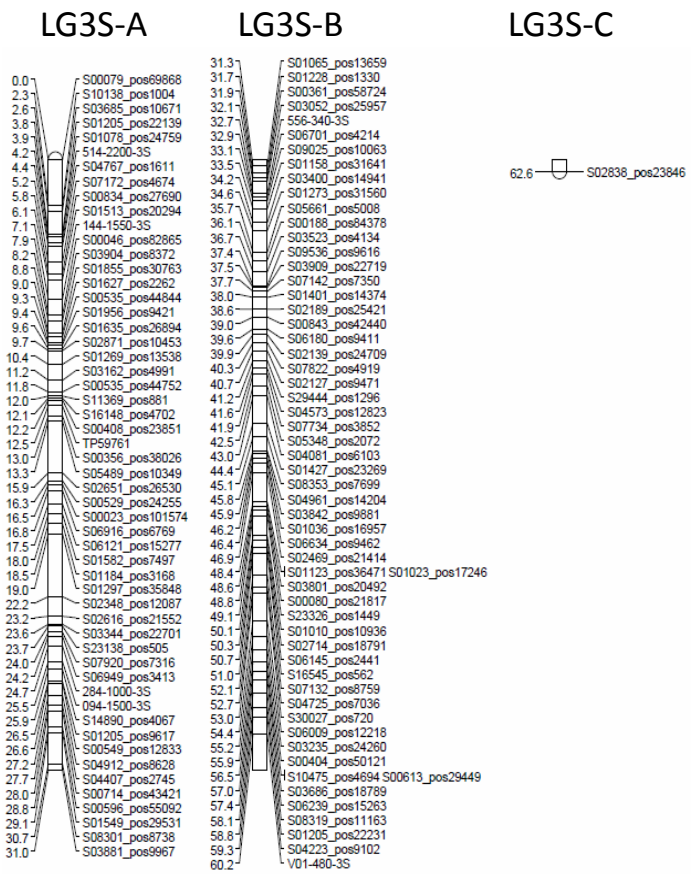
(C)


LG2S-A
LG2S-B
LG2S-C
LG2S-D
LG2S-E
0.0 S03730_pos21241
4.9 S05091_pos12842
7.8 S02834_pos10637
8.3 S00836_pos46320 TP98049
8.4 S00307_pos63705
8.6 S06819_pos10392 S02561_pos2729
8.9 S00530_pos1734
10.4 S14718_pos4059
11.4 S01922_pos9601
11.6 S04245_pos17613
11.8 S00100_pos17384
12.0 S00585_pos26602
12.3 S00004_pos14103
12.5 S00384_pos57871
13.1 420-530-2SB
13.8 S00236_pos68414
14.6 S00760_pos1046
15.5 S07872_pos6217
15.8 S00425_pos21039 S00212_pos6270
17.6 036-800-2SB
19.8 S02491_pos27387
21.3 S09813_pos544
21.4 S04109_pos3149
21.5 S01443_pos25666
21.7 S01933_pos13358
22.1 S00507_pos34516
22.2 S05540_pos3663
22.3 S00214_pos54864 TP113582
23.4 S00934_pos47779
23.8 S07990_pos4444
24.1 S00843_pos38324
24.4 S12905_pos6109
24.5 D06-1250-2SB
24.6 S01876_pos16774
24.8 S06066_pos11641
24.9 AD14-950-2SB
25.0 S01857_pos31846
25.1 227-1600-2SB S03867_pos22114
25.2 S09290_pos1588
25.3 S00440_pos2379 S00299_pos47622
25.7 457-1600-2SB
25.9 S14542_pos4277
26.1 S00932_pos10078
26.4 S19861_pos3447
26.7 S00537_pos29244
27.0 S05352_pos6575 S01430_pos9857
28.1 S03228_pos1955 S03238_pos19546
28.2 S00021_pos86894
29.9 S04363_pos2140 TP8203
30.1 S02546_pos13163 S09019_pos4940
30.7 654-900-2SB
30.9 S04941_pos17022
31.4 S00410_pos39255
31.7 S01193_pos33763
34.3 S01564_pos35573
35.0 442-1100-2SB
35.5 S04584_pos4931
35.7 S05345_pos1810
35.9 S02007_pos18935
38.1 485-700-2SB
39.7 S02315_pos20307
50.6 S11854_pos1228
59.6 S00003_pos239841
60.8 S00843_pos38260
63.8 S00362_pos14775
64.9 S07328_pos4236
65.5 S03616_pos17997
67.7 S09228_pos5980
68.8 S21507_pos472
68.9 S03771_pos10969
70.2 F08-600-2SB
71.2 S08353_pos7631
72.3 S00190_pos7787
72.6 662-1520-2SB
72.8 127-1250-2SB
72.9 127-500-2SB
73.3 586-900-2SB
73.4 204-1300-2SB
73.8 S06928_pos1218
74.1 S27952_pos85
74.3 S00966_pos30360
74.4 S07851_pos12014
74.5 S34538_pos831
74.9 S03410_pos19130 S08553
75.0 S05289_pos8357
75.1 S02170_pos23700 S00435
75.2 S04098_pos4910
75.3 S01621_pos37877
75.5 S01732_pos31151
75.6 S07352_pos11217
75.9 S00012_pos52443
76.2 S00483_pos54382
76.4 S01399_pos11658
76.7 S01014_pos14412
77.2 S00711_pos34073
77.4 S05039_pos4082
77.5 S00189_pos65013
77.9 S01081_pos5422
78.7 S01889_pos16929
79.3 S06572_pos2229
S03865_pos186
80.1 S09409_pos6863
82.5 S04407_pos8292

(D)


LG2R-A
LG2R-B
LG2R-C
LG2R-D
LG2R-E
S17502_pos1646
S06957_pos9436
S05153_pos951
S05225_pos9013
S05533_pos5757
S05533_pos4597
S05494_pos12256
S07277_pos301
S08949_pos5007
S21424_pos390
S04332_pos681
660-775-2R
S07952_pos6487
S15875_pos4076
S02561_pos5734
S02561_pos5622 425-1050-2R
S10066_pos7899
S10066_pos7901
S02055_pos7960
S08448_pos186
S20336_pos1594
S19077_pos3694
C20-825-2R
S09679_pos6482
S20766_pos1574
S08340_pos4268
S03944_pos5283
S00491_pos1986
S08150_pos5455
S05680_pos12978
S10137_pos7123
S04450_pos1101
S04816_pos7588
S03330_pos22101
S01830_pos30916
S01601_pos8520
S01881_pos28027
S05621_pos14064
S12235_pos5067
S03448_pos15192
S14676_pos1712
S14676_pos1786
S00224_pos39056
S01846_pos31910
S01159_pos8880
AD02-1200-2R
S02014_pos11775
S05626_pos3728
S02259_pos30255 S02061_pos3589
S08937_pos1904
S02920_pos21113
S08733_pos4094 S00329_pos54992
S02693_pos2892 S12140_pos4545
S00546_pos32529
S01543_pos8323
S21558_pos1919 S10415_pos4885
S18195_pos2653 S01063_pos39773
S01285_pos22612 S10068_pos8276
H12-1000-2R O03-1550-2R
S06913_pos2508 S06635_pos13111
S03330_pos22167
S05343_pos2465
S01097_pos22488
S01069_pos23420
S01483_pos25951
TP5528
S03819_pos7764
S05142_pos14768
S00044_pos92588 459-440-2R
S00172_pos20449
S03010_pos2678
S00318_pos12754
S26430_pos835
S00141_pos8061
S25283_pos1361
S00554_pos46773
S04293_pos13872
S16433_pos3451
S01048_pos1163
TP125299
S03509_pos20624
S23005_pos2190
S04267_pos8335
S01326_pos4722
S01997_pos28316
S23005_pos2325

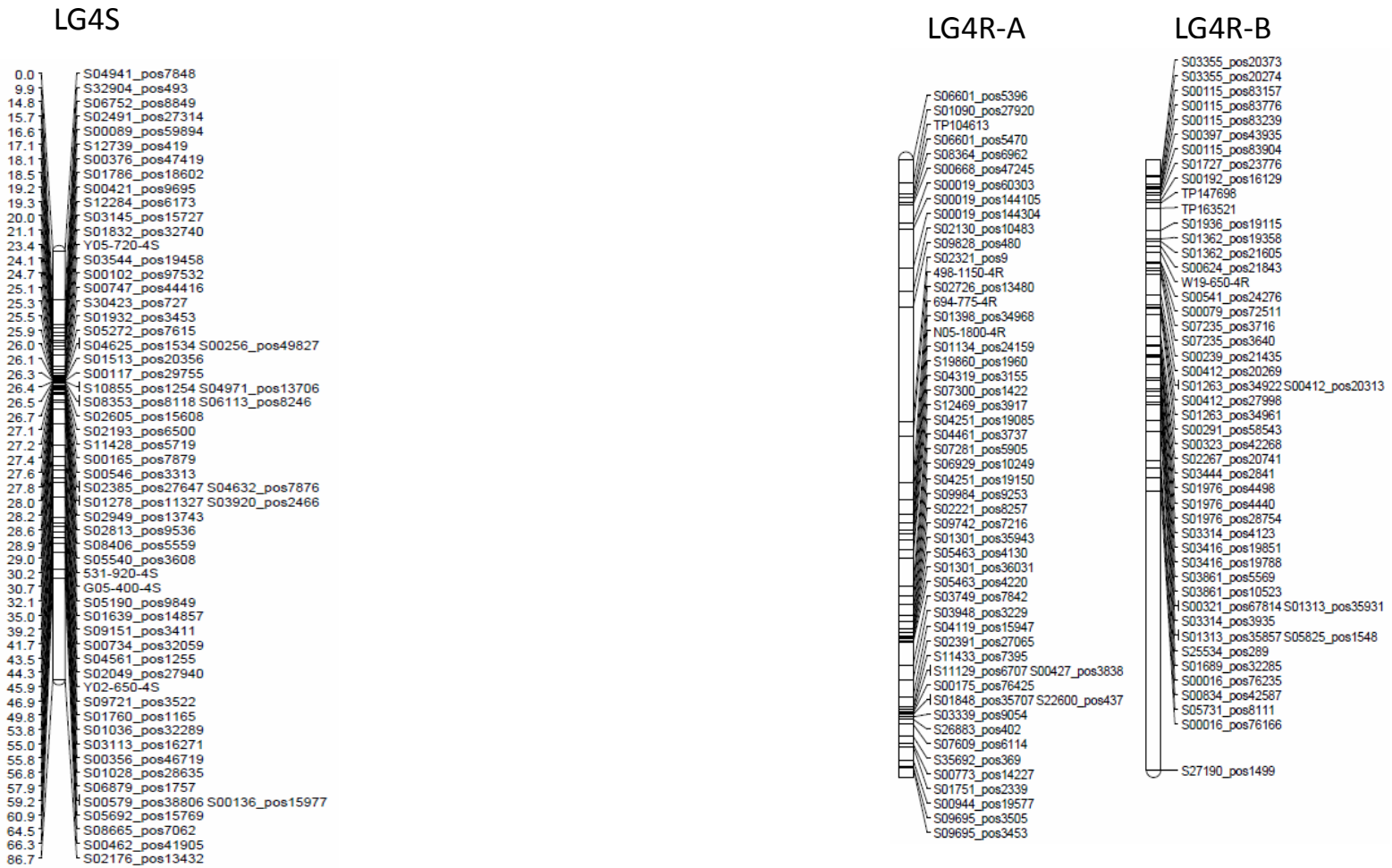
(E)


LG3S-A
LG3S-B
LG3S-C
0.0 S00079_pos69868
2.3 S10138_pos1004
2.6 S03685_pos10671
3.8 S01205_pos22139
3.9 S01078_pos24759
4.2 514-2200-3S
4.4 S04767_pos1611
5.2 S07172_pos4674
5.8 S00834_pos27690
6.1 S01513_pos20294
7.1 144-1550-3S
7.9 S00046_pos82865
8.2 S03904_pos8372
8.8 S01855_pos30763
9.0 S01627_pos2262
9.3 S00535_pos44844
9.4 S01956_pos9421
9.6 S01635_pos26894
9.7 S02871_pos10453
10.4 S01269_pos13538
11.2 S03162_pos4991
11.8 S00535_pos44752
12.0 S11369_pos881
12.1 S16148_pos4702
12.2 S00408_pos23851
12.5 TP59761
13.0 S00356_pos38026
13.3 S05489_pos10349
15.9 S02651_pos26530
16.3 S00529_pos24255
16.5 S00023_pos101574
16.8 S06916_pos6769
17.5 S06121_pos15277
18.0 S01582_pos7497
18.5 S01184_pos3168
19.0 S01297_pos35848
22.2 S02348_pos12087
23.2 S02616_pos21552
23.6 S03344_pos22701
23.7 S23138_pos505
24.0 S07920_pos7316
24.2 S06949_pos3413
24.7 284-1000-3S
25.5 094-1500-3S
25.9 S14890_pos4067
26.5 S01205_pos9617
26.6 S00549_pos12833
27.2 S04912_pos8628
27.7 S04407_pos2745
28.0 S00714_pos43421
28.8 S00596_pos55092
29.1 S01549_pos29531
30.7 S08301_pos8738
31.0 S03881_pos9967
31.3 S01065_pos13659
31.7 S01228_pos1330
31.9 S00361_pos58724
32.1 S03052_pos25957
32.7 556-340-3S
32.9 S06701_pos4214
33.1 S09025_pos10063
33.5 S01158_pos31641
34.2 S03400_pos14941
34.6 S01273_pos31560
35.7 S05661_pos5008
36.1 S00188_pos84378
36.7 S03523_pos4134
37.4 S09536_pos9616
37.5 S03909_pos22719
37.7 S07142_pos7350
38.0 S01401_pos14374
38.6 S02189_pos25421
39.0 S00843_pos42440
39.6 S06180_pos9411
39.9 S02139_pos24709
40.3 S07822_pos4919
40.7 S02127_pos9471
41.2 S29444_pos1296
41.6 S04573_pos12823
41.9 S07734_pos3852
42.5 S05348_pos2072
43.0 S04081_pos6103
44.4 S01427_pos23269
45.1 S08353_pos7699
45.8 S04961_pos14204
45.9 S03842_pos9881
46.2 S01036_pos16957
46.4 S06634_pos9462
46.9 S02469_pos21414
48.4 S01123_pos36471 S01023_pos17246
48.6 S03801_pos20492
48.8 S00080_pos21817
49.1 S23326_pos1449
50.1 S01010_pos10936
50.3 S02714_pos18791
50.7 S06145_pos2441
51.0 S16545_pos562
52.1 S07132_pos8759
52.7 S04725_pos7036
53.0 S30027_pos720
54.4 S06009_pos12218
55.2 S03235_pos24260
55.9 S00404_pos50121
56.5 S10475_pos4694 S00613_pos29449
57.0 S03686_pos18789
57.4 S06239_pos15263
58.1 S08319_pos11163
58.8 S01205_pos22231
59.3 S04223_pos9102
60.2 V01-480-3S
62.6 S02838_pos23846

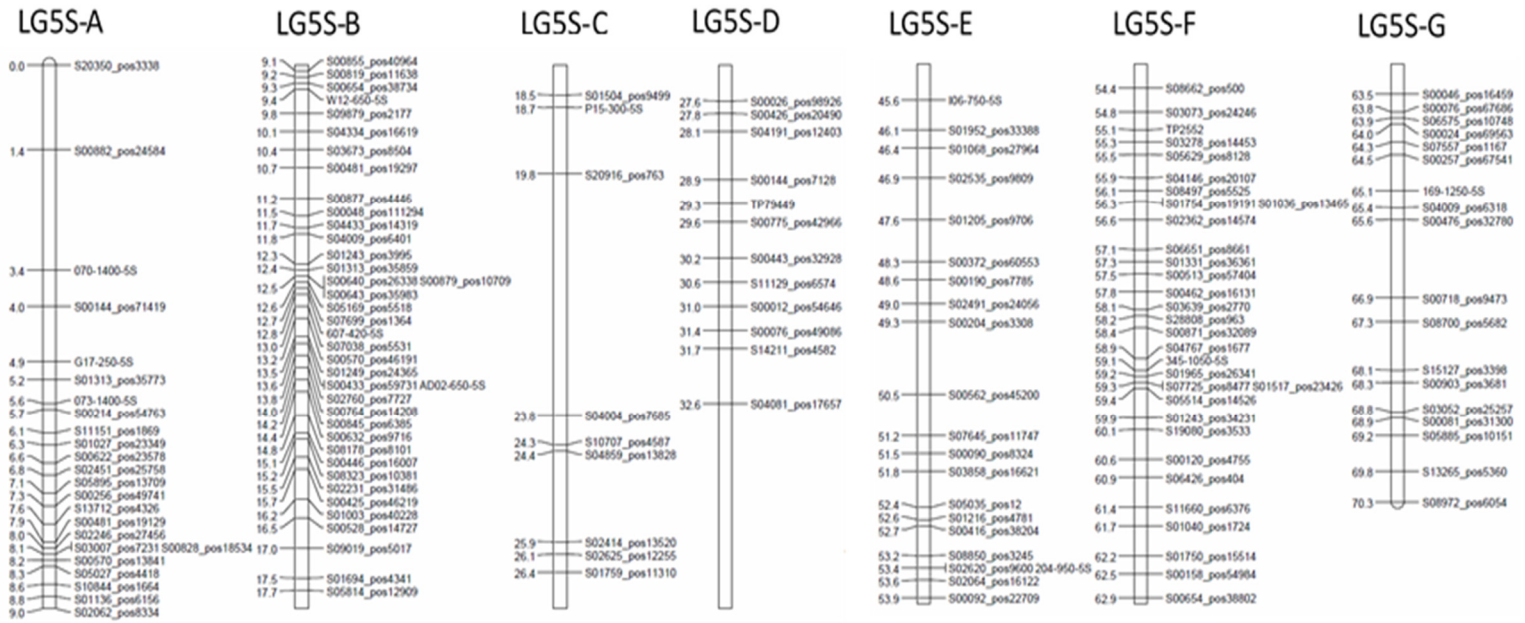
(F)


LG3R
S00617_pos49625
S16505_pos542
S03272_pos9110
S05942_pos2293
S00191_pos70683
S00099_pos82185
S22181_pos2123
S02328_pos17991
S14570_pos4500
S01541_pos37535
S00299_pos9817
S08262_pos7455
S02710_pos12808
S04386_pos8291
S02119_pos28757
S00540_pos7927
S02710_pos12740
S08554_pos483
S03470_pos3200
S02119_pos28630
S01535_pos21487
S02695_pos4523
S32604_pos376
S12517_pos4033
S00523_pos3452
S09687_pos4828
S03872_pos16495
S09637_pos5618
S09637_pos5563
S09637_pos5696
S06113_pos8360
S06113_pos9029
S02461_pos13808
S03872_pos16497 S26627_pos75
S03825_pos370
S03872_pos16393
S00413_pos59484 S04581_pos17269
S17093_pos1593 S02836_pos15880
276-900-3R
S12628_pos1436
M02-850-3R
S17093_pos1674
S01983_pos14495
S02530_pos24837
S19706_pos3231
S01172_pos19403
S08537_pos3689
S04799_pos9720
S02573_pos14704
S03067_pos4515
S21712_pos1505

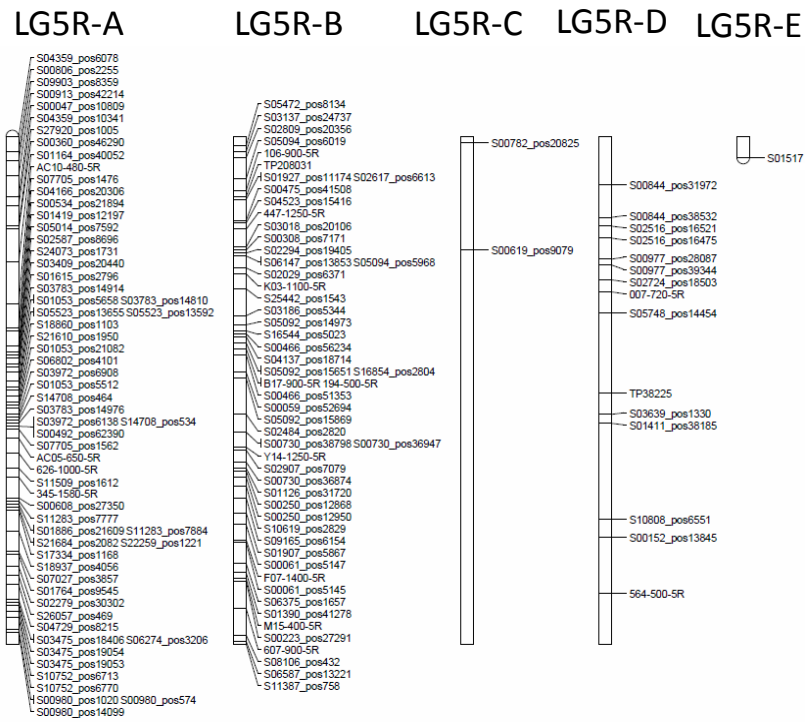
(G)


LG4S
0.0 S04941_pos7848
9.9 S32904_pos493
14.8 S06752_pos8849
15.7 S02491_pos27314
16.6 S00089_pos59894
17.1 S12739_pos419
18.1 S00376_pos47419
18.5 S01786_pos18602
19.2 S00421_pos9695
19.3 S12284_pos6173
20.0 S03145_pos15727
21.1 S01832_pos32740
23.4 Y05-720-4S
24.1 S03544_pos19458
24.7 S00102_pos97532
25.1 S00747_pos44416
25.3 S30423_pos727
25.5 S01932_pos3453
25.9 S05272_pos7615
26.0 S04625_pos1534 S00256_pos49827
26.1 S01513_pos20356
26.3 S00117_pos29755
26.4 S10855_pos1254 S04971_pos13706
26.5 S08353_pos8118 S06113_pos8246
26.7 S02605_pos15608
27.1 S02193_pos6500
27.2 S11428_pos5719
27.4 S00165_pos7879
27.6 S00546_pos3313
27.8 S02385_pos27647 S04632_pos7876
28.0 S01278_pos11327 S03920_pos2466
28.2 S02949_pos13743
28.6 S02813_pos9536
28.9 S08406_pos5559
29.0 S05540_pos3608
30.2 531-920-4S
30.7 G05-400-4S
32.1 S05190_pos9849
35.0 S01639_pos14857
39.2 S09151_pos3411
41.7 S00734_pos32059
43.5 S04561_pos1255
44.3 S02049_pos27940
45.9 Y02-650-4S
46.9 S09721_pos3522
49.8 S01760_pos1165
53.8 S01036_pos32289
55.0 S03113_pos16271
55.8 S00356_pos46719
56.8 S01028_pos28635
57.9 S06879_pos1757
59.2 S00579_pos38806 S00136_pos15977
60.9 S05692_pos15769
64.5 S08665_pos7062
66.3 S00462_pos41905
86.7 S02176_pos13432
LG4R-A
S06601_pos5396
S01090_pos27920
TP104613
S06601_pos5470
S08364_pos6962
S00668_pos47245
S00019_pos60303
S00019_pos144105
S00019_pos144304
S02130_pos10483
S09828_pos480
S02321_pos9
498-1150-4R
S02726_pos13480
694-775-4R
S01398_pos34968
N05-1800-4R
S01134_pos24159
S19860_pos1960
S04319_pos3155
S07300_pos1422
S12469_pos3917
S04251_pos19085
S04461_pos3737
S07281_pos5905
S06929_pos10249
S04251_pos19150
S09984_pos9253
S02221_pos8257
S09742_pos7216
S01301_pos35943
S05463_pos4130
S01301_pos36031
S05463_pos4220
S03749_pos7842
S03948_pos3229
S04119_pos15947
S02391_pos27065
S11433_pos7395
S11129_pos6707 S00427_pos3838
S00175_pos76425
S01848_pos35707 S22600_pos437
S03339_pos9054
S26883_pos402
S07609_pos6114
S35692_pos369
S00773_pos14227
S01751_pos2339
S00944_pos19577
S09695_pos3505
S09695_pos3453
LG4R-B
S03355_pos20373
S03355_pos20274
S00115_pos83157
S00115_pos83776
S00115_pos83239
S00397_pos43935
S00115_pos83904
S01727_pos23776
S00192_pos16129
TP147698
TP163521
S01936_pos19115
S01362_pos19358
S01362_pos21605
S00624_pos21843
W19-650-4R
S00541_pos24276
S00079_pos72511
S07235_pos3716
S07235_pos3640
S00239_pos21435
S00412_pos20269
S01263_pos34922 S00412_pos20313
S00412_pos27998
S01263_pos34961
S00291_pos58543
S00323_pos42268
S02267_pos20741
S03444_pos2841
S01976_pos4498
S01976_pos4440
S01976_pos28754
S03314_pos4123
S03416_pos19851
S03416_pos19788
S03861_pos5569
S03861_pos10523
S00321_pos67814 S01313_pos35931
S03314_pos3935
S01313_pos35857 S05825_pos1548
S25534_pos289
S01689_pos32285
S00016_pos76235
S00834_pos42587
S05731_pos8111
S00016_pos76166
S27190_pos1499

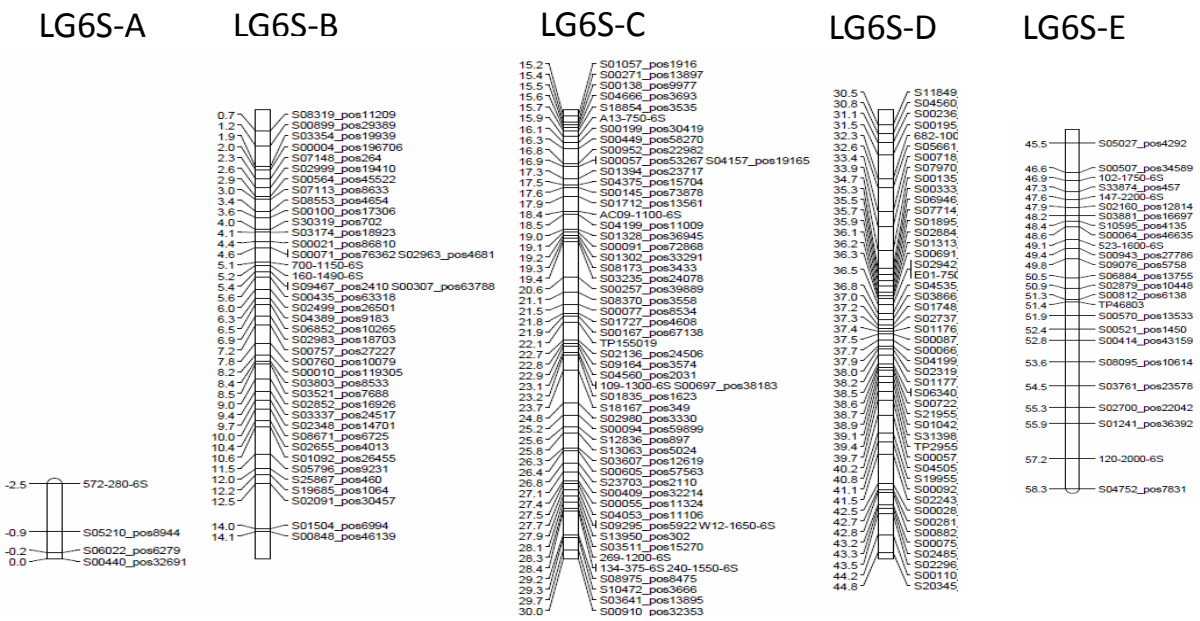
(H)



(I)

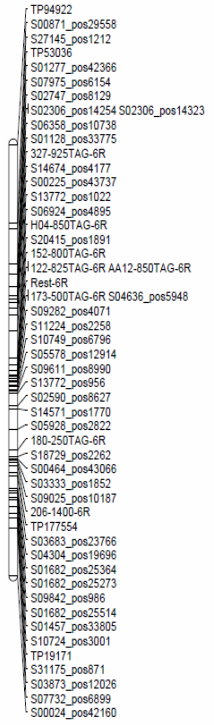

LG5R-A
LG5R-B
LG5R-C
LG5R-D
LG5R-E
S04359_pos6078
S00806_pos2255
S09903_pos8359
S00913_pos42214
S00047_pos10809
S04359_pos10341
S27920_pos1005
S00360_pos46290
S01164_pos40052
AC10-480-5R
S07705_pos1476
S04166_pos20306
S00534_pos21894
S01419_pos12197
S05014_pos7592
S02587_pos8696
S24073_pos1731
S03409_pos20440
S01615_pos2796
S03783_pos14914
S01053_pos5658 S03783_pos14810
S05523_pos13655 S05523_pos13592
S18860_pos1103
S21610_pos1950
S01053_pos21082
S06802_pos4101
S03972_pos6908
S01053_pos5512
S14708_pos464
S03783_pos14976
S03972_pos6138 S14708_pos534
S00492_pos62390
S07705_pos1562
AC05-650-5R
626-1000-5R
S11509_pos1612
345-1580-5R
S00608_pos27350
S11283_pos7777
S01886_pos21609 S11283_pos7884
S21684_pos2082 S22259_pos1221
S17334_pos1168
S18937_pos4056
S07027_pos3857
S01764_pos9545
S02279_pos30302
S26057_pos469
S04729_pos8215
S03475_pos18406 S06274_pos3206
S03475_pos19054
S03475_pos19053
S10752_pos6713
S10752_pos6770
S00980_pos1020 S00980_pos574
S00980_pos14099
S05472_pos8134
S03137_pos24737
S02809_pos20356
S05094_pos6019
106-900-5R
TP208031
S01927_pos11174 S02617_pos6613
S00475_pos41508
S04523_pos15416
447-1250-5R
S03018_pos20106
S00308_pos7171
S02294_pos19405
S06147_pos13853 S05094_pos5968
S02029_pos6371
K03-1100-5R
S25442_pos1543
S03186_pos5344
S05092_pos14973
S16544_pos5023
S00466_pos56234
S04137_pos18714
S05092_pos15651 S16854_pos2804
B17-900-5R 194-500-5R
S00466_pos51353
S00059_pos52694
S05092_pos15869
S02484_pos2820
S00730_pos38798 S00730_pos36947
Y14-1250-5R
S02907_pos7079
S00730_pos36874
S01126_pos31720
S00250_pos12868
S00250_pos12950
S10619_pos2829
S09165_pos6154
S01907_pos5867
S00061_pos5147
F07-1400-5R
S00061_pos5145
S06375_pos1657
S01390_pos41278
M15-400-5R
S00223_pos27291
607-900-5R
S08106_pos432
S06587_pos13221
S11387_pos758
S00782_pos20825
S00619_pos9079
S01517
S00844_pos31972
S00844_pos38532
S02516_pos16521
S02516_pos16475
S00977_pos28087
S00977_pos39344
S02724_pos18503
007-720-5R
S05748_pos14454
TP38225
S03639_pos1330
S01411_pos38185
S10808_pos6551
S00152_pos13845
564-500-5R

(J)

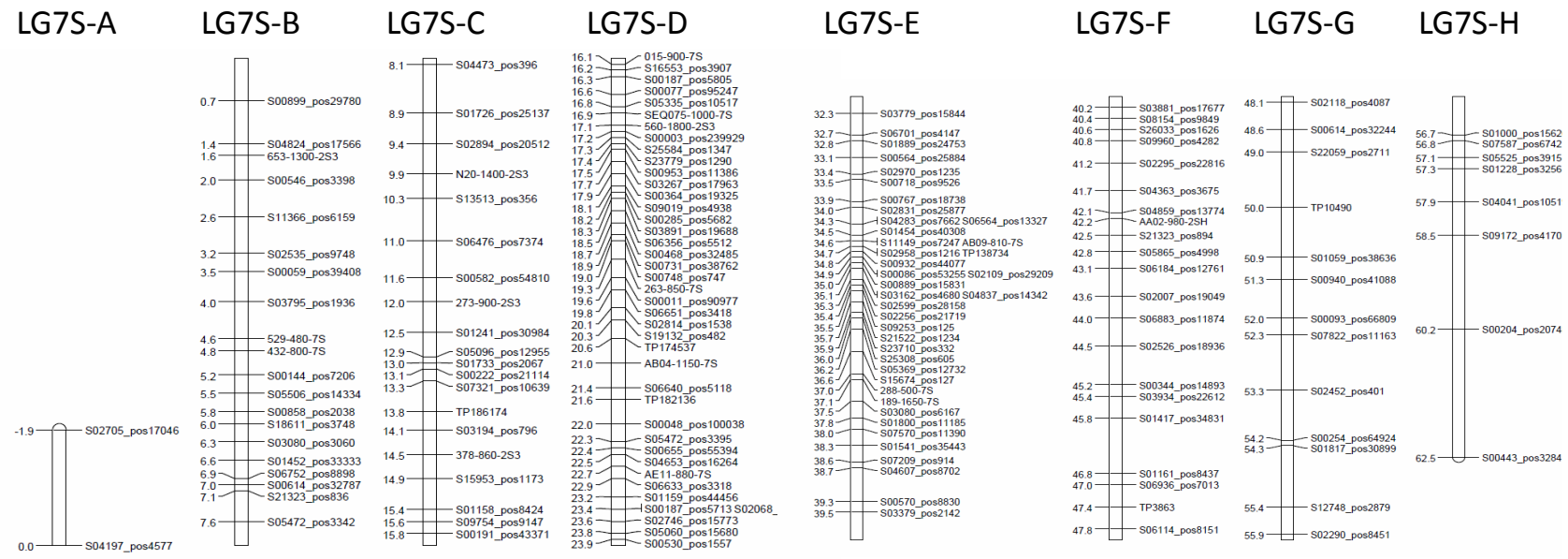

LG6S-A
LG6S-B
LG6S-C
LG6S-D
LG6S-E
-2.5 572-280-6S
-0.9 S05210_pos8944
-0.2 S06022_pos6279
0.0 S00440_pos32691
0.7 S08319_pos11209
1.2 S00899_pos29389
1.9 S03354_pos19939
2.0 S00004_pos196706
2.3 S07148_pos264
2.6 S02999_pos19410
2.9 S00564_pos45522
3.0 S07113_pos8633
3.4 S08553_pos4654
3.6 S00100_pos17306
4.0 S30319_pos702
4.1 S03174_pos18923
4.4 S00021_pos86810
4.6 S00071_pos76362 S02963_pos4681
5.1 700-1150-6S
5.2 160-1490-6S
5.4 S09467_pos2410 S00307_pos63788
5.6 S00435_pos63318
6.0 S02499_pos26501
6.3 S04389_pos9183
6.5 S06852_pos10265
6.9 S02983_pos18703
7.2 S00757_pos27227
7.8 S00760_pos10079
8.2 S00010_pos119305
8.4 S03803_pos8533
8.5 S03521_pos7688
9.0 S02852_pos16926
9.4 S03337_pos24517
9.7 S02348_pos14701
10.0 S08671_pos6725
10.4 S02655_pos4013
10.6 S01092_pos26455
11.5 S05796_pos9231
12.0 S25867_pos460
12.2 S19685_pos1064
12.5 S02091_pos30457
14.0 S01504_pos6994
14.1 S00848_pos46139
15.2 S01057_pos1916
15.4 S00271_pos13897
15.5 S00138_pos9977
15.6 S04686_pos3693
15.7 S18854_pos3535
15.9 A13-750-6S
16.1 S00199_pos30419
16.3 S00449_pos58270
16.8 S00952_pos22982
16.9 S00057_pos53267 S04157_pos19165
17.3 S01394_pos23717
17.5 S04375_pos15704
17.6 S00145_pos73878
17.9 S01712_pos13561
18.4 AC09-1100-6S
18.5 S04199_pos11009
19.0 S01328_pos36945
19.1 S00091_pos72868
19.2 S01302_pos33291
19.3 S08173_pos3433
19.4 S03236_pos24078
20.6 S00257_pos39889
21.1 S08370_pos3558
21.5 S00077_pos8534
21.8 S01727_pos4608
21.9 S00167_pos67138
22.1 TP155019
22.7 S02136_pos24506
22.8 S09164_pos3574
22.9 S04560_pos2031
23.1 109-1300-6S S00697_pos38183
23.2 S01835_pos1623
23.7 S18167_pos349
24.8 S02980_pos3330
25.2 S00094_pos59899
25.6 S12836_pos897
25.8 S13063_pos5024
26.3 S03607_pos12619
26.4 S00605_pos57563
26.8 S23703_pos2110
27.1 S00409_pos32214
27.4 S00055_pos11324
27.5 S04053_pos11106
27.7 S09295_pos5922 W12-1650-6S
27.9 S13950_pos302
28.1 S03511_pos15270
28.3 269-1200-6S
28.4 134-375-6S 240-1550-6S
29.2 S08975_pos8475
29.3 S10472_pos3666
29.7 S03641_pos13895
30.0 S00910_pos32353
30.5 S11849
30.8 S04560
31.1 S00236
31.5 S00195
32.3 682-100
32.6 S05661
33.4 S00716
33.9 S07970
34.7 S00135
35.3 S00333
35.5 S06946
35.7 S07714
35.9 S01895
36.1 S02884
36.2 S01313
36.3 S00691
36.5 S02942
E01-750
36.8 S04535
37.0 S03866
37.2 S01748
37.3 S02737
37.4 S01176
37.5 S00087
37.7 S00066
37.9 S04199
38.0 S02319
38.2 S01177
38.5 S06340
38.6 S00722
38.7 S21955
38.9 S01042
39.1 S31398
39.4 TP2955
39.7 S00057
40.2 S04505
40.8 S19955
41.1 S00092
41.5 S02243
42.5 S00028
42.7 S00281
42.8 S00882
43.2 S00075
43.3 S02485
43.5 S02296
44.2 S00110
44.8 S20345
45.5 S05027_pos4292
46.6 S00507_pos34589
46.9 102-1750-6S
47.3 S33874_pos457
47.6 147-2200-6S
47.9 S02160_pos12814
48.2 S03881_pos16697
48.4 S10595_pos4135
48.6 S00064_pos46635
49.1 523-1600-6S
49.4 S00943_pos27786
49.8 S09076_pos5758
50.5 S06884_pos13755
50.9 S02879_pos10448
51.3 S00812_pos6138
51.4 TP46803
51.9 S00570_pos13533
52.4 S00521_pos1450
52.8 S00414_pos43159
53.6 S08095_pos10614
54.5 S03761_pos23578
55.3 S02700_pos22042
55.9 S01241_pos36392
57.2 120-2000-6S
58.3 S04752_pos7831

(K)

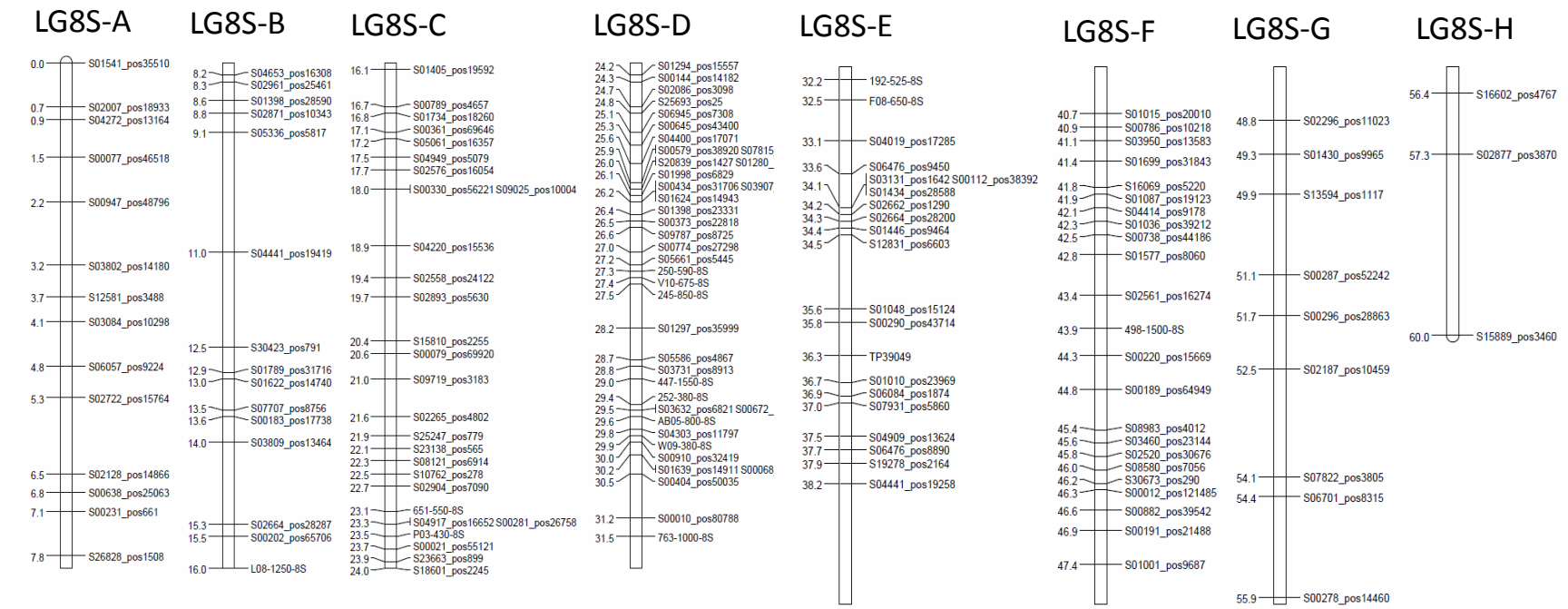
LG6R


TP94922
S00871_pos29558
S27145_pos1212
TP53036
S01277_pos42366
S07975_pos6154
S02747_pos8129
S02306_pos14254 S02306_pos14323
S06358_pos10738
S01128_pos33775
327-925TAG-6R
S14674_pos4177
S00225_pos43737
S13772_pos1022
S06924_pos4895
H04-850TAG-6R
S20415_pos1891
152-800TAG-6R
122-825TAG-6R AA12-850TAG-6R
Rest-6R
173-500TAG-6R S04636_pos5948
S09282_pos4071
S11224_pos2258
S10749_pos6796
S05578_pos12914
S09611_pos8990
S13772_pos956
S02590_pos8627
S14571_pos1770
S05926_pos2822
180-250TAG-6R
S18729_pos2262
S00464_pos43066
S03333_pos1852
S09025_pos10187
206-1400-6R
TP177554
S03683_pos23766
S04304_pos19696
S01682_pos25364
S01682_pos25273
S09842_pos986
S01682_pos25514
S01457_pos33805
S10724_pos3001
TP19171
S31175_pos871
S03873_pos12026
S07732_pos6899
S00024_pos42160

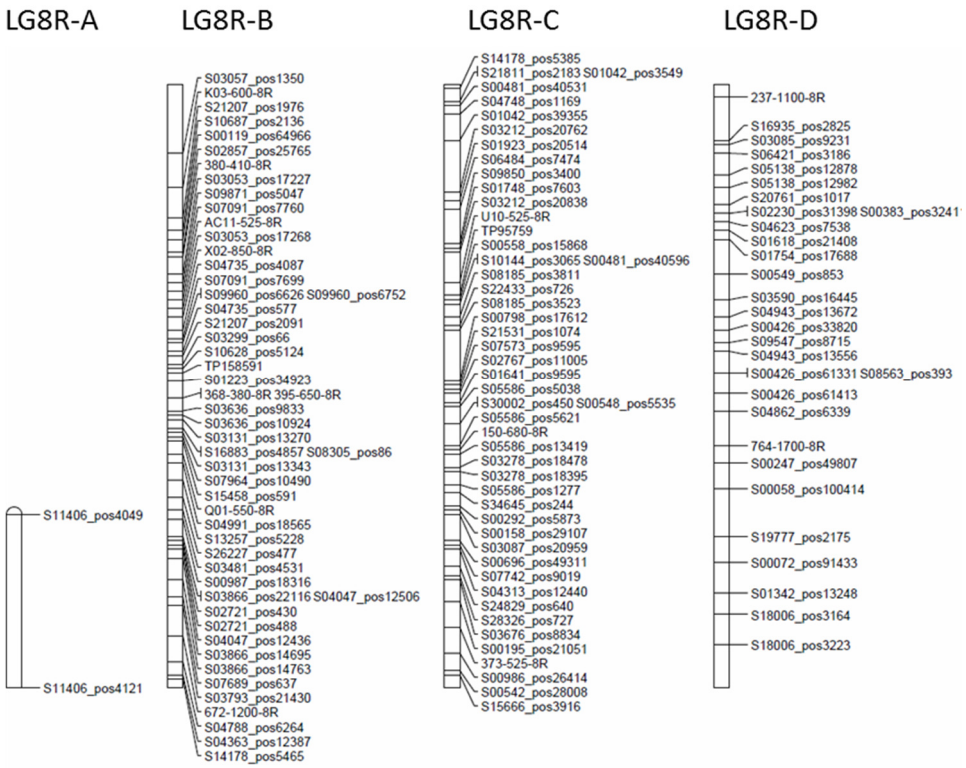
(L)


LG7S-A
LG7S-B
LG7S-C
LG7S-D
LG7S-E
LG7S-F
LG7S-G
LG7S-H
-1.9 S02705_pos17046
0.0 S04197_pos4577
0.7 S00899_pos29780
1.4 S04824_pos17566
1.6 653-1300-2S3
2.0 S00546_pos3398
2.6 S11366_pos6159
3.2 S02535_pos9748
3.5 S00059_pos39408
4.0 S03795_pos1936
4.6 529-480-7S
4.8 432-800-7S
5.2 S00144_pos7206
5.5 S05506_pos14334
5.8 S00858_pos2038
6.0 S18611_pos3748
6.3 S03080_pos3060
6.6 S01452_pos33333
6.9 S06752_pos8898
7.0 S00614_pos32787
7.1 S21323_pos836
7.6 S05472_pos3342
8.1 S04473_pos396
8.9 S01726_pos25137
9.4 S02894_pos20512
9.9 N20-1400-2S3
10.3 S13513_pos356
11.0 S06476_pos7374
11.6 S00582_pos54810
12.0 273-900-2S3
12.5 S01241_pos30984
12.9 S05096_pos12955
13.0 S01733_pos2067
13.1 S00222_pos21114
13.3 S07321_pos10639
13.8 TP186174
14.1 S03194_pos796
14.5 378-860-2S3
14.9 S15953_pos1173
15.4 S01158_pos8424
15.6 S09754_pos9147
15.8 S00191_pos43371
16.1 015-900-7S
16.2 S16553_pos3907
16.3 S00187_pos5805
16.6 S00077_pos95247
16.8 S05335_pos10517
16.9 SEQ075-1000-7S
17.1 560-1800-2S3
17.2 S00003_pos239929
17.3 S25584_pos1347
17.4 S23779_pos1290
17.5 S00953_pos11386
17.7 S03267_pos17963
17.9 S00364_pos19325
18.1 S09019_pos4938
18.2 S00285_pos5682
18.3 S03891_pos19688
18.5 S06356_pos5512
18.7 S00468_pos32485
18.9 S00731_pos38762
19.0 S00748_pos747
19.3 263-850-7S
19.6 S00011_pos90977
19.8 S06651_pos3418
20.1 S02814_pos1538
20.3 S19132_pos482
20.6 TP174537
21.0 AB04-1150-7S
21.4 S06640_pos5118
21.6 TP182136
22.0 S00048_pos100038
22.3 S05472_pos3395
22.4 S00655_pos55394
22.5 S04653_pos16264
22.7 AE11-880-7S
22.9 S06633_pos3318
23.2 S01159_pos44456
23.4 S00187_pos5713 S02068_
23.6 S02746_pos15773
23.8 S05060_pos15680
23.9 S00530_pos1557
32.3 S03779_pos15844
32.7 S06701_pos4147
32.8 S01889_pos24753
33.1 S00564_pos25884
33.4 S02970_pos1235
33.5 S00718_pos9526
33.9 S00767_pos18738
34.0 S02831_pos25877
34.3 S04283_pos7662 S06564_pos13327
34.5 S01454_pos40308
34.6 S11149_pos7247 AB09-810-7S
34.7 S02958_pos1216 TP138734
34.8 S00932_pos44077
34.9 S00086_pos53255 S02109_pos29209
35.0 S00889_pos15831
35.1 S03162_pos4680 S04837_pos14342
35.3 S02599_pos28158
35.4 S02256_pos21719
35.5 S09253_pos125
35.7 S21522_pos1234
35.9 S23710_pos332
36.0 S25308_pos605
36.2 S05369_pos12732
36.6 S15674_pos127
37.0 288-500-7S
37.1 189-1650-7S
37.5 S03080_pos6167
37.8 S01800_pos11185
38.0 S07570_pos11390
38.3 S01541_pos35443
38.6 S07209_pos914
38.7 S04607_pos8702
39.3 S00570_pos8830
39.5 S03379_pos2142
40.2 S03881_pos17677
40.4 S08154_pos9849
40.6 S26033_pos1626
40.8 S09960_pos4282
41.2 S02295_pos22816
41.7 S04363_pos3675
42.1 S04859_pos13774
42.2 AA02-980-2SH
42.5 S21323_pos894
42.8 S05865_pos4998
43.1 S06184_pos12761
43.6 S02007_pos19049
44.0 S06883_pos11874
44.5 S02526_pos18936
45.2 S00344_pos14893
45.4 S03934_pos22612
45.8 S01417_pos34831
46.8 S01161_pos8437
47.0 S06936_pos7013
47.4 TP3863
47.8 S06114_pos8151
48.1 S02118_pos4087
48.6 S00614_pos32244
49.0 S22059_pos2711
50.0 TP10490
50.9 S01059_pos38636
51.3 S00940_pos41088
52.0 S00093_pos66809
52.3 S07822_pos11163
53.3 S02452_pos401
54.2 S00254_pos64924
54.3 S01817_pos30899
55.4 S12748_pos2879
55.9 S02290_pos8451
56.7 S01000_pos1562
56.8 S07587_pos6742
57.1 S05525_pos3915
57.3 S01228_pos3256
57.9 S04041_pos1051
58.5 S09172_pos4170
60.2 S00204_pos2074
62.5 S00443_pos3284

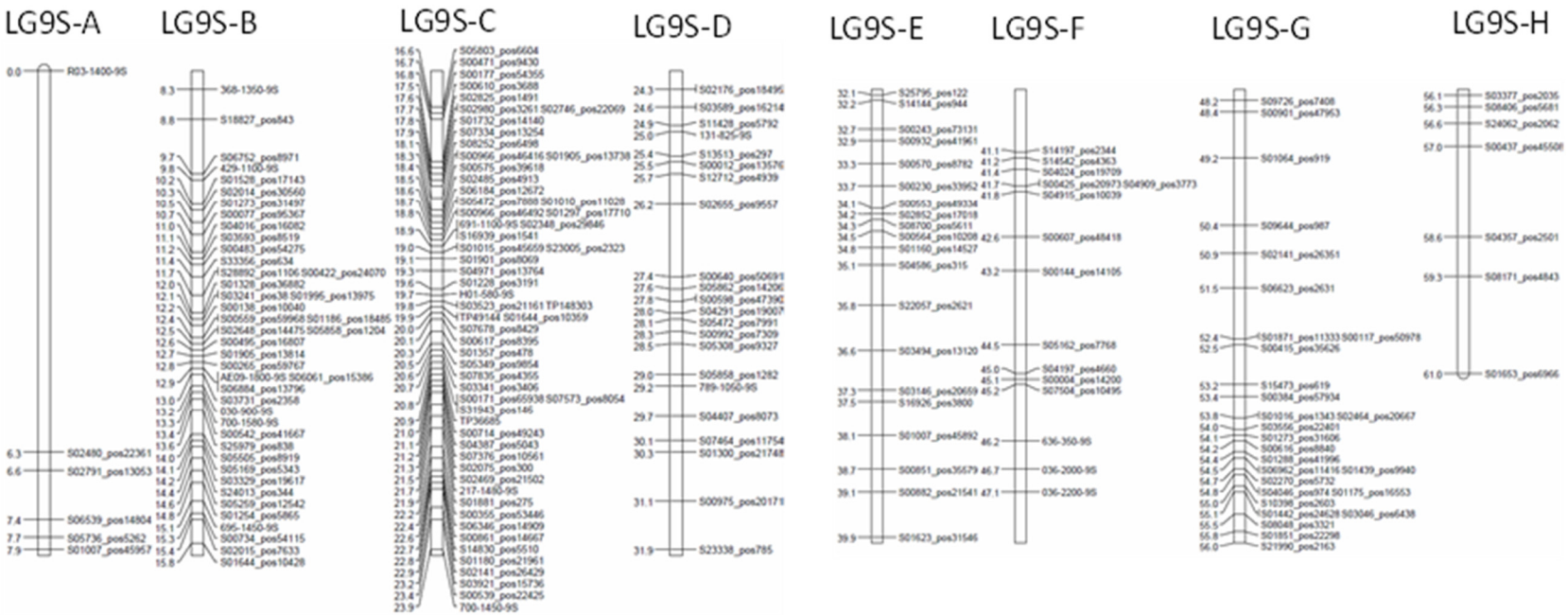
(M)


LG8S-A
0.0 S01541_pos35510
0.7 S02007_pos18933
0.9 S04272_pos13164
1.5 S00077_pos46518
2.2 S00947_pos48796
3.2 S03802_pos14180
3.7 S12581_pos3488
4.1 S03084_pos10298
4.8 S06057_pos9224
5.3 S02722_pos15764
6.5 S02128_pos14866
6.8 S00638_pos25063
7.1 S00231_pos661
7.8 S26828_pos1508
LG8S-B
8.2 S04653_pos16308
8.3 S02961_pos25461
8.6 S01398_pos28590
8.8 S02871_pos10343
9.1 S05336_pos5817
11.0 S04441_pos19419
12.5 S30423_pos791
12.9 S01789_pos31716
13.0 S01622_pos14740
13.5 S07707_pos8756
13.6 S00183_pos17738
14.0 S03809_pos13464
15.3 S02664_pos28287
15.5 S00202_pos65706
16.0 L08-1250-8S
LG8S-C
16.1 S01405_pos19592
16.7 S00789_pos4657
16.8 S01734_pos18260
17.1 S00361_pos69646
17.2 S05061_pos16357
17.5 S04949_pos5079
17.7 S02576_pos16054
18.0 S00330_pos56221 S09025_pos10004
18.9 S04220_pos15536
19.4 S02558_pos24122
19.7 S02893_pos5630
20.4 S15810_pos2255
20.6 S00079_pos69920
21.0 S09719_pos3183
21.6 S02265_pos4802
21.9 S25247_pos779
22.1 S23138_pos565
22.3 S08121_pos6914
22.5 S10762_pos278
22.7 S02904_pos7090
23.1 651-550-8S
23.3 S04917_pos16652 S00281_pos26758
23.5 P03-430-8S
23.7 S00021_pos55121
23.9 S23663_pos899
24.0 S18601_pos2245
LG8S-D
24.2 S01294_pos15557
24.3 S00144_pos14182
24.7 S02086_pos3098
24.8 S25693_pos25
25.1 S06945_pos7308
25.3 S00645_pos43400
25.6 S04400_pos17071
25.9 S00579_pos38920 S07815
26.0 S20839_pos1427 S01280_
26.1 S01998_pos6829
S00434_pos31706 S03907
26.2 S01624_pos14943
26.4 S01398_pos23331
26.5 S00373_pos22818
26.6 S09787_pos8725
27.0 S00774_pos27298
27.2 S05661_pos5445
27.3 250-590-8S
27.4 V10-675-8S
27.5 245-850-8S
28.2 S01297_pos35999
28.7 S05586_pos4867
28.8 S03731_pos8913
29.0 447-1550-8S
29.4 252-380-8S
29.5 S03632_pos6821 S00672_
29.6 AB05-800-8S
29.8 S04303_pos11797
29.9 W09-380-8S
30.0 S00910_pos32419
30.2 S01639_pos14911 S00068
30.5 S00404_pos50035
31.2 S00010_pos80788
31.5 763-1000-8S
LG8S-E
32.2 192-525-8S
32.5 F08-650-8S
33.1 S04019_pos17285
33.6 S06476_pos9450
S03131_pos1642 S00112_pos38392
34.1 S01434_pos28588
34.2 S02662_pos1290
34.3 S02664_pos28200
34.4 S01446_pos9464
34.5 S12831_pos6603
35.6 S01048_pos15124
35.8 S00290_pos43714
36.3 TP39049
36.7 S01010_pos23969
36.9 S06084_pos1874
37.0 S07931_pos5860
37.5 S04909_pos13624
37.7 S06476_pos8890
37.9 S19278_pos2164
38.2 S04441_pos19258
LG8S-F
40.7 S01015_pos20010
40.9 S00786_pos10218
41.1 S03950_pos13583
41.4 S01699_pos31843
41.8 S16069_pos5220
41.9 S01087_pos19123
42.1 S04414_pos9178
42.3 S01036_pos39212
42.5 S00738_pos44186
42.8 S01577_pos8060
43.4 S02561_pos16274
43.9 498-1500-8S
44.3 S00220_pos15669
44.8 S00189_pos64949
45.4 S08983_pos4012
45.6 S03460_pos23144
45.8 S02520_pos30676
46.0 S08580_pos7056
46.2 S30673_pos290
46.3 S00012_pos121485
46.6 S00882_pos39542
46.9 S00191_pos21488
47.4 S01001_pos9687
LG8S-G
48.8 S02296_pos11023
49.3 S01430_pos9965
49.9 S13594_pos1117
51.1 S00287_pos52242
51.7 S00296_pos28863
52.5 S02187_pos10459
54.1 S07822_pos3805
54.4 S06701_pos8315
55.9 S00278_pos14460
LG8S-H
56.4 S16602_pos4767
57.3 S02877_pos3870
60.0 S15889_pos3460

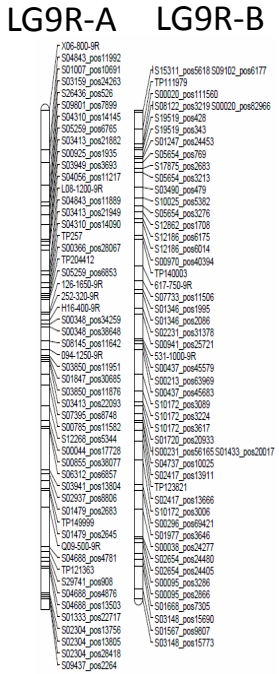
(N)


LG8R-A
LG8R-B
LG8R-C
LG8R-D
S11406_pos4049
S11406_pos4121
S03057_pos1350
K03-600-8R
S21207_pos1976
S10687_pos2136
S00119_pos64966
S02857_pos25765
380-410-8R
S03053_pos17227
S09871_pos5047
S07091_pos7760
AC11-525-8R
S03053_pos17268
X02-850-8R
S04735_pos4087
S07091_pos7699
S09960_pos6626 S09960_pos6752
S04735_pos577
S21207_pos2091
S03299_pos66
S10628_pos5124
TP158591
S01223_pos34923
368-380-8R 395-650-8R
S03636_pos9833
S03636_pos10924
S03131_pos13270
S16883_pos4857 S08305_pos86
S03131_pos13343
S07964_pos10490
S15458_pos591
Q01-550-8R
S04991_pos18565
S13257_pos5228
S26227_pos477
S03481_pos4531
S00987_pos18316
S03866_pos22116 S04047_pos12506
S02721_pos430
S02721_pos488
S04047_pos12436
S03866_pos14695
S03866_pos14763
S07689_pos637
S03793_pos21430
672-1200-8R
S04788_pos6264
S04363_pos12387
S14178_pos5465
S14178_pos5385
S21811_pos2183 S01042_pos3549
S00481_pos40531
S04748_pos1169
S01042_pos39355
S03212_pos20762
S01923_pos20514
S06484_pos7474
S09850_pos3400
S01748_pos7603
S03212_pos20838
U10-525-8R
TP95759
S00558_pos15868
S10144_pos3065 S00481_pos40596
S08185_pos3811
S22433_pos726
S08185_pos3523
S00798_pos17612
S21531_pos1074
S07573_pos9595
S02767_pos11005
S01641_pos9595
S05586_pos5038
S30002_pos450 S00548_pos5535
S05586_pos5621
150-680-8R
S05586_pos13419
S03278_pos18478
S03278_pos18395
S05586_pos1277
S34645_pos244
S00292_pos5873
S00158_pos29107
S03087_pos20959
S00696_pos49311
S07742_pos9019
S04313_pos12440
S24829_pos640
S28326_pos727
S03676_pos8834
S00195_pos21051
373-525-8R
S00986_pos26414
S00542_pos28008
S15666_pos3916
237-1100-8R
S16935_pos2825
S03085_pos9231
S06421_pos3186
S05138_pos12878
S05138_pos12982
S20761_pos1017
S02230_pos31398 S00383_pos3241
S04623_pos7538
S01618_pos21408
S01754_pos17688
S00549_pos853
S03590_pos16445
S04943_pos13672
S00426_pos33820
S09547_pos8715
S04943_pos13556
S00426_pos61331 S08563_pos393
S00426_pos61413
S04862_pos6339
764-1700-8R
S00247_pos49807
S00058_pos100414
S19777_pos2175
S00072_pos91433
S01342_pos13248
S18006_pos3164
S18006_pos3223

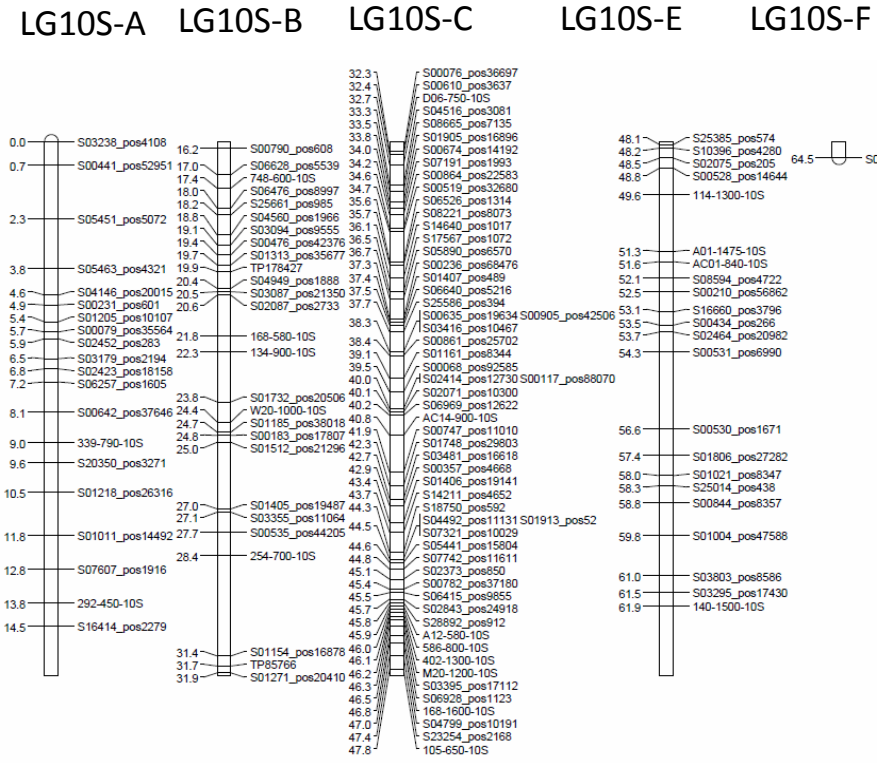
(O)



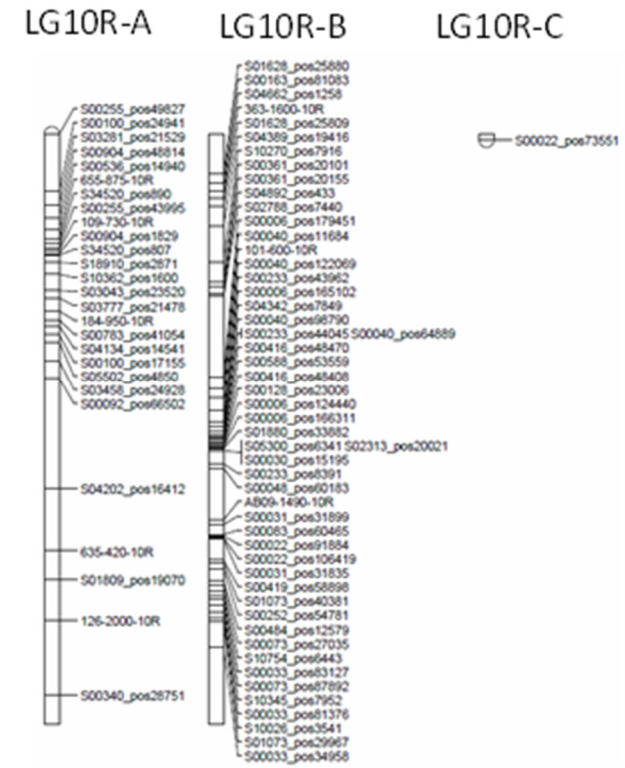
(P)


LG9R-A
LG9R-B
X06-800-9R
S04843_pos11992
S01007_pos10691
S03159_pos24263
S26436_pos526
S09801_pos7699
S04310_pos14145
S05259_pos6765
S03413_pos21882
S00925_pos1935
S03949_pos3693
S04056_pos11217
L08-1200-9R
S04843_pos11889
S03413_pos21949
S04310_pos14090
TP257
S00366_pos28067
TP204412
S05259_pos6853
126-1650-9R
252-320-9R
H16-400-9R
S00348_pos34259
S00348_pos38648
S08145_pos11642
094-1255-9R
S03650_pos11951
S01847_pos30685
S03850_pos11876
S03413_pos22093
S07395_pos8748
S00785_pos11582
S12286_pos5344
S00044_pos17728
S00855_pos38077
S06312_pos6857
S03941_pos13804
S02937_pos8806
S01479_pos2683
TP149999
S01479_pos2645
Q09-500-9R
S04688_pos4781
TP121363
S29741_pos608
S04686_pos4876
S04688_pos13503
S01333_pos22717
S02304_pos13756
S02304_pos13605
S02304_pos28418
S09437_pos2264
S15311_pos5618 S09102_pos6177
TP111979
S00020_pos111560
S08122_pos3219 S00020_pos82966
S19519_pos428
S19519_pos343
S01247_pos24453
S05654_pos769
S17875_pos2683
S05654_pos3213
S03490_pos479
S10025_pos5382
S05654_pos3276
S12862_pos1708
S12186_pos6175
S12186_pos6014
S00970_pos40394
TP140003
617-750-9R
S07733_pos11506
S01346_pos1995
S01346_pos2086
S02231_pos31378
S00941_pos25721
531-1000-9R
S00437_pos45579
S00213_pos63969
S00437_pos45683
S10172_pos3089
S10172_pos3224
S10172_pos3617
S01720_pos20933
S00231_pos56165 S01433_pos20017
S04737_pos10025
S02417_pos13911
TP123821
S02417_pos13666
S10172_pos3006
S00296_pos69421
S01977_pos3646
S00038_pos24277
S02654_pos24480
S02654_pos24405
S00095_pos3266
S00095_pos2866
S01668_pos7305
S03148_pos15690
S01567_pos9807
S03148_pos15773

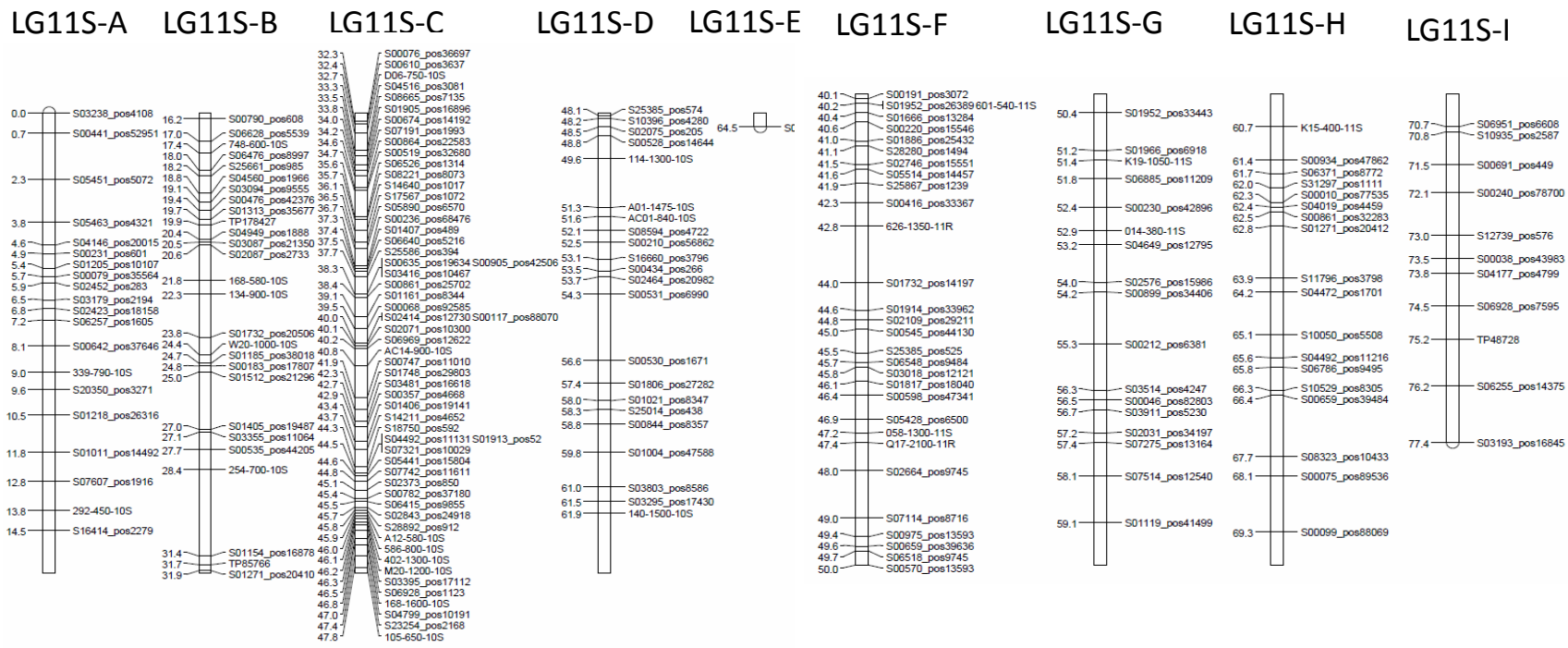
(Q)


LG10S-A
LG10S-B
LG10S-C
LG10S-E
LG10S-F
0.0 S03238_pos4108
0.7 S00441_pos52951
2.3 S05451_pos5072
3.8 S05463_pos4321
4.6 S04146_pos20015
4.9 S00231_pos601
5.4 S01205_pos10107
5.7 S00079_pos35564
5.9 S02452_pos283
6.5 S03179_pos2194
6.8 S02423_pos18158
7.2 S06257_pos1605
8.1 S00642_pos37646
9.0 339-790-10S
9.6 S20350_pos3271
10.5 S01218_pos26316
11.8 S01011_pos14492
12.8 S07607_pos1916
13.8 292-450-10S
14.5 S16414_pos2279
16.2 S00790_pos608
17.0 S06628_pos5539
17.4 748-600-10S
18.0 S06476_pos8997
18.2 S25661_pos985
18.8 S04560_pos1966
19.1 S03094_pos9555
19.4 S00476_pos42376
19.7 S01313_pos35677
19.9 TP178427
20.4 S04949_pos1888
20.5 S03087_pos21350
20.6 S02087_pos2733
21.8 168-580-10S
22.3 134-900-10S
23.8 S01732_pos20506
24.4 W20-1000-10S
24.7 S01185_pos38018
24.8 S00183_pos17807
25.0 S01512_pos21296
27.0 S01405_pos19487
27.1 S03355_pos11064
27.7 S00535_pos44205
28.4 254-700-10S
31.4 S01154_pos16878
31.7 TP85766
31.9 S01271_pos20410
32.3 S00076_pos36697
32.4 S00610_pos3637
32.7 D06-750-10S
33.3 S04516_pos3081
33.5 S08665_pos7135
33.8 S01905_pos16896
34.0 S00674_pos14192
34.2 S07191_pos1993
34.6 S00864_pos22583
34.7 S00519_pos32680
35.6 S06526_pos1314
35.7 S08221_pos8073
36.1 S14640_pos1017
36.5 S17567_pos1072
36.7 S05890_pos6570
37.3 S00236_pos68476
37.4 S01407_pos489
37.5 S06640_pos5216
37.7 S25586_pos394
38.3 S00635_pos19634 S00905_pos42506
S03416_pos10467
38.4 S00861_pos25702
39.1 S01161_pos8344
39.5 S00068_pos92585
40.0 S02414_pos12730 S00117_pos88070
40.1 S02071_pos10300
40.2 S06969_pos12622
40.8 AC14-900-10S
41.9 S00747_pos11010
42.3 S01748_pos29603
42.7 S03481_pos16618
42.9 S00357_pos4668
43.4 S01406_pos19141
43.7 S14211_pos4652
44.3 S18750_pos592
44.5 S04492_pos11131 S01913_pos52
S07321_pos10029
44.6 S05441_pos15804
44.8 S07742_pos11611
45.1 S02373_pos850
45.4 S00782_pos37180
45.5 S06415_pos9855
45.7 S02843_pos24918
45.8 S28892_pos912
45.9 A12-580-10S
46.0 586-800-10S
46.1 402-1300-10S
46.2 M20-1200-10S
46.3 S03395_pos17112
46.5 S06928_pos1123
46.8 168-1600-10S
47.0 S04799_pos10191
47.4 S23254_pos2168
47.8 105-650-10S
48.1 S25385_pos574
48.2 S10396_pos4280
48.5 S02075_pos205
48.8 S00528_pos14644
49.6 114-1300-10S
51.3 A01-1475-10S
51.6 AC01-840-10S
52.1 S08594_pos4722
52.5 S00210_pos56862
53.1 S16660_pos3796
53.5 S00434_pos266
53.7 S02464_pos20982
54.3 S00531_pos6990
56.6 S00530_pos1671
57.4 S01806_pos27282
58.0 S01021_pos8347
58.3 S25014_pos438
58.8 S00844_pos8357
59.8 S01004_pos47588
61.0 S03803_pos8586
61.5 S03295_pos17430
61.9 140-1500-10S
64.5

(R)


LG10R-A
LG10R-B
LG10R-C
S00255_pos49827
S00100_pos24941
S03281_pos21529
S00904_pos48814
S00536_pos14940
655-875-10R
S34520_pos890
S00255_pos43995
109-730-10R
S00904_pos1829
S34520_pos807
S18910_pos2871
S10362_pos1600
S03043_pos23520
S03777_pos21478
184-950-10R
S00783_pos41054
S04134_pos14541
S00100_pos17155
S05502_pos4850
S03458_pos24928
S00092_pos66502
S04202_pos16412
635-420-10R
S01809_pos19070
126-2000-10R
S00340_pos28751
S01628_pos25880
S00163_pos81083
S04662_pos1258
363-1600-10R
S01628_pos25809
S04389_pos19416
S10270_pos7916
S00361_pos20101
S00361_pos20155
S04892_pos433
S02788_pos7440
S00006_pos179451
S00040_pos11684
101-600-10R
S00040_pos122069
S00233_pos43962
S00006_pos165102
S04342_pos7849
S00040_pos98790
S00233_pos44045 S00040_pos64889
S00416_pos48470
S00588_pos53559
S00416_pos48408
S00128_pos23006
S00006_pos124440
S00006_pos166311
S01880_pos33882
S05300_pos6341 S02313_pos20021
S00030_pos15195
S00233_pos8391
S00048_pos60183
AB09-1490-10R
S00031_pos31899
S00083_pos60465
S00022_pos91884
S00022_pos106419
S00031_pos31835
S00419_pos58898
S01073_pos40381
S00252_pos54781
S00484_pos12579
S00073_pos27035
S10754_pos6443
S00033_pos83127
S00073_pos87892
S10345_pos7952
S00033_pos81376
S10026_pos3541
S01073_pos29967
S00033_pos34958
S00022_pos73551

(R)


LG11S-A
LG11S-B
LG11S-C
LG11S-D
LG11S-E
LG11S-F
LG11S-G
LG11S-H
LG11S-I

(T)

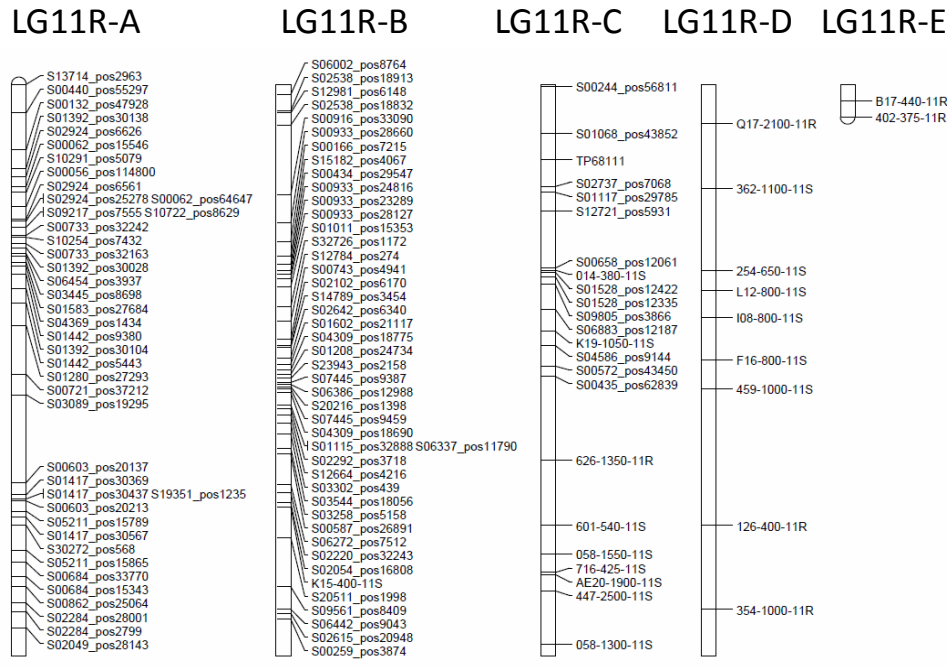


Fig. 3.2. Improved genetic linkage map for hazelnut. Linkage groups (LG) are numbered from 1 to 11. Groups from the EFB-susceptable maternal parent (OSU 252.146), are indicated with an S; those from the EFB-resistant paternal parent (OSU 252.146) are indicated with an R.

The maternal parent, OSU 252.146, is susceptible to EFB, and the paternal parent, OSU 414.062, is resistant to blight. A total of 21,379 high-confidence SNPs were genotyped across the 138 offspring including 2,198 heterozygous SNPs in OSU 252.146 and 1,966 SNPs in OSU 414.146. Maps were constructed using JoinMap 4.1 with an independence LOD score of 10.0. The maternal genetic map (OSU 252.146) has 1,741 GBS markers and 270 SSR/RAPD markers spanning 762 cM across 11 linkage groups (**Table 3.2**). These 11 linkage groups correspond to the haploid chromosome number in hazelnut. Markers are distributed relatively evenly across the map and the number of markers ranges from 118 in LG3 to 296 in LG4. Marker density ranges from an average distance of 0.26 cM in LG4 to 0.53 cM in LG3, and the average distance between markers across the map is 0.38 cM.

The paternal genetic map (OSU 414.146) has 1,368 GBS markers and 178 SSR/RAPD markers spanning 795 cM across 10 linkage groups (**Table 3.3**). This is one less than the haploid chromosome number suggesting that the chromosome corresponding to LG7R and are merged with other linkage groups. LG7R has low SSR marker density in in previously reported maps. Markers are distributed relatively evenly across the paternal map, and the number of markers ranges from 95 in LG9R to 254 in LG4R. The average distance between adjacent markers is 0.51 cM, and the distances range from 0.63 cM in LG2R to 0.47 cM in LG10R. EFB-resistance marker, rest-6R, maps to linkage group 6R between SSR marker AA12-850TAG-6R and a GBS marker on scaffold C.avellana_Jefferson_10749.

Table 3.2. Summary of the hazelnut maternal genetic map statistics

Linkage Group	Total Markers	GBS Markers	SSR Markers	Map Size (cM)	Average Distance (cM)
LG1S	208	180	28	87.273	0.42
LG2S	201	163	38	83.704	0.42
LG3S	118	107	11	62.711	0.53
LG4S	296	264	32	72.79	0.26
LG5S	163	144	19	79.358	0.48
LG6S	174	155	19	74.735	0.43
LG7S	194	167	27	62.438	0.32
LG8S	163	138	25	60.182	0.37
LG9S	175	151	24	56.386	0.33
LG10S	135	109	26	55.462	0.41
LG11S	184	163	21	67.433	0.37
Total	**2011**	**1741**	**270**	**762.472**	**0.38**

Table 3.3. Summary of the hazelnut paternal genetic map statistics

Linkage Group	Total Markers	GBS Markers	SSR Markers	Map Size (cM)	Average Distance (cM)
LG1R	254	227	27	135.146	0.53
LG2R	201	175	26	126.728	0.63
LG3R	178	160	18	85.492	0.48
LG4R	167	149	18	94.656	0.56
LG5R	167	146	21	45.36	0.27
LG6R	150	130	20	84.612	0.56
LG8R	138	121	17	81.735	0.59
LG9R	111	95	16	56.573	0.51
LG10R	180	165	15	84.785	0.47
LG11R	121	100	21	133.553	0.93
Total	**1667**	**1468**	**199**	**908.64**	**0.55**

The previously reported SSR markers are largely collinear in both maps, verifying mapping accuracy. A total of 1,298 scaffolds are represented in the genetic map, but the proportion of mapped scaffolds is too small to produce a chromosome-scale assembly. The high-density genetic map will be useful for

marker-assisted breeding and for the identification of new, desirable traits in hazelnut. The full map is available for download and use on the FTP server hazelnut.mocklerlab.org.

Construction of the hazelnut physical map and genome anchoring. Hazelnut is self-incompatible and cultivars are clones with high levels of residual within genome heterozygosity. Heterozygosity is a major challenge for genome assembly and most sequenced heterozygous genomes are highly fragmented as a result of contig breaks due to mismatches between haplotypes (Michael and VanBuren 2015). The within genome heterozygosity of the hazelnut genome may be as high as ~2% and is at least partially responsible for the low overall contiguity of the assembly. We incorporated a physical map generated by Dr. Shawn Mehlenbacher to improve the genome assembly and overcome issues associated with heterozygosity. The 'Jefferson' hazelnut BAC library consists of 26,752 BACs with an average insert size of 115kb and represents 12x genome coverage (Sathuvalli and Mehlenbacher 2011). The physical map was constructed using capillary electrophoresis based fingerprinting with five restriction enzymes (Mehlenbacher, SA, unpublished data). The estimated map length is 549 Mb in 1,673 contigs and 4,237 singltetons. The map length is 1.4X larger than the estimated genome size, which is likely an artifact of the heterozygosity. The minimum tiling path (MTP) required to span each contig with the least overlap is 5,232 BACs. We used the sequence of each BAC end in the minimum tiling path generated by Chris Saski at Clemson University (USA) to facilitate integration of the physical map with the genome sequence and the genetic map. The 10,464 Sanger based BAC end sequences (BES) have an average read length of 668 bp and ~96% (10,032) match the genome sequence. The BES anchor 6,864 scaffolds collectively span 47% (161 Mb) of the assembly. The mean length of scaffolds matching the BES is 23.5kb preventing merging of the physical map with the draft assembly.

***Comparative genomics within the Rosids*.** Hazelnut belongs to the order Fagales, a large and diverse group with several economically important species including walnut (Juglandaceae family), beech and oak (Fagaceae family), and birch (Betulaceae family). Hazelnut has the first sequenced genome among the Fagales; this map will be useful for comparative genomics in the rosids (Verde *et al.* 2013). Markers from the genetic map were used to assess macro-synteny between hazelnut and peach, which currently has the most complete genome among the rosids (Verde *et al.* 2013). A total of 367 markers from the hazelnut genetic map uniquely to peach and the remaining markers either have ambiguous matches or no matches to peach **(Figure 3.3)**.

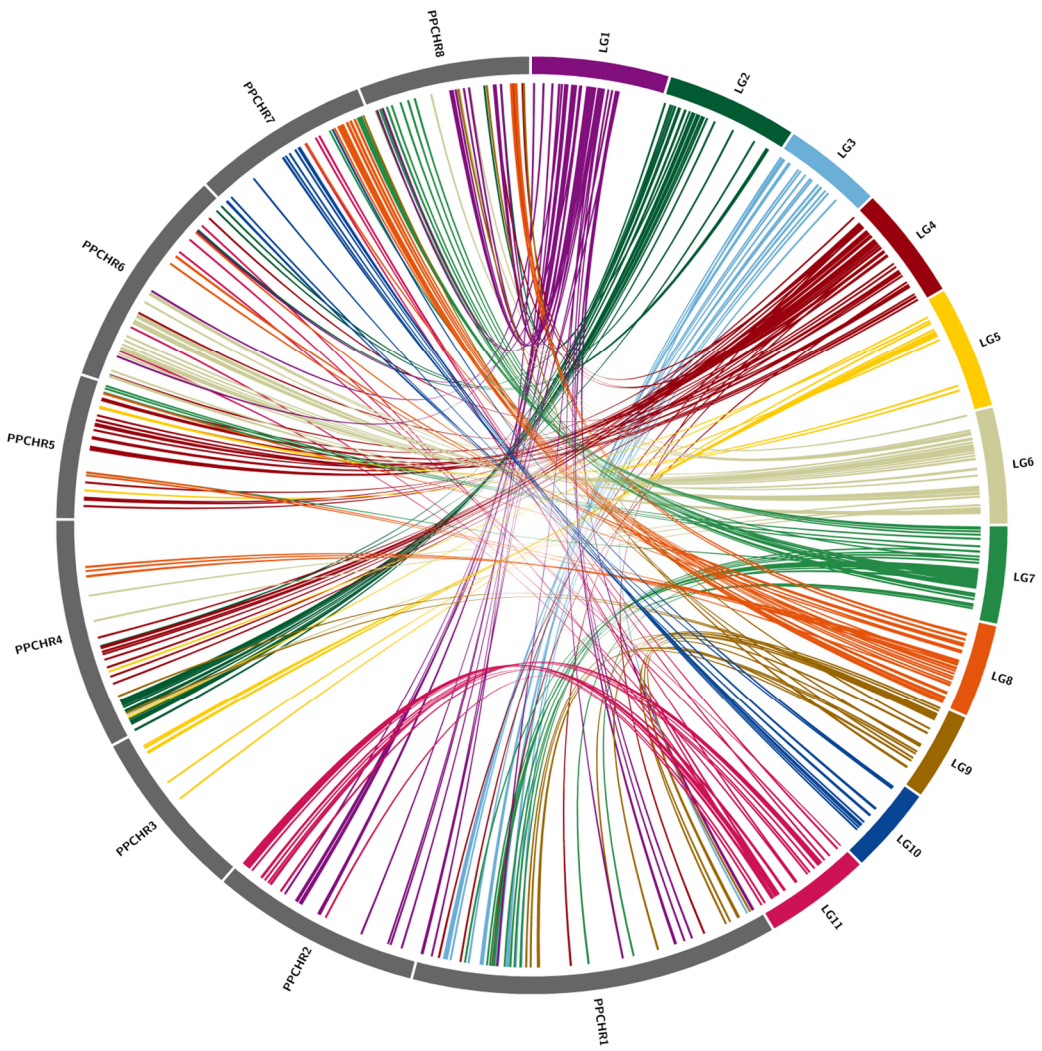


Figure 3.3. Synteny between the peach and hazelnut genomes based on markers from the genetic map. Links are based on the sequences of flanking markers in the hazelnut genetic map and their orthologs in peach. Connections are based on position in the genetic map for hazelnut and physical position in the peach pseudochromosomes.

Hazelnut and peach have 1:1 synteny with no evidence of lineage-specific whole-genome duplication in hazelnut. Numerous structural rearrangements are apparent when the peach and hazelnut genomes are compared. Peach chromosome 3 has 1:1 colinearity with LG5, but the remaining chromosomes have a more complex syntenic relationship reflective of chromosome fusions, breakage, and rearrangements that have occurred since the shared ancestral

karyotype of nine chromosomes (Potter *et al.* 2007). For instance, peach chromosome 1 is collinear to LG3 but also contains regions of LG9 and LG1 from hazelnut.

***Discovery and characterization of putative protein coding genes*.** To detect protein coding loci and assign biological functions to genomic regions, we masked all regions of the 'Jefferson' genome assembly that were repetitive or low complexity using the program RepeatMasker (Smit, AFA, Hubley, R. *RepeatModeler Open-1.0*. 2008-2010 http://www.repeatmasker.org). We then used the *ab initio* gene prediction program AUGUSTUS (Stanke *et al.* 2004) to detect putative protein coding loci within the 36,641 scaffolds and contigs of the assembly. We trained AUGUSTUS with the known gene features of *Arabidopsis thaliana*; both genera are in the Rosid clade, and the *Arabidopsis* genome is extremely well characterized (*Arabidopsis* Genome Initiative 2000). Gene predictions were restricted to whole models only in order to avoid partial gene calls resulting from alternative splicing, incomplete transcripts, and incomplete genes resulting from mis-assembly. The 28,255 ESTs from the European hazelnut 'Jefferson' transcriptome (**Chapter 2**) were used as empirical evidence to guide the predictions, and 94% of the assembled transcripts aligned over 75% of their median length to the assembly. In total, 36,090 putative coding loci were identified and named Corav_g1.t1 through Corav_g36090.t1. Sequence homology has been used to assign gene products to putative amino acid sequences in novel *de novo* assembled genome assemblies; therefore, we aligned the amino acid sequences of the AUGUSTUS output to the NCBI non-redundant protein database using the BLASTP tool (Altschul *et al.* 1997) with an E-value cut-off of $1x10^{-10}$. This homology-based query allowed functional annotation of 23,652 (65.5% of the 36,090 putative loci).

We then identified all loci annotated as transposable elements, retroelements, and gag-polymerases. This removed 2,398 (1,178 unique) loci from both the

annotations and from the putative protein file. These sequences are available for visualization on the hazelnut JBrowse portal (hazelnut.mocklerlab.org/JBrowse) as a separate track, and for query on the BLAST portal (hazelnut.mocklerlab.org). C.avellana_Jefferson_1 was discarded post-filtering, resulting in 34,912 putative protein-coding regions. Of the 23,652 protein sequences mapped to the non-redundant database, 82.5% are represented in the well-annotated, phylogenetically related plant genomes *Vitis*, *Prunus*, *Populus*, and *Ricinus* (**Figure 3.4**).

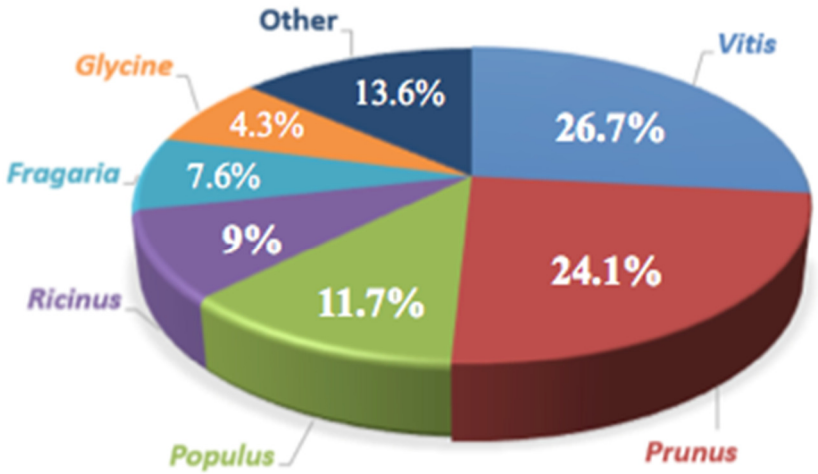


Figure 3.4. Percent of protein-coding sequences in the hazelnut assembly present in the related genera *Vitus*, *Prunus*, *Populus*, and *Ricinus* as determined using BLASTP

We performed functional classification using gene ontology (GO) analysis (Ashburner *et al.* 2000) to survey, categorize, and define the potential properties of gene products with respect to their predicted biological contexts. We used the program BLAST2GO (Conesa *et al.* 2005), a tool for assigning GO terms to unknown sequences, to functionally annotate the filtered set of putative amino acid sequences from the AUGUSTUS output,

resulting in GO functional classifications for 11,221 unique loci (31.8% of total) comprising 65,536 GO terms, broadly grouped by GO component classifications. Among sequences classified by BLAST2GO, 26.9% were assigned to Molecular Function ontology, 18.8% were assigned to Cellular Component ontology, and 51% were assigned to Biological Process ontology, with only 3% lacking assignment to GO classifications (**Figure 3.5**).

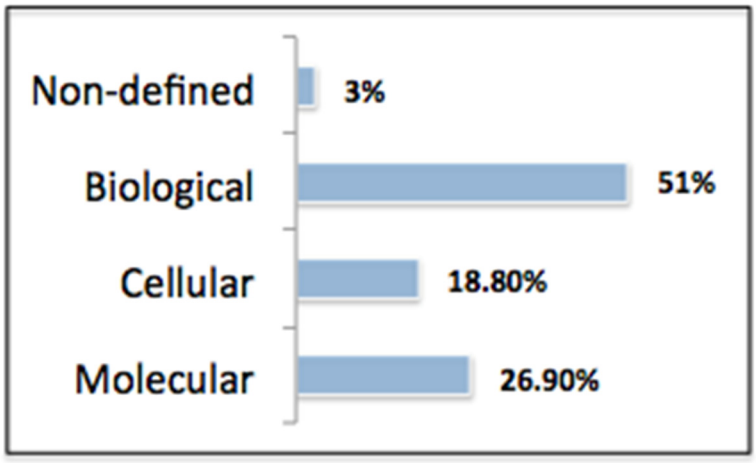


Figure 3.5. Functional gene categories derived from the Gene Ontology (GO) database.

These GO functional annotations are available for download at hazelnut.mocklerlab.org. These resources will be useful for the identification of candidate genes underlying traits of interest, for homology-based comparisons to other tree crops, and for the design of molecular markers to enhance breeding efforts. Here we highlight usefulness of the annotations by profiling several interesting gene families with emphasis on their importance to hazelnut breeding.

***Disease-resistance genes*.** Plants utilize various defense mechanisms to resist attack by pathogens. Resistance genes, known as R genes, control these

mechanisms. R genes confer resistance to specific pathogens, expressing matching avirulence genes in a "gene-for-gene" manner (Flor 1971). The largest class of resistance genes in plants is the NBS-LRR family (McHale *et al.* 2006); the only known functions of the proteins encoded by these genes is in pathogen recognition and defense. The number of NBS-LRR genes varies by species from 57 in cucumber (Wan *et al.* 2013) to ~200 in *Arabidopsis thaliana* (Meyers *et al.* 2003) to over 500 in rice (Bai *et al.* 2002) and in *Medicago truncatula* (Oa *et al.* 2008). There are two functionally distinct subfamilies of NBS-LRR proteins that possessing either a TIR domain or a CC domain upstream of their nucleotide-binding NBS domain. The TIR and CC domains induce distinct downstream response pathways (McHale *et al.* 2006).

In the current annotation of the hazelnut genome there are 115 putative NBS-LRR proteins; 35 (30.4%) contain TIR-domains, 50 (43%) contain CC domains, and 31 (26%) have no subfamily designation. In *A. thaliana* the NBS-LRR sequences (and those of most R genes) occur in clusters of closely related sequences around a parent locus (Baumgarten *et al.* 2003; Leister 2004). Similarly, the hazelnut genome contains occurrences of NBS-LRR genes arranged in clusters on the same contig, as evidenced in **Figure 3.6**.

Figure 3.6. Cluster of loci encoding TIR-domain containing NBS-LRR disease resistance proteins, characteristic of many R genes.

In addition to members of the NBS-LRR family of resistance proteins, there are many other loci encoding annotated proteins implicated in disease resistance. For example, 27 loci encode four members of the 17-member pathogenesis-related protein (PR) family of defense proteins. The PR proteins are induced upon pathogen attack and function as an integral layer of plant systemic acquired resistance (van Loon *et al.* 2008). Many of these proteins have been demonstrated to possess antimicrobial activities. The numbers of these genes and their distribution varies depending on the plant species; these genes have been identified in tobacco (Stintzi *et al.* 1993), tomato (Jongedijk *et al.* 1995), *Arabidopsis* (Kus *et al.* 2002), cucumber (Alkahrani *et al.* 2011), parsley (Somssich *et al.* 1986), radish (Terras *et al.* 1995), and barley (Christensen *et al.* 2002) . Of the 26 putative PR loci in hazelnut, 11 are members of the PR-1 (antifungal) family, 12 are members of the PR-3 (chitinase type I, II, IV, V, VI, VII) family, two belong to the PR-5-like (thaumatin-like) family, one belongs to the PR-10 family, and one belongs to the yet uncharacterized PR-17 family, which shows similarity to zinc metalloproteinases.

Many loci are implicated in fungal resistance. These transcribe genes encoding pti6 transcription factors, which have been shown to bind to the promoters of PR genes, activating expression during fungal attack. Five loci have homology to the blight-associated protein p12, a 12-kDa protein of unknown function that is associated with citrus blight. Hundreds of proteins have homology to proteins implicated in resistance to verticillium wilt, nematode resistance, and other known pathogens. The annotation of the hazelnut genome assembly will be useful in identifying candidate disease resistance genes for future genetic improvement studies. An example is the

'Gasaway' gene, which gives 'Jefferson' resistance to EFB. The current annotation is available for BLAST query at hazelnut.mocklerlab.org.

***Pollen incompatibility S-loci*.** Hazelnuts are monoecious with separate male and female flowers on the same tree. In order to prevent inbreeding and encourage genetic diversity, flowering plants have evolved mechanisms, known as self-incompatibility (SI), to promote outcrossing. In dicots, self-incompatibility is inherited as a segregating unit, with pairs of two or more linked genes that encode the male and female incompatibility determinants known as S-haplotypes (reviewed in Takayama and Isogai 2005).

The family in which hazelnuts reside, Betulacea, exhibits sporophytic incompatibility (SSI), whereby pollen incompatibility is determined by the parents diploid genotype, inherited as a single locus with multiple alleles (Takayama and Isogai 2005). In Brassica there are three genes comprising the core region of the S-locus; the determinant male pollen component is either dominant or codominant and has been shown to be a cysteine-rich pollen coat protein (Doughty et al 1999) encoded by the S-locus cysteine-rich (SCR/SP11). This allele is dominantly expressed in the pollen coat (Shiba *et al.* 2001) and transmitted to SRK via the pollen tube (Kemp and Doughty 2007). The female determinant is codominant and encoded by two inherited polymorphic loci: the stigma localized glycoprotein (SLG), and the S-locus kinase (SRK). SRK is a transmembrane receptor for the male pollen coat protein (Stein *et al.* 1991), with SLG acting as a co-receptor for the male determinant (Takasaki et al 2000).

In angiosperms SI not only provides beneficial self/non-self-determination, it also limits the individuals that may be used in hybrid crosses. For example, if the stigma and pollen express the same allele, SI would render the cross incompatible. This hampers hazelnut-breeding efforts by limiting the parents that can successfully be used for crosses. Fluorescence microscopy is used to determine the compatibility of pollinations, and to identify S-alleles in

hazelnut cultivars used for crosses (Mehlenbacher 1997). To date 33 S-alleles have been identified among hazelnut varieties from several geographical origins (Mehlenbacher 2014).

The current 'Jefferson' genome assembly contains 17 candidates that are annotated as encoding self-incompatibility or S-locus-linked proteins. Of these 17 loci, 5 are annotated as encoding the stigma localized SLG family (**Table 3.4A**), while 2 encode the transmembrane-binding SRK (**Table 3.4B**). The annotation contains 3 loci (**Table 3.4C**) predicted to encode pollen specific S-locus related proteins, which function as the male determinants. The hazelnut annotation also contains an additional set of 9 loci which are annotated as S-locus related proteins (**Table 3.4D**) which may participate in hazelnut SI as well but are currently on different contigs due to the fragmented nature of the assembly.

Table 3.4A. Putative stigma localized glycoprotein (SGL) encoding loci

Locus	**Annotation**
g6132.t1	S-locus-specific glycoprotein s6
g10567.t1	S-locus-specific glycoprotein s6
g25571.t1	S-locus-specific glycoprotein s13
g34002.t1	S-locus-specific glycoprotein s6
g35241.t1	S-locus-specific glycoprotein s6

Table 3.4B. Putative S-locus kinase (SRK) encoding loci

Locus	**Annotation**
g26148.t1	S-locus receptor partial
g33469.t1	S-locus lectin protein kinase family

Table 3.4C. Putative pollen expressed self-incompatibility loci

Locus	Annotation
g4284.t1	S1 self-incompatibility locus-linked pollen protein
g6376.t1	S3 self-incompatibility locus-linked pollen expressed
g28441.t1	S3 self-incompatibility locus-linked pollen 3.15 protein isoform 1

Table 3.4D. Putative self-Incompatibility and related loci

Locus	Annotation
g13703.t1	Self-incompatibility protein
g15816.t1	Self-incompatibility s-locus f-box partial
g16995.t1	S-locus f-box brothers-like protein
g28580.t1	Self-incompatibility protein
g34070.t1	S-locus f-box brothers-like protein
g35699.t1	S-locus linked f-box protein type-5

The identification of the S-alleles that controls pollen-stigma incompatibility would allow development of PCR primers to quickly screen new hazelnut accessions for compatibility prior to crossing, thus avoiding incompatible crosses, enhance marker assisted selection, identifying the s-locus in hazelnut.

***Paclitaxel biosynthesis precursor molecules*.** Paclitaxel is a plant alkaloid used in the treatment of various types of cancers. It acts as an antimicrotubule agent, slowing cell division to prevent chromosomal segregation and induce apoptosis (Yvon, Wadsworth, and Jordan 1999). Paclitaxel was originally identified in the bark of Pacific yew (Taxus brevifolia) and is marketed as a chemotherapeutic compound under the trade name Taxol. Paclitaxel has historically been laborious and costly to produce; initially developed synthetic routes required precursor molecules from bark of Taxus, a genus that is slow to mature and difficult to cultivate. Currently paclitaxel is largely produced through cell culture methods, which, although vastly superior to

procedures requiring Taxus bark, are time and labor intensive (Tabata 2004). Recently, organisms other than yew, such as endophytic fungi, have been found to synthesize paclitaxel (Yang *et al.* 2014). The presence of paclitaxel and related taxanes was discovered in the stems of the *C. avellana* cultivar 'Gasaway' (Hoffman *et al.* 1998), as well as in extracts from the leaves and shells of the cultivar 'Tombul' (Hoffman and Shahidi 2009). A similar biochemical pathway likely exists in hazelnut, perhaps through varying intermediates, which results in the synthesis of the active form of the terpenoid paclitaxel. A recent study suggests that the non-pollen allergic reaction to hazelnut may be a reaction to the natural occurring taxolic components native to some hazelnut cultivars rather than a nut-protein allergy, and a cancer patient with a pre-existing reaction to hazelnuts displayed similar symptoms upon treatment with taxol (Bukacel, D. G., Bander, R., and Ibrahim 2007).

Given that hazelnut trees mature more rapidly than yew and are cultivated worldwide, it is possible that certain hazelnut cultivars could be used to produce the needed precursor molecules for paclitaxel production in a more cost-effective manner than those relying in *Taxus* sp. The renewable nature of hazelnut also makes it an attractive candidate for paclitaxel production, possibly through bioengineering of the pathway. Such a feat requires a greater understanding of how paclitaxel synthesis is achieved in the hazelnut system. The first step is identification of genes encoding the necessary precursor molecules for the pathway. Genes encoding known paclitaxel precursor molecules are present in the 'Jefferson' genome assembly. An example is geranylgeranyl diphosphate (GGPP). GGPP is the initial metabolite in the paclitaxel biosynthetic pathway; GGPP biosynthesis is catalyzed by geranylgeranyl diphosphate synthase (GGPPS) (Hefner *et al.* 2010). Recently a cDNA encoding GGPPS (CgGGPPS) was cloned from the hazelnut cultivar 'Gasaway' and demonstrated to have high levels of tissue-specific expression in leaves (Wang *et al.*2010). Primer sequences used to

amplify CgGGPPS aligned to the 'Jefferson' assembly with 100% identity to genomic regions flanking Corav_g32121.t1, a putative geranylgeranyl pyrophosphate synthase (a synonym for GGPPS). A query of the NCBI non-redundant protein database using the BLASTP tool with an E-value cut-off of $1x10^{-10}$, demonstrated 100% alignment of g32121.t1 to GenBank accession ABW06960.1, the 373 amino acid long CgGGPPS. This analysis validates both the gene predictions and the functional annotations of the hazelnut genome. Southern blot analysis (Wang *et al.* 2010) revealed that CgGGPPS is part of a larger gene family in hazelnut of at least three members. This empirical determination was enhanced by our annotation of the 'Jefferson' genome, which contains four putative genes annotated as members of the GGPPS family (**Table 3.5A)**.

Another key molecule necessary for the synthesis of taxol is *N*-debenzoyl-2'-deoxypaclitaxel *N*-benzoyltransferase (NDTBT). NDTBT catalyzes the final step in the paclitaxel biosynthetic pathway, and has the potential to be engineered to produce paclitaxel more efficiently (Nevarez *et al.* 2009; Walker *et al.* 2002). There are six loci in the 'Jefferson' genome whose gene products are annotated via homology as members of the NDTBT family (**Table 3.5B**).

Table 3.5A. Geranylgeranyl Pyrophosphate Synthase encoding (GGPPS) encoding loci

Locus	**Annotation**
g32121.t1	geranylgeranyl pyrophosphate synthase
g33309.t1	geranylgeranyl pyrophosphate synthase-related protein
g35263.t1	geranylgeranyl diphosphate synthase
g35966.t1	geranyl-diphosphate synthase

Table 3.5B. *N*-debenzoyl-2-deoxypaclitaxel *N*-benzoyltransferase (NDTBT) encoding loci

Locus	Annotation
g4324.t1	3-n-debenzoyl-2-deoxytaxol n-
g14144.t1	3-n-debenzoyl-2-deoxytaxol n-benzoyltransferase-like
g28826.t1	3-n-debenzoyl-2-deoxytaxol n-benzoyltransferase-like
g30539.t1	3-n-debenzoyl-2-deoxytaxol n-
g31890.t1	3-n-debenzoyl-2-deoxytaxol n-
g33250.t1	3-n-debenzoyl-2-deoxytaxol n-

Thus, the draft 'Jefferson' genome assembly establishes the necessary resources for future research and ultimate elucidation of the paclitaxel biosynthetic pathway in hazelnut. These findings highlight the usefulness of the current annotation in facilitating the identification of candidate genes involved in the synthesis of economically important molecules such as chemotherapeutics.

Resequencing and polymorphism detection in additional hazelnut cultivars. An overarching goal in plant breeding is to correlate variations in genomic sequence with traits of interest. Variations in DNA sequence can occur via single-nucleotide substitutions (SNPs) or by small-scale insertions and deletions (INDELs), ranging in size from a single to hundreds of nucleotides. The vast majority of polymorphisms occur in the less-conserved and non-coding (intergenic) regions rather than protein-coding exonic regions (Castle 2011). Although these polymorphisms do not affect coding potential, both occurrences have been shown to be under selective pressure and may be related to agronomic traits (Ometto *et al.* 2005).

SNPs and INDELs are powerful molecular markers, due in part to their abundance and relative ease of detection in a genome-wide high-throughput experiment (Mammadov *et al.* 2012). In simple sequence repeat (SSR) marker discovery, an allele is scored as polymorphic by resolving the

amplified sequence on an capillary electrophoresis (Powell *et al.* 1996). SNPs and INDELs are biallelic and co-dominant; however, the probability of two independent base changes occurring at a single position is quite low (Vignal *et al.* 2002). Quantitation of SNP and INDEL abundance allows construction of genetic maps and functional markers (Mammadov *et al.* 2012; Miller *et al.* 2011; Ryynanen *et al.* 2007; Xiong and Jin 1999).

We implemented this method to identify polymorphisms and explore sequence diversity among seven hazelnut cultivars representing four major geographical regions (Boccacci et al 2006, Gökirmak *et al.* 2009) of hazelnut production (**Table 3.6**).

Table 3.6. 7 European hazelnut cultivars chosen for re-sequencing

Cultivar	Origin	EFB Resistance	Description
Barcelona	Spain	Intermediate	Important in Oregon
Ratoli	Spain	Resistant (Single Gene)	EFB Resistant
Daviana	England	Susceptible	Pollinizer for Barcelona
Tonda di Giffoni	Italy	Resistant (quantitative)	Excellent kernel quality
Tonda Gentile delle Langhe	Italy	Susceptible	Excellent kernel quality
Tombul (Extra Ghiaghli)	Black Sea	Susceptible	Important in Turkey
Halls Giant	Germany / France	Intermediate	Pollinizer for Barcelona

From the Spanish-Italian group: ‘Barcelona’ from Spain; accounting for 60% of the hazelnut trees in Oregon and moderately susceptible to EFB, 'Tonda Gentile delle Langhe' from northern Italy, with excellent kernel quality and high susceptibility to EFB, ‘Tonda di Giffoni’ from southern Italy, which also has excellent kernel quality and high quantitative resistance to EFB, and ‘Ratoli’ from eastern Spain, which is highly resistant to EFB. ‘Daviana’ represents the English group, a pollinizer for ‘Barcelona’ and highly susceptible to EFB. Representing the Central European group is ‘Halls Giant’, a universal pollinizer which is cold-hardy and moderately resistant to EFB. Finally, from the Black Sea region: ‘Tombul (Extra Ghiaghli)’, a clone of the Turkish cultivar 'Tombul' which is susceptible to EFB.

We collected whole leaf tissue from individual trees of each of the seven chosen cultivars grown at the Smith Horticulture Research Farm in Corvallis, Oregon. High-quality gDNA was extracted and 250-bp PE libraries were constructed and uniquely barcoded. The resulting unique reads from each cultivar were aligned to the ‘Jefferson’ reference genome with BWA (Li and Durbin 2009) using default parameters and polymorphisms were identified using the GATK pipeline (DePristo *et al.* 2011). In order to increase confidence in polymorphism calls, the SNP and INDEL predictions from GATK were further filtered by implementing the SnpSift component of the SnpEff package (Cingolani *et al.* 2012) requiring 15 overlapping reads and a quality score (Q) of 30 for the SNPs and 20 for the INDELS. We used the program SnpEff for genome-wide variant annotation and to predict the effects of each polymorphism on the coding potential of the putative loci.

Functional consequences of polymorphisms in protein coding loci. The resequencing of seven hazelnut cultivars and alignment to the ‘Jefferson’ reference has allowed for the genome-wide association of millions of polymorphisms with the protein coding potential of hundreds of loci (**Tables 3.7 and 3.8**).

Table 3.7. Total SNPs and predicted effect on coding potential

Cultivar	High	Moderate	Low	UTR	None	Homozygous	Heterozygous
Barcelona	9,892	9,697	139,197	84,394	2,199,855	359,083	2,078,254
Ratoli	10,129	9,924	140,181	87,094	2,285,443	501,851	2,023,803
Daviana	11,839	10,848	159,328	95,462	2,506,522	461,186	2,058,973
Tonda di Giffoni	9,823	9,787	135,493	86,395	2,286,425	751,549	2,023,136
Tonda Gentile delle Langhe	7,464	7,874	110,584	71,156	1,804,165	379,112	1,615,409
Tombul (Extra Ghiaghli)	3,723	4,351	57,213	39,730	1,067,644	189,527	979,701
Halls Giant	11,475	11,045	157,625	96,485	2,545,048	763,661	2,047,536

Table 3.8. Total INDELs and predicted effect on coding potential

Cultivar	High	Moderate	Low	UTR	None	Homozygous	Heterozygous
Barcelona	294	1,310	0	15,356	321,115	45,612	285,901
Ratoli	330	1,287	0	15,801	337,077	66,222	280,171
Daviana	385	1,527	0	17,675	375,356	102,357	281,699
Tonda di Giffoni	305	1,314	0	15,634	336,697	58,957	285,876
Tonda Gentile delle Langhe	227	1,067	0	12,530	260,172	46,596	221,003
Tombul (Extra Ghiaghli)	117	596	0	6,845	154,162	24,768	134,123
Halls Giant	354	1,530	0	17,812	379,194	98,933	287,029

Table 3.9. Unique loci containing SNPs and predicted effect on coding potential

Cultivar	High	Moderate	Low	UTR	None	Total Loci Hit	Percent of Loci Hit	Unique High
Barcelona	6,942	4,224	14,160	4,610	2,026	31,962	94.6	513
Ratoli	6,970	4,163	13,559	4,716	2,171	31,579	93.5	563
Daviana	7,822	4,292	13,955	4,266	1,853	32,188	95.3	424
Tonda di Giffoni	6,757	4,232	13,270	5,173	2,099	31,531	93.3	494
Tonda Gentile delle Langhe	5,422	3,841	13,055	5,430	2,711	30,459	90.2	940
Tombul (Extra Ghiaghli)	2,838	2,489	9,153	6,131	5,039	25,650	75.9	923
Halls Giant	7,595	4,399	13,701	4,554	1,749	31,998	94.7	216

Table 3.10. Unique loci containing INDELs and predicted effect on coding potential

Cultivar	High	Moderate	Low	UTR	None	Total Loci Hit	Percent of Loci Hit	Unique High
Barcelona	285	1,198	0	9,845	10,857	22,185	65.7	67
Ratoli	321	1,177	0	10,015	10,728	22,241	65.8	83
Daviana	371	1,374	0	10,823	10,706	23,274	68.9	58
Tonda di Giffoni	299	1,205	0	9,958	10,653	22,115	65.5	79
Tonda Gentile delle Langhe	223	988	0	8,364	10,610	20,185	59.8	126
Tombul (Extra Ghiaghli)	115	569	0	4,963	9,071	14,718	43.6	124
Halls Giant	346	1,388	0	10,824	10,635	23,193	68.7	34

Hundreds of protein-coding loci with functions potentially altered or disrupted by the introduction of either a SNP or INDEL were identified **(Tables 3.9 and 3.10)**.

The discussion of several of these polymorphisms below seeks to underscore the value of these predictions in candidate discovery and hypothesis generation for future experiments. The complete annotated variant files for each cultivar used in this study are hosted on the FTP server at hazelnut.mocklerlab.org, available for both download and query.

Each of the seven resequenced cultivars contains dozens of polymorphisms in putative disease resistance genes; here we focus on annotated variants unique to each cultivar. As mentioned previously, defined protein sequences display homology to the citrus blight associated "blight-associated protein p12" family, currently of unknown function (Ceccardi *et al.*1998). The English cultivar 'Daviana', which is susceptible to EFB, has a 1-bp INDEL that introduces a premature stop codon (PSC) in the putative gene Corav_g17836.t1, an annotated member of the "blight-associated protein p12" family. Inoculation of *Populus* leaf tissue with fungal rust showed an increase in apoplast transcript levels of "blight-associated protein p12 family" homologues, suggesting an role for this unclassified protein family in response to biotic attack (Pechanova *et al.* 2010). 'Daviana' also has a SNP in the "putative class III acidic chitinase" encoding locus Corav_g5616.t1 **(Figure 3.7)**.

Figure 3.7. SNP in "putative class III acidic chitinase" in the EFB-susceptible 'Daviana' that introduces a mutation in the splice site acceptor region, possibly increasing susceptibility to fungal attack.

Plant chitinases play a role in defense from fungal pathogens by interrupting vegetative growth of fungal hyphae (Punja and Zhang 1993). Overexpression of plant chitinases, often in combination with PR proteins (Cletus *et al.* 2013), have led to improved fungal resistance in many crop systems. The polymorphism in locus Corav_g5616.t1 introduces a mutation into a splice site acceptor region relative to contig C.avellana_Jefferson_00575; this may result in a mis-spliced transcript destined for degradation prior to translation. It is possible that these variants in resistance genes increase the susceptibility of 'Daviana' to fungal attack, such as infection with EFB.

Transcript levels of thaumatin-like proteins, which have been previously linked to host resistance to pathogens, are also increased in apoplasts of *Populus* after fungal attack (Pechanova *et al.* 2010). There are 22 loci predicted to encode thaumatin-like proteins in the 'Jefferson' genome; four of these contain polymorphisms that are annotated as having HIGH effects in both of the hazelnut cultivar with intermediate levels of resistance to EFB. The Spanish accession 'Barcelona', the heirloom accession grown throughout much of Oregon, has an INDEL resulting in the gain of a PSC in the thaumatin-like protein 1-like encoding locus g19010.t1, located on contig C.avellana_Jefferson_03950 and in the putative thaumatin-like protein encoding locus Corav_g34507.t1. The French cultivar 'Halls Giant' has SNPs that introduce PSCs in Corav_g19009.t1 and Corav_g25197.t1, both are putative genes annotated as encoding thaumatin-like proteins. It is possible the levels of resistance (or susceptibility) to EFB shown by the resequenced hazelnut cultivars from unique geographical origins (**Table 5**) are due in part to a combination of loci. These computational predictions will be useful in selecting candidate genes for future molecular experiments.

Three of the resequenced cultivars are fully susceptible to EFB: the English cultivar 'Daviana', the Italian cultivar 'Tonda Gentile delle Langhe', and the Turkish cultivar 'Tombul (Extra Ghiaghli)'. Uniquely shared among these

cultivar are 12 loci containing SNPs that introduce HIGH effects. Two of these loci are known disease resistance genes: Corav_g5237.t1, a TIR-NBS-LRR resistance protein and Corav_g34429.t1, a member of the "blight-associated protein p12 precursor" proteins discussed above. Two of the five members of this family in the hazelnut annotation have polymorphisms in the EFB-susceptible cultivars that are predicted to alter coding potential.

Detrimental variants also are predicted to occur in protein coding regions homologous to those yielding agriculturally important traits. In the cultivars 'Daviana' and 'Tonda Gentile delle Langhe' there are SNPs in four putative cellulose synthase genes; these SNPs are predicted to result in truncated proteins. Cellulose fibers are broken down into glucose, which may be later converted into biofuels such as ethanol. The digestion of cellulose is one of the many hurdles that must be overcome before biofuel production is efficient, and understanding the mechanisms of cellulose synthesis and degradation will guide production enhancement efforts (Sticklen 2008). It is possible that the nucleotide variations in these cultivars of hazelnut result in reduced amounts of cellulose accumulation within cell walls, which may lead to more effective digestion of cellulose.

The primary cause of the allergic reaction to hazelnut is tree pollen; this immune response is known as oral allergy syndrome (OAS) (Schocker *et al.* 2000; Jiménez-lópez *et al.* 2010). Proteins in pollen that contain the Ole domain are the major class of plant-derived allergens that affect humans (Jiménez-lópez *et al.* 2010; Florido *et al.* 2002). Originally identified in olives (*Olea europaea* L.), these proteins share epitopes with proteins in many other plant orders, including Betulaceae, and are strongly implicated as the causal agent of allergies to pollen, latex, and certain fruit (Barral *et al.* 2015). Proteins containing the Ole domains are involved in the formation of the pollen tube (Barral *et al.* 2015). There are eight members of the "Pollen Ole e1 allergen and extension" family, the major allergen from olive

(Jiménez-lópez *et al.* 2010), in the current annotation of the hazelnut genome. One of these, Corav_g2909.tl (located on the contig C.avellana_Jefferson_00226), has a 31-bp deletion in the resequenced Turkish cultivar 'Tombul'. This deletion leads to the gain of a stop codon, resulting in a truncated protein. Characterization of allergens in hazelnut may lead to greater understanding and of intra-species cross-reactions and further the development of new diagnostic tools for OAS in humans.

Perhaps one of the most interesting groups of genes with annotated variants is that of the putative SI genes. Of the 17 loci in the hazelnut annotation annotated as participating in SI, five of the SSI related loci have SNPs that are highly likely to impact protein production (**Table 3.11**).

Table 3.11. Variants in S-loci annotated as affecting coding potential

Locus	Annotation	Barcelona	Daviana	Halls Giant	Ratoli	Tonda GdL	Tombul	Tonda di Giffoni
g6132.t1	S-locus-specific glycoprotein s6	SNP	SNP	SNP	SNP	SNP	SNP	SNP
g6376.t1	S3 self-incompatibility locus-linked pollen expressed	SNP	SNP	SNP	SNP	SNP	SNP	SNP
g10567.t1	S-locus-specific glycoprotein s6	SNP	SNP	SNP			SNP	
g28441.t1	S3 self-incompatibility locus-linked pollen 3.15 protein isoform 1		SNP	SNP	SNP	SNP	SNP	

These variants may underlie the genetic SI mechanism, which leads to the incompatibility phenotype observed in hazelnut. For example, Corav_g28441.t1, a putative male determinant "S3 self-incompatibility locus-linked pollen 3.15 protein isoform 1" encoding locus, has SNPs in the cultivars 'Daviana', 'Halls Giant', 'Ratoli', and Tonda Gentile de Langhe' **(Figure 3.8)**, leading to a PSC.

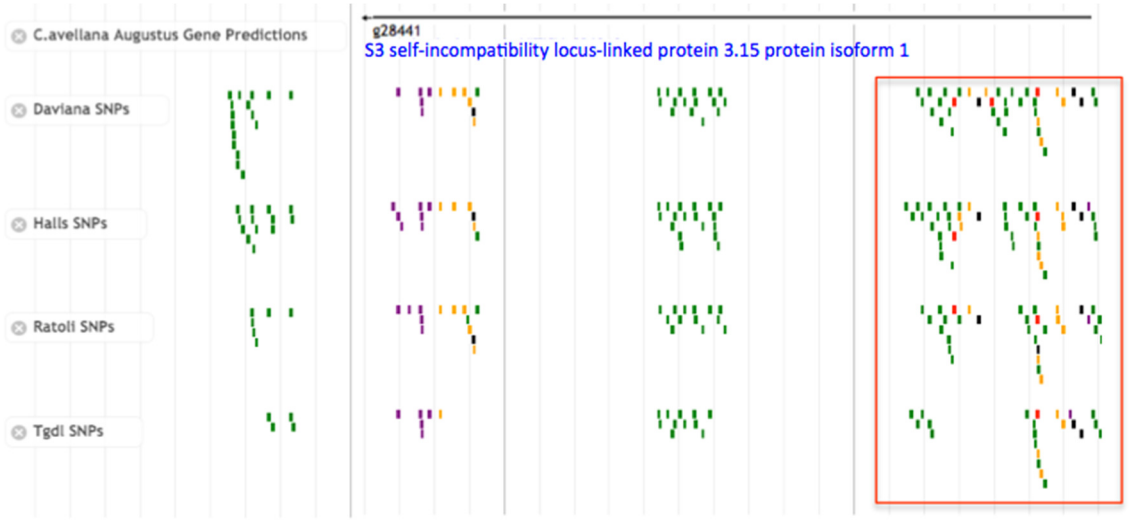


Figure 3.8. 'Daviana', 'Hall's Giant', 'Ratoli', and 'Tonda Gentile de Langhe' contain PSC-introducing SNPs in a putative pollen-expressed male determinant SI-related locus.

Discussion:

The association of genes with desirable traits has long been a goal of plant breeding, and having an annotated and characterized genome assembly is the first step in realizing this goal. We have sequenced, assembled, and characterized the draft genome of the European hazelnut (Corylus avellana L.) cultivar 'Jefferson' and resequenced seven additional cultivars. Hazelnut is the first sequenced plant in the order Fagales which includes the nut trees walnut and pecan and several economically important wood species including oak, beech and hickory. Gene prediction and functional annotation of protein-coding loci allowed identification of agriculturally relevant loci and will be extremely useful for future molecular characterization and marker discovery. Variations in DNA sequence between 'Jefferson' and resequenced cultivars range in size from single to hundreds of nucleotides. Predicting the effects of these variations on the coding potential of gene loci are an integral step in the identification of genes underlying traits of interest and hypothesis generation for future molecular experiments. The establishment of PCR markers corresponding to genes of interest, such as those controlling the pollen-stigma incompatibility, can be used to rapidly screen additional cultivars to determine the presence of the gene of interest prior to using as parents in breeding experiments. In addition, progeny of these crosses can also be screened via PCR to determine whether they possess the desired trait.

Disease resistance is often involves several loci and pathways; the resistance and susceptibility to EFB displayed in certain cultivars (Table 5) is highly unlikely to involve the same genes. For example, the 'Gasaway' gene is not present in the resequenced EFB resistant Spanish cultivar 'Ratoli', and the sources of resistance map to different linkage groups (Mehlenbacher *et al.* 2006; V. R. Sathuvalli *et al.* 2011). Novel sources of resistance likely arose independently in response to pathogen attack in the cultivars native regions;

establishment of ‘Jefferson’ as the reference hazelnut cultivar will aid in unraveling the genes responsible for disease resistance. Once new sources of resistance are identified, it will be possible to further enhance resistance by stacking the traits of multiple cultivars through several rounds of selective breeding. It may be possible to utilize wild accessions as parents in crosses; the discovery of new sources of EFB resistance in American hazelnut would allow the creation of American x European hybrids.

As ‘Jefferson’ is the F1 progeny of a clonal cross, it is heterozygous throughout its genome and within alleles. This, along with the within genome heterozygosity demonstrated by the bimodal K-mer distribution, contributed to the fragmented nature of the assembly; the short-read sequencing technology and programs used for genome assembly were unable to resolve the more heterozygous regions of the genome, resulting in breaks in the assembly and generation of many smaller contigs. Despite the fragmented contigs, the initial assembly of the hazelnut genome captures over 90% of the hazelnut genome and has been used to identify over 700 new SSR markers (S.A. Mehlenbacher, personal communication, April 2016). The current draft assembly will be useful for the alignment of future physical markers and sequenced BAC libraries, the identification of additional molecular markers, and alignment of RNA-seq reads from experiments that, for example, survey differential gene expression from incompatible and compatible pollinations over time.

At the writing of this manuscript, the ‘Gasaway’ allele conferring resistance to EFB has not yet been identified. However, Sathuvalli *et al.* (2011, 2013) generated BAC libraries and identified polymorphic RAPD markers linked to EFB resistance. Alignment of these BAC end markers to the ‘Jefferson’ assembly resulted in matches of 100% identify across their length (data not shown), but the fragmented nature of the assembly and small contigs did not allow for the full resolution of the downstream sequence nor positive

identification of the gene. This is not surprising given the fact markers may be located megabases away from the loci they segregate with; such resolution is not possible given the current fragmented assembly.

GBS analysis of the hazelnut F1 mapping population resulting from the cross of the EFB-susceptible maternal parent OSU 252.146 and the resistant paternal parent OSU 414.062 added an additional 3,209 markers to the existing genetic linkage map (Mehlenbacher *et al.* 2006). This genetic linkage map represents a 5-fold improvement over the existing map. The average distance between markers in the current map is 0.38 cM. This resolution allows genome-wide analyses of molecular variation and additional marker discovery to accelerate molecular breeding in hazelnut. For example, the EFB-resistance in 'Gasaway' was previously assigned to LG6 (Mehlenbacher *et al.* 2006), and the separate gene responsible for resistance in the Spanish cultivar 'Ratoli' segregates on LG7 (Sathuvalli *et al.* 2011a). In addition, the recently characterized accession OSU 759.010 transmits resistance via a locus on LG2 (Sathuvalli *et al.* 2011b) different from locations of other identified resistance loci. The variants discovered and characterized in this study, including the new GBS-derived markers placed on the genetic linkage map, will be useful for identifying candidate genes responsible for EFB resistance and other genetic factors that reduce yield. The 'Jefferson' reference hazelnut genome, annotations, and other resources are available to the public as a queryable BLAST portal (hazelnut.mocklerlab.org). An interactive Jbrowse interface (www.hazelnut.mocklerlab/org/Jbrowse) allows visualization of the annotations, alignments, and polymorphisms for each cultivar in separate overlayable tracks. Variant calls for each of the seven resequenced cultivar annotated with their predicted functional effect are available for download on the FTP server at hazelnut.mocklerlab.org. This server also hosts the current high-density GBS-derived genetic linkage map and GO terms.

Author contributions:

ERR and SAM collected the hazelnut tissues. ERR extracted the DNA. DWB conducted the sequence assembly. ERR conducted the sequence analysis and wrote the manuscript. RV conducted the GBS analysis and contributed to writing the manuscript. HDP developed the web interfaces and conducted the data hosting. SAM and TCM conceived of the study, participated in its design and coordination, and provided funding. All authors read and approved the final manuscript.

Acknowledgements:

We would like to thank the Georgia Genomics Facility at the University of Georgia for the preparation of Illumina libraries; Anne-Marie Girard, Caprice Rosato, Mark Dasenko, and Mathew Peterson for qualitative assessment of the libraries, Illumina cluster generation, and computational support (Center for Genome Research and Biocomputing, Oregon State University); and scientists at MOgene (St. Louis) for construction and sequencing of the GBS libraries; and Chris Saski at the Clemson University Genomics Unit for providing information on the minimal tiling path of BACs and BAC-end sequences. This work was supported by the Oregon State University Agricultural Research Foundation, the Oregon Hazelnut Commission, and the Donald Danforth Plant Science Center.

Literature cited:

Altschul, S., Madden T, Schaffer A., Zhang J, Zhang J, Miller, W, Lipmann 1997. Gapped BLAST and PSI- BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

Ameline-torregrosa, C, Wang B, O'Bleness, M. S., Deshpande, S., Zhu, H., Roe, B., Young, N.D., Canon, S.B. 2008. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiol 146:5–21.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815.

Ashburner, M.C., Ball, J., Blake, D., Botstein, H., Butler, J.M., Cherry, P., Davis. 2000. Gene Ontology: Tool for the Unification of Biology. Nature Genetics 25:25–29.

Bai, J., Pennill, L.A., Ning, J., Lee, S.W., Ramalingam, J., Webb, C.A., Zhao, B. 2002. Diversity in nucleotide binding site leucine-rich repeat genes in cereals. Genome Res 12:1871–84.

Barral, P., Batanero, E., Palomares, O. 2015. A major allergen from pollen defines a novel family of plant proteins and shows intra- and interspecie cross-reactivity. J Immunol 172:3644-51.

Baumann, P., Baumann, L., Lai, C.Y., Rouhbakhsh, D., Moran, N.A., Clark, M.A. 1995. Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. Annual Review of Microbiology 49:55–94.

Baumgarten, A., Cannon, S., Spangler, S., May, G. 2003. Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics 319:309–319.

Boetzer, M., Henkel, C.V., Jansen, H., Butler, D. 2011. Scaffolding pre-assembled contigs using SSPACE summary. Bioinformatics 27:578-579.

Bukacel, D.G., Bander, R., Ibrahim, R.B. 2007. Cross-reactivity between paclitaxel and hazelnut: A case report. J Oncol Pharm Pract 13:53–55.

Castle, J.C. 2011. SNPs occur in regions with less genomic sequence conservation. PLoS ONE. dx.doi.org/10.1371/journal.pone.0020660.

Ceccardi, T.L., Barthe, G.A., Derrick, K.S. 1998. A novel protein associated with citrus blight has sequence similarities to expansin. Plant Mol Biol 38:775-83.

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., Lu, X. 2012. Using Drosophila melanogaster as a model for genotoxic chemical mutationalstudies with a new program, SnpSift. Frontiers in Genetics 3:1–9.

Cletus, J., Balasubramanian, V., Vashisht, D., Sakthivel, N. 2013. Transgenic expression of plant chitinases to enhance disease resistance. Biotechnol Lett 35:1719–1732.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–7366.

DePristo, M., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A., Angel, G., Rivas, M.A., Hanna, M., McKenna, A.,

Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J. 2011. A framework for variation discovery and genotyping using next-deneration DNA sequencing data. Nature Genetics 43:491.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E. 2011.) A robust, simple genotyping-by-bequencing (GBS) approach for high diversity species. PLoS ONE 6:1–10.

Flor, H.H. 1971. Current status of the gene-for-gene concept. Annu. Rev. Phytopathol 9:275–76.

Florido, L., Enriquez, J.Q., Arias de Saavedra, J.M., Saenz de San, Pedro, B., Casañez, E.M. 2002. An allergen from *Olea europaea pollen* (Ole E 7) is associated with plant-derived food nnaphylaxis. Allergy 57:53–59.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., Buckler, E.S. 2014. TASSEL-GBS : A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:e90346.

Gökirmak, T., Mehlenbacher, S.A., Bassil, N.V. 2008. Characterization of European hazelnut (*Corylus avellana*) cultivars using SSR markers. Genetic Resources and Crop Evolution 56:147–72.

Gürcan, K., Mehlenbacher, S.A., Erdoğan, V. 2010. Genetic diversity in hazelnut (*Corylus avellana* L.) cultivars from Black Sea countries assessed using SSR markers. Plant Breeding 129:422–34.

Hefner, J., Ketchum, R.E., Croteau, R. 1998. Cloning and functional expression of a cDNA encoding geranylgeranyl diphosphate synthase from *Taxus canadensis* and assessment of the role of this prenyltransferase in cells induced for taxol production. Arch Biochem Biophys 360:62–74.

Hoffman, A., Khan, W., Worapong, J., Strobel, G., Griffin, D., Arbogast, B., Barofsky, D., Ning, L., Zheng, P., Daley, L. 1998. Bioprospecting for taxol in angiosperm plant extracts. Spectroscopy 13:22–32.

Hoffman, A. and Shahidi, F. 2009. Paclitaxel and other taxanes in hazelnut. Journal of Functional Foods. 1:33–37.

Jiménez-lópez, J.C., Carlos, J., Rodríguez-garcía, M.I., Alché, J.D. 2010. Systematic and phylogenetic analysis of the Ole E 1 pollen protein family members in plants. Systems and Computational Biology - Bioinformatics and Computational Modeling, InTech publishing ISBN:978-953-307-875-5.

Johnson, K.B., Pinkerton, J.N., Mehlenbacher, S.A., Stone, J.K., Pscheidt, J.W. 1996. Eastern filbert blight of European hazelnut: It's becoming a manageable disease. Plant Dis. 80:1308–16.

Jongedijk, E., Tigelaar, H., van Roekel, J.S.C., Bres-Vloemans, S.A., Dekker, I., van den ElzenBen, P.J.M., Cornelissen, J.C., Melchers, L.S. 1995. Synergistic activity of chitinases and β-1,3-glucanases enhances fungal resistance in transgenic tomato plants. Euphytica 85:173–80.

Julian, J., Seavert, C.F., Olsen, J.L. 2009. An economic evaluation of the impact of eastern filbert blight resistant hazelnut cultivars in Oregon, USA. Acta Hort 845:725–32.

Leister, D. 2004. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet 20:116–122.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760.

Mammadov, J., Aggarwal, R., Buyyarapu, R., Kumpatla, S. 2012. SNP markers and their impact on plant breeding. International Journal of Plant Genomics vol. 2012: Article ID 728398.

Marone, D., Russo, M.A., Laidò, G., De Leonardis, A.M., Mastrangelo, A.M. 2013. Plant nucleotide binding site – leucine-rich repeat (NBS-LRR) genes : Active guardians in host defense responses. Int J Mol Sci 14:7302-7326.

McHale, L., Tan, X., Koehl, P., Michelmore, R.W. 2006. Plant NBS-LRR proteins: Adaptable guards. Genome Biology 7: 212.

Mehlenbacher, S.A., Brown, R.N., Nouhra, E.R., Gökirmak, T., Bassil, N.V., Kubisiak, T.L. 2006. A genetic linkage map for hazelnut (*Corylus avellana* L.) based on RAPD and SSR markers. Genome 49:122–133.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., Michelmore, R.W. 2003. Genome-wide analysis of NBS-LRR – encoding genes in *Arabidopsis*. Plant Cell 15: 809–834.

Michael, T.P., VanBuren, R. 2015. Progress, challenges and the future of crop genomes. Curr Opin Plant Biol. 2015 24:71-81.

Miller, J.M., Poissant, J., Kijas, J.W., Coltman, D.W. 2011. A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Molecular Ecology Resources 11:314–322.

Moser, B.R. 2012. Preparation of fatty acid methyl esters from hazelnut, high-oleic peanut and walnut oils and evaluation as biodiesel. Fuel 92:231–238.

Nas, M.N., and Read, P. 2004. Improved rooting and acclimatization of micropropagated hazelnut shoots. HortScience 39:1688-1690.

Neale, D.B., Wegrzyn, J.L., Stevens, K.A., Zimin, A.V., Puiu, D., Crepeau, M.W., ris Cardeno, *et al.* 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol 15:R59.

Nevarez, D., Mengistu, M., Nawarathne, Y.A., Walker, K.D. 2009. An N-Aroyltransferase of the BAHD superfamily has broad aroyl CoA specificity in vitro with analogues of N-dearoylpaclitaxel. J Am Chem Soc 131:5994–6002.

Ometto, L., Stephan, W., Lorenzo, D. 2005. Insertion / Deletion and nucleotide polymorphism data reveal constraints in *Drosophila melanogaster* introns and intergenic regions. Genetics 169:1521-1527.

Pechanova, O., Hsu, C., Adams, J.P., Pechan, T., Vandervelde, L., Drnevich, J., Jawdy, S., *et al.* 2010. Apoplast proteome reveals that extracellular matrix contributes to multistress response in Poplar. BMC Genomics 11:674.

Peterschmidt, B.C. 2013. DNA markers and characterization of novel sources of eastern filbert blight iesistance in European hazelnut (*Corylus avellana* L.). Oregon State University MS thesis https://ir.library.oregonstate.edu/xmlui/handle/1957/37973.

Potter, D., Eriksson, T., Evans, R.C., Oh, S., Smedmark, J., Morgan, D.R., Kerr, M., Robertson, K.R., Arsenault, M., Dickinson, T.A. 2007. Phylogeny and classification of Rosaceae. Plant Systematics and Evolution 266:5–43.

Powell, W., Machyay, G.C., Provan, J. 1996. Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1:215-222.

Punja, Z.K., Zhang, Y. 1993. Plant chitinases and their roles in resistance to fungal diseases. Journal of Nematology 25:526–540.

Rowley, E.R., Fox, S.E., Bryant, D.W., Sullivan, C.M., Priest, H.D., Givan, S.A., Mehlenbacher, S.A., Mockler, T.C. 2012. Assembly and characterization of the European hazelnut 'Jefferson' transcriptome. Crop Science 52: 2679-2686.

Ryynanen, H.J., Tonteri, A., Vasemagi, A., Primmer, C.R. 2007. A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). Journal of Heredity 98:692–704.

Sathuvalli, V.R., Chen, S., Mehlenbacher, S.A., Smith, D.C. 2011. DNA markers linked to eastern filbert blight resistance in 'Ratoli' hazelnut (*Corylus avellana* L.) Tree Genetics and Genomes 7:337–345.

Sathuvalli, V.R., Mehlenbacher, S.A. 2013. De novo sequencing of hazelnut bacterial artificial chromosomes (BACs) using multiplex Illumina sequencing and targeted marker development for eastern filbert blight resistance. Tree Genetics and Genomes 9:1109–111.

Sathuvalli, V.R., Mehlenbacher, S.A., Smith, D.C. 2011. DNA markers linked to eastern filbert blight resistance from a hazelnut selection from the Republic of Georgia. J Amer Soc Hort Sci 136:350-357.

Schocker, F., Lüttkopf, D., Müller, U., Thomas, P., Vieths, S., Becker, W.M. 2000. IgE binding to unique hazelnut allergens: Identification of non pollen-related and heat-stable hazelnut allergens eliciting severe allergic reactions. Eur J Nutr 39:172–180.

Seo, P.J., Lee, A., Xiang, F., Park, C. 2008. Molecular and functional profiling of Arabidopsis pathogenesis-related genes : Insights into their roles in salt response of seed germination. Plant Cell Physiol 49:334–344.

Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B. 2004. AUGUSTUS: A web server for gene finding in eukaryotes. Nucleic Acids Research 32:309–312.

Sticklen, M.B. 2008. Plant genetic engineering for biofuel production: Towards affordable cellulosic ethanol. Nature Reviews Genetics 9: 433–443.

Stintzi, A., Heitz, T., Prasad, V., Wiedemann-Merdinoglu, S., Kauffmann, S., Geoffroy P., Legrand, M., Fritig, B. 1993. Plant 'pathogenesis-related' proteins and their role in defense against pathogens." Biochimie 75:687–706.

Tabata, H. 2004. Paclitaxel production by plant-cell-culture technology. Adv Biochem Eng Biotechnol 87:1–23.

Takayama, S., Isogai, A. 2005. Self-incompatibility in plants. Annu Rev Plant Biol 56:467-489.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., *et al.* 2006. The Genome of Black Cottonwood *Populus trichocarpa* (Torr. & Gray). Science 313:1596-1604.

VanBuren, R., Bryant, D., Edger, P.P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., Lyons, E., Freeling, M., Bartes, D., Hallers, B.T., Hastie, A., Michael, T.P., Mockler, T.C. 2015. Single-molecule sequencing of the desiccation tolerant grass *Oropetium thomaeum*. Nature (in press).

Van der Auwera, G.A., Carneiro, M., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K., Altshuler, D., Gabriel, S., DePristo, M. 2013. From FastQ data to high-confidence variant calls: The Genome Analysis Tool kit best practices pipeline. Curr Protoc Bioinformatics 11:11.10.1-11.10.33.

Van Loon, L.C., Rep, M., Pieterse, C.M.J. 2006. Significance of inducible defense-related proteins in infected plants. Annual Review of Phytopathology 44:135–162.

Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J., Cattonaro, F. 2013. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of geneticdiversity, domestication and genome evolution. Nature Genetics 45:487–494.

Vignal, A., Milan, D., SanCristobal, M., Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol 34:275–305.

Walker, K., Long, R., Croteau, R. 2002. The final acylation step in taxol biosynthesis : Cloning of the taxoid C13-side-chain N-benzoyltransferase from taxus. PNAS 99:9166–9171.

Wan, H., Yuan, W., Bo, K., Shen, J., Pang, X., Chen J. 2013. Genome-wide analysis of NBS-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. BMC Genomics 14:109.

Wang, Y., Miao, Z., Tang, K. 2010. Molecular cloning and functional expression analysis of a new gene encoding geranylgeranyl diphosphate synthase from hazel (*Corylus avellana* L. 'Gasaway'). Molecular Biology Reports 37:3439–3544.

Xiong, M., Jin, L. 1999. Comparison of the power and accuracy of biallelic and microsatellite markers in population-based gene-mapping methods. American Journal of Human Genetics 64:629–640.

Yang, Y., Zhao, H., Barrero, R.A., Zhang, B., Sun, G., Wilson, I.W., Xie, F., *et al.* 2014. Genome sequencing and analysis of the paclitaxel-producing endophytic fungus *Penicillium aurantiogriseum* NRRL 62431. BMC Genomics 15:69.

Yu, X., Reed, B.M. 1995. A micropropagation system for hazelnuts (Corylus species). HortScience 30:120-123.

Yvon, A.M., Wadsworth, P., Jordan, M.A. 1999. Taxol suppresses dynamics of individual microtubules in living human tumor cells. Molecular Biology of the Cell 10: 947–959.

Zerbino, D.R., Birney, E. 2008. Velvet : algorithms for de novo short read assembly using de Bruijn Graphs. Genome Res18:821-829.

Chapter 4:

Conclusion

Erik R Rowley

We established additional genomic resources for the diploid hazelnut cultivar ‘Jefferson’ (OSU 703.007), allowing for the identification of novel molecular markers and genes of interest to hazelnut breeders. This represents the first sequenced genome among the order Fagales, and will enhance future breeding efforts by serving as a tool for gene discovery and functional studies. Previously established resources include a genetic map (Mehlenbacher et al 2006), and a BAC library (Sathuvalli *et al.*, 2013). The addition of characterized genome and transcriptome sequences for hazelnut will allow for the integration of the existing resources, to ultimately determine the loci underlying agriculturally important traits.

In order to develop ‘Jefferson’ into a genome-enable model system, we first established transcriptome sequence resources by sequencing the RNA content in four distinct hazelnut tissue types: leaves, catkins, whole seedlings, and bark. The filtered RNA-seq dataset from these tissues was pooled together and assembled *de novo* using the short read assemblers Velvet (Zerbino and Birney, 2008) and MIRA (Chevreux *et al.*, 2004). The resulting transcriptome assembly for ‘Jefferson’ comprises 28,255 transcript contigs, having an average length of 532bp, and an N50 of 961bp. Descriptive functional annotations using BLASTX protein homology were made for 21,202 (~75%) of these putative encoded proteins, with ~81% of the predicted proteins having high conservation with the most closely related plant sequences of *Vitus*, *Populus*, and *Ricinus*. Through gene ontology (GO) classification we were able to functionally characterize 11,221

unique loci (31.8% of total) comprising 65,536 GO terms Through differential expression analysis of these sequences, genes were identified as being enriched in particular tissues. For example, the leaf and seedling libraries (which contained emerging leaves) were enriched for chloroplast-associated GO categories, which are expected within leaf tissues. Such results serve to validate the transcript assemblies and offer insight into gene regulation within these tissues.

This resource promises to offer great insights to the hazelnut genetics community, such as future gene expression studies and homology comparisons. Since initial publication, the transcriptome assembly has been used to develop hundreds of new polymorphic SSR markers (Peterschmidt 2013) providing another resource to the hazelnut genetics community. Additionally, this resource was used in a comparison study against the Chinese hazelnut (*Corylus mandshurica*) to identify orthologous sequences and study local adaptation (Ma *et al.* 2013).

In addition to the transcriptomic resources, we also generated a *de novo* draft genome assembly for 'Jefferson'. Genomic DNA was sequenced on the Illumina HiSeq 2000 platform, representing ~93x coverage. We assembled the 'Jefferson' genome using the sequence assembler Velvet (Zerbino et al 2008) to generate initial assemblies of these genomic data. SSPACE v2.0 (Boetzer *et al.* 2011) was then applied, to merge and extend scaffolds where possible. The 'Jefferson' assembly is comprised of 36,641 contigs and scaffolds; half of the assembly is contained in scaffolds and contigs greater than 21.5 Kb, with the largest scaffold

comprising 274.5 Kb and captures 91% (345 Mb) of the flow-cytometry-determined genome size. To detect protein coding loci within the assembly we used the *ab initio* gene prediction program AUGUSTUS (Stanke *et al.* 2004), trained with gene features of *Arabidopsis thaliana* and using the ESTs from 'Jefferson' transcriptome (**Chapter 2**) as empirical evidence. This analysis identified 34,910 putative gene loci, named Corav_g1.t1 through Corav_g36090.t1, which were functionally annotated via homology to the NCBI non-redundant protein database, and assigned Gene Ontology (GO) terms. Of these predicted loci, 22,474 (64%) have homology to an entry in the NCBI non-redundant protein database, and 82.5% of the annotated genes are presented in the best annotated and closest related genera *Vitus, Prunus, Populus,* and *Ricinus*. We identified numerous candidate genes for future molecular validation such as those annotated as participating in the self-incompatability phenotype, which at the time of writing this thesis has not yet been identified.

We improved the existing hazelnut genetic linkage map (Mehlenbacher *et al.* 2006) by placing additional markers derived from genotyping by sequence (GBS) technology. This existing map offers a chromosomal level view of DNA markers that are linked to the aforementioned genes responsible for agriculturally important phenotypes and useful for screening hazelnut germplasm.

Each individual has unique polymorphic sites within its genome, which will be cut in a different pattern by a restriction enzyme. When these reads are aligned

back to the reference genome, variation between individuals of a population can be identified. We extracted high-quality gDNA from the F1 mapping population from Corvallis, Oregon resulting from the cross of the hazelnut accessions OSU 252.146 and OSU 414.062. GBS libraries were constructed using the restriction enzyme *ApeK1* and pooled into 72 barcoded sets prior to sequencing on an Illumina HiSeq 2500. We identified polymorphisms using the UNEAK package of the TASSLE-GBS pipeline (Glaubitz *et al.* 2014) and the SNPs that segregated 1:1 were used for map construction. Markers were assigned to linkage groups using JOINMAP 4.1 (Van Ooijen 2006). We added 3,209 additional GBS-derived markers to the existing hazelnut genetic linkage map, a five-fold increase over the previous map. Many of these new markers are closer to the loci of interest and may have a higher segregation percentage. These offer new possibilities for breeding and future molecular validation.

Associating polymorphisms in DNA sequence to coding potential allows the identification of causal mutations in candidate genes. Identification of these genes is of huge benefit for breeding purposes; markers can easily be designed around these regions and parental germplasm can be rapidly screened via PCR to determine the presence of these polymorphisms. Current high-throughput sequencing technologies enable rapid resequencing of whole genomes. We resequenced seven European hazelnut cultivars, (‘Barcelona’, 'Tonda Gentile delle Langhe', ‘Tonda di Giffoni’,‘Ratoli’, ‘Daviana’, ‘Halls Giant’, and ‘Tombul

(Extra Ghiaghli)') important in the Mediterranean region and Europe, at ~ 20x coverage on the Illumina HiSeq 2000 platform.

The resulting cultivar-specific reads were filtered as described previously and aligned to the 'Jefferson' reference genome with BWA (Li and Durbin 2009). We used the GATK pipeline (DePristo *et al.* 2011) to identify SNPs and INDELs relative to 'Jefferson' In order to reduce false positives due to ambiguity and sequence error we required a cutoff of 15 overlapping reads (representing ~75% of total) and a quality score (Q) of 30 for the SNPs and 20 for the INDELS in order for each variant to be reported. The SnpEff package (Cingolani *et al.* 2012) was used to annotation each variant and predict the impact on the coding potential of each affected gene model. Possible variant annotations include: upstream, downstream, splice site, untranslated region, intronic, or intergenic. The program also predicts the effect of the polymorphism on potential protein product. These effects include frame shifts, synonymous or non-synonymous amino acid replacement, start codon gains or losses, and stop codon gains or losses. These estimations are characterized as HIGH (for example the introduction of a pre-mature stop codon), MODERATE (a missense mutation), LOW (synonymous mutations), or MODIFIER (polymorphisms in UTRs). On average, 2 million SNPs and 300,000 INDELs were discovered between one or more of these genomes and that of 'Jefferson'. We restricted our analyses to those variants unique between each cultivar and the 'Jefferson' reference, of which ~6,000 SNPs and 300 INDELs are annotated as having a deleterious effect (HIGH) on coding

potential. This subset still includes a large set of genome-wide polymorphisms between seven resequenced hazelnut cultivars and the ‘Jefferson’ reference genome.

Examples of positional effects on coding potential include those shown to participate in disease resistance, such as the SNP in the English cultivar ‘Daviana’ located in Corav_g5616.t1, which is predicted to encode a “putative class III acidic chitinase”, a gene family participating in defense from fungal pathogens (Punja and Zhang 1993). Research has demonstrated that overexpression of chitinases can lead to improved fungal resistance in many crop systems (Cletus *et al.* 2013). Inversely, the loss of expression may increase susceptibility. The SNP in Corav_g5616.t1 introduces a mutation into a putative splice site acceptor region, which may result in a mis-spliced and ultimately degraded transcript. This SNP occurs in the EFB-susceptible ‘Daviana’, and may contribute to this cultivars susceptibility to the fungal born disease.

One method to validate this prediction is by use of site specific gene editing. Targeted gene editing, specifically using the CRISPR/Cas9 system, has recently emerged as a powerful tool to make precise changes to the sequence of an allele (Bortesi and Fischer 2015). CRISPR stands for “clustered regularly interspaced short palindromic repeats” and Cas (CRISPR-associated) is a protein family that cleaves double-stranded DNA. The CRISPR/Cas9 system is part of the native immune response of bacteria (Barrangou *et al.* 2007), and has been used with

success to stably modify a variety of traits in many plant species such as rice (Shan *et al.* 2013), sorghum (Jiang *et al.* 2014) and *Arabidopsis* (Schiml *et al.* 2014). A rapid and straightforward process, the CRISPR/Cas9 system makes targeted double-stranded DNA breaks at a specific position in the genome (ie, the genetic locus of interest) which are repaired via homologous recombination or nonhomologous end-joining to integrate changes to the sequence (Sander and Joung 2014, Bortesi and Fischer 2015). A guide RNA strand guides the Cas complex to the target DNA to be cleaved. These changes have the ability to change the coding potential of the targeted locus, and can be made to multiple genomic positions simultaneously allowing for trait stacking, to guarantee co-segregation (Wang *et al.* 2014). Companies such as Thermo Scientific offer services for designing the sequences to carry a targeted CRISPR/Cas9 gene editing experiment, which is introduced to the plant via A*grobacterium*-mediated transformation (Shan *et al.* 2013, Jiang *et al.* 2014, Schiml *et al.* 2014). Since hazelnut is amenable to transformation via *Agrobacterium* infiltration in tissue culture, it is possible to use CRISPR technology to modify loci containing deleterious mutations. For example, the SNP in "putative class III acidic chitinase" locus in the EFB susceptible cultivar 'Daviana' discussed above could be restored to the functional form and the resulting transformant tested for susceptibility to EFB. Given the ability for this technology to make targeted edits to multiple genomic positions simultaneously, and number of deleterious variants could be made functional again in order to biologically access the function of each protein product of phenotype. These datasets contain many agriculturally

significant loci and will be of tremendous value in selecting candidate genes for future molecular experiments and empirical characterization.

Due to the heterozygosity of 'Jefferson' genome as an F1 progeny, the genome assembly is quite fragmented, currently preventing assembly into chromosome level scaffolds, however hundreds of SSR markers have currently been developed (S.A. Mehlenbacher, personal communication, April 2016). Given the agricultural significance of hazelnut, this is a genome that would benefit from the incorporation of third generation sequencing strategies such as Pacific Biosciences (PacBio) ultra-long read single molecule sequencing. This technology enables generation of sequencing reads up to >25 Kb in length (http://www.pacificbiosciences.com/) and would allow resolution of repetitive regions and joining of contigs. This is evident from the recent genome sequencing of the resurrection plant *Oropetium thomaeum* using the new P6-C4 chemistry on the PacBio RS II platform, which resulted in 72X coverage with an average read length of 16 kb. The *Oropetium* assembly covers 98% of the genome (244 Mb) with a contig N50 of 2.4 Mb (VanBuren *et al.* 2015). There are efforts underway to conduct PacBio sequencing in the later part of 2016 (S.A. Mehlenbacher, personal communication, April 2016), which will incorporate all of the hazelnut genomic data collected thus far, in an effort to construct a chromosomal level assembly.

The hazelnut genomic and transcriptomic sequence resources we have developed will allow breeders the opportunity to exploit the wealth of genetic diversity when choosing parents from among the hundreds of accessions worldwide. We provide these resources to the community via the European hazelnut genomic resource portal, offering a BLAST interface for homology searches against hazelnut genome and transcriptome sequences, and a JBrowse genome browser for visualization of genomic features, annotations, and variant affects between 7 re-sequenced hazelnut accessions and 'Jefferson'. Additionally, the FTP site hosts: FASTA files of the 'Jefferson' genome and transcriptome assemblies and their associated annotations, variant call format (.vcf) files of polymorphisms between 7 re-sequenced hazelnut cultivars and 'Jefferson, and the improved genetic linkage map derived from GBS markers among a the individuals from the F1 mapping population. The variants discovered and characterized in this study, including the new GBS-derived markers placed on the genetic linkage map, will be useful for identifying candidate genes responsible for EFB resistance and other genetic factors that affect yield.

Hazelnut genome sequencing has provided new resources to the scientific community and promises to accelerate trait discovery and enhance future breeding efforts. This resource will serve as a tool for gene discovery and functional studies, for the development of polymorphic DNA markers and other genomic tools, and will allow future integration of the genome sequence with genetic and physical maps and the incorporation of new sequencing technologies.

Literature cited:

Boetzer, M., Henkel, C.V., Jansen, H., Butler, D. 2011. Scaffolding pre-assembled contigs using SSPACE summary. Bioinformatics 27:578-579.

Bortesi, L., and Fischer, R. 2015. The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol Adv. 1:41-52.

Cingolani, .P, Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., Lu, X. 2012. Using Drosophila melanogaster as a model for genotoxic chemical mutationalstudies with a new program, SnpSift. Frontiers in Genetics 3:1–9.

Chevreux, B., T. Pfisterer, B. Drescher, A.J. Driesel, W.E. Müller, T. Wetter, and S. Suhai. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 14:1147– 1159.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., Buckler, E.S. 2014. TASSEL-GBS : A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:e90346.

Ma, H., Lu. Z., Liu, B., Qiu, Q., Liu, J. 2013. Transcriptome analyses of a Chinese hazelnut species Corylus mandshurica. BMC Plant Biol. 5:13:152.

Mehlenbacher, S.A. 2014. Geographic distribution of incompatibility alleles in cultivars and selections of European hazelnut. J Amer Soc Hort Sci. 139. 191–212.

Peterschmidt, B.C. 2013. DNA markers and characterization of novel sources of Eastern filbert blight iesistance in European hazelnut (*Corylus avellana* L.). Oregon State University MS thesis https://ir.library.oregonstate.edu/xmlui/handle/1957/37973

Punja, Z.K., Zhang, Y. 1993. Plant chitinases and their roles in resistance to fungal diseases. Journal of Nematology 25:526–540

Sander, J. D., and Joung, K., K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 32:347–355.

Sathuvalli, V.R., Mehlenbacher, S.A. 2013. De novo sequencing of hazelnut bacterial artificial chromosomes (BACs) using multiplex Illumina sequencing and targeted marker development for eastern filbert blight resistance. Tree Genetics and Genomes 9:1109–111.

Schiml S, Fauser F, Puchta H. 2014. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. Plant J. 6:1139-50.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J., Qiu, J.L., Gao, C. 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol. 8:686-8.

Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B. 2004. AUGUSTUS: A web server for gene finding in eukaryotes. Nucleic Acids Research 32:309–312.

Van Ooijen, J.W., 2006. JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., Qiu, J. 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat Biotechnol. 32:947-951.

Wenzhi Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., Weeks, D. 2013. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. Nucl. Acids Res. doi:10.1093/nar/gkt780

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

Appendices

Chapter 1:

Rapid Synthesis of a long double-stranded oligonucleotide from a single-stranded nucleotide using magnetic beads and an oligo library

Sumate Pengpumkiat, Myra Koesdjojo, Erik R. Rowley, Todd C. Mockler,

Vincent T. Remcho

PLoS ONE
1160 Battery Street
Koshland Building East, Suite 100
San Francisco, CA 94111, USA
11(3): e0149774 (2016)
DOI: 10.1371/journal.pone.0149774

Abstract:

Chemical synthesis of oligonucleotides is a widely used tool in the field of biochemistry. Several methods for gene synthesis have been introduced in the growing area of genomics. In this paper, a novel method of constructing dsDNA is proposed. Short (28-mer) oligo fragments from a library were assembled through successive annealing and ligation processes, followed by PCR. First, two oligo fragments annealed to form a dsDNA molecule. The double-stranded oligo was immobilized onto magnetic beads (solid support) via streptavidin-biotin binding. Next, single-stranded oligo fragments were added successively through ligation to form the complete DNA molecule. The synthesized DNA was amplified through PCR and gel electrophoresis was used to characterize the product. Sanger sequencing showed that more than 97% of the nucleotides matched the expected sequence. Extending the length of the DNA molecule by adding single-stranded oligonucleotides from a basis set (library) via ligation enables a more convenient and rapid mechanism for the design and synthesis of oligonucleotides on the go. Coupled with an automated dispensing system and libraries of short oligo fragments, this novel DNA synthesis method would offer an efficient and cost-effective method for producing dsDNA.

Introduction:

Oligonucleotide synthesis, the chemical synthesis of nucleic acids, has become an important tool in the field of molecular biology. Synthetic oligonucleotides have been utilized in numerous applications such as diagnosis of genetic and infectious diseases, new drug discovery, and disease treatment. Oligonucleotide synthesis has a long history, starting with a synthetic approach developed by Marvin Caruthers in early 1980s (Beaucage and Caruthers 1981). Solid-phase phosphoramidite chemistry is a well- established 4-step process, which elongates a chain of nucleotide from the 3'end to the 5'end, and is used by many commercial DNA synthesizers. The phosphoramidite chemistry has enabled routine synthesis oligos up to 100 nt with error rates of 1 in 200 nt or better (Kosuri and Church 2014), yet provides short oligonucleotides owing to the fact that the method adds one base at a time to the growing oligonucleotide chain. Each step in the synthetic cycle must have very high yield in order to obtain a final product in the required amount with a very low accumulated error rate. For example, for 200 nt oligo synthesis, 99% yield from each cycle will result in 13% yield of the desired final product. The longer the desired oligomer, the lower they yield that can be obtained from the synthesis process.

To assemble longer DNA strands, a set of pre-synthesized oligonucleotides can be used as building blocks and assembled using enzymatic methods. Ligation-based assembly is a method to join overlapping oligonucleotides using DNA ligase to form a longer gene. Here, all of the oligos are mixed together with DNA ligase

and are thermocycled for annealing and ligation to build the gene product. The Polymerase Chain Reaction (PCR) is then used to amplify the full-length product. The drawbacks of the method are a relatively high probability of generating ligation by-products and the need for multiple overlapping oligos along the entire length of both strands.

Most gene synthesis today is conducted by service companies, and typically employs one or more of four different general approaches: ligation (Borodina *et al.* 2003, Juhn *et al.* 2005), PCR-mediated assembly (Wu *et al.* 2006), convergent assembly (Horspool *et al* .2001), or solid-phase assembly (Caruthers 1985, Caruthers 1991, Sierzchala *et al.* 2003, Kumar *et al.* 2003)). These approaches are relatively simple and published protocols are available. Recently, technologies and applications of DNA synthesis have been reviewed (Engles *et al.* 1989, Vaijayanthi *et al.* 2003, Koruir and Church 2014). However, while these approaches generally succeed in producing usable oligos, they each have limitations than can lead to errors in assembly (LeProust *et al.* 2010). Additionally, some sequences are impossible to synthesize using these approaches, and the methodologies, which rely on laboriously constructed large (up to 200-base) oligos, are time consuming and error-prone.

As implemented by service companies, gene synthesis is still being conducted via a large-scale, semi-automated process using custom-synthesized individual DNA oligonucleotides. Because every gene is different, each is synthesized anew for

each individual purchase order. Attempts have been made to automate the gene synthesis approaches described above by using DNA microarrays or microfluidic devices (Engles 1989, Kim *et al.* 2011) in synthesis platforms that use ligation or PCR-mediated assembly. However, these technologies and approaches required custom synthesis, purification and amplification of relatively long oligonucleotides that are then joined into longer DNA molecules. In the microarray format, these approaches also tend to produce DNA products of lower-than-ideal purity. Accumulated errors and truncated DNA products can be significant problems as there is no purification following each synthetic step (Baker 2011). This is exacerbated by decreasing product yields with longer oligonucleotides due to the fact that chains may either stop growing or incorporate undesired nucleotides. For DNA synthesis in microfluidic systems, the materials used for fabrication must have excellent chemical resistance given the variety of chemicals and organic solvents used in standard DNA synthesis (Tian *et al.* 2009).

In this paper, we demonstrated enzymatic DNA synthesis using magnetic beads as solid supports. Magnetic beads have been widely utilized as solid supports for biomolecule separation, as they facilitate washing and easy manipulation of the sample. Typically, streptavidin is coupled to magnetic beads to specifically capture a target of interest such as a nucleic acid, protein, or antibody that has been biotinylated. Recently, chemical gene synthesis using magnetic beads as the solid support was presented (Horspool *et al.* 2010). Short double stranded (ds)

oligo fragments 8 bp long were assembled to form longer *32 bp intermediate single stranded fragments* through ligation. The target DNA sequence of 128 bp was constructed from four sets of intermediate fragments in a pair-wise manner**.** Herein, we describe a proof of principle study for the construction of a *full-length double-stranded DNA molecule from short fragments of single-stranded nucleotides* (ss-oligonucleotides). 28-mer oligo fragments were used as the building blocks to assemble the full length DNA. The gene was constructed by adding one ssDNA fragment at a time to extend the length of a target sequence through repeated ligation reactions. T4 DNA ligase was used to catalyze the formation of phosphodiester bonds between the adjacent 3'-hydroxy and 5'-phosphate termini in the double-stranded DNA. Streptavidin-coated magnetic beads were utilized as the solid supports for easy means of washing excess reagents at the completion of each ligation process. Without magnetic bead separation for each ligation step in the process, it is impossible to clean or eliminate the unreacted ssDNA from the enzymatic reaction. The magnetic bead purification approach provides each ligation reaction with only main chain DNA on the building block and the fragment ssDNA to be annealed and ligated to the main chain. Compared to conventional ligation-based assembly, our protocol generates more pure and accurate DNA final products, and affords an easy path to full automation through the use of automatic liquid dispenser, work that is ongoing in our laboratory. The overall procedure is shown in Fig 1. In this paper, we demonstrated the successful synthesis of a long dsDNA from a 28-mer library through successive ligation reactions. Combined with an automated system, this

novel strategy will provide a powerful and rapid method for synthesizing custom dsDNA molecules in a standard laboratory.

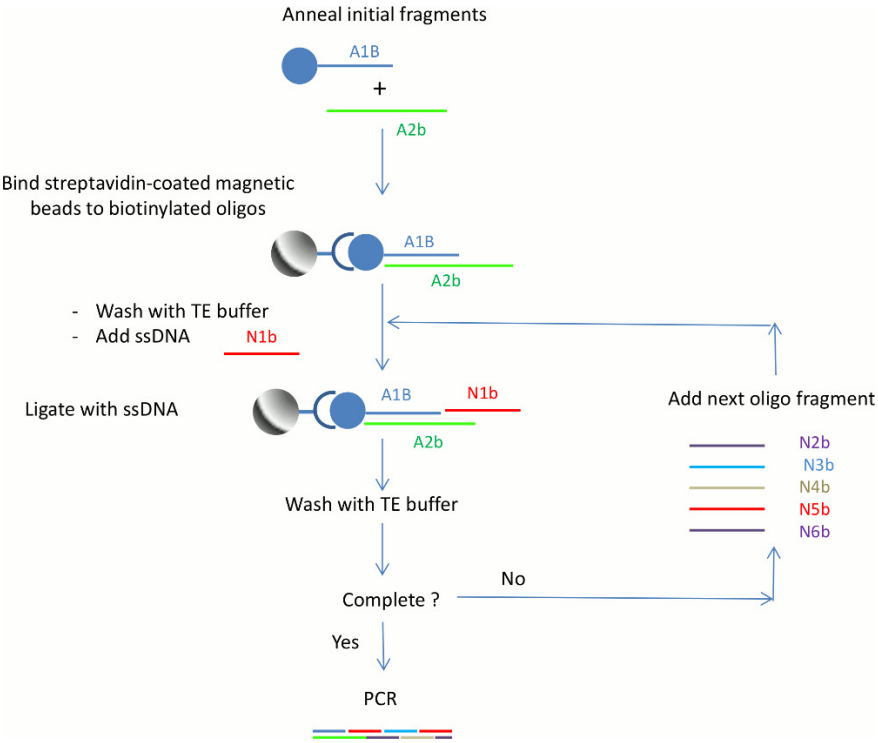


Fig 1.1. The overall procedure for dsDNA synthesis is composed of three processes: annealing, binding of streptavidin coated magnetic beads to biotinylated oligos, and ligation.

Materials and Methods:

***Materials*.** All single-stranded oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) with the 5' ends of the oligos phosphorylated. A schematic of the building blocks for construction of the target DNA molecule is shown in **Fig 2**. The building blocks (N1b-N6b) were synthesized as 28 nucleotide-long fragments that overlapped the next fragment by 14 bases. The first two fragments (A1B and A2b) were designed to be 29 and 43 nucleotides

long, respectively. All oligonucleotides were prepared at 100 µM (Table 1) and were stored at 4°C throughout the study. Streptavidin-coated magnetic beads, Dynabeads M-270, were obtained from Invitrogen™/Life Technologies (Grand Island, NY). Quick Ligation™ kits were obtained from New England Biolabs (Ipswich, MA).

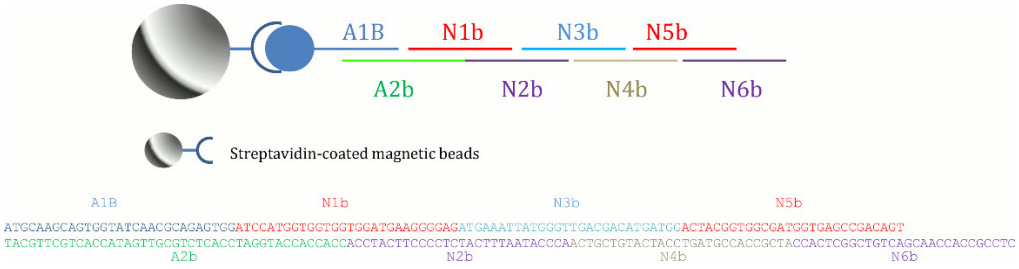


Fig 1.2. Streptavidin-coated magnetic beads were used as solid support for dsDNA synthesis and oligo fragments were ligated to the building block one at a time.

Table 1.1. Sequence of the oligo fragments used to form the complete dsDNA.

Sequence name	Sequence	Number of nucleotides
Biotin-A1B	Biotin -5'-ATGCAAGCAGTGGTATCAACGCAGAGTGG-3'	29
A2b	3'-TACGTTCGTCACCATAGTTGCGTCTCACCTAGGTACCACCACC-5'	43
N1b	5'-ATCCATGGTGGTGGTGGATGAAGGGGAG-3'	28
N2b	3'-ACCTACTTCCCCTCTACTTTAATACCCA-5'	28
N3b	5'-ATGAAATTATGGGTTGACGACATGATGG-3'	28
N4b	3'-ACTGCTGTACTACCTGATGCCACCGCTA-5'	28
N5b	5'-ACTACGGTGGCGATGGTGAGCCGACAGT-3'	28
N6b	3'CCACTCGGCTGTCAGCAACCACCGCCTC-5'	28

doi:10.1371/journal.pone.0149774.t001

Annealing buffer (10×) consisted of 100mM Tris-HCl, 500 mM NaCl, 10mM EDTA, was adjusted to pH 7.4 with NaOH. The ligation kit contains quick ligation buffer (2× QLB) (132 mM Tris-HCl, 20 mM $MgCl_2$, 2mM dithiothreitol, 2mM ATP, 15% PEG6000, pH 7.6 at 25°C) and T4 ligase. EmeraldAmp® GT

PCR Master Mix was obtained from Takara Bio Company (Mountain View, CA). DNA Ladder (O'RangeRuler 20 bp) was obtained from Life Technologies (Grand Island, NY).

Binding and washing buffer (2× B&W) buffer contains 10mM Tris-HCl, 1mM EDTA, 2M NaCl, and 0.05% Tween20, pH 7.5. Binding and washing 1× buffer (B&W) was prepared by diluting 2× B&W buffer with TE buffer in equal proportions. The TE buffer contains 10mM Tris-HCl, 1mM EDTA, and 0.05% Tween20.

Tris-acetate-EDTA (TAE) 50× buffer was obtained from Bio-Rad (Hercules, CA). It was diluted to 1× buffer with deionized water. The composition of 1× TAE buffer was 40mM Tris (pH 7.6), 20 mM acetic acid, and 1 mM EDTA.

Preparation of Streptavidin-coated magnetic beads. 50 μL of magnetic beads were washed with 50 μL of 2× B&W buffer (10mM Tris-HCl pH 7.5, 1mM EDTA, 2M NaCl, and 0.05% Tween20) for three times to remove excess sodium azide bacteriostatic agent and resuspended in 50 μL, 1X B&W buffer.

Annealing process. 5 μL of stock solutions (100 μM) of the first (A1B) and second (A2b) single-stranded oligo fragments were mixed with 80 μL of deionized water and 10 μL of 10× annealing buffer (100mM Tris-HCl, 500 mM NaCl, 10mM EDTA, pH 7.4). The mixture was vortexed and heated in a water

bath at 95°C for 10 min, and was slowly cooled to room temperature. The final concentration of each oligonucleotide was 5 μM.

Binding streptavidin coated-magnetic beads and biotinylated oligos. The annealed products from the first two fragments (Biotin A1B-A2b) were incubated with the magnetic beads by gentle-rotation at room temperature for 30 min and then washed for 2 times with TE buffer to eliminate excess amount of biotinylated oligos from the beads.

Ligation. The magnetic beads (product of 2.2.3) were added to a PCR tube along with 8 μL of deionized water, 2 μL of 100 μM subsequent single-stranded oligo (N1b), 10 μL of Quick Ligase Buffer, and 1 μL of T4 ligase. The PCR tube was vortexed and incubated at room temperature (25°C) for 15 min. Phosphodiester bonds were formed between the two fragments at the 5' phosphate and 3' hydroxyl groups.

These steps were repeated for the remainder of the oligo fragments (N2b, N3b, N4b, N5b, and N6b). Excess oligos and T4 ligase enzyme were removed by washing with 2×100 μL of TE buffer after each ligation step. At the end of the ligation process, the magnetic beads were resuspended in 20 μL deionized water.

PCR. The final product of the synthesis was amplified by PCR (Hybaid PCR Express HBPX gradient Thermal Cycler). The forward and reverse primers were

5’-TGCAAGCAGTGGTATCAACG-3’ and 5’-ACCATCGCCACCGTAGTC-3’, respectively. The PCR solution contained 2 μL of ligation product, 22 μL deionized water, 25 μL Emerald master mix, 1 μL DMSO and 0.5 μL of each primer. The negative control was prepared without the addition of the template DNA. The steps for the PCR program were: 95°C for 30 s, followed by 30 amplification cycles (95°C for 15 s, 65°C for 30 s, and 72°C for 20 s) and 72°C for 5 min.

***Gel electrophoresis*.** PCR products were run on a 3% agarose gel in 1×TAE buffer. The horizontal electrophoresis system and power supply were obtained from Bio-Rad (Hercules, CA). The potential was set at 75 V.

Sanger DNA sequencing. Bands were cut from the gel, extracted, and purified using the GeneJet PCR purification kit from Fermentas (Thermo Scientific, Pittsburgh, PA). The solutions were stored in a freezer (-20°C) prior to DNA sequencing.

***One-pot dsDNA synthesis*.** To perform a one-pot gene synthesis experiment, 5 μL of each of the annealed products of N1b-N2b, N3b-N4b, N5b-N6b was mixed in a tube with 10 μL Quick Ligation Buffer and 1 μL T4 Ligase. The tube was incubated at room temperature for 15 min. The final product was analyzed using gel electrophoresis.

Results and Discussion:

We have demonstrated a novel strategy to synthesize a full-length dsDNA molecule accurately and efficiently. A 101-bp long dsDNA product was synthesized by successive ligation of 28-mer fragments. The synthesis process involves annealing, binding of biotinylated oligo fragments to magnetic beads (solid support), and consecutive ligation reactions. In this study, we tested each of these steps separately to ensure the optimization of all conditions prior to synthesizing the full-length dsDNA. Oligo fragments N1b-N6b shown in Table 1 were used to test and optimize the annealing and ligation processes.

Annealing. Deoxyribonucleic acids recognize their complementary oligonucleotides by base pairing: A-T and G-C. Single-stranded DNA will hybridize with its complement via the formation of hydrogen bonds. The annealing process was carried out at 95° for 10 min in an annealing buffer that promotes hydrogen bonding between the two complementary ssDNA molecules. The annealing process was tested by pairing N1b with N2b, N3b with N4b, and N5b with N6b. The annealed product was characterized using gel electrophoresis. The lengths of all the annealed products were expected to be 42 bp (28+14). Lanes 2, 3, and 4 in Fig 3 show the expected bands around 40 bp, which suggests that the annealing process was successful.

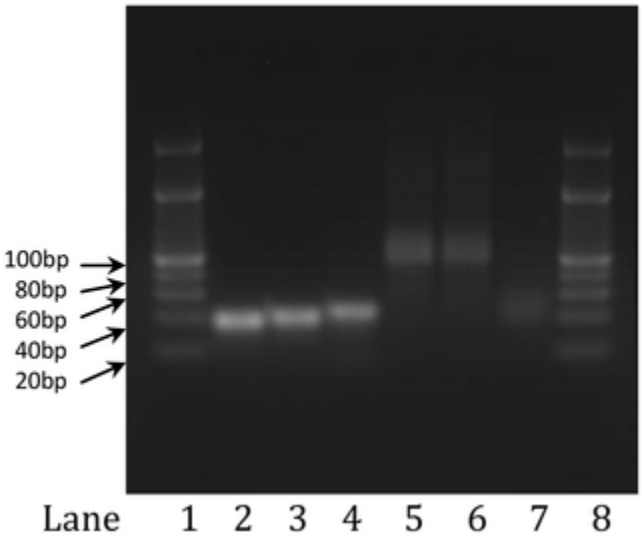


Fig 1.3. Agarose gel electrophoresis of annealed and ligation products**.**

Lane 1 and 8, ladder; Lane 2, annealed product of N1b and N2b; Lane 3, annealed product of N3b and N4b; Lane 4, annealed product of N5b and N6b; Lane 5, "one-pot" ligation product of [(N1b-N2b)+(N3b-N4b)+(N5b-N6b)]; Lane 6, sequential ligation product of [(N1b-N2b)+(N3b-N4b)]+(N5b-N6b); Lane 7, ligation product of [(N1b-N2b)+(N5b-N6b)].

doi:10.1371/journal.pone.0149774.g003

***Ligation*.** In this study, T4 DNA ligase was utilized as the catalyst for the formation of phosphodiester bonds between two ssDNA oligo fragments. This enzyme has been shown to effectively join both blunt and sticky ends of dsDNA. In this part of the study, we wanted to evaluate whether the T4 ligase has a favorable reaction towards dsDNA with sticky ends compared to ssDNA with blunt end. T4 ligase is most commonly used to join dsDNA fragments. In our

approach we used this enzyme to join ssDNA fragments with the complementary sticky ends of dsDNA molecules (Fig 1).

To better understand the efficiency of the T4 ligase in joining our dsDNA fragments with sticky ends, we first performed a sequential ligation reaction using the annealed products of N1b-N2b, N3b-N4b, and N5b-N6b described in section 3.1. The first ligation was carried out between the N1b-N2b and N3b-N4b dsDNA fragments and was allowed to incubate at room temperature for 15 min. The next dsDNA fragment, N5b-N6b, was then added into the solution mixture and incubated under the same conditions as the first ligation. The resulting approximately 100 bp product is shown in **Fig 3,** Lane 6, and is consistent with our expected product size of 98 bp (28+28+28+14 bp). This confirms that the sequential ligation of dsDNA fragments with complementary sequences can be accomplished at room temperature in as fast as 15 minutes.

We also investigated whether or not the T4 ligase enzyme was able to effectively join dsDNA fragments at their complementary sites in the presence of impurities in the form of other dsDNA fragments. A “one-pot” reaction was performed by mixing all of the annealed products with the T4 ligase in a tube, which was incubated at room temperature (25°C) for 15 min. The resulting product was run on a gel and showed a similar band at ~100 bp (Fig 3, Lane 5). To determine whether or not ligation occurs between the N1b-N2b and N5b-N6b fragments, we ran a similar test in which only these two dsDNA fragments were mixed with the

T4 ligase and incubated at 25°C for 15 min. The product was again analyzed on a gel and the result showed a band around 40 bp that corresponded to the sizes of the annealed products themselves (42 bp expected) and not the product of successful ligation **(Fig 3, Lane 7).** These two annealed products were not designed with complementary sequences in their overhang portions. This result indicated that ligation of dsDNA fragments with non-complementary sticky ends was unfavorable under the specified reaction conditions with T4 ligase. This feature plays an important role in the proposed gene synthesis approach, as it shows that ligation is only favorable in the presence of DNA fragments with complementary sticky ends. This significantly reduced the chance of misalignments or sequence errors in the final gene product as a result of non-complementary oligo fragments that may be present in solution during the ligation process.

Oligonucleotide synthesis using streptavidin-coated magnetic beads

The proposed gene synthesis strategy began with the annealing process, where the biotinylated A1B fragment was annealed with the A2b fragment to form the initial dsDNA molecule. The annealed product of A1B/A2b was immobilized on magnetic beads via streptavidin-biotin binding. Magnetic beads allow for convenient washing and separation of the desired products from excess reagents. Holmberg *et al.* demonstrated that dissociation of biotin and streptavidin molecules occurs when a solution is heated to 70°C. Thus, the annealing of the first dsDNA fragment was performed prior to its immobilization on the magnetic

beads to avoid exposing the immobilized products to the elevated temperature of the annealing process (95°C).

The results from section 3.2 demonstrated that T4 DNA ligase was capable of joining dsDNA fragments with complementary sticky ends. In this experiment, the ligation process was repeated by adding the annealed products of N1b-N2b, N3b-N4b and N5b-N6b one at a time to the main chain. At the completion of the ligation reactions, the full-length dsDNA was amplified by PCR. The result is illustrated in Fig 4, lane 2, as a band of approximately 100 bp, which corresponds to the expected length of the final product (101 bp).

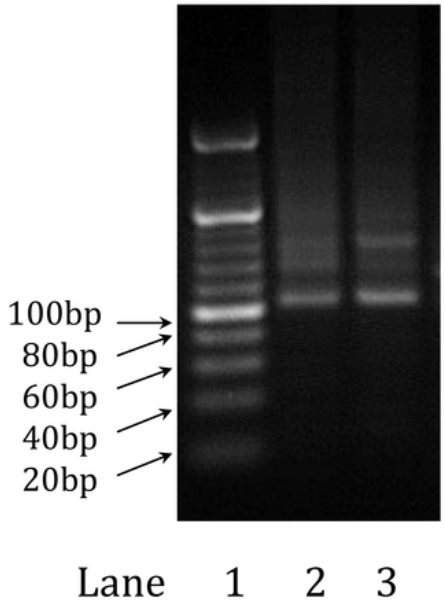


Fig 1.4. Agarose gel electrophoresis of the final product.Lane 1, ladder; Lane 2, ligation product from double-stranded oligo fragments; Lane3, ligation product from single-stranded oligo fragments.

We also investigated whether or not it was possible to ligate single-stranded oligonucleotides sequentially to the main dsDNA chain using the T4 DNA ligase under the same experimental conditions described in section 3.2. To test this, 28-mer ssDNA fragments were used as building blocks (rather than the double-stranded oligos), and were ligated one at a time to extend the length of the main chain. Thorough washing was performed at the completion of each ligation step to minimize interference from excess oligos that may still be present in solution. Once all of the ligation steps were completed to form the desired dsDNA, the product was amplified by PCR and analyzed using gel electrophoresis. The result of the PCR product is illustrated in Fig 4 lane 3 and clearly shows a product of approximately 100 bp. We have successfully demonstrated an efficient approach to synthesizing and extending the length of dsDNA through repetitive ligation of short ssDNA oligo fragments. Each cycle of the DNA synthesis step takes approximately 20 minutes, including the cleaning and washing steps. Our proposed method would take only about 12 hours to synthesize a DNA molecule of 500 bp, whereas today's commercial gene synthesis operations generally require multiple days to synthesize DNA molecules of this size.

***Sequencing of the target DNA*.** PCR products were cut from the agarose gels, extracted, and purified using the GeneJet PCR purification kit (ThermoScientific) for sequencing analysis. Sanger sequencing was performed to characterize the final products. Since the quality of Sanger sequencing was normally poor for the first 20 bp of the DNA sequence, sequencing of the DNA products was performed

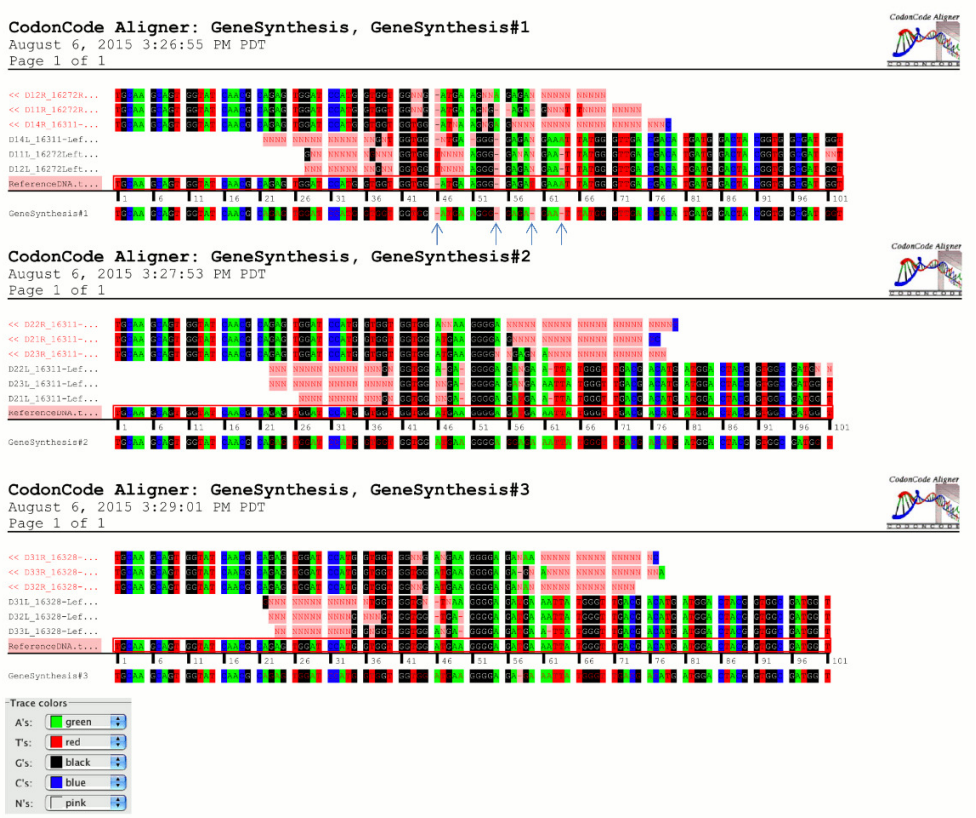
in both the forward and reverse directions to obtain the whole sequence data. The reproducibility of the gene synthesis approach and the accuracy of the DNA sequencing results were determined by repeating the entire process in triplicate. The resulting PCR product for each process was sequenced 3 times. Sequencing data were aligned and compared to the expected sequence by CodonCode Aligner (v. 5.1.5 CodonCode Corp., MA). The sequence validation by Sanger sequencing revealed that gene synthesis product was nearly identical (97%) to the expected sequence. DNA sequencing alignment is shown in Supporting Information.

Conclusions

In this work, we have demonstrated a novel approach of synthesizing a long double stranded oligonucleotide through repetitive ligation of short ssDNA oligo fragments. The method is simple and straightforward, using streptavidin coated magnetic beads as solid supports for ease of washing. The magnetic bead purification for each ligation-based assembly helps promote the assembly of a pure and accurate dsDNA product and enables complete automation of the synthetic approach–work that is underway in our lab. The accuracy of the synthesis method was validated by Sanger sequencing, and the results showed we were able to generate DNA products with precision of more than 97%. With this approach, it is possible to create custom DNA with rapid turnaround time. Combined with automation technology and access to universal libraries of short oligo fragments, this approach would provide a powerful gene synthesis solution with significant time and cost savings.

Supporting Information

Fig 1.5. DNA sequencing alignment of the final product.



The data obtained via Sanger sequencing was aligned using CodonCode Aligner (v. 5.1.5 CodonCode Corp., MA). This figure illustrates the quality of the actual product relative to the intended product. Errors are indicated by the blue arrows.

Acknowledgments:

We would like to thank Dr. Sergei Filichkin, Department of Botany and Plant Pathology, Oregon State University for mentoring; Yuanyuan Wu, Nicole J. Hams, Sara Townsend for technical discussion; the Oregon State University Center for Genome Research & Biocomputing (CGRB lab) for DNA sequencing, and Dr. Genevieve Weber for editorial review.

Author Contributions:

Conceived and designed the experiments: VTR TCM SP MK. Performed the experiments: SP MK ERR. Analyzed the data: VTR TCM SP MK ERR. Contributed reagents/materials/analysis tools: VTR TCM. Wrote the paper: SP MK ERR TCM VTR.

Literature cited:

Baker M. Microarrays, megasynthesis. Nat Methods. 2011 Jun;8(6):457–60.

Beaucage SL, Caruthers MH. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. 1981;22(20):1859–62.

Borodina TA, Lehrach H, Soldatov AV. Ligation-based synthesis of oligonucleotides with block structure. Anal Biochem. 2003 Jul 15;318(2):309–13.

Caruthers MH. Gene synthesis machines: DNA chemistry and its uses. Science. 1985 Oct 18;230(4723):281–5.

Caruthers MH. Chemical synthesis of DNA and DNA analogs. Acc Chem Res. 1991 Sep 1;24(9):278–84.

Engels JW, Uhlmann E. Gene Synthesis [New Synthetic Methods (77)]. Angew Chem Int Ed Engl. 1989 Jun 1;28(6):716–34. doi: 10.1002/anie.198907161

Engels JW. Gene Synthesis on Microchips. Angew Chem Int Ed. 2005 Nov 11;44(44):7166–9.

Holmberg A, Blomstergren A, Nord O, *et al.* The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. Electrophoresis. 2005 Feb;26(3):501–10.

Horspool DR, Coope RJ, Holt RA. Efficient assembly of very short oligonucleotides using T4 DNA Ligase. BMC Res Notes. 2010 Nov 9;3(1):291.

Kim H, Jeong J, Bang D. Hierarchical gene synthesis using DNA microchip oligonucleotides. J Biotechnol. 2011 Feb 20;151(4):319–24.

Kosuri S, Church GM. Large-scale de novo DNA synthesis: technologies and applications. Nat Methods. 2014 May;11(5):499–507.

Kuhn H, Frank-Kamenetskii MD. Template-independent ligation of single-stranded DNA by T4 DNA ligase. FEBS J. 2005 Dec;272(23):5991–6000.

Kumar P, Gupta KC. A Rapid Method for the Construction of Oligonucleotide Arrays. Bioconjug Chem. 2003 May 1;14(3):507–12

LeProust EM, Peck BJ, *et al.* Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. Nucleic Acids Res. 2010 May;38(8):2522–40.

Sierzchala AB, Dellinger DJ, Betley JR, *et al.* Solid-phase oligodeoxynucleotide synthesis: a two-step cycle using peroxy anion deprotection. J Am Chem Soc. 2003 Nov 5;125(44):13427–41.

Tian J, Ma K, Saaem I. Advancing high-throughput gene synthesis technology. Mol Biosyst. 2009 Jul;5(7):714–22.

Vaijayanthi B, Kumar P, Ghosh PK, Gupta KC. Recent advances in oligonucleotide synthesis and their applications. Indian J Biochem Biophys. 2003 Dec;40(6):377–91

Wu G, Wolf JB, Ibrahim AF, Vadasz S, *et al.* Simplified gene synthesis: A one-step approach to PCR-based gene construction. J Biotechnol. 2006 Jul 25;124(3):496–503.

Chapter 2:

Genome Sequencing and Resource Development for European Hazelnut

Erik. R. Rowley, Doug. W. Bryant, Sam. E. Fox, Scott.A. Givan, Shawn. A. Mehlenbacher and Todd. C. Mockler

Acta Horticulturae
ISHS Secretariat
PO Box 500
3001 Leuven 1, Belgium
1052, 75-78 (2014)
DOI: 10.17660/ActaHortic.2014.1052.8

Abstract:

European hazelnut (*Corylus avellana* L.) is of global agricultural and economic significance, with genetic diversity existing in hundreds of accessions. Breeding efforts are focused on maximizing nut yield and quality, while reducing susceptibility to diseases such as Eastern Filbert Blight (EFB). We are establishing genomic tools to aide in breeding efforts, including a genome and transcriptome assembly of an EFB resistant cultivar, 'Jefferson' (OSU 703.007), and characterizing sequence diversity among 7 additional accessions. Polymorphisms associated with gene coding regions will be useful in the establishment of molecular markers for breeding efforts.

Introduction:

European Hazelnut (*Corylus avellana* L.*)* is well adapted to the Pacific Northwest (USA). Oregon's Willamette Valley produces nearly 99% of the U.S. crop, which accounts for ~3–5% of global hazelnut production. The USDA and Oregon State University (OSU) have collected nearly 700 accessions of *C. avellana* L., and representatives of other *Corylus* species, and preserve them in Corvallis OR (USA) for use in genetic improvement. We are developing genomic tools to assist genetic improvement efforts. 'Jefferson' (OSU 703.007), released by OSU in 2009, and used for BAC library construction, was selected from the mapping population. 'Jefferson' was also chosen for Illumina next-generation sequencing, assembly, and annotation. We sequenced the 'Jefferson' genome at ~63x coverage and have completed a draft genome assembly, with continuing sequencing efforts aimed at improving the assembly. This resource will serve as a tool for gene discovery and functional studies, for the development of DNA markers and other genomic tools for breeders, and will allow integration of the genome sequences and the genetic and physical maps. Other genomic resources include the 'Jefferson' transcriptome assembly (Rowley *et al.*, 2012), and draft genome assemblies of seven additional diverse hazelnut accessions (**Table 2.1**) used to investigate genetic diversity and identify DNA markers for traits of interest.

Table 2.1 Statistics for 'Jefferson' reference genome assembly.

	Number	**Avg Length**	**N50**
Contigs	212,714	1831 bp	3,823

This re-sequencing effort has discovered millions of SNPs between the accessions and the 'Jefferson' reference genome. Hazelnut genome and transcriptome sequencing has provided new insights and promises many applications in accelerating or enhancing breeding efforts.

Materials and methods

Plant Materials. Tissues for DNA isolation were collected from field grown examples of 'Jefferson' and each of the 7 additional accessions used in this study. Leaves were collected in the early spring and frozen in liquid nitrogen. Tissues for RNA isolation were collected from both nursery and field grown trees: leaves from both nursery and field grown 'Jefferson' (OSU 252.146 x OSU 414.062) trees and bark from nursery grown trees. Catkins were collected from a field grown example of the "Barcelona' accession (PI 557037), while the whole seedlings were the progeny of the cross OSU 954.076 x OSU 976.091. Tissues were frozen in liquid nitrogen prior to RNA extraction.

Nucleic Acid Extraction and Preparation of Libraries for Illumina Sequencing. Total RNA was extracted, purified, and cDNA libraries were prepared for sequencing on the Illumina Genome Analyzer IIx as described previously (Rowley *et al.* 2012). Genomic DNA was extracted from all accessions using the DNeasy Mini kit (Qiagen) according to the manufacturer's instructions and sent to the Georgia Genomics facility at the University of Georgia for the construction of Illumina paired end sequencing libraries. Illumina cluster generation on the Illumina HiSeq 2000 was performed in the Oregon State University CGRB core facility using a standard Illumina protocol.

Assembly of Hazelnut Genomes and Transcriptome. Illumina reads representing the 'Jefferson' genome were initially filtered to remove reads having greater than 5% ambiguous basecalls (Ns) as called by the Illumina CASAVA pipeline, and further filtered to remove reads mapping to chloroplast, mitochondria, or aphid. The ~ 36 Gb of filtered 'Jefferson' PE reads were *de novo* assembled into contigs and initial scaffolds using the

assembly package SOAPdenovo (Li *et al.*, 2010), and further scaffolded using SSPACE (Boetzer *et al.* 2011), resulting in 212,714 scaffolds and contigs. The 7 additional accessions were assembled in the same manner (**Table 2.2**).

Table 2.2. Assembly statistics for 7 additional European hazelnut accessions sequenced.

Accession	Coverage	Avg. Length	# Contigs / Scaffolds	N50	% Reads Assembled	Origin
Barcelona	23.3x	381 bp	1.08 million	567	84.30%	Spain
Ratoli	36.5x	344 bp	1.37 million	482	81.40%	Spain
Tonda Gentile delle Langhe	23.9x	360 bp	1.17 million	520	84.80%	Italy
Tonda di Giffoni	31.6x	367 bp	1.22 million	543	84.40%	Italy
Daviana	38.5x	402 bp	1.07 million	639	86.60%	England
Halls Giant	34.7x	412 bp	1.05 million	673	86.80%	Germany
Tombul (Extra Ghiaghli)	23.3x	381 bp	1.08 million	567	84.30%	Turkey

The ‘Jefferson’ transcriptome was assembled as described previously (Rowley *et al.* 2012), resulting in 28,255 hazelnut transcript contigs.

Functional Annotation and SNP Discovery. Single nucleotide polymorphisms (SNPs) were detected between the accessions and ‘Jefferson’ using BWA (Li *et al.* 2009a) and Samtools (Li *et al.* 2009b). Gene prediction for the ‘Jefferson’ genome assembly was conducted using the program AUGUSTUS (Stanke *et al.*, 2003). Functional annotation of the transcriptome assembly was conducted as described previously (Rowley *et al.* 2012).

Results and Dissuction

Assembly and SNP Discovery Among European Hazelnut Genomes. The ‘Jefferson’ reference genome assembly is represented by 212,714 contigs and scaffolds with an average length of 1831 bp, a maximum length of 422,505 bp and an N50 of 3,823 bp (**Table 2.1**). Re-alignment of the ‘Jefferson’ genomic reads to the assembly demonstrated that 89% of the reads re-align, for an average sequencing depth of ~63x. Alignment of the transcriptome assembly to the genome demonstrated that 94.3% of the transcript assemblies align over 75% of their median length. A preliminary estimate using the gene prediction program AUGUSTUS (Stanke *et al.*, 2003) predicts 34,898 gene models. The genomic reads from 7 additional accessions were used to predict SNPs relative to ‘Jefferson’ (**Table 2.3**), resulting in the discovery of ~ 2 million SNPs within each accession that will be useful in the future establishment of molecular markers to aide in breeding efforts.

Table 2.3. Summary of SNPs and Indels detected between hazelnut accessions and ‘Jefferson’.

Accession	SNPs vs. ‘Jefferson’	# Indels
Barcelona	1,743,435	181,029
Ratoli	1,930,763	204,453
Tonda Gentile delle Langhe	1,923,231	190,675
Tonda di Giffoni	2,018,859	211,971
Daviana	2,188,995	239,250
Halls Giant	2,419,321	259,223
Tombul (Extra Ghiaghli)	2,037,429	214,267

In addition, relative to Jefferson each accession contains ~200,000 short insertions or deletions (Indels). A comparison between 3 accessions: 'Barcelona', 'Halls Giant', and 'Daviana' using BWA (Li *et al.*, 2009a) and Samtools (Li *et al.*, 2009b) demonstrate 547,828 SNPs conserved between them (data not shown).

Transcriptome Assembly. The transcriptome assembly, comprising 28,255 transcript contigs displaying high homology to available sequenced plant transcripts, has been functionally annotated with gene ontology (GO) terms as described (Rowley *et al.*, 2012). The assembly is available to the community as a public web-based database (http://hazelnut.cgrb.oregonstate.edu/), and will benefit breeders in the discovery of the genes responsible agronomic traits of interest, as well as facilitate the empirical annotation of future hazelnut genome assemblies, and assist comparative genomics efforts among cultivars.

Acknowledgements

We would like to thank the Georgia Genomics Facility at the University of Georgia for the preparation of additional Illumina libraries; Anne-Marie Girard, Caprice Rosato, Mark Dasenko, and Mathew Peterson for qualitative assessment, Illumina cluster generation, and computational support (Center for Genome Research and Biocomputing, Oregon State University); and Henry Priest for the filtering of the raw Illumina read data. This work was supported by grants from the Oregon State University Agricultural Research Foundation and the Oregon Hazelnut Commission.

Literature Cited:

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics., 27, 578-579.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754-60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics., 25, 2078-2079.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li,y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., and Wang, J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res., 20, 265-272.

Rowley, E.R, Fox, S.E., Bryant, D.W., Sullivan, C.M., Givan, S.A., Mehlenbacher, S.A., Mockler, T.C. 2012. Assembly and characterization of the European hazelnut (*Corylus avellana* L.) 'Jefferson' transcriptome. Crop Science., 52, 2679-2686.

Stanke, M., and Waak, S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics., 19, 215-225.

Chapter 3:

Analysis of global gene expression in *Brachypodium distachyon* reveals extensive network plasticity in response to abiotic stress

Henry D. Priest, Samuel E. Fox, Erik R. Rowley, Jessica R. Murray, Todd P. Michael, and Todd C. Mockler

PLoS ONE
1160 Battery Street
Koshland Building East, Suite 100
San Francisco, CA 94111, USA
9(1): e87499 (2014)
DOI: 10.1371/journal.pone.0087499

Abstract:

Brachypodium distachyon is a close relative of many important cereal crops. Abiotic stress tolerance has a significant impact on productivity of foodstocks. Analysis of the transcriptome of *Brachypodium* after chilling, high-salinity, drought, and heat stresses revealed diverse differential expression of many transcripts. Weighted Gene Co-Expression Network Analysis (WGCNA) revealed 22 distinct gene modules with specific profiles of expression under each stress. Promoter analysis implicated short DNA sequences directly upstream of module members in the regulation of 21 of 22 modules. Functional analysis of module members revealed enrichment in functional terms for 10 of 22 network modules. Analysis of condition-specific correlations between differentially expressed gene pairs revealed extensive plasticity in the expression relationships of gene pairs. Photosynthesis, cell cycle, and cell wall expression modules were down-regulated by all abiotic stresses. The modules up-regulated by salt and drought fell into unique gene ontology GO categories, whereas cold and heat up-regulated transcription factor expression and protein folding chaperone expression, respectively. This study provides genomics resources and improves our understanding of abiotic stress responses of *Brachypodium*.

Introduction:

Plants are sessile organisms that have evolved an exceptional ability to perceive, respond, and adapt to their environment. Environmental stresses are a major limiting factor in agricultural productivity (Wang *et al.*, 2003; Witcombe *et al.*, 2008), as plant growth is severely affected by environmental conditions such as cold, salt, drought, and heat (Hirayama and Shinozaki, 2010; Mahajan and Tuteja, 2005). In comparison to *Arabidopsis thaliana* and *Oryza sativa*, relatively little is known about how agriculturally important cereals (e.g., wheat, corn, barley) respond to abiotic stresses (Kilian *et al.*, 2007; Matsui *et al.*, 2008; Zeller *et al.*, 2009; Zhou *et al.*, 2007). The stress-induced transcriptomic responses of plants reveal the molecular mechanisms underlying the abiotic stress response. An understanding of these mechanisms will allow researchers to improve stress tolerance of food crops to enhance agricultural productivity under imperfect growing conditions to ensure the world's long-term food security (Araus, 2002; Ashraf *et al.*, 2010; Chew and Halliday, 2011).

The abiotic stress response occurs in two stages, an initial sensory/activation stage followed by a physiological stage during which the plant responds to the perceived stress (Mahajan and Tuteja, 2005; Zhu, 2001; Rowley and Mockler, 2011). Once a stress cue is perceived, secondary messengers such as calcium and inositol phosphates (Parre *et al.*, 2007) and reactive oxygen species (ROS) are produced. The increase in Ca^{2+} is sensed by various calcium-binding proteins that initiate phosphorylation cascades that subsequently activate transcription factors (Tuteja, 2007; Doherty *et al.*, 2009). Transcription factors in turn activate expression of stress responsive genes. This begins the second phase and elicits physiological changes necessary to survive the particular environmental stress (reviewed in Rowley and Mockler, 2010). The genes expressed and subsequent physiological changes induced during the second phase are dependent upon the particular abiotic stress encountered. These changes can include modifications to cell membrane components – resulting in changes in membrane fluidity –

(Moellering *et al.*, 2010), stomatal closure (Lopushinsky, 1969), decreased photosynthetic activity (Oliveira and Peñuelas, 2004; Brinker *et al.*, 2010), and increased production of heat shock proteins (HSPs) or dehydrin cryoprotectants (Mahajan and Tuteja, 2005).

Brachypodium distachyon is a temperate monocot grass with close evolutionary relationships to economically important species including temperate cereals, forage and turf grasses, and other species potentially useful as biofuel feedstocks. Genome projects for several monocot species, including rice, maize, sorghum, and wheat, have recently completed or are ongoing, but these species lack many attributes expected of model plants. In comparison to these other grass systems, *Brachypodium* has simpler growth requirements, a smaller genome (~272 Mbp), a shorter life cycle, and smaller physical stature. It is self-fertilizing and a large collection of naturally occurring diploid accessions are available. These characteristics make *Brachypodium* an excellent model grass system for global transcriptome profiling studies (Bevan *et al.*, 2010; Brkljacic *et al.*, 2011). Elucidation of the transcriptomic responses of *Brachypodium* to various abiotic stresses will provide essential information that may ultimately lead to significant advances in agricultural production including greater crop yields and more efficient production of biofuels (Chew and Halliday, 2011).

Previous work in monocot stress responses has been completed in rice (*Oryza sativa* ssp. *japonica* cv. 'Nipponbare' and ssp. *indica* cv. 'Minghui 63'). Expression levels of 20,500 transcriptional units in rice callus treated with abscisic acid (ABA) and gibberellin were evaluated using oligonucleotide arrays (Yazaki *et al.*, 2004). A more comprehensive approach using a microarray querying 36,926 genes was used to profile expression responses of rice to drought and high-salinity stresses in three tissues (Zhou *et al.*, 2007). Recently, profiling of transcriptional responses to cold stresses in winter barley was performed using a microarray-based approach (Janská *et al.*, 2011), and the transcriptional

responses of three wheat cultivars to cold stress was explored in a separate study using microarray-based approaches (Winfield *et al.*, 2010).

Here, we present a genome-wide survey of *Brachypodium* transcript-level gene expression responses to four abiotic stresses: heat, high salinity, drought, and cold. We found significant differences in responses of the *Brachypodium* transcriptome to the four abiotic stresses in terms of timing and magnitude. We were able to identify 22 modules, 10 of which defined clear biological processes. As expected from studies of other plant model systems, photosynthesis, cell cycle and cell wall expression modules were down-regulated under abiotic stress, whereas surprisingly, we found that the modules up-regulated by salt and drought fell into unique gene ontology (GO) categories, whereas cold and heat up-regulated transcription factor (TF) expression and expression of genes involved in stabilizing protein folding, respectively. The response of *Brachypodium* to heat, high salinity, drought, and cold stress was profiled over twenty-four hours after the onset of stress conditions. This study represents a significant development in genomics resources for *Brachypodium*, a close relative of many agriculturally and economically important cereal crop species.

Material and Methods:

Experimental Growth Conditions and Tissue Sampling. *Brachypodium distachyon* control plants were grown at 22 °C with 16 hours light and 8 hours dark in a controlled environment growth room. Abiotic stress conditions included cold, heat, salt, and drought. All treatments were conducted with a light intensity of 200 µmol photons $m^{-2}s^{-1}$. For the heat experiments, *Brachypodium* plants were placed in a Conviron PGR 15 growth chamber at 42 °C. Cold treatments were conducted in a walk-in cold room maintained at 4 °C. Salt stress (soil saturation with 500 mM NaCl) and drought (simulated by removing plants from soil and placing them on paper towels to desiccate) treatments were conducted under the same light and temperature as the control samples. Three-week-old *Brachypodium* plants were placed under the respective conditions two hours after dawn (10

a.m.). Leaves and stems (total above ground tissues) from individual plants were collected at 1, 2, 5, 10, and 24 hours after exposure to the abiotic stress.

RNA Preparation, Labeled cDNA Synthesis, and Microarray Hybridization. Leaf tissues were pulverized in liquid nitrogen, total cellular RNA was extracted using the RNA Plant reagent (Invitrogen), and RNA was treated with RNase-free DNase (essentially as described in Filichkin *et al.* 2010). DNase-treated RNA integrity was evaluated using an Agilent Bioanalyzer. Labeled target cDNA was prepared from 125 ng of *Brachypodium* leaf total RNA samples using the NuGen Applause WT-Amp Plus ST RNA amplification system kit (Cat# 5510-24) and Encore Biotin module V2 (Cat# 4200-12). All samples for this study were processed at the same time. Approximately 4.55 µg of fragmented cDNA from each sample was hybridized for 18 hours to an Affymetrix Brachypodium Genome Array (BradiAR1b520742) using a GeneChip® Fluidics Station 450. The Affymetrix eukaryotic hybridization control kit and Poly-A RNA control kit were used to ensure efficiency of hybridization and cRNA amplification. All three replicates of all times and treatments were hybridized concurrently. Microarray chips were scanned using GeneChip® Scanner 3000 with autoloader at 570 nm. Microarrays were quality-controlled according to the standard Affymetrix protocols (Affymetrix GeneChip® Expression Analysis Technical Manual, 701021 Rev. 5; http://www.affymetrix.com) at the Oregon State University Center for Genome Research and Biocomputing, Central Service Laboratory (detailed protocols are available at http://www.cgrb.oregonstate.edu/). Image processing and data extraction were performed using AGCC software version 3.0. Each array image was visually screened to discount for signal artifacts, scratches, or debris. One array – heat-stress hour 5 replicate C – did not pass quality control and was discarded.

***Mapping of Probes*.** Probes on the Affymetrix BradiAR1b520742 array were mapped to the Bd21 v1.0 assembly using the Burrows-Wheeler Aligner (BWA; Li & Durbin 2009). The Bd21 Brachypodium Array contains 6,503,526 non-control

probes. Of these, 99.81% (6,491,341 probes) map to a single location in the genome. Most of the probes (6,491,341) match their target sequences unambiguously with no mismatches in alignment. Only 12,183 probes align with mismatches. All probe sequences represented on the array are entirely distinct from each other. For the probe-set level analysis, probes were associated with annotated genic features. Probes that associated with a single gene's exonic features were collected into strand-specific probe-sets. Only those probe sets associated with the forward strand of a target gene were retained for analysis in differential expression or network prediction. If a probe was associated with exonic features of two genes (if two genes overlap, for instance), that probe was not assigned to any probe set. If a probe was associated with both intronic and exonic features (if a gene has multiple transcripts, or a probe spanned an exon/intron boundary), the probe was not assigned to a probe set. In the 47,960 genic probe sets, each gene was detected by, on average, 31.5 probes. The median number of probes per set was 22.

***Microarray Data Analysis*.** Probeset level expression values were obtained utilizing the Robust Multi-array Average (RMA; Irizarry *et al.* 2003) technique via the Affymetrix Power Tools (APT) software package (http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx). Probe set summarization and expression estimates for each gene were conducted using the apt-probeset-summarize tool (version 1.15.0) from Affymetrix. Data manipulations were performed using Perl scripts. From the resulting signal intensities, differentially expressed genes were calculated using the Significance Analysis of Microarrays (SAM; Tusher *et al.* 2001) R package in conjunction with Microsoft Excel.

SAM uses permutations of repeated measurements to estimate the percentage of genes that are identified by chance, representing the false discovery rate. SAM was run with default settings, using 100 permutations, using the 'two class unpaired' response type. The S_0 factor was estimated automatically and no fold-

change cutoff was applied. The Delta value was selected such that the median false-discovery rate was below 0.01. In every case, control and stress RMA expression values were compared in a pairwise fashion within a single stress and time point combination.

***Heatmap and Principle Component Analysis*.** Heatmap and PCA analyses were conducted in R. RMA expression differences between the average expression value per stress time point per treatment were set to saturate at a difference of 4 RMA (such that the maximum value reported in the heatmap was +/- 4 RMA). These expression differences were graphed using the 'heatmap.2' function of the gplots package of R. For principle component analysis, the average RMA expression value of each stress time-point, without the above saturation, was used as input for the 'PCA' function of the R package 'factominer' (http://factominer.free.fr/; Le *et al.* 2008).

***GO Analysis and Transcription Factor Annotation*.** Over-represented GO terms were identified using the AgriGO: GO analysis toolkit (http://bioinfo.cau.edu.cn/agriGO/; Du *et al.* 2010). Analysis was done by comparing the number of GO terms in the test sample to the number of GO terms within a background reference. Over-represented GO terms had a FDR corrected *P*-value of less than 0.05 and more than 5 mapping entries with a particular GO term. GO-terms were assigned to genes based first on InterProScan (Zdobnov and Apweiler, 2001) results for the entire predicted proteome of the *Brachypodium distachyon* MIPS version 1.2 annotation (The International Brachypodium Initiative, 2010). Approximately 40% of genes did not have any GO-terms associated with them. Gene products from this set that had high-quality BLASTP matches to *Arabidopsis thaliana* gene products were assigned the same set of GO terms that their *Arabidopsis* homolog possessed. This decreased the set of un-annotated *Brachypodium* genes to only about 14% of the transcriptome. *Brachypodium* loci were classified as TFs if they were identified by InterProScan

as containing a DNA-binding domain or were orthologous to *Arabidopsis* loci that were annotated as TFs.

***Network Analysis*.** Normalized RMA expression values for 9,496 differentially expressed genes were loaded into the R package WGCNA (Langfelder and Horvath, 2008). An adjacency matrix was calculated using B=23. Distance metrics between profiles were calculated using the TOMdist function using an un-signed TOM type. Hierarchical tree solution was calculated using the flashClust (Langfelder and Horvath, 2012) function with the 'method' option set to 'average'. Modules were called using the moduleNumber function, cutHeight=0.91, and minimum module size was set to 25. Module colors were set using labels2colors. These modules were merged, using mergeCloseModules, a cut height of 0.1, iteration set to 'true', and enabling re-labeling. Final module colors were set as a result of this merging. Modules were exported for visualization in Cytoscape (Smoot *et al.*, 2011) using the "exportNetworkToCytoscape" function in the WGCNA R package and an adjacency threshold of 0.35. Once imported to Cytoscape, edges were filtered for a minimum value of 0.45, and the final network layout was obtained using the "Force Directed" in-built Cytoscape layout method. Cytoscape-layout and edge filtering caused some modules to not be connected by edges. These were not included in final Cytoscape layout; however, their mutual connectivities in the adjacency matrix served to allow WGCNA to call them as modules so they were analyzed as such for AgriGO-mediated GO enrichment and for Element-mediated promoter analysis. Only those modules that were graphed in Cytoscape as being interconnected with edges above the 0.45 cutoff were included in the final figures.

Promoter Analysis. Genes were grouped based on module membership. Based on the MIPS version 1.2 *Brachypodium distachyon* annotation, the 500 nucleotides directly upstream of each gene was extracted from the *Brachypodium* genome. The promoters for the genes in each module were analyzed on a module-by-module basis using Element (Michael *et al.*, 2008). The set of all predicted

promoters in the genome were analyzed using the 'bground' command using all possible 5 to 8 nucleotide sequences as the set for analysis. The set of 5 to 8 nucleotide motifs comprise 43,520 unique sequences (each of which has a reverse complement), which served as a substrate for this analysis. Motifs shorter than 5 nucleotides in length are expected to fall into one of two categories – background false-discoveries or true-positives that will be contained within larger, also significant motifs. Transcription factor binding sites longer than 8 nucleotides in length are expected to either overlap or be multi-partite motifs, both of which will generate significant sub-motifs in this analysis. In some cases, for specific examples, membership lists from two modules were combined for analysis by Element. Element was run using default cutoffs for significance (FDR<0.01), on 16 processors ('-t 16').

Network Plasticity Analysis. Network plasticity was determined by comparing the correlation of gene pairs between conditions. Between two conditions, every gene that was called by SAM as being differentially expressed in both conditions was segregated into one of two groups – the TF group or the non-TF group. The list of putative *Brachypodium* transcription factors was obtained from gene annotation queries and BLASTP comparisons to rice (*Oryza sativa*) transcription factors obtained from Plant Transcription Factor Database (http://plntfdb.bio.uni-potsdam.de/v3.0/; Pérez-Rodríguez *et al.* 2010). All pairwise Pearson's correlation values were calculated between groups in each of the conditions. This yielded two correlation values for each gene pair – one value corresponding to each condition. The order of the values of each gene expression profile across all assayed stress conditions was then randomly shuffled via the Fisher-Yates Shuffle procedure (Fisher and Yates, 1963) creating 7,200 random permutations of the data. In each permutation, two subsets of equal size (N=15) were selected. Each permutation therefore was a random permutation of a gene's total expression data profile from which two independent samples of size N=15 were selected. The pairwise Pearson's correlations between all TF-TG pairs were calculated in each permutation. In order to determine significance of correlation change across

conditions, a cutoff was chosen such that the average number of genes pairs that had correlation changes exceeding that cutoff in each random permutation (average number of false discoveries per permutations) was an appropriately small ratio of the number of gene pairs that had correlation changes exceeding that threshold in the true dataset (number of positives). This process is similar to SAM (Tusher *et al.*, 2001). In all comparisons, the threshold was chosen such that the FDR was less than or equal to 0.05.

Accession Number. The raw data is available at the Plant Expression Database (www.plexdb.org) under PLEXDB accession number 'BD2'.

Results:

Overall differential gene expression. Drought, high-salinity, cold, and heat are four important abiotic stresses that adversely affect the productivity of plants. We surveyed *Brachypodium* transcript-level gene expression responses to these stresses using the Affymetrix *Brachypodium* Genome Array (BradiAR1b520742). This microarray queries all annotated genes in the *Brachypodium* genome with multiple individual probes targeting each gene. The response of *Brachypodium* to heat, high salinity, drought, and cold stress was profiled in an asymmetric time-course over the twenty-four hours immediately following onset of stress conditions. This allowed us to monitor the transcriptional responses of the plant to stress rather than the endogenous circadian or diurnal rhythms. Biological triplicate samples were taken from control and stressed plants at each time point.

Significance Analysis of Microarrays (SAM; Tusher *et al.* 2001) was used to define genes differentially expressed in response to drought, salt, cold, and heat relative to unstressed plants. After analysis, genes were surveyed for fold-change levels. Overall, 3,105 genes were up-regulated by more than 2 fold, and 6,763 genes were down-regulated by more than 2 fold in response to at least one abiotic stress. In response to cold, heat, salt, and drought stresses 40, 1,621, 1,137, and 5,790 genes were down-regulated, respectively. In contrast, 447, 458, 1,565, and

2,290 genes were up-regulated in response to cold, heat, salt, and drought stress, respectively.

Not surprisingly, the overall number of genes differentially expressed in each stress condition increased over time (**Figure 3.1**); however, the directionality of differential expression differed strikingly with the type of stress. The cold stress response consisted almost entirely of up-regulated genes; very few genes were down-regulated at twenty-four hours (**Figure 3.2,** top left).

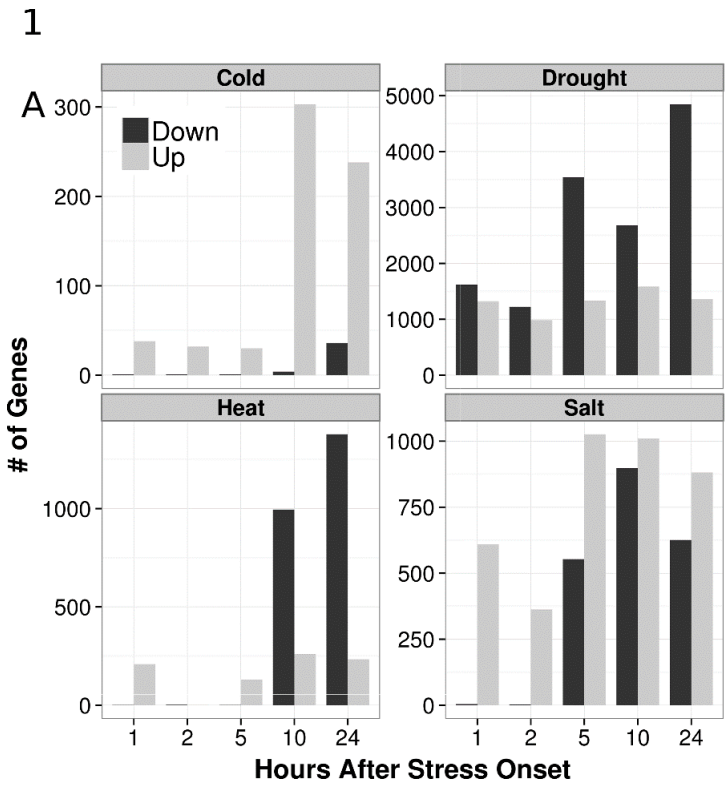


Figure 3.1. Numbers of genes up-regulated and down-regulatedshown as a function of time after stress onset

In contrast, the response to heat stress was primarily down regulation (**Figure 3.2,** bottom left). Up-regulation of certain genes in response to heat stress response was observed after 1 hour, but no significant differential expression was observed

at 2 hours after onset of stress. After 10 and 24 hours of heat treatment, more than 1,000 genes were down-regulated. Between 1,000 and 2,000 genes were up-regulated at all time points of drought treatment (**Figure 3.2**, top right). Down-regulation of genes was low in the early phases of drought response and increased drastically as the treatment continued beyond 2 hours. More than 2,500 genes were differentially expressed 5, 10, and 24 hours after drought onset. Early in the response to salt stress, only up-regulation of genes was observed. At 5 hours post-onset, down-regulation was observed in conjunction with up-regulation with neither dominant as was seen in the other three stresses (**Figure 3.2,** bottom right).

Principal Component Analysis (PCA) is a method of transforming a large dataset onto a coordinate system that defines the common trends of variation present in the data. PCA of the array dataset showed that the transcriptional responses of *Brachypodium distachyon* to the four assayed stresses were fundamentally distinct (**Supplemental Figure 1**). Whereas a significant portion of variation in the transcriptional responses to salt and drought stresses could be attributed to the first principal component, heat stress was most strongly represented by the second principal component, and the cold stress response was nearly completely orthogonal to the other three stresses. Overall, 64% and 12% of the total variance in the expression data were explained by the first two principal components, respectively. Drought and salt stresses yielded the most similar patterns of variance, whereas the cold and heat stress responses differed strongly from each of the other two stresses and from each other. Similarities are observed in the heatmap depicting the hierarchical clustering of the expression data (**Figure 3.2.**) in which the Robust Multi-array Average (RMA; Irizarry *et al.* 2003) expression value differences between mRNA abundances in control and stress-treated plants are plotted for all stress conditions. The overall similarity between the salt and drought stress responses can also be seen in this hatmap and is reflected in the PCA results (**Supplemental Figure 1**).

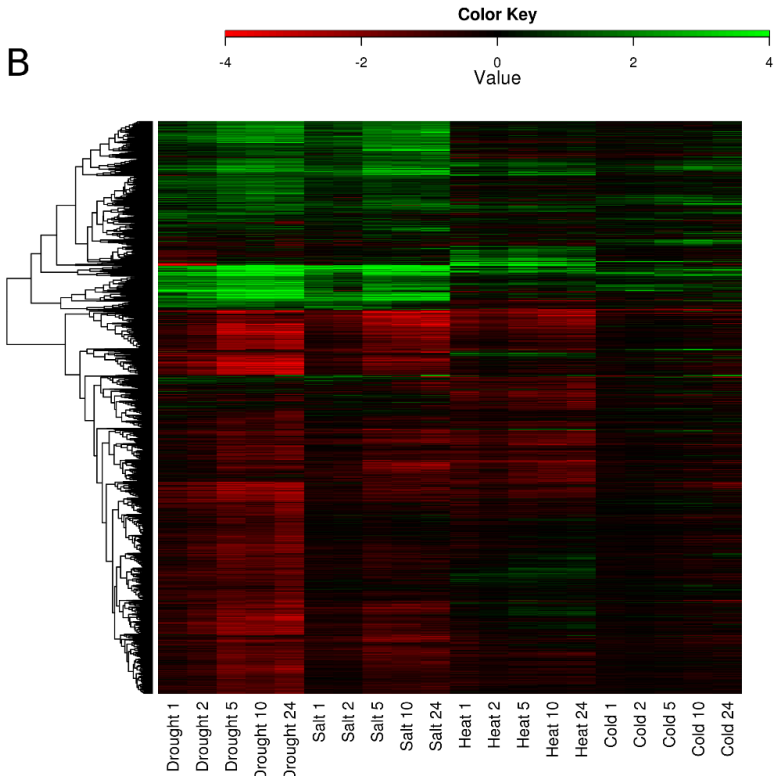


Figure 3.2. Heatmap of expression differences between control and indicated stress assays

A large number of genes are differentially expressed only under drought stress (purple ovals, **Figures 3.3 and 3.4**).

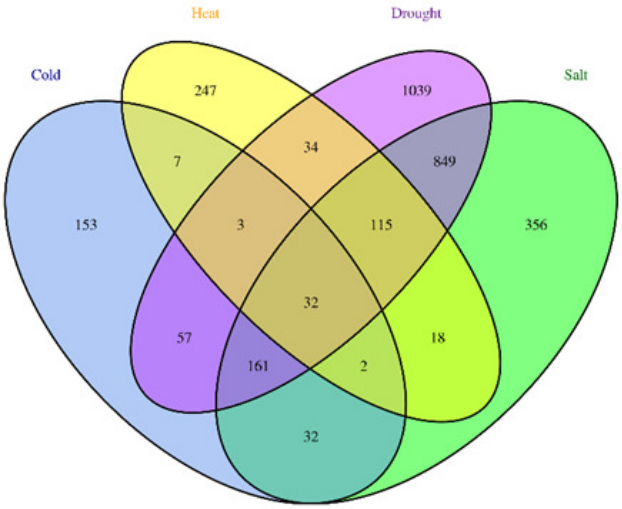


Figure 3.3. Venn diagram showing overlap of up-regulated genes in response to the four assayed abiotic stresses

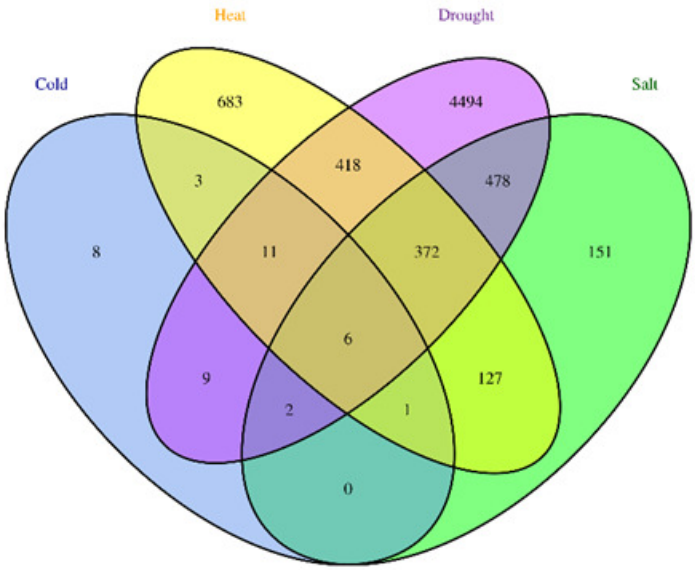


Figure 3.4. Venn diagram showing overlap of down-regulated genes in response to the four assayed abiotic stresses

In response to drought treatment, 1,039 genes were up-regulated and 4,494 were down-regulated. Only about half of the genes differentially expressed in the heat treatment were also responsive to drought (1,088 of 2,079 genes responsive to heat were also responsive to drought). Further, 44.7% of all genes differentially expressed in response to heat stress were unique to that response (930 of 2,079,

compare yellow to purple ovals in **Figures 3.3** and **3.4**). Only about 25% of genes differentially expressed upon salt treatment were independent of the drought response (687 of 2,702), and even fewer were unique to salt (507 of 2,702, 18.8%; compare green to purple ovals in **Figures 3.3** and **3.4**). The response to extended cold treatment had strong overlap with the drought response as well. Only 206 genes were responsive to cold stress and not to drought treatment (206 of 487, 42.3%), and 161 genes (of 487 differentially regulated by cold relative to unstressed plants) were uniquely regulated by cold stress (compare blue to purple ovals, **Figures 3.3** and **3.4**). From these analyses, the complex nature of the timing of gene regulation in response to stresses (**Figure 3.1**), the differences in intensities of differential expression in response to stresses (**Figure 3.2**), and the extensive overlap among genes regulated during stress responses (**Figures 3.3** and **3.4**) are apparent.

***Network Analysis of Stress Response in* Brachypodium.** In order to further analyze the systematic transcriptional responses of *Brachypodium* to abiotic stresses, we performed weighted gene co-expression network analysis (WGCNA) on data collected on the 9,496 differentially expressed genes using the WGCNA package in R (Langfelder and Horvath, 2008). Gene co-expression network analysis reveals gene pairs that share similar expression profiles. Gene modules are composed of genes that share similar profiles and have high correlations with each other. The weighted interaction network is shown in **Figure 3.5**. Nodes (genes) are connected by edges (co-expression relationships).

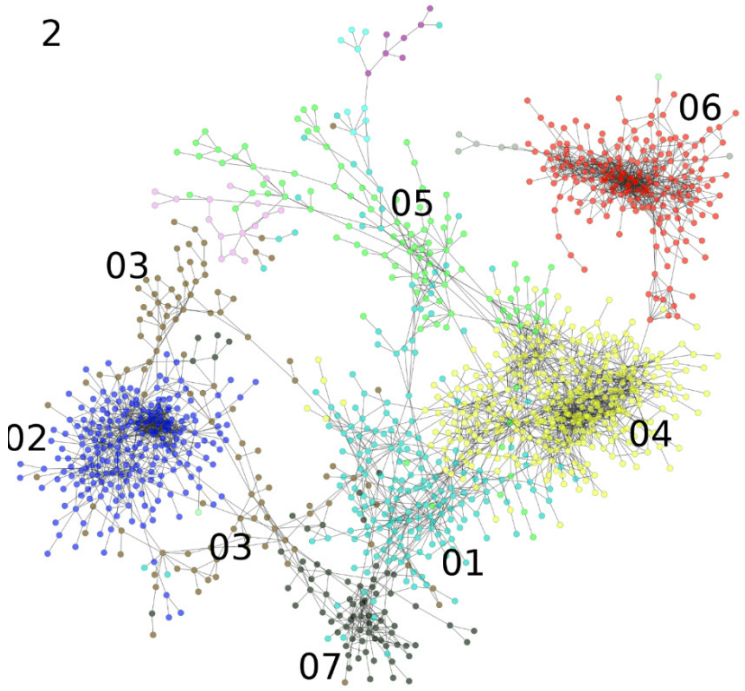


Figure 3.5. Weighted co-expression network of Brachypodium stress responsive genes

The connection between two nodes was determined by the correlation between the expression levels of the genes those nodes represent across all experiments used in the analysis. Correlation values were modified by a 'soft-thresholding' process that weights edge values dependent on their correlations (see Langfelder & Horvath 2008 for details) to generate adjacency values. Adjacency values vary between 0 and 1. The dendrogram depicting the hierarchical clustering and module assignment of all genes is shown in **Supplemental Figure 2**. The spatial layout depends on the adjacency value of the connection between genes, with higher adjacency values giving stronger connections and less spatial distance between two genes. Groups of nodes that are highly interconnected can still be called modules without any of those edges between the nodes themselves being particularly strong. Alternatively, the nodes may be connected by strong edges but have few, or zero, strong edges outside the module, which results in an isolated graph unconnected to the larger network.

This analysis resulted in a network that grouped 6,399 genes into 22 modules, the most strongly interconnected of which are shown in **Figure 3.2**. The expression profile of each module is shown in **Figure 3.6** as the average difference in RMA expression level between treatment and control arrays.

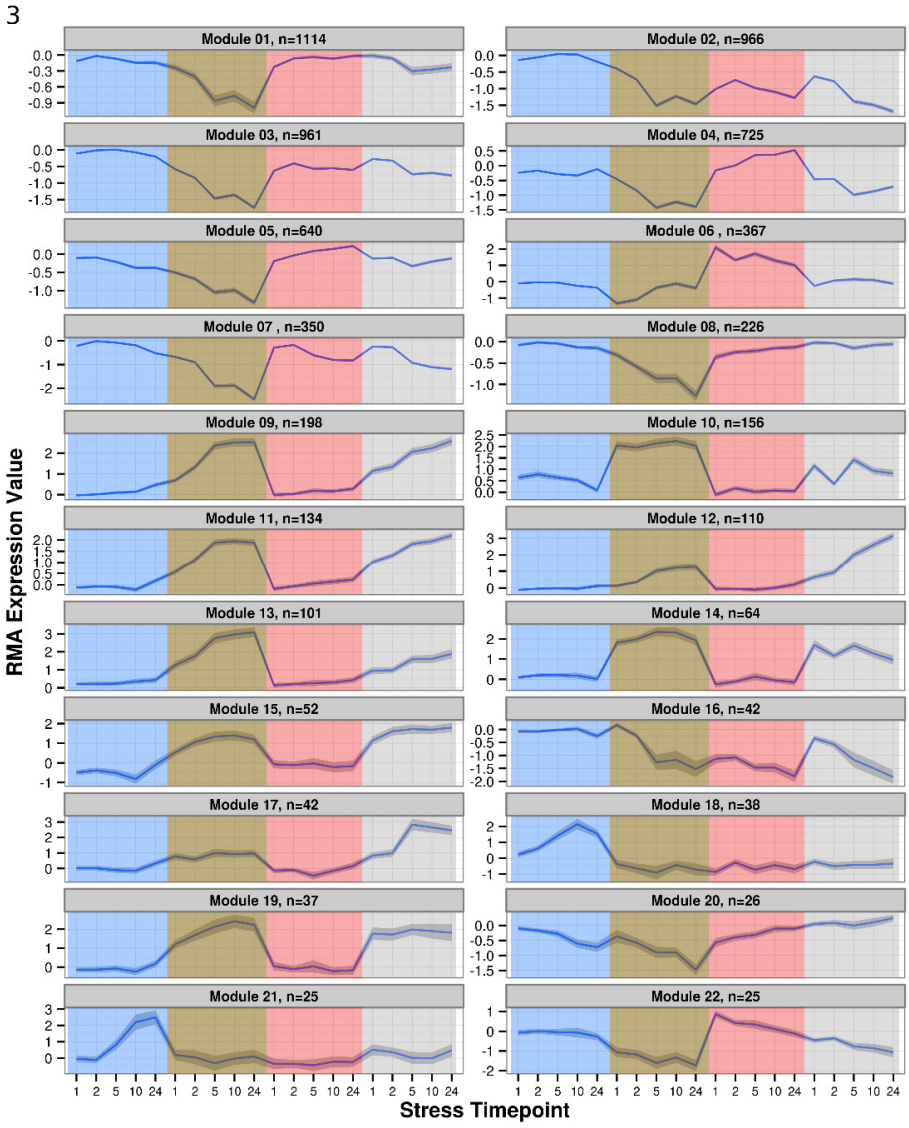


Figure 3.6. Expression profiles of modules as a function of time

The modular response of *Brachypodium* to abiotic stress was dominated by expression changes in response to the drought stress (**Figure 3.6**).

Differential expression of modules in response to stress was determined by a requirement that an average expression profile must differ from that of the control by one RMA-normalized expression value at one time point under the given stress. Using this criterion, only one module was not responsive to drought stress (module 21; **Figure 3.6**, lower left). Nineteen of the 22 modules were either stress-specific in their response or responded to only one other stress in addition to drought stress. The remaining three modules are module 16, module 02, and module 07, which were all down-regulated in response to heat, high salinity, and drought stresses. No module was responsive to all four abiotic stresses. The complete lists of all genes in each of the 22 modules may be found in **Supplemental File 1**.

***Functional annotation and promoter analysis*.** The combination of the functional annotations of the genes that comprise these modules with their expression profiles shed light on how the plant responds to abiotic stress conditions, but co-regulation is undoubtedly achieved through a combination of transcriptional and post-transcriptional regulation. The grouping of genes facilitated direct analysis of promoters to identify condition-specific over-represented *cis*-regulatory DNA elements. To assign functions to the modules, the module gene lists were analyzed using AgriGO (http://bioinfo.cau.edu.cn/agriGO/analysis.php; Du *et al.* 2010). We also analyzed 500 nucleotides from the promoter regions of each of the genes in each module using the Element software package to identified over-represented DNA elements (Michael *et al.*, 2008). The module wide enrichment of GO terms and DNA sequences contained in promoters is shown in **Table 3.1**

Table 3.1. Module wide enrichment of GO terms and DNA sequences

Module	N	Unique GO terms	Total GO terms	Unique DNA Elements	Total DNA Elements
Module 01	1114	20	81	60	235
Module 02	966	59	75	299	441
Module 03	961	27	53	56	208
Module 04	725	55	101	323	504
Module 05	640	0	0	90	225
Module 06	367	11	13	107	151
Module 07	350	54	110	97	145
Module 08	226	0	0	5	24
Module 09	198	1	7	12	45
Module 10	156	0	15	190	354
Module 11	134	3	4	0	8
Module 12	110	0	0	32	69
Module 13	101	0	0	8	50
Module 14	64	0	0	9	37
Module 15	52	0	0	4	12
Module 16	42	1	2	3	17
Module 17	42	0	0	8	13
Module 18	38	4	25	1	15
Module 19	37	0	0	1	26
Module 20	26	0	0	0	0
Module 21	25	0	0	6	12
Module 22	25	0	0	1	1

There was a moderate correlation between the number of genes in the module with both the number of GO terms and with the number of DNA sequence elements found to be enriched within that module (Pearson's r: 0.616 and 0.755, respectively). This general correlation between module size and enrichment discovery is expected; however, there were exceptions to this general trend. For example, module 05 is ranked fifth in module size, with 640 member genes, but was not enriched for any GO terms (**Table 3.1**), although most (585) genes were associated with at least one GO term. Eleven modules were not enriched for any GO terms, and twelve were not uniquely enriched for any GO terms. The modules with no GO-term enrichment varied in size from the minimum size (N=25) to 640 members (module 05) (**Table 3.1,** column 'N'). The complete set of all AgriGO output for all 22 modules may be found in **Supplemental File 2**. Upon examination of the GO-terms enriched in each particular module, a pattern of enrichment was often apparent. A selection of the GO-terms enriched in each module, along with the relevant statistics, is shown in **Table 3.2**.

Table 3.2. Specific GO terms uniquely enriched in a selection of modules.

Module	GO-term	Description	FDR
Module 01	GO:0031969	chloroplast membrane	0.0083
	GO:0006418	tRNA aminoacylation for protein translation	8.80E-06
	GO:0006800	oxygen and reactive oxygen species metabolic process	0.022
	GO:0005525	GTP binding	0.039
	GO:0016875	ligase activity, forming carbon-oxygen bonds	1.90E-05
Module 02	GO:0007049	cell cycle	0.0059
	GO:0006260	DNA replication	3.30E-5
	GO:0034728	nucleosome organization	0.00045
	GO:0009832	plant-type cell wall biogenesis	0.00063
	GO:0000271	polysaccharide biosynthetic process	0.016

Module 04	GO:0003899	DNA-directed RNA polymerase activity	7.8E-07
	GO:0006281	DNA repair	0.00082
	GO:0033279	Ribosomal subunit	3.40E-13
	GO:0006364	rRNA processing	1.60E-9
	GO:0008026	ATP-dependent helicase activity	0.00091
Module 06	GO:0031072	heat shock protein binding	0.0012
	GO:0006457	protein folding	2.00E-21
	GO:0009408	response to heat	4.40E-19
	GO:0050896	response to stimulus	4.70E-04
	GO:0010035	response to inorganic substance	0.0043
Module 07	GO:0015979	Photosynthesis	3.20E-45
	GO:0033014	Tetrapyrrole biosynthetic process	1.9E-10
	GO:0006091	generation of precursor metabolites and energy	2.60E-21
	GO:0009765	photosynthesis, light harvesting	2.90E-18
	GO:0010114	response to red light	1.9E-06
Module 09	GO:0009415	response to water	0.0094
Module 11	GO:0009072	aromatic amino acid family metabolic process	0.0062
	GO:0022804	active transmembrane transporter activity	0.038
Module 18	GO:0006351	transcription, DNA-dependent	0.0018
	GO:0016070	RNA metabolic process	0.0076
	GO:0065007	biological regulation	0.0084
Module 16	GO:0016740	transferase activity	0.0088

Even in small modules with the minimum number of genes and no GO-term enrichment, we found over-representation of certain DNA sequences in member gene promoter sequences. Only module 20 was not enriched for any GO terms

and had no over-represented DNA elements (**Table 3.1**). The over-representation of short regions of DNA sequence in the promoters of module member genes may provide insight into the transcriptional circuitry that mediates the regulation of the module. Twenty-one modules had at least one significantly over-represented DNA element (FDR-corrected p-value <0.01). Only two modules had no unique significantly over-represented DNA elements (**Table 3.1**, modules 11 and 20). Nine of the 22 modules had at least 32 unique elements over-represented in the promoters of their member genes (**Table 3.1**, column 'Unique DNA Elements'). Especially in conjunction with the functional annotation of modules via GO-term enrichment, the specific DNA elements which were uniquely enriched show how the transcriptomic responses of *Brachypodium* to abiotic stress compare to other plant systems (**Table 3.3**).

Table 3.3. Specific short DNA sequences found to be statistically enriched in the promoters of module member genes.

Module	DNA Element	Number of Hits	Number of Promoters	FDR
Module 01	TTAAAAA	346	267	4.94E-08
	TTTAAAA	301	197	1.71E-07
	CTCGTC	423	342	3.52E-05
	ACGTGGGC	139	120	6.03E-05
	CGGCC	380	299	4.80E-05
Module 02	CAACGGTC	57	48	3.79E-17
	AACGGCT	90	79	1.02E-09
	AGCCGTTG	47	39	2.43E-09
	CCAACGG	121	104	2.43E-08
	CAACGGC	115	98	5.38E-05
Module 04	AAACCCT	311	248	2.02E-69
	AGCCCAA	161	134	1.86E-14
	AGGCCCA	211	169	1.02E-28
	AAGCCCAT	57	50	2.57E-11
	GCCCAAC	115	100	1.86E-08
Module 05	ACAAAA	550	345	2.00E-05
	CAATA	617	368	7.05E-08
	ACAATA	197	151	4.04E-05
	ACAATAA	80	71	6.02E-06
	AATAA	1078	463	1.71E-05
Module 06	GAACCTTC	33	30	3.47E-15
	CTAGAAG	55	46	9.78E-11
	CTTCCAGA	28	26	3.98E-10

	AAGCTTC	61	40	1.01E-07
	GAAGCTTC	20	20	1.04E-06
Module 07	ACGTGGC	69	55	4.83E-12
	CCACGTC	59	53	1.39E-07
	GACGTGGC	25	21	5.88E-06
	CACGTGGC	26	20	1.27E-06
	CCTATC	92	81	1.12E-09
	GGGATA	83	78	7.11E-07
	AGATAA	126	105	0.00026
Module 09	ACGTAT	50	32	3.91E-05
	ACGTATA	23	14	1.14E-05
	ACACGTA	31	28	1.38E-06
	CACGTAC	36	28	1.29E-05
	CGTAA	118	83	0.000276
Module 10	CGATCG	47	35	0.00227
	CCGATCG	28	18	0.00049
	ATCGC	122	83	0.00424
Module 12	GTACGTA	27	13	6.08E-06
	GTACAC	41	36	1.44E-05
	ACGTACG	27	14	2.08E-05

In total, 1,312 elements of 5 to 8 nucleotides long were uniquely associated with specific modules (**Supplemental File 3**). Element output pertaining to significant DNA motifs can be found in **Supplemental File 3**.

***Network plasticity*.** Plasticity of gene regulatory circuits is an expected property of biological systems. There are multiple methods by which the expression relationship between a regulator gene and a target gene may change in response to conditions. The regulatory relationship between such gene pairs may change as a result of chromatin rearrangement or DNA methylation (Bai and Morozov, 2010; Lauria and Rossi, 2011), both of which have been shown to be responsive to stress in plant species (Zhong *et al.*, 2009; Mukhopadhyay *et al.*, 2013). It is also conceivable that the abundance of the mRNA encoding a particular regulator could be detached from the target expression levels by protein modifications that alter either the activity or degradation rate of the protein in question (Marino *et al.*, 2013; Lindemose *et al.*, 2013). The expectation that a transcription factor and

target gene pair which interacts will generate correlated expression measurements may not reflect biological reality in all cases.

Figure 3.4 shows heatmap-scatterplots of transcription factor/target gene (TF-TG) pairs in correlation space.

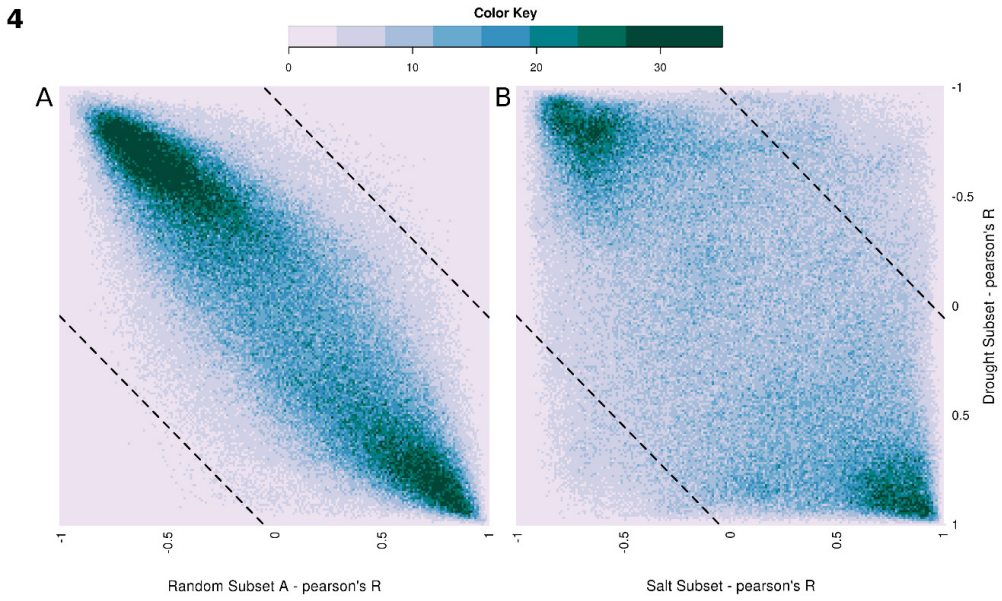


Figure 3.4 Heatmap-scatterplots of transcription factor/target gene (TF-TG) pairs in correlation space.

TF-TG pairs are plotted according to their pairwise correlations in each of the shown conditions. Transcription factor/target gene pairs are defined as all possible pairings of genes differentially expressed in the two conditions of interest. Transcription factors are defined via a combination of sequence homology and InterProScan results (see Methods; Zdobnov & Apweiler 2001). The x-coordinate of a TF-TG pair is determined by the pairwise Pearson's correlation between that TF-TG pair in the indicated subset of stress data. The y-coordinate of that TF-TG pair is determined by the pairwise Pearson's correlation of that pair in the subset of stress data drawn from the drought experiment. The heatmap value is determined by the total number of TF-TG pairings with any particular combination of correlations. **Figure 3.4A** shows the distribution of pairwise TF-TG correlation changes between a random subset of the stress data and the subset of data drawn from the drought experiment, as an indication of what would be expected based on random changes of expression patterns. **Figure 3.4B** shows the distribution of pairwise TF-TG correlation changes between salt and drought stress data subsets.

In the salt-drought comparison, 146 TFs and 1910 non-TF genes were differentially expressed under both stress conditions. Based on the calculated threshold of $\Delta r = 0.97$ for the salt and drought comparison (see Methods), 27,916 of 276,950 TF-TG pairings (10.1%, **Table 3.4**) showed significant differential correlation across conditions, indicating possible plasticity in the relationship between the TF and TG of the pair (**Figure 3.4B**, top right and bottom left).

Table 3.4. Putative network plasticity present between all pairwise conditional comparisons**.**

Stress A	Stress B	Gene Pairings	Plastic Pairs	Average False Positives	FDR	Δr cutoff
Drought	Salt	276,950	27,916 (10.1%)	1368.1	0.049	0.97
Drought	Cold	16,665	2,921 (17.5%)	144.9	0.049	0.96
Drought	Heat	70,434	4,890 (6.9%)	239.9	0.049	0.98
Salt	Heat	26,562	241 (0.9%)	11.9	0.049	1.35
Salt	Cold	8,132	2,027 (24.9%)	94.8	0.047	0.94
Heat	Cold	522	128 (24.5%)	6.0	0.047	0.88

The remaining 249,034 gene pairings showed less than significant changes in correlation across conditions. **Figure 3.4A** shows a representative distribution of correlation changes between gene pairs populated by a random permutation of the same data underlying **Figure 3.4B**. In distributions created by random permutation, an average of 1368.1 gene pairs per permutation were found to have significant changes in correlation based on the threshold of $\Delta r = 0.97$ for the same salt-drought comparison, corresponding to the targeted maximum FDR of 0.05 or less (**Table 3.4**). In all pairwise stress condition comparisons, between 0.9% and 24.9% of gene pairings were found to have potentially plastic relationships (salt/heat and salt/cold, respectively, **Table 3.4**).

Discussion:

***Stress responsive modules in* Brachypodium *transcriptional circuitry*.** The motivations behind linking groups of genes to specific expression profiles in response to stress are multifold. First, modules represent regulatory relationships, indicating how *Brachypodium* reacts in a transcriptional and post-transcriptional manner to abiotic stresses. Second, the expression profiles themselves allow interrogation of the transcriptional regulatory circuitry that allows *Brachypodium* to achieve steady-state levels of stress-responsive transcripts at the appropriate time. This provides links between specific sequences present in the upstream regions of genes, key regulators (e.g. transcription factors), and traits of agricultural and economic interest.

Of all differentially expressed genes, 3,097 (32.6%) were not associated with a module. Different applications of stress, stress treatment severity, temporal distribution of sampling, and temporal density of sampling may enable association of many of these genes with these or other modules to more completely describe the stress response system of *Brachypodium*. Here, four abiotic stress treatments were used: hear, drought, salt, and cold. We did not examine abiotic stresses such as high intensity light, UV, or chemical inducers of

reactive oxygen species (ROS). With data on additional stresses, we will be able to associate more genes over-arching modes of stress response.

Photosynthesis. Several sub-systems in plants are affected by multiple stresses. Photosynthetic activity (either capacity or efficiency) is known to be down-regulated or depressed upon heat stress (Salvucci *et al.*, 2001), drought stress (Aranjuelo *et al.*, 2011), salt stress (Brinker *et al.*, 2010), and cold stress (Oliveira and Peñuelas, 2004). One of the modules we identified, module 07 (**Figure 3.3,** top left), is comprised of 350 genes that are very strongly enriched for genes annotated with GO-categories related to photosynthesis, chlorophyll biosynthesis, light response and harvesting, and the chloroplast (**Table 2**, **Supplemental File 2**). For example, of the 143 genes in *Brachypodium* annotated with GO:0015979 'Photosynthesis', 50 are present in this module (a significant enrichment with FDR-corrected p-value of 3.2×10^{-45}). This module was down-regulated in drought, heat, and salt stresses (**Figure 3.6**). This indicates that under abiotic stress *Brachypodium* down-regulates photosynthesis as observed in several other plant systems (Oliveira and Peñuelas, 2004; Brinker *et al.*, 2010; Aranjuelo *et al.*, 2011; Salvucci *et al.*, 2001). As these genes associated with photosynthesis are affected by several stresses in a coordinated manner, these stresses likely modulate a common transcriptional circuit.

The ABRE (ACGT-containing abscisic acid response element) is a known *cis*-regulatory motif in *Arabidopsis thaliana* that contains an ACGT core and is responsive to drought (Fujita *et al.*, 2011; Hattori *et al.*, 2002). This sequence was found in the promoter regions of many genes in the photosynthesis module (module 07), the water-response module (module 09, **Table 2**) and a transcription factor enriched module (module 10, **Table 2**, **Supplemental File 2**). Notably, even though the photosynthesis module and the signaling module (module 03) share highly similar expression profiles, this core sequence was not significantly enriched in the promoters of genes in the signaling module. The photosynthesis module is down-regulated under drought stress, whereas modules 09 and 10 are

up-regulated under the same stress (**Figure 3.6**). Thirteen variations of the ABRE (including the ACGT core with differing flanking regions) were found in the photosynthesis module (**Table 3**, **Supplemental File 2**). Negative regulation of the photosynthesis module by the ABRE in response to drought stress was expected based on previous studies (Kim and Portis, 2005; Flexas *et al.*, 2007; Chaves *et al.*, 2009). Forms of the ABRE were also over-represented in the promoters of genes in modules 11, 12, 13, 14, 15, and 19. These modules were not found to be over-represented for any GO-terms. However, these modules were up-regulated by both salt and drought stresses. The functional roles of these modules remain to be explored.

Late embryogenesis abundant (LEA) proteins have long been studied in connection with abiotic stress (Hong-Bo *et al.*, 2005) and are implicated in abiotic stress in many plant systems, principally in response to elevated ABA levels (Liu *et al.*, 2013; Shinde *et al.*, 2012). Seventy-seven genes are annotated due to gene sequence homology (BLAST; Altschul *et al.* 1990; Camacho *et al.* 2009) or protein domains/motifs (InterProScan; Zdobnov & Apweiler 2001) as encoding LEA-proteins. The RMA-expression value differences were plotted for these 77 genes in all conditions (**Supplemental Figure 3**). This analysis revealed a set of 15 putative LEA loci that are strongly up-regulated in response to drought and salt stress (**Supplemental Figure 4A**). The promoters of these genes were analyzed using Element. This analysis revealed statistical enrichment for five sequences of 5 to 8 nucleotides in length, all with FDR-corrected p-value $< 1.44 \times 10^{-7}$ (**Supplemental Figure 4B**). These sequence motifs have been previously described as the 'coupling element' CE3 (Shen *et al.*, 1996; Gómez-Porras *et al.*, 2007). The ABRE-core containing 'ACGTG' was enriched in 11 parent sequence motifs in the promoters of the module 10 (**Supplemental File 3**). The sequence 'ACGCG' was enriched in 18 parent sequences in the promoters of the same module. These results indicate that the previously described nature of the ABRE, including the requirement for a CE3 motif (Shen *et al.*, 1996; Gómez-Porras *et al.*, 2007), is conserved in *Brachypodium*. Further experimentation is needed to

investigate the specific function of the CE3 motif and to identify transcription factors associated with it.

The photosynthesis module (**Figure 3.2, Table 3.2**) is strongly enriched for genes related to photosynthesis and was severely down-regulated in drought and moderately down-regulated in heat and salt stresses. These genes were not down-regulated in cold stress, but the overall depression of photosynthesis-related genes appears to be conserved in *Brachypodium* (**Figure 3.3**, top left). The relative strength of the stress conditions applied no doubt plays a role in the relative levels of regulation observed for this module.

Plant growth. Plant growth is severely affected by environmental conditions such as cold, salt, drought, and heat abiotic stresses (Hirayama and Shinozaki, 2010; Mahajan and Tuteja, 2005). Module 02 (**Figure 3.6**) is characterized by an expression profile similar to the photosynthesis module (module 07), though it shows larger negative changes in expression under both salt and heat stress treatments. Module 02 is enriched for genes annotated with GO-terms related to DNA replication, chromatin and nucleosome assembly, the cell cycle, and cell wall biogenesis (**Table 2**, **Supplemental File 2**). The down-regulation of these genes suggests that an early response of *Brachypodium* to abiotic stresses is to suppress cell growth, DNA replication, and the cell cycle.

The Mitosis-Specific Activator (MSA) motif includes the core sequence 'AACGG' and is associated with G2/mitosis specific genes in *Arabidopsis* (Haga *et al.*, 2011). *At*MYB3R4 has been shown to directly bind to this motif *in vitro* (Haga *et al.*, 2011). Module 02 is enriched for GO categories related to DNA replication, microtubule-based processes, chromatin, and nucleosome assembly. Thus, the 'cell-cycle' module is down-regulated under stress, indicating a suppression of these systems, which may result in a lengthened G2 phase and a slowed cell cycle. The promoters of the cell-cycle module are heavily enriched with the 'AACGG' core of the MSA motif, as well as its reverse complement (**Table 3.3**). Notably, the sequence 'AACGG' was found 907 times in 540 of the

966 gene promoters in this module (FDR-corrected p-value = 0.00043). Six distinct 8-nucleotide sequences containing this core were found 275 times (all six with FDR-corrected p-value <3.94 x 10^{-5}, **Table 3.3**). This core was also enriched in module 10; we observed this sequence 168 times in 95 of the 156 promoters (FDR-corrected p-value = 0.001, **Supplemental File 3**). Small plant stature and decreased yield are a major consequence of abiotic stress in plants (Hirayama and Shinozaki, 2010; Mahajan and Tuteja, 2005). A decrease in expression of genes activated by the MSA motif could conceivably result in a much slower or completely suspended cell cycle in the G2 phase. *Arabidopsis* plants deficient in TFs associated with the MSA showed pleiotropic dwarfism and other developmental and growth defects (Haga *et al.*, 2011). The putative ortholog of *At*MYB3R4, *Bradi2g31887*, is a member of the signaling module (module 03). The signaling module is also enriched for microtubule related GO-terms, as well as many signaling-related GO-terms. However, none of the unique significantly enriched DNA sequence elements present in the promoters of module 03 contain the MSA core nor is the MSA core itself enriched in gene promoters from this module (**Table 3, Supplemental File 3**). Elucidation of the relationship between the MSA and TFs such as that encoded by *Bradi2g31887* that may bind the MSA and suppression of the cell cycle by down-regulation of MSA-controlled genes will require further study.

Cold response. Families of ice-recrystallization inhibition proteins (IRIPs) and C-repeat binding factors (CBFs) have been previously identified in *Brachypodium* (Li *et al.*, 2012). Of the seven *Brachypodium* CBFs genes, five were up-regulated in response to cold (*Bradi4g35630*, *Bradi1g57970*, *Bradi3g35570*, *Bradi4g35600*, and *Bradi1g77120*), *Bradi3g35570* was up-regulated in response to salt stress, and *Bradi1g77120* was down-regulated in response to drought stress. Only *Bradi1g57970* was assigned to a network module – module 18. The three IRIP loci were strongly responsive to stress. All three (*Bradi5g27330*, *Bradi5g27340*, and *Bradi5g27350*) were up-regulated in response to cold stress, and two (*Bradi5g27330* and *Bradi5g27340*) were up-regulated in drought and salt stresses

and down-regulated in heat stress. Only one gene (*Bradi5g27330*) was a module member; in this case member of module 21 (no unique enrichment, **Table 3.1**). Although the expression profiles of modules 18 and 21 appear very similar, the peak expression of module 18 was reached fourteen hours earlier than that of the module 21. Modules such as these, which are responsive to a single abiotic stress in our assays, are especially fascinating in that they may indicate genes that play a key role in mediating stress-specific responses.

Modules 18 and 21 contained 25 and 38 member genes, respectively. The promoters of these genes were re-analyzed together in order to determine whether the promoters of their member genes contain over-represented motifs that may confer cold-responsive transcriptional regulation. These modules collectively contained nine suspected TF genes. The promoters of these 63 genes might be expected to contain the CRT/DRE DNA TF-binding site, RCCGAC (Yamaguchi-Shinozaki and Shinozaki, 1994; Stockinger *et al.*, 1997), a well-characterized binding site for the *Arabidopsis* CBF/DREB family of TFs. Genes in modules 18 and 21 contained 15 sequences that are present more often than expected (FDR-corrected p-value < 3.67×10^{-5} in all cases). Over-represented elements contained either a C-repeat (i.e., 5-8 cytosines), the sequence GTCGAC, or a substring of that sequence (8 of 15 sequences, **Supplemental File 3**). The Element pipeline correctly accounts for the palindromic nature of sequences, so the statistical significance of this motif is not confounded by errors in counting. This sequence is similar to the canonical CRT/DRE sequence, with a pyrimidine swap at the second position. A guanine was always observed in the first position, making this sequence a close derivative of the CRT/DRE sequence.

Heat response. Module 06 contains 367 member genes and was up-regulated early in response to heat stress (**Figure 3.5**). This module is enriched for the GO-terms “protein folding”, and responses to heat, temperature stimulus, stress, and abiotic stimulus (GO:0006457, GO:0009408, GO:0009266, GO:0006950, GO:0009628, respectively, FDR <2×10^{-6} in all cases, **Table 3.2**). Promoter

analysis revealed that 107 of 151 significant motifs were unique to this module (**Table 3.1**). Of these, 27 contained a CTTC or GAAG core (**Table 3.3**), known in both *Lycopersicon esculentum* and *Helianthus annuus* to be related to heat-shock responses (Sun *et al.*, 1996; Coca *et al.*, 1996). Only 28 of the enriched elements did not include a variant of this core (GAAG, GTTC, GAA, or their reverse complements). This motif and larger sequences such as GAAGCTTC (FDR-corrected p-value= 1.03795×10^{-6}, **Table 3.3**) can be used to identify potential substrates of one or more heat stress-responsive putative TFs in *Brachypodium*.

Calcium-mediated stress response. Calcium receptors and calcium-binding proteins are important components of plant abiotic stress response. Calcium levels increase early in the cellular response to cold stress (Knight *et al.*, 1996), and a link exists between calcium binding proteins and the cold-response CBF pathway in *Arabidopsis*. A model was recently proposed linking an increase in cellular Ca^{2+} levels with positive transcriptional control of CBF/DREB loci in *Arabidopsis* (Doherty *et al.*, 2009). Calcium levels also play a key role in drought and salt stress responses. *At*CBL1 is an *Arabidopsis* calcium sensor that is up-regulated in response to salt, drought, and cold stresses (Cheong *et al.*, 2003). Evidence suggests that calcium sensing plays a role in heat-stress response in monocot species as well (Liu *et al.*, 2003; Qin *et al.*, 2008; Zhao and Tan, 2005).

Using homology to other model systems combined with annotation via InterProScan, 359 genes were associated with GO:0005509 ('calcium ion binding') or were associated with the phrase 'calcium binding'. Expression data for these genes was hierarchically clustered and plotted in a heatmap (**Supplemental Figure 5**) that shows the expression of calcium ion binding genes in *Brachypodium* in response to the four assayed stresses. The expression levels of calcium ion binding loci were strongly affected by abiotic stress (**Figure 1B and 1C**) and were highly-correlated in drought and salt responses, although were independent in heat and cold stress responses. Principal component analysis of the expression data of the 359 genes annotated with GO:0005509 (**Supplemental**

Figure 6) revealed that trends in expression of the 359 genes were highly similar to the trends in expression of differentially expressed genes overall. The first principle component was the strongest factor in later hours of drought and salt stress and explained 65.44% of the total variance of the expression data associated with the 359 putative calcium ion binding loci.

Of the 359 putative calcium binding loci, 88 genes were part of a module. This is significantly fewer than would be expected by chance alone (average expected overlap: 242 genes, Z-score -18.1). Sixteen of the 22 modules contained at least one putative calcium-binding locus. No module was enriched for GO:0005509 ('calcium ion binding'). The large distribution of calcium responses to abiotic stress (**Supplemental Figure 5**) indicate that there are multiple regulatory pathways that trigger calcium ion binding protein expression and that these loci play a role in mediating the response of *Brachypodium* to the four assayed stresses. Further, their significant under-representation among modular loci suggests that the general response of differentially expressed calcium loci does not conform to the major modes of stress response. The regulatory circuits that control calcium binding loci appear to be specific to these individual genes. Prior studies provided evidence that calcium ion levels, calcium binding protein levels, and abiotic stress responses are linked in multiple plant systems (Doherty *et al.*, 2009; Cheong *et al.*, 2003; Qin *et al.*, 2008). Our analysis confirms that calcium sensing and calcium binding loci are responsive to abiotic stress in *Brachypodium*. We found no evidence of a centralized calcium response system.

Translation. Module 01 is strongly interconnected with modules 03 and 07, which are in turn strongly connected to module 02 (**Figure 3.6**). Module 01 is more divergent, in terms of expression profiles, than modules 02, 03, and 07 (**Figure 3.6**). Module 01 shares the down-regulation under drought stress characteristic of modules 02, 03, and 07, lacks differential regulation under heat stress, and shows minor down-regulation under salt stress. Module 01 is enriched for genes annotated with GO-terms related to translation (**Table 3.2**). The most

strongly enriched biological process GO-term is GO:0006412 “translation” (FDR-corrected p-value $1.7x10^{-10}$). This module is also strongly enriched for GO:004812 “aminoacyl-tRNA ligase activity” (FDR-corrected p-value $1.9x10^{-5}$; **Table 3.2**).

Module 04 genes (**Figure 3.5**) were down-regulated under drought stress in a manner similar to the translation module (module 01). However, module 04 is more severely down-regulated under salt stress than are module 1 genes (**Figure 3.6**). The annotations of genes making up module 04 are similar to those of the translation module, with some key differences. Both modules are populated with genes enriched with GO-terms relating to translation. Module 01 is enriched for tRNA amino acylation genes, whereas module 04 is enriched for rRNA processing and ribosome biogenesis genes (GO:0042254: “ribosome biogenesis”, FDR $1.5x10^{-13}$; GO:0006364, “rRNA processing”, $1.6x10^{-9}$) (**Table 3.2**).

Comparisons of the promoters of the rRNA processing and translation modules revealed some differences. The core GCCCA was found in 64 motifs that were significantly over-represented in promoters of the rRNA-processing module, 39 of which were uniquely enriched in genes from this module (**Table 3.3**). The GCCCA core was present in genes from six modules: the cell cycle (module 02), signal transduction (module 03), heat response (module 06, see **Table 3.2**), module 05, the translation module, and the rRNA processing module (**Supplemental File 3**). In modules 02, 03, and 06, two, one, and six motifs containing this core were over-represented, respectively. In contrast, 17, 21, and 64 motifs containing this core were enriched in the promoters of modules 05, 01, and 04, respectively. The motif GCCCR is necessary for high-level expression of ribosomal proteins. Multiple copies of the motif serve as a binding site for *At*TCP20, a class I TCP transcription factor that promotes ribosomal gene expression (a category of genes for which both the translation and rRNA processing modules are enriched to varying degrees) and the cell cycle (Gutie *et al.*, 2005). The putative *Brachypodium* ortholog of *AtTCP20* is *Bradi2g59240*;

this gene was not assigned to a module but was down-regulated in drought (drought hour 5, 2.69-fold down-regulated). Many annotations in *Brachypodium* are taken from homology and *Bradi2g59240* is annotated as a TCP-family transcription factor.

Also found to be enriched only in the rRNA processing module (module 04) are the telo-box (Regad *et al.*, 1994) conserved motifs AAACCC (16 occurrences) and AAACCCT (four occurrences). AAACCCC also occurred four times, but only one telo-box sequence with a terminal adenine, and no occurrences of a telo-box core with a terminal guanine were over-represented (**Table 3.3**). Related to the telo-box is the 'starch-box' motif AAGCCC (Michael *et al.*, 2008; Kim and Guiltinan, 1999). This sequence occurred in 21 motifs that were significantly enriched: four times in module 05 (no GO enrichment), four in the translation module (module 01), and thirteen in the rRNA processing module (**Table 3.3, Supplemental File 3**). It is not surprising that similar sequences were found to be over-represented in the promoters of the modules 01, 04, and 05, as their expression profiles are similar (**Figure 3.6**).

Signaling. Module 03 has a similar expression profile to module 07 (**Figure 3.6**) but a very different functional makeup (**Table 3.2**). The strongest contrast in expression profiles between modules 03 and 07 was observed under heat stress: Module 03 showed consistent expression levels, whereas module 07 genes were down-regulated (**Figure 3.6,** top left). Enriched in this module are genes annotated with GO-categories related to signaling and signal transduction including the regulation of Rab/Ras GTPase signal transduction (**Table 3.2**). These families of GTPases influence the composition of the cell membrane, a critical response to abiotic stress in *Arabidopsis* (Harb *et al.*, 2010). It is therefore not surprising that module 03, the 'signaling' module, is similarly regulated across drought, heat, and salt stresses.

Novel and uncharacterized modules. Module 05 is down-regulated under drought stress but not differentially expressed under any of the other three stresses.

Module 05 was not enriched for any GO terms (**Table 3.1**). Of the 640 genes in the module, 585 genes were annotated with at least one GO-term. The promoter regions of the genes in this module were enriched for 225 specific conserved motifs; of these, 90 are uniquely enriched in module 05 (**Table 3.1**). These include the core CAATA (FDR-corrected p-value 7.05×10^{-8}) and the variant ACAAAA (FDR-corrected p-value 2×10^{-5}). The PlantCARE (Lescot *et al.*, 2002) database lists the core CAATA as part of an Auxin Response Element (ARE) in *Glycine max.*

Like module 05, module 08 is down-regulated only in drought. This module has 226 member genes and is not enriched for any GO terms. Twenty-four DNA sequence motifs were significantly enriched in promoters of module 08. Uniquely significant motifs included TCCTTCA, CCCGAC, and CCGAAA. These motifs are similar to the CRT/DRE DNA TF-binding site, RCCGAC (Yamaguchi-Shinozaki and Shinozaki, 1994; Stockinger *et al.*, 1997). Conserved *cis*-acting elements similar to those found in the promoters of modules 05 and 08 have been observed in other species, lending weight to the hypothesis that these DNA sequences could be responsible for driving the module-wise expression profiles observed here. No enriched functional terms could be associated with modules 05 and 08. An extended examination of gene expression responses to abiotic stress – especially stretching into the days after stress onset – may reveal the functional roles these modules play.

Network plasticity. Analysis of differential correlations for transcription factor/target gene pairs in various conditions revealed a high degree of plasticity in these relationships. The proportion of potentially plastic relationships varied greatly depending on the conditions compared. Neither the conditional comparison with the lowest ratio of potentially plastic gene pair relationships (salt/heat, 241 plastic TF-TG pairs, **Table 3.4**) nor the comparison with the highest ratio of potentially plastic relationships (salt/cold, 2,027 plastic TF-TG pairs, **Table 3.4**) were the comparison with the most extreme number of total

possible pairings. Of particular interest is the great diversity of differential correlations between salt and drought stresses. There are a large segment of gene pairs that experience very large changes in correlation. More than 11,000 genes pairings had large negative correlations under drought stress and very large positive correlations under salt stress (top right, **Figure 3.7B**). Conversely, more than 16,000 gene pairings had large positive correlations under drought stress and large negative correlations under salt stress (bottom left, **Figure 3.7B**). Comparisons between the differential correlations observed between salt and drought stresses and the differential correlations observed between random subsets of the stress data indicate that the differential correlations between salt and drought stresses are unlikely to arise by chance (**Figure 3.7A**).

The basic underlying assumption of gene co-expression network analysis is that two genes, when co-expressed, can reliably be expected to be co-expressed if there is a biological relationship between them. The stronger the biological relationship between two genes – either due to co-regulation or from necessary co-expression born of functional relatedness, the higher the correlation in expression between the two genes. The relationships between transcription factor/target gene pairs across conditions are plastic due to dependence on methylation and chromatin status, among many other factors. This highlights the importance of inclusion of epigenomic data in any large genomic discovery endeavor. Based on the dataset used here, we cannot assign cause to the large changes in expression correlation across conditions. It is clear that a full understanding of the abiotic stress response of *Brachypodium* requires epigenomic analysis. With increasing throughput and decreasing costs, full integration of multi-type sequence data waits only on development of novel bioinformatic methods that can take full advantage of rich datasets. The high degree of plasticity observed in the stress response of *Brachypodium* also has implications for whole-genome gene co-expression network reconstruction. Current state-of-the-art software packages, such as WGCNA (Langfelder and Horvath, 2008), may be made even more powerful by accounting for the changing

relationship between gene pairs across conditions in meta-data enhanced expression datasets. Adopting a 'regulator-target' dichotomous view of gene loci – as is common in applications designed for smaller networks – may further improve large network reconstruction efforts.

Weighted gene co-expression analysis of the *Brachypodium* transcriptome under normal growth and four abiotic stress conditions identified 22 modules of genes. Over-expression, knock-down, and knock-out experiments will elucidate the roles of these genes in abiotic stress responses and may guide genetic changes that confer stress tolerance in *Brachypodium*. Homology between *Brachypodium* and the closely related important crop species will allow identification of homologous genes in cereal and biofuel feedstock crops, enabling improved stress tolerance in plants critical to serving the needs of society.

We have identified numerous potential transcription factor binding site sequences that are associated with specific expression profiles under abiotic stresses. In addition to linking these motifs to specific gene expression profiles, we have linked these DNA sequence motifs to specific endogenous plant systems. These candidate *cis*-regulatory sequences may represent key components of the transcriptional circuitries that define the plant's gene regulatory networks. Systems and synthetic biology approaches may take advantage of these circuits to place a gene of interest under the control of existing stress response pathways to achieve a desired phenotype of stress tolerance in agriculturally or economically important crops.

Web Resources. All microarray datasets are accessible through the *Brachypodium* web genome browser (http://jbrowse.brachypodium.org). The module membership lists, AgriGo GO-enrichment analysis output, and Element promoter content analysis output may be found as supplemental files and are available for download on the Brachypodium.org FTP website (ftp://brachypodium.org/brachypodium.org/Stress/). Individual gene RMA

expression profiles for each assayed stress condition may be viewed at the Mockler Lab's plant stress response web portal (http://stress.mocklerlab.org/).

Acknowledgements:

We are grateful to Anne-Marie Girard and Caprice Rosato for the qualitative assessment of RNA, preparation of labeled targets, and executing array hybridizations and scanning. We thank the Donald Danforth Plant Science Center's Biocomputing Core for computational support. We thank Malia Gehan for helpful suggestions on the manuscript.

Funding:

This work was supported by a grant to TCM (DE-FG02-08ER64630) from the DOE - Plant Feedstock Genomics for Bioenergy Program.

Author contributions:

TCM, TPM, and SEF conceived the experimental design. SEF and JRM collected tissues and prepared RNAs for microarray analysis. HDP developed bioinformatics tools and conducted analyses. HDP, SEF, ERR, and TCM wrote the paper. All authors read and approved the final manuscript.

Literature cited:

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.

Aranjuelo, I., Molero, G., Erice, G., Avice, J.C., and Nogués, S. 2011. Plant physiology and proteomics reveals the leaf response to drought in alfalfa (*Medicago sativa* L.). Journal of Experimental Botany 62: 111–23.

Araus, J.L. 2002. Plant Breeding and Drought in C3 Cereals: What Should We Breed For? Annals of Botany 89: 925–940.

Ashraf, M., Hayat, M.Q., Jabeen, S., Shaheen, N., Ajab, M., and Yasmin, G. 2010. *Artemisia* L. species recognized by the local community of northern areas of Pakistan as folk therapeutic plants. Journal of Medicinal Plants Research 4: 112–119.

Bai, L. and Morozov, A. V 2010. Gene regulation by nucleosome positioning. TRENDS in Genetics 26: 476–83.

Bevan, M.W., Garvin, D.F., and Vogel, J.P. 2010. Brachypodium distachyon genomics for sustainable food and fuel production. Current Opinion in Biotechnology 21: 211–7.

Brinker, M., Brosché, M., Vinocur, B., Abo-Ogiala, A., Fayyaz, P., Janz, D., Ottow, E.A., Cullmann, A.D., Saborowski, J., Kangasjärvi, J., Altman, A., and Polle, A.(2010. Linking the salt transcriptome with physiological responses of a salt-resistant *Populus* species as a strategy to identify genes important for stress acclimation. Plant Physiology 154: 1697–709.

Brkljacic, J., Grotewold, E., Scholl, R., Mockler, T., Garvin, D.F., Vain, P., Brutnell, T., Sibout, R., Bevan, M., Budak, H., Caicedo, A.L., Gao, C., Gu, Y., Hazen, S.P., Holt, B.F., Hong, S.-Y., Jordan, M., Manzaneda, A.J., Mitchell-Olds, T., Mochida, K., *et al.* 2011. Brachypodium as a model for the grasses: today and the future. Plant Physiology 157: 3–13.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Chaves, M.M., Flexas, J., and Pinheiro, C. 2009. Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. Annals of Botany 103: 551–60.

Cheong, Y.H., Kim, K., Pandey, G.K., Gupta, R., Grant, J.J., and Luan, S. 2003. CBL1 , a Calcium Sensor That Differentially Regulates Salt , Drought , and Cold Responses in *Arabidopsis*. The Plant Cell 15: 1833–1845.

Chew, Y.H. and Halliday, K.J. 2011. A stress-free walk from *Arabidopsis* to crops. Current Opinion in Biotechnology 22: 281–6.

Coca, M.A., Almoguera, C., Thomas, T.L., and Jordano, J. 1996. Differential regulation of small heat-shock genes in plants: analysis of a water-stress-inducible and developmentally activated sunflower promoter. Plant Molecular Biology 31: 863–876.

Doherty, C.J., Van Buskirk, H. a, Myers, S.J., and Thomashow, M.F. 2009. Roles for *Arabidopsis* CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. The Plant Cell 21: 972–84.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Research 38: W64–70.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.-K., and Mockler, T.C. 2010. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Research 20: 45–58.

Fisher, R.A. and Yates, F. 1963. Statistical Tables: For Biological, Agricultural and Medical Research 6th ed. (Hafner Press: New York).

Flexas, J., Diaz-Espejo, A., Galmés, J., Kaldenhoff, R., Medrano, H., and Ribas-Carbo, M. 2007. Rapid variations of mesophyll conductance in response to changes in CO2 concentration around leaves. Plant, Cell & Environment 30: 1284–98.

Fujita, Y., Fujita, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. 2011. ABA-mediated transcriptional regulation in response to osmotic stress in plants. Journal of Plant Research 124: 509–25.

Gómez-Porras, J.L., Riaño-Pachón, D.M., Dreyer, I., Mayer, J.E., and Mueller-Roeber, B. 2007. Genome-wide analysis of ABA-responsive elements ABRE and CE3 reveals divergent patterns in *Arabidopsis* and rice. BMC Genomics 8: 260.

Gutie, R.A., Doerner, P., Li, C., Colon-Carmona, A., and Potuschak, T. 2005. *Arabidopsis* TCP20 links regulation of growth and cell division control pathways. Proceedings of the National Academy of Sciences of the United States of America 102.

Haga, N., Kobayashi, K., Suzuki, T., Maeo, K., Kubo, M., Ohtani, M., Mitsuda, N., Demura, T., Nakamura, K., Jürgens, G., and Ito, M. 2011. Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in *Arabidopsis*. Plant Physiology 157: 706–17.

Harb, A., Krishnan, A., Ambavaram, M.M.R., and Pereira, A. 2010. Molecular and Physiological Analysis of Drought Stress in *Arabidopsis* Reveals Early Responses Leading to Acclimation in Plant Growth. Plant Physiology 154: 1254–1271.

Hattori, T., Totsuka, M., Hobo, T., Kagaya, Y., and Yamamoto-Toyoda, A. 2002. Experimentally determined sequence requirement of ACGT-containing abscisic acid response element. Plant & Cell Physiology 43: 136–40.

Hirayama, T. and Shinozaki, K. 2010. Research on plant abiotic stress responses in the post-genome era: past, present and future. The Plant Journal 61: 1041–52.

Hong-Bo, S., Zong-Suo, L., and Ming-An, S. 2005. LEA proteins in higher plants: structure, function, gene expression and regulation. Colloids and surfaces. B, Biointerfaces 45: 131–5.

Irizarry, R. a, Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–64.

Janská, A., Aprile, A., Zámečník, J., Cattivelli, L., and Ovesná, J. 2011. Transcriptional responses of winter barley to cold indicate nucleosome remodelling as a specific feature of crown tissues. Functional & Integrative Genomics 11: 307–25.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. 2007. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. The Plant Journal 50: 347–63.

Kim, K. and Portis, A.R. 2005. Temperature dependence of photosynthesis in *Arabidopsis* plants with modifications in Rubisco activase and membrane fluidity. Plant & Cell Physiology 46: 522–30.

Kim, K.N. and Guiltinan, M.J. 1999. Identification of cis-acting elements important for expression of the starch-branching enzyme I gene in maize endosperm. Plant Physiology 121: 225–36.

Knight, H., Trewavas, a J., and Knight, M.R. 1996. Cold calcium signaling in *Arabidopsis* involves two cellular pools and a change in calcium signature after acclimation. The Plant Cell 8: 489–503.

Langfelder, P. and Horvath, S. 2012. Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of Statistical Software 46.

Langfelder, P. and Horvath, S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559.

Lauria, M. and Rossi, V. 2011. Epigenetic control of gene regulation in plants. Biochimica et Biophysica Acta 1809: 369–78.

Le, S., Josse, J., and Husson, F. 2008. FactoMineR : An R Package for Multivariate Analysis. Journal of Statistical Software 25: 1–18.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., and Rombauts, S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 30: 325–7.

Li, C., Rudi, H., Stockinger, E.J., Cheng, H., Cao, M., Fox, S.E., Mockler, T.C., Westereng, B., Fjellheim, S., Rognli, O.A., and Sandve, S.R. 2012. Comparative analyses reveal potential uses of *Brachypodium distachyon* as a model for cold stress responses in temperate grasses. BMC Plant Biology 12: 65.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25: 1754–1760.

Lindemose, S., O'Shea, C., Jensen, M.K., and Skriver, K. 2013. Structure, function and networks of transcription factors involved in abiotic stress responses. International Journal of Molecular Sciences 14: 5842–78.

Liu, H., Li, B., Shang, Z., Li, X., Mu, R., Sun, D., and Zhou, R. 2003. Calmodulin Is Involved in Heat Shock Signal Transduction in Wheat 1. Plant Physiology 132: 1186–1195.

Liu, Y., Wang, L., Xing, X., Sun, L., Pan, J., Kong, X., Zhang, M., and Li, D. 2013. ZmLEA3, a multifunctional group 3 LEA protein from maize (Zea mays L.), is involved in biotic and abiotic stresses. Plant & cell physiology 54: 944–59.

Lopushinsky, W. 1969. Stomatal Closure in Conifer Seedlings in Response to Leaf Moisture Stress. The University of Chicago Press 130: 258–263.

Mahajan, S. and Tuteja, N. 2005. Cold, salinity and drought stresses: an overview. Archives of Biochemistry and Biophysics 444: 139–58.

Marino, D., Froidure, S., Canonne, J., Ben Khaled, S., Khafif, M., Pouzet, C., Jauneau, A., Roby, D., and Rivas, S. 2013. *Arabidopsis* ubiquitin ligase

MIEL1 mediates degradation of the transcription factor MYB30 weakening plant defence. Nature Communications 4: 1476.

Matsui, A., Ishida, J., Morosawa, T., Mochizuki, Y., Kaminuma, E., Endo, T. a, Okamoto, M., Nambara, E., Nakajima, M., Kawashima, M., Satou, M., Kim, J.-M., Kobayashi, N., Toyoda, T., Shinozaki, K., and Seki, M. 2008. *Arabidopsis* transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. Plant & Cell Physiology 49: 1135–49.

Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., Hazen, S.P., Shen, R., Priest, H.D., Sullivan, C.M., Givan, S.A., Yanovsky, M., Hong, F., Kay, S.A., and Chory, J. 2008. Network Discovery Pipeline Elucidates Conserved Time-of-Day–Specific cis-Regulatory Modules. PLoS Genetics 4: 17.

Moellering, E.R., Muthan, B., and Benning, C. 2010. Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. Science 330: 226–8.

Mukhopadhyay, P., Singla-pareek, S.L., Reddy, M.K., and Sopory, S.K. 2013. Stress-Mediated Alterations in Chromatin Architecture Correlate with Down-Regulation of a Gene Encoding 60S rpL32 in Rice. Plant & Cell Physiology 54: 528–540.

Oliveira, G. and Peñuelas, J. 2004. Effects of winter cold stress on photosynthesis and photochemical efficiency of PSII of the Mediterranean Cistus albidus L . and Quercus ilex L . Plant Ecology 175: 179–191.

Parre, E., Ghars, M.A., Leprince, A.-S., Thiery, L., Lefebvre, D., Bordenave, M., Richard, L., Mazars, C., Abdelly, C., and Savouré, A. 2007. Calcium signaling via phospholipase C is essential for proline accumulation upon ionic but not nonionic hyperosmotic stresses in *Arabidopsis*. Plant Physiology 144: 503–12.

Pérez-Rodríguez, P., Riaño-Pachón, D.M., Corrêa, L.G.G., Rensing, S.A., Kersten, B., and Mueller-Roeber, B. 2010. PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Research 38: D822–D827.

Qin, D., Wu, H., Peng, H., Yao, Y., Ni, Z., Li, Z., Zhou, C., and Sun, Q. 2008. Heat stress-responsive transcriptome analysis in heat susceptible and

tolerant wheat (*Triticum aestivum* L.) by using Wheat Genome Array. BMC Genomics 9: 432.

Regad, F., Lebas, M., and Lescure, B. 1994. Interstitial Telomeric Repeats within the *Arabidopsis* thaliana Genome. Journal of Molecular Biology: 163–169.

Rowley, E.R. and Mockler, T.C. 2011. Plant Abiotic Stress : Insights from the Genomics Era. In Abiotic Stress Response in Plants - Physiological, Biochemical and Genetic Perspectives. (Intech Open Access Journals).

Salvucci, M.E., Osteryoung, K.W., Crafts-brandner, S.J., and Vierling, E. 2001. Exceptional Sensitivity of Rubisco Activase to Thermal Denaturation in Vitro and in Vivo 1. Plant Physiology 127: 1053–1064.

Shen, Q., Zhang, P., and Ho, T.-H.D. 1996. Modular Nature of Abscisic Acid (ABA) Response Complexes: Composite Promoter Units That Are Necessary and Sufficient for ABA lnduction of Gene Expression in Barley. The Plant Cell 8: 1107–1119.

Shinde, S., Nurul Islam, M., and Ng, C.K.-Y. 2012. Dehydration stress-induced oscillations in LEA protein transcripts involves abscisic acid in the moss, Physcomitrella patens. The New Phytologist 195: 321–8.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27: 431–2.

Stockinger, E.J., Gilmour, S.J., and Thomashow, M.F. 1997. *Arabidopsis* thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proceedings of the National Academy of Sciences of the United States of America 94: 1035–40.

Sun, S., Chung, M., and Lin, T. 1996. The structure and expression of an hsc70 gene from Lycopersicon esculentum. Gene 170: 237–241.

The International Brachypodium Initiative 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463: 763–8.

Tusher, V.G., Tibshirani, R., and Chu, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the

National Academy of Sciences of the United States of America 98: 5116–21.

Tuteja, N. 2007. Mechanisms of high salinity tolerance in plants. Methods in Enzymology 428: 419–38.

Wang, W., Vinocur, B., and Altman, A. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218: 1–14.

Winfield, M.O., Lu, C., Wilson, I.D., Coghill, J. a, and Edwards, K.J. 2010. Plant responses to cold: Transcriptome analysis of wheat. Plant Biotechnology Journal 8: 749–71.

Witcombe, J.R., Hollington, P. a, Howarth, C.J., Reader, S., and Steele, K. a 2008. Breeding for abiotic stresses for sustainable agriculture. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 363: 703–16.

Yamaguchi-Shinozaki, K. and Shinozaki, K. 1994. A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. The Plant Cell 6: 251–64.

Yazaki, J., Shimatani, Z., Hashimoto, A., Nagata, Y., Fujii, F., Kojima, K., Suzuki, K., Taya, T., Tonouchi, M., Nelson, C., Nakagawa, A., Otomo, Y., Murakami, K., Matsubara, K., Kawai, J., Carninci, P., Hayashizaki, Y., and Kikuchi, S. 2004. Transcriptional profiling of genes responsive to abscisic acid and gibberellin in rice: phenotyping and comparative analysis between rice and *Arabidopsis*. Physiological Genomics 17: 87–100.

Zdobnov, E.M. and Apweiler, R. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847–848.

Zeller, G., Henz, S.R., Widmer, C.K., Sachsenberg, T., Rätsch, G., Weigel, D., and Laubinger, S. 2009. Stress-induced changes in the *Arabidopsis* thaliana transcriptome analyzed using whole-genome tiling arrays. The Plant Journal 58: 1068–82.

Zhao, H.-J. and Tan, J.-F. 2005. Role of calcium ion in protection against heat and high irradiance stress-induced oxidative damage to photosynthesis of wheat leaves. Photosynthetica 43: 473–476.

Zhong, L., Xu, Y., and Wang, J. 2009. DNA-methylation changes induced by salt stress in wheat Triticum aestivum. African Journal of Biotechnology 8: 6201–6207.

Zhou, Z.S., Huang, S.Q., Guo, K., Mehta, S.K., Zhang, P.C., and Yang, Z.M. 2007. Metabolic adaptations to mercury-induced oxidative stress in roots of Medicago sativa L. Journal of Inorganic Biochemistry 101: 1–9.

Zhu, J. 2001. Cell signaling under salt , water and cold stresses. Current Opinion in Plant Biology 4: 401–406.

Chapter 4:

Plant Abiotic Stress: Insights from the Genomics Era

Erik R. Rowley and Todd C. Mockler

INTECH
Janeza Trdine 9
51000 Rijeka, (Croatia)
ISBN 978-953-307-672-0 (2011)
DOI: 10.5772/23215

Introduction:

Agricultural crop plants make up a large proportion of the world's economy and in many countries constitute the main sustenance for humans. Therefore maximizing crop yield is of extreme importance and interest. There are many factors that can limit the yield of a crop; however the main causes of crop failure are abiotic stresses such as salinity, drought, extremes in temperature, intense light, and oxidative stress caused by reactive oxygen species. Plants have evolved mechanisms and pathways allowing them to cope with the environment by modifying their physiological and cellular states. For example, plants living in colder regions undergo a phenomenon known as cold acclimation, resulting in cell membrane composition and protein concentration changes to reduce intracellular ice crystal formation and dehydration due to freezing (Thomashow 1998).

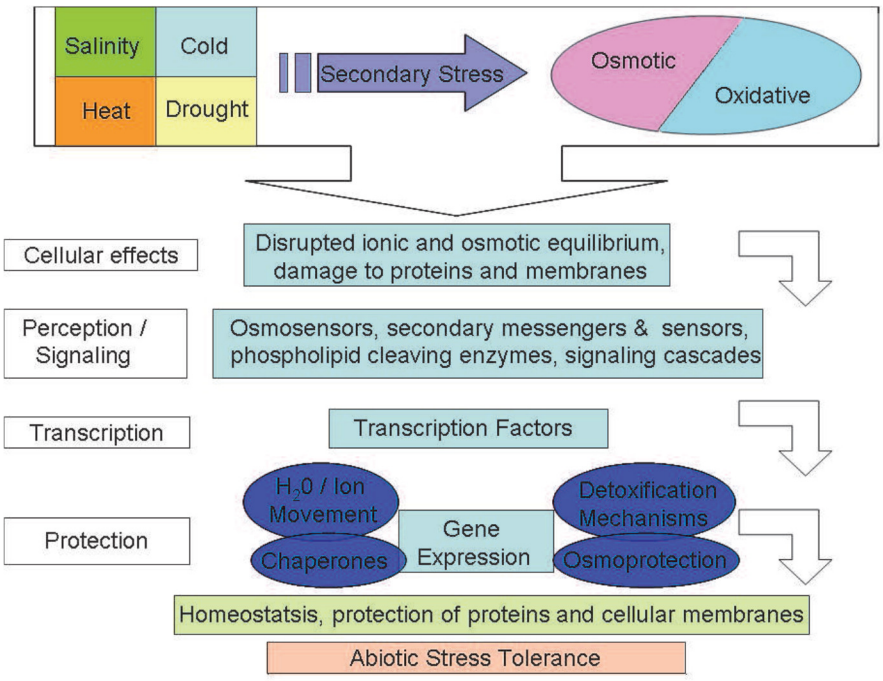
Abiotic stresses alter gene networks and signaling cascades in an effort to restore cellular homeostasis. It has been demonstrated (Reddy, 2007, Palusa *et al.*, 2007) that abiotic stress conditions alter the alternative splicing of a number of genes. Alternative pre-mRNA splicing in higher eukaryotes is a highly regulated mechanism, often allowing for many proteins (isoforms) to be derived from a single gene, thereby increasing overall proteome diversity. These alternative transcripts often result in functionally and structurally distinct proteins (Biamonti, 2009) with their own functions in development, cellular localization, and responses to the environment (Tanabe *et al.*, 2006).

Recently genome and transcriptome-wide surveys (Rensink *et al.*, 2005, Kreps *et al.*, 2002, Egawa *et al.*, 2006, Filichkin *et al.*, 2010) have offered glimpses into transcript abundance profiles under abiotic stresses, demonstrating dramatic shifts in alternative splicing patterns, as well as upregulation of key transcription factors controlling stress-induced signaling cascades. This research offers the potential for discovery of candidate genes

that, through genetic engineering, may confer increased tolerance to abiotic stresses, several examples of which will be discussed in this chapter.

Recent research has demonstrated the possibility of pre-disposing plants to stress tolerance by overexpressing a gene known to be upregulated in response to a certain stress (for example: Haake *et al.*, 2002, Forment *et al.*, 2002, Kim *et al.*, 2010) often acting upstream or in conjunction with a longer signaling cascade, such as the mitogen-activated protein (MAP) kinase cascade (Xiong *et al.* 2003), salt overly sensitive (SOS) pathway (Liu and Zhu, 1998; Ishitani *et al.*, 2000, Qiu *et al.*, 2004), or C-repeat-binding factor/dehydration-responsive element (CBF/DREB) pathway (Thomashow, 2010).

A high degree of crosstalk exists between these pathways, as often the plant's physiological and cellular responses to different abiotic stresses are similar (**Figure 4.1**).



In drought and cold stress for example, two types of molecular responses occur simultaneously: those protecting cells from acute dehydration and those protein factors involved in further regulation of gene expression and signal transduction functioning in overall stress response (Shinozaki *et al.*, 2000 and 2007). Other examples include crosstalk between cold and osmotic signaling pathways, as well as cold and abscisic acid (Ishitani *et al.*, 1997). Abscisic acid regulates stomatal aperture changes and is a crucial signaling molecule in stress plant responses along with changes in metabolite concentrations.

Understanding the genes and gene networks that underlie abiotic stress resistance is integral not only in improving the world's long term food production and security, but also in piecing together the web of abiotic stress induced global gene regulation including post-transcriptional regulation such as alternative splicing and regulation by miRNAs. Recent insights into genes conferring abiotic stress tolerance, particularly salt tolerance, have come from the study of plants naturally adapted for growth in extreme conditions such as the halophytes, which thrive in areas of elevated salt levels. Differential gene expression in seedlings of the salt marsh halophyte Kosteletzkya virginica was recently analyzed (Guo *et al.*, 2009), identifying genes necessary for re-establishing ion homeostasis and protecting the plant from stress damage, as well as those involved in metabolism and plant development under salt stress. Also demonstrated was the ability of K. virginica seedlings to sequester sodium, primarily in the roots. In another example, a dehydration and abscisic acid (ABA) induced transcription factor was functionally characterized in *Craterostigma plantagineum*, a plant possessing the ability to survive complete dehydration.

The knowledge gained from high-throughput sequencing (HTS) technologies and comparative studies of halophytes, coupled with our continually

expanding knowledge of metabolites, and the molecular and physiological responses to abiotic stresses that are profiled within this chapter, will allow a plethora of opportunities for directed genetic engineering and breeding strategies that will allow us to meet the worlds demand for food despite a growing population. This chapter aims to offer insights from the past decade of plant abiotic stress research, and give an overview of the wealth of knowledge generated by the genomics era, such as advances from global gene expression surveys and differential gene expression between stresses.

Salt. Increasing salinity of soil leads to reduction of crop yields, and while soil salinity is not uncommon (Flowers *et al.*, 1997), secondary effects such as irrigation-induced salination affects about 20% of the world's cultivated fields (Yeo, 1998) with 50% of lands predicted to be affected in the coming decades. These irrigated lands currently produce an estimated one-third of the world's food (Munns 2002). Irrigation water typically contains a variety of ions, such as Ca+, Mg+, as well as Na+ in the form of NaCl. It is when the water evaporates and the Ca+ and Mg+ precipitate that the Na+ ions begin to become dominate in the soil (Serrano *et al.*, 1999). Plants do not have specific mechanisms for the uptake of Na+ ions; however several pathways exist for passive entry. For example, root cells uptake Na+ ions via cation channels, of which there are two main classifications. Voltage dependent channels, namely the K+ transporter HKT1, import Na+ ions into root cells (Na+ leakage), ultimately leading to higher concentrations of intracellular sodium. Excess salt in the soil now begins to present an issue due to this osmotic gradient, as elevated Na+ levels in the soil begins to drive water out of the cell. Initial plant response to salt stress, the osmotic stress component, share metabolic similarities with drought however long-term exposure introduces the ion toxicity component, the displacement of K+ ions with Na+ and Cl- ions. Plants differ in their methods of coping with Na+ entry: some prevent or minimize entry altogether (at the roots), while others reduce the

cytoplasmic Na+ concentration by compartmentalization in the vacuoles, thus avoiding toxic effects on photosynthesis and other key metabolic processes (Chaves *et al.*, 2009).

Na+ transport from roots to stem is quite rapid due to the transpiration stream in the xylem, and can only be returned to the roots via the phloem. The roots are able to regulate Na+ levels by export to either the stem of back to the soil; however once in the xylem the Na+ ions will accumulate as the leaves age and the water evaporates (Tester *et al.*, 2003). This rapid accumulation of sodium ions has several detrimental metabolic effects to the plant cell.

Turgor pressure is necessary in order to stretch the cells walls during growth. When faced with the initial sudden influx of Na+ ions, the plant cell is able to sequester these ions in the vacuole, in effect reducing the osmotic potential in an attempt to restore homeostasis and equally importantly reducing degradation of cytosolic enzymes. Research (Carden *et al.*, 2003) comparing ion concentrations in the roots of two barley (Hordeum vulgare) varieties differing in NaCl tolerance indicated the cytosolic Na+ concentration requirements to be quite low; around 10-30mM. Na+ concentration within the vacuoles however, may be much higher.

During the initial osmotic phase of salt stress response, the expansion rate of growing leaves is reduced, along with stomatal aperture in response to leaf turgor decline, leading to decreased emergence of new leaves and therefore fewer branches. Among the cereals, barley is the most tolerant with rice (Oryza sativa) being the least tolerant.

Na+ ions also compete with K+ ions for binding sites, often required for crucial cellular and metabolic processes. Protein stability is also coupled with K+ concentration, as it is a cofactor for many enzymes, and the tRNA

binding to ribosomes is also dependent on high K+ concentration (Zhu, 2002, Tester *et al.*, 2003). Nutrient uptake from the roots is affected due to K+ ion channels being disrupted, and Na+ induced osmotic changes inhibit root growth.

Another way plants reduce osmotic stress is by the accumulation of cytoplasmic osmolytes such as proline and glyceinebetaine, shown to stabilize the chloroplasts and cellular membranes, as well as play a role in maintaining cell volume and fluid balance (Bohnert *et al.*, 1996). These osmolytes also serve to protect proteins from degradation by reacting oxygen species (ROS). Salt stress (among other abiotic stresses) leads to the accumulation of high levels of ROS. When present at low levels, ROS may act to signal upregulation of the defense-responsive genes. Typically however, excessive production and accumulation of ROS such as hydrogen peroxide (H2O2), superoxide (O2-) and hydroxyl radicals (OH-) can perturb the cellular redox homeostasis leading to oxidative injuries. There is also growing evidence that the cell's downstream ability to repair damage, and scavenge damaging reactive oxygen species (ROS) is equally as important as Na+ uptake and vacuolar sequestration itself. Expression of ROS defense transcripts was found to be elevated in *Arabidopsis* plants constitutively expressing the zinc finger protein Zat10 (Mittler *et al.*, 2006). The plants displayed enhanced tolerance of salinity, heat and osmotic stress. Surprisingly, knockout and RNAi mutants of Zat10 were also more tolerant to osmotic and salinity stress suggesting that Zat10 plays a key role as both a positive and a negative regulator of plant defenses. Readers are directed toward a recent review (Miller *et al.*, 2008) for a discussion on how ROS integrate cellular signals generated from abiotic stress.

Studies of salt tolerant clones of Eucalyptus camaldulensis, an important crop in Australia due to the use of its oils, have demonstrated a significant

increase in shoot proline levels when exposed to 100mM NaCl (Woodward *et al.*, 2005). Proline accumulation is known to be mediated by both ABA-dependent and ABA-independent signaling pathways (Hare *et al.*, 1999). Assays on in vitro shoot cultures of Populus euphratica suggest accumulated proline and sugars promote both osmotic and salt tolerance (Watanabe *et al.*, 2000). Another study using sugar beet cultivars (Ghoulam *et al.*, 2001) report a positive trend with proline levels increasing with relation to salt tolerance; however the quantitative contribution of proline to osmotic adjustment in a salt tolerant variety was weak. It was determined the elevation of proline in *Arabidopsis*, acting as an osmoprotectant during salt stress adaptation, led to the enhancement of the enzymes scavenging reactive oxygen species (Abraham *et al.*, 2003). Recently proline and glycine betaine were shown to improve salt stress in cultured tobacco cells via scavenging of hydrogen peroxide and methylglyoxal (Banu *et al.*, 2010). Glycine betaine was shown to be in induced in the burning bush, Kochia scoparia, (Kern *et al.*, 2004), and the Mediterranean shrub Atriplex halimus (Martinex *et al.*, 2004). The cyanobacterium Synechococcus also displays increased tolerance to both salt and cold stress following induction of glycine betaine (Ohnishi *et al.*, 2006).

The second major constraint, besides the osmotic stress of Na+ surrounding the roots, is intracellular Na+ toxicity. Because of the similarity in physicochemical properties between Na+ and K+ (i.e. ionic radius and ion hydration energy), the former competes with K+ for major binding sites in key metabolic processes in the cytoplasm, such as enzymatic reactions, protein synthesis and ribosome functions (Shabala *et al.*, 2008). Increased concentrations of Na+ ions in the soil reduce the activity of many essential nutrients (including K+), making them less available. Secondly, Na+ competes with K+ for uptake sites at the plasma membrane. Recent evidence indicates the K+/Na+ intracellular ratio is a key determinant of salt tolerance. The optimal cytosolic K+/Na+ ratio can be maintained by either restricting

Na+ accumulation in plant tissues or by preventing K+ loss from the cell. At the cellular level, restricted Na+ uptake, active Na+ exclusion back to the soil solution (via the plasma membrane salt overly sensitive (SOS1) Na+/H+ antiporter; (Zhu, 2002) and compartmentalization of excessive Na+ in the vacuole by the tonoplast Na+/H1+exchanger (Zhang and Blumwald 2001) are considered central to salt tolerance.

The past decade of research into the SOS (Salt Overly Sensitive) pathway, utilizing *Arabidopsis* knock-out mutants and the plants basal tolerance to NaCl stress as a background concentration for screening, has elucidated key steps in the salt stress signaling pathway (Liu and Zhu, 1998; Ishitani *et al.*, 2000, Qiu *et al.*, 2004). A transient Ca2+ signal, an important secondary messenger for many cellular processes, is ultimately propagated by the secondary messenger IP3 and is the crucial first step in restoring cellular homeostasis. This process involves the sensing of the Ca+ ion by SOS3 (also known as AtCBL4: calcineurin B-like protein) followed by interaction with SOS2, a serine/threonine protein kinase, resulting in its activation (Halfter *et al.*, 2000). These work in conjunction to phosphorylate and activate the transport activity of the plasma membrane Na+/H+ antiporter SOS1. SOS1 also has a large cytoplasmic domain predicted to act as potential novel Na+ sensor (Zhu 2002), which may act in feedback regulation. Recent research has indicated the C-terminal region of SOS1 interacts with RCD1 under salt and oxidative stresses (Katiyar-Agarwat *et al.*, 2006). Typically a nuclear protein, RCD1 is found both in the nucleus and in the cytoplasm near the cell periphery during salt and oxidative stresses and demonstrated similar expression and tissue localization as SOS1, perhaps regulating transport of ROS across the cell membrane and oxidative-stress signaling.

There are likely more components to the SOS pathway, the function of which are the focus of current research. For example, there is a family of 9 SO3-like

Ca+ binding proteins (the SCaBP's) present in *Arabidopsis* and 24 SOS2-like protein kinases. One of the SCaBP's, the putative calcium sensor SCABP8/CBL10 was shown to interact with the protein kinase SOS2 to protect *Arabidopsis* shoots from salt stress (Xie *et al.*, 2009). Further screens under more stringent (100mM NaCl) conditions for salt-hypersensitive mutants have yielded more members of the SOS pathway. SOS4 encodes a pyridoxal kinase that is involved in the biosynthesis of pyridoxal-5-phosphate, an active form of vitamin B6, which is often found in roots and necessary for growth (Mahajan *et al.*, 2005). Knock-out mutants for SOS4 are defective in root hair formation and root tip growth, perhaps acting as an integral upstream regulator of root hair development (Zhu *et al.* 2002). Another component, SOS5, has been shown to be a putative cell surface (Shi *et al.* 2003) adhesion protein that is required for normal cell. For a detailed overview into current insights into the SOS pathway, the reader is directed to the 2008 review by Mahajan.

The SOS pathway is far from a unique response in *Arabidopsis*, or for that matter glycophytes in general, as conservation of SOS pathway components have been identified in halophytes as well as cereals and also woody periennals. For example in halophytes, salt treatment of *Thellungiella halophila* led to increased expression of an AtSOS1 homologue in the plasma membrane and increased H+ transport and hydrolytic activity of the H+-ATPase was observed in both the plasma membrane as well the tonoplast (Vera-Estrella *et al.*, 2005). *Chenopodium quinoa*, a halophyte native to the Andes Mountains, was found to contain 2 AtSOS1 homologs (Maughan *et al.*, 2009), with future work to include complementation of a mutant sos1 *Arabidopsis* line with the homologues from *C. quinoa*. Homologues of AtSOS1 have also been identified for multiple glycophyte plant species such as rice (*Oryza sativa*), the seagrass *Cymodocea nodosa*, and *Populus*

trichocarpa, the woody perennial poplar tree (Martínez-Atienza *et al.*, 2007, Garciadeblás *et al.*, 2007, Tang *et al.*, 2007, respectively).

Not surprisingly, also conserved are genes controlling sodium entry, such as the previously mentioned K+ channel HKT1 and also genes controlling vacuole compartmentalization, of which the following discussion will focus primarily on AtNHX1, a gene encoding a vacuolar Na+/H+ exchanger. Shi *et al.*, 2002 demonstrated AtNHX1 transcript up-regulation following treatment with NaCl, KCl or ABA, as well as detecting strong expression in guard cells and root hairs, suggesting AtNHX1 plays a role in pH regulation and/ K+ homeostasis along with storing Na+ in the enlarged vacuoles in root hair cells, respectively. As previously mentioned, one way to remove harmful Na+ ions from the cytosol and maintain osmotic balance within the cells is by compartmentalization in the vacuoles, and this aspect of Na+ tolerance has been the focus of much current research, with very encouraging results.

In 1999 Apse *et al.* demonstrated increased Na+ tolerance from overexpression of the AtNHX1 Na+/H+ antiporter in *Arabidopsis*, and also that salinity tolerance was correlated with higher-than-normal levels of AtNHX1transcripts, protein, and vacuolar Na+/H+(sodium/proton) antiport activity. In 2007 tomato (*Lycopersicon esculentum* cv. Moneymaker) was successfully transformed with an overexpressed AtHKT1, demonstrating not only the ability to grow in 200mM Na+ concentrations, but an accumulation of sodium ions in the leaves rather than the fruit (Zhang *et al.*, 2001). This discovery was quite exciting in two ways: firstly this yielded the potential for agriculturally relevant crop, as the fruit quality was not adversely affected, and secondly it demonstrated an increased resistance to salt tolerance in an agriculturally important crop plant resulting from the modification of a single trait. Then an AtNHX homologue from a monocot halophyte, Aeluropus littoralis, was identified (AlNHX) and cloned (Zhang *et al.* 2008). This gene

was then transformed into tobacco, which displayed the ability to grow in MS media containing 250mM NaCl, and survived 400mM NaCl in pots for one month. Na+ ions were found to be sequestered primarily in the roots rather than stem tissue, with the leaves maintaining a higher K+ level than the WT control plants. Notably the results indicate the halophyte AlNHX may play a role in root rather than shoot Na+ levels, which was different than observations in overexpressed OsNHX1 in transgenic rice (Fukuda *et al.*, 2004).

In fact there have been many examples of increased salt tolerance resulting from overexpression of the NHX family of Na+/H+ antiporters from various plant species, selected examples being: perennial ryegrass transformed with OsNHX1 (Wu *et al.*, 2005), wheat (Xue *et al.*, 2004), *Petunia hybrida* with AtNHX1 (Xu *et al.*, 2009) demonstrated increased salt and drought tolerance. Recently the SsNHX1 gene from the halophyte *Salsola soda* (Li *et al.*, 2010) conferred salt tolerance when overexpressed in transgenic alfalfa (Medicago sativa). These transgenic alfalfa plants had the ability to grow normally for 50 days under Na+ treatment, with no apparent difference in growth detectable between transgenic plants and wild-type plants under normal conditions, likely due to the use of the stress inducible promoter rd29A rather than the typical constitutively active Ca35S promoter. Clearly the NHX family of Na+/H+ of antiporters are able to confer increased Na+ tolerance across a wide range of plant species, and aside from being a single trait, may be even more relevant from a genetic engineering standpoint as sixth-generation soybean plants expressing AtNHX1 proved to be just as resistant (Li *et al.*, 2010) to salt stress as the first generation transgenic plants, indicating this single trait change in heritable.

Although much research has been conducted into the NHX family of Na+/H+ of antiporters, several others have also shown promise for genetic

engineering. The plasma membrane Na"/H+ antiporter SOS1, activated it response to salt stresses by the SOS pathway reviewed above, has been shown to be critical for Na+ partitioning in plant organs as well as the ability for the plants to partition Na+ in the stems, preventing the ions to reach photosynthetic tissues (Olias *et al.*, 2009). Ca+ antiporters, such as *Arabidopsis* H+/Ca2+ Antiporter CAX1 were found to confer increased Ca+ transport and salt tolerance (Chen et al 2004). Both salt and drought tolerance can be significantly increased in *Arabidopsis* plants by overexpressing AtAVP1, the gene encoding a vacuolar pyrophosphatase which acts as a vacuolar membrane proton pump (Gaxiola *et al.*, 2001), moving more H+ into the vacuoles to create a higher electrochemical gradient. In addition to *Arabidopsis*, overexpression of AtAVP1 in tomato also enhances drought tolerance (Park *et al.*, 2005), due the increased osmotic adjustment ability conferred by the increased vacuolar H+ concentration.

Gene expression studies in halophytes have yielded fascinating candidate genes for future study; root and leaf tissue collected from *Kosteletzkya virginica* seedlings (Guo *et al.*, 2008) identified 34 differentially expressed gene fragments homologous to known genes from other species and 4 of novel function. The differentially expressed genes were classified into four groups: those necessary for re-establishing ion homeostasis those involved in metabolism or energy and resuming plant growth and development under salt stress, those involved in regulation of gene expression, and those responsible for signal transduction (Guo *et al.*, 2008).

The halophyte *Craterostigma plantagineum*, known as the resurrection plant, has the ability to survive complete dehydration. In an attempt to further understand desiccation tolerance in this plant, the CpMYB10 transcription factor gene was functionally characterized (Villalobos *et al.*, 2004) and found to be rapidly induced by dehydration and abscisic ABA treatments in leaves

and roots, with no expression detected in fully hydrated tissues. Its subsequent overexpression in *Arabidopsis* also leads to salt tolerance of the transgenic lines. However, it also was found that plants overexpressing CpMYB10 also exhibited glucose-insensitive and ABA hypersensitive phenotypes. This finding exemplifies an issue in studies with model organisms in short-term laboratory settings: is there overlap between the molecular mechanisms to cope with stress in *Arabidopsis*, crops plants, and halophytes? Are there overlaps between gene regulation, transcriptional activators, and their tissue-specificity? These distinctions are essential in order for genetic engineering to beneficially be used in crop species for trait selection. One useful tool for candidate gene discovery is genome-wide profiling of both stress-induced expression and post-transcriptional events occurring as a result of stress exposure.

Recent microarray studies have provided sets of candidate genes for further investigation in order to define the transcriptome profile under salt stress. Tomato (Zhou *et al.*, 2007) gene expression was profiled under salt stress, discovering several key enzyme genes in the metabolic pathways of carbohydrates, amino acids, and fatty acids to be initiated. Also higher transcript levels were detected for antioxidant enzymes, ion transporters, and genes known to be involved with numerous signal transduction pathways.

The Euphrat poplar tree (*Populus euphratica*) that thrives in a saline and arid environment is expanding our understanding of stress induced gene networks in trees, which spend a much greater amount of time in soils due to their longer life and therefore must possess robust systems for dealing with abiotic stresses. Acclimation to increasing levels of Na+ requires adjustment to the osmotic pressure of leaves, achieved by accumulation of Na+ and compensatory decreases in Ca+ and soluble carbohydrates. The primary strategy of *P. euphratica* to protect the cytosol against sodium toxicity is

apoplastic, instead of vacuolar, salt accumulation, suggesting that Na+ adaptation requires suppression of Ca+ related signaling pathways. Evidence also points to shifts in carbohydrate metabolism and suppression of reactive oxygen species in mitochondria under salt stress (Ottow *et al.*, 2005). Overexpression of a single Ca+ dependent protein kinase in rice increases salt tolerance (Saijo *et al.*, 2000), with levels of tolerance correlation to levels of protein.

A recent microarray study of *P. euphratica* by Brinker *et al.* (2010) noted three distinct transcriptome phase changes associated with salt stress, with the duration and intensity of these phases differing between the leaf and root tissues sampled. Key factors initially involved with salinity-stress are molecular chaperones, namely the dehydrins and osmotin, which assist with protein stabilization. Leaves initially suffered from dehydration stress, resulting in transcript level shifts of mitochondrial and photosynthetic genes, indicating adjustment of energy metabolism. Initially a decrease in known stress-associated genes occurs, with induction occurring later, after excessive sodium concentrations accumulate in the leaves. In roots a decrease in aquaporins occurs, potentially reducing water loss. Roots and leaves perceive physiologically different stress situations, and therefore activate unique stress responses; however sucrose synthase and chaperones from leaves were also found upregulated in roots as the only overlapping salt-responsive genes in roots and leaves. To identify the stress-specific genes within the poplar salt-stress responsive transcriptome Brinker *et al.*, used in silico analyses with *Arabidopsis* orthologs to reduce the number of candidate genes for functional analysis. Ultimately two genes, a lipocalin-like gene and a gene encoding a protein with previously unknown functions were identified and shown to display salt-sensitive phenotypes in *Arabidopsis* knockout mutants, suggesting these genes play roles in salt tolerance. These results are quite exciting, since they demonstrate salt-susceptible plants harbor genes

important for salt tolerance that cannot be identified by conventional salt screens relying on differential gene expression (Brinker *et al.* 2010).

Foxtail millet (*Setaria italica*) is a food and fodder grain crop grown in arid and semi-arid regions (Puranik *et al.*, 2011) and is a self-pollinating, diploid, C4 grass. Comparative transcriptome analyses between two cultivars differing in response to short-term salinity stress identified 81 differentially expressed novel transcripts. These transcripts represent an "untapped genetic resource" (Puranik *et al.*, 2011), in a model crop with natural increased resistance to abiotic stress.

High-throughput Illumina based RNA-seq experiments are allowing for genome-wide glimpses into transcript abundance and transcriptional regulation, having the benefit of not requiring previously annotated genes or being limited to specific probes present on a microarray. Genome-wide mapping of alternative splicing in *Arabidopsis* under abiotic stresses (Filichkin *et al.*, 2010) have identified different types of stress differentially regulating known genes implemented in various pathways and cellular responses. For example, a splicing factor in the SR (serine/arginine rich) family, SRP30/SR30 (*At*1g09140), displays upregulation of the reference isoform under salt stress. This makes SR30 a candidate for further study in order to elucidate salt-stress responses from a splicing factor, rather than a transcriptional angle.

Another Illumina-based RNA-seq experiment using rice (*Oryza sativa* L. 'Nipponbare) cDNAs focused on the identification of salt-responsive unannotated transcripts derived from root and shoot mRNAs in rice and those transcripts encoding putative functional proteins (Mizuno *et al.*, 2010). 7-day old rice seedlings were transferred to either 150mM NaCl solution or water (control) for 1hr. Of the total unannotated transcripts discovered, 1,525 in

shoot and 1,659 in root were novel transcripts. Of these transcripts, 213 (shoot) and 436 (root) were differentially expressed in response to salinity stress. The predicted encoded proteins were associated with amino acid metabolism in response to abiotic stresses, and mechanosensitive ion channel function. These responses are gated directly by physical stimuli such as osmotic shock and known to transduce these stimuli into electrical signals. Also captured were previously identified genes involved in salinity tolerance; those associated with trehalose synthesis, dehydrin, ABA synthesis sugar transport, glycerol transferase, and transcription factors similar to those of the DREB family (Mizuno *et al.*, 2010). The DREB transcription activators are involved in ABA-independent and abiotic stress response, binding to the consensus dehydration-responsive element (DRE), present in promoter regions of genes induced by osmotic, saline, and cold stresses (Stockinger *et al.*, 1997). As a substantial number of transcripts were exclusively upregulated only in the root, being directly exposed to 1 hour of salinity stress, it was hypothesized it may take longer exposure time to induce a greater network of genes (Mizuno *et al.*, 2010).

Cold. Low temperatures, both sudden and for sustained periods, cause dramatic decreases in crop sustainability and yield by affecting the germination and reproductive rate of plants. Low temperature induced cold stress leads to reduced cell expansion and consequently reduced leaf growth, with the loss in turgor pressure causing severe wilting of leaves, ultimately leading to plant death.

Plants differ in their abilities to survive both freezing (temperatures below 0°C) and chilling (0°C to around 20°C) conditions by modifying their physiological and cellular states. The seeds of plants native to latitudes undergoing a freezing winter period such as the winter cereals (certain barley and wheat cultivars, rye, and oats, among others) require a period of cold

temperature, called vernalization, prior to germination. This epigenetic response alters the chromatin structure of a flowering repressor gene, in effect allowing the seedlings to "remember" the period of cold preceding the warmth of the growing season (Sung & Amasino, 2009). The vernalization period is necessary to prevent premature transition to the reproductive phase before the winter freezing threat has ended, however this does not continue past onset of the vegetative phase (Chinnusamy *et al.*, 2007). During the warm growing season, these temperate region plants have little ability to withstand freezing, however as the temperatures gradually fall in the time preceding winter, they are able to increase their freezing tolerance by undergoing a phenomenon known as cold acclimation (Thomashow 1999). This results in cell membrane composition and protein concentration changes to reduce intracellular ice crystal formation and dehydration due to freezing (Thomashow, 1998). Plants that do not undergo this gradual acclimation phase have drastically reduced tolerance to freezing. Temperatures of -5°C kill non-acclimated rye yet after a period of gradual exposure to low nonfreezing temperatures the plant is able to survive freezing down to -30°C (Thomashow 1999). Plants native to warmer regions such as the tropics are much more sensitive to chilling and generally lack the ability to acclimatize. Several of these plants are agriculturally important; such as rice, tomato, soybean, grapes, and maize. To this end, efforts have been made to increase freezing tolerance of these plants by combinations of transgenic and conventional breeding approaches, which will be discussed in more detail later.

Much research in the past decade has been directed towards dissecting the mechanisms by which plants initially sense low temperature to subsequently activate the cold-acclimation response, along with regulation by transcription factors, post-transcriptional modifications, secondary messengers, and cross-talk with other stress responses at stress response "nodes". Research has

focused on the identification of freezing-tolerance genes through microarray, high-throughput sequencing, and genetic approaches such as comparative studies of freezing-tolerant cultivars. Much information has been yielded thus far, however the story is far from complete. The following section of this chapter will initially provide an overview of the physiology and mechanisms causing freezing injury to the plant, work through our current understanding of the subsequent response pathway(s) and the players involved, before concluding with examples of genetic engineering for improved freezing tolerance and how the genomics era will continue to yield further insight into this multifaceted field. Our understanding of the cold response pathway and the roles of the genes involved is continually improving, and ultimately this will allow for directed single gene modification at multiple steps of the pathway, allowing for enhanced crop improvement.

There are two types of physiological changes a plant must confront upon the onset of cold temperatures: osmotic stress from low non-freezing temperatures and severe membrane damage from freezing. Chilling stress results in ratio changes between fatty acids and proteins as well as decreased membrane fluidity, due to fatty acid unsaturation in membrane lipids (Wang *et al.*, 2006). Chilling also promotes dehydration due to the impairment of water uptake from the roots and a reduction in stomatal closure. Yet by far the largest cause of cold-associated crop loss is membrane damage as a result of freezing, along with the associated intracellular ice crystal formation leading to further dehydration.

Initially ice crystals form in the cell walls and intracellular spaces, decreasing the water potential outside the cell. The unfrozen water within the cell then travels the down the potential gradient, moving out of the cell and towards the intercellular spaces. This dehydration is what leads to the wilting phenotypes of leaf tissue after exposure to freezing temperatures, or in crops

as a result of a “cold snap”. Colder temperatures result in greater water loss: at -10°C 90% of the osmotically active water will move out of the cell into intercellular spaces (Thomashow, 1998). Freeze-induced cellular dehydration also results in a barrage of membrane damage: expansion-induced-lysis, lamellar to hexagonal - II phase transitions and fracture jump lesions (Uemura *et al.*, 1995, Steponkus *et al.*, 1993). Expansion-induced lysis occurs at temperatures around -2° to -4°C and is a result of the mechanical damage due to multiple freeze/thaw cycles, where the expansion and contraction of the plasma membrane leads to rupturing (lysing) of the cellular membrane.

Injury from fracture jump lesions is associated with the occurrence of localized deviations of the plasma membrane fracture plane to closely appressed lamellae (Webb *et al.*, 1994). The cold acclimation process has been shown (Uemura *et al.*, 1995, Steponkus *et al.*, 1993) to prevent both expansion induced lysis and the formation of hexagonal II phase lipids in rye and other plants.

Multiple mechanisms are involved in the stabilization of the plant cell membrane. The content and composition of polar lipids and fatty acids in tomato (*Lycopersicon esculentum*) at 6°C suggests maintenance of high levels of chloroplast membrane lipids play an important role in the survival of cold-tolerant plants (Novitskaya *et al.*, 2000). The *Arabidopsis* dSFR2 protein also compensates for changes in organelle volume and stabilizes the chloroplast membranes during freezing (Moellering *et al.*, 2010).

The accumulation of sucrose and related simple sugars correspond with cold acclimation, quite likely contributing in part to the stabilization of plant plasma membranes. Investigation of sucrose metabolizing enzyme activity and sugar content during cold acclimation of perennial ryegrass (Lolium

perenne) found that the sucrose metabolizing enzymes: phosphate synthase, sucrose synthase and sucrose phosphate synthase are similarly regulated by cold acclimation (Bhowmik *et al.*, 2006). In *Arabidopsis*, sucrose was found to have a regulatory role in the acclimation of whole plants to cold, likely also playing an important role during diurnal dark periods (Rekarte-Cowie *et al.*, 2008).

In addition there is emerging evidence that certain novel hydrophilic and LEA (late embryogenesis abundant) polypeptides also participate in the stabilization of membranes against freeze-induced injury, where disordered plant LEA proteins act as molecular chaperones (Kovacs *et al.*, 2008). The level of expression of the winter barley LEA abscisic acid-regulated gene HVA1 accumulates upon cold acclimation, before disappearing 2 hours post exposure, with greater expression in the lesser of freezing-resistant cultivars (Sutton *et al.*, 1992). Accumulation of chloroplast LEA proteins is correlated with the capacity of different wheat and rye cultivars to develop freezing tolerance (Dong *et al.*, 2002). Transgenic *Arabidopsis* expressing a wheat LEA gene displays significant increases in freezing tolerance in cold-acclimated plants. *Arabidopsis* Cor15am is a late embryogenesis abundant (LEA) related protein shown to exhibit cryoprotective activity in vitro, likely by preventing protein aggregation (Nakayama *et al.*, 2008). Global expression profiles of rice genes under abiotic stresses (Rabbani *et al.*, 2003) using microarrays found an upregulation of LEA proteins post-stress. Genome-wide analysis of LEA proteins in *Arabidopsis* identified 51 LEA protein encoding genes in the having ABA and/or low temperature response elements in their promoters and, thus induced by ABA, cold, or drought (Hundertmark *et al.*, 2008). The majority of LEA proteins were predicted to be highly hydrophilic and natively unstructured, but some were predicted to be folded. This comprehensive analysis will be an important starting point for

future efforts to elucidate the functional role of these proteins (Hundertmark *et al.*, 2008).

Plant cells initially sense cold stress resulting from the change in the fluidity of the cellular membrane. Cellular membranes are inherently dynamic, and cytoskeleton re-organization is an integral component in low-temperature signal transduction. The cold acclimation process is associated with gene expression requiring a transient influx of Ca+ from the cytosol. Under normal conditions the influx of Ca+ at 4C is nearly 15 times greater than at 25C, but when treated with chemical agents causing an increased Ca+ influx, cold acclimatization-specific genes are expressed at higher temperatures (Monroy & Dhindsa, 1995). When alfalfa (*Medicado sativa*) cells are treated with chemicals blocking this influx (Ovar *et al.*, 2000), they are unable to cold-acclimatize. Furthermore Ovar *et al.* demonstrated the activation of cold-acclimation genes, Ca+ influx, and freezing tolerance at 4C to all be prevented by membrane stabilization, yet induced at 25C by the addition of an actin microfilament destabilizer, thereby linking the membrane rigidification process to the influx of Ca+ necessary to signal cold acclimation genes. Calcium sensing and sequestering proteins (Komatsu *et al.*, 2007), and phosphoinositides also play roles as signaling molecules in the cold-stress pathway. Phosphoinositides are signaling molecules that regulate cellular events including vesicle targeting and interactions between membrane and cytoskeleton, and accumulate in salt, cold, and osmotically stressed plants (Williams *et al.*, 2005). Mutations in the *Arabidopsis* phosphoinositide phosphatase gene SAC9 lead to overaccumulation of phosphoinositides and confer the characteristics of a constitutive stress response, including dwarfism, closed stomata. The mutations also upregulate stress-induced genes and overaccumulate ROS.

Accumulation of ROS (O2-, H2O2, and HO) as secondary signals has strong impacts on plants ability to withstand cold. Once thought to only be an un-wanted byproduct of aerobic metabolism upregulated under biotic and abiotic stresses, ROS are now known to act as key regulators in numerous biological processes (Miller *et al.*, 2008). An *Arabidopsis* mutant defective in the respiratory electron chain of mitochondria (frostbite1) constitutively produces ROS and displays reduced cold induction of stress-responsive genes such as RD29A, KIN1, COR15A, and COR47. The leaves also have a reduced capacity for cold acclimation, appear water-soaked, and leak electrolytes (Lee *et al.*, 2002).

Hormones are also implicated in the response of plants to environmental stresses. The polyamine putrescine also progressively increases upon cold stress treatment and likely acts as regulator of hormone biosynthesis (Cuevas *et al.*, 2008). Loss of function mutants and reverse complementation tests indicated that putrescine also modulates ABA biosynthesis at the transcriptional level in response to low temperature. Hormonal levels drive cell division and expansion and the plant hormone auxin is a key regulator of virtually every aspect of plant growth and development. Auxin plays a major role in cell expansion and growth, as well as being quite sensitive to temperature changes (Gray *et al.* 1998). Root growth and gravity response of *Arabidopsis* after cold stress suggests that cold stress affects auxin transport rather than auxin signaling (Shibasaki *et al.*, 2009). Additionally, cold stress differentially affects various protein trafficking pathways, independently of cellular actin organization and membrane fluidity. Taken together, these results suggest that the effect of cold stress on auxin is linked to the inhibition of intracellular movement of auxin efflux carriers (Shibasaki et al 2009).

In 1991 Johnson-Flanagan *et al.* demonstrated increased freezing tolerance of Brassica napus suspension-cultured cells by the addition of the herbicide mefluidide or ABA to the culture medium. In 2000 Llorente *et al.* showed that ABA is required for full development of freezing tolerance in cold-acclimated *Arabidopsis*, and plays a role in mediating constitutive freezing tolerance. The *Arabidopsis* mutant frs1 (freezing sensitive 1) is deficient in an allele of the ABA3 locus, displaying reduced constitutive freezing tolerance as well as tolerance post cold acclimation, producing the wilty phenotype corresponding with excessive water loss. Upon receiving an exogenous ABA treatment, frs1 plants recover both their wild-type phenotype and capability to tolerate freezing temperatures and retain water. Gene expression in the frs1 mutants was also altered in response to dehydration, suggesting dependence on ABA-regulated proteins allowing plants to cope with freeze-induced cellular dehydration (Llorente *et al.*, 2000). Not all genes induced by low temperature are ABA-dependent, as evidenced by some of transcriptional regulators mentioned in the following section and indicative of the complexity and crosstalk of the regulatory network. Recent genome-wide profiling studies have begun to identify further downstream transcription factors and gene targets resulting from both pathways. In the ABA-dependent pathway, ABA likely activates the bZIP (basic leucine zipper) transcription factors, which regulate ABA dependent COR (COld Regulated) genes through ABA-responsive elements (ABRE) promoters. In the ABA-independent pathway, low temperature triggers the expression of the CBF family of transcription factors, which in turn activate downstream COR genes with other specific motifs in their promoters (Thomashow, 1999). There is also evidence (Knight *et al.*, 2004, Talanova *et al.*, 2008) of ABA initiating CBF expression although at lower levels than those resulting from cold acclimation. Both of these pathways confer or enhance freezing tolerance in plants and are described in more detail below.

The CBF cold response pathway plays a central role in cold acclimation and has been the focus of intense research for the past 2 decades. The CBF/DREB (C-repeat-binding factor/dehydration responsive element-binding factor) genes encode a small family of transcriptional activators that play an important role in freezing tolerance and cold acclimation (Thomashow 1999). In *Arabidopsis* there are three members CBF1-3 (also known as DREB1-B, C, and A, respectively), with transcripts beginning to accumulate within 15 minutes after exposure to cold temperatures. A microarray experiment to determine the core set of cold-induced genes in *Arabidopsis* (Vogel *et al.*, 2005) found 302 genes to be upregulated upon cold stress, with 85% of these assigned to the CBF2 regulon and also induced upon CBF2 overexpression. The CBF proteins bind to the CRT/DRE motif (CCGAC) present in the promoters of a number of COR genes named the CBF regulon, which imparts freezing tolerance by activating the COR genes, with CBF induction occurring by ICE1 (Inducer of CBF Expression 1).

ICE1 was identified (Chinnusamy *et al.*, 2003) as an upstream transcription factor regulating transcription of CBF genes in the cold. ICE1 encodes a MYC-like bHLH transcriptional activator that binds the CBF3 promoter. In *Arabidopsis* the ice1 mutation blocks the expression of CBF3 as well as decreases the expression of genes downstream of CBFs, leading to a significant reduction in plant chilling and freezing tolerance. It is also constitutively expressed at low levels, and its overexpression in wild-type plants enhances the expression of the CBF regulon in the cold and improves freezing tolerance of the transgenic plants. ICE2, another bHLH transcription factor and homologue to ICE1, confers decreased levels of carbohydrate and increased levels of lipids when overexpressed in *Arabidopsis* (Fursova *et al.*, 2008). CBF1 displayed differential expression in transgenic plants compared to wild-type control plants, suggesting a regulatory role provided by ICE2. HOS1 is negative regulator of ICE1, mediating its ubiquitination and

subsequent degradation both *in vitro* and *in vivo* (Dong *et al.*, 2006). Overexpression of HOS1 represses expression of the CBFs and their downstream genes, conferring increased sensitivity to freezing stress.

SIZ1, a SUMO E3 ligase, is a positive regulator of ICE1 and the sumoylation of ICE1 may activate and/or stabilize the protein, facilitating expression of CBF3/DREB1A and repression of MYB15, leading to low temperature tolerance (Miura *et al.*, 2007). *Arabidopsis* knockouts siz1-2 and siz1-3 cause freezing and chilling sensitivities indicating that the SIZ1 is a controller of low temperature adaptation in plants. Interestingly a protein associated with stomatal differentiation, SCREAM, was shown to in fact be ICE1 (Kanaoka *et al.*, 2008). This creates a potential link between cold acclimation and stomatal differentiation and a basis for future research.

All three CBF genes do not play the same roles in freezing tolerance. The function of CBF2 was not only demonstrated as having a distinct function from CBF1 and CBF3, but was shown to be a negative regulator of their activity. Reverse genetic approaches using an *Arabidopsis* knockout mutant for CBF2 displayed an increased capacity to tolerate freezing both before and after cold acclimation, and the plants displayed increased tolerance to dehydration and salt stresses (Novillo *et al.*, 2004). The mutants also had stronger and more sustained expression of CBF/DREB1-regulated genes, resulting from increased expression of CBF1 and CBF3 in the cbf2 plants, with the authors suggesting CBF1/CBF3 induction to precede CBF2. Indeed, DNA motifs for the calmodulin binding transcription activator (CAMTA) family of transcription factors have been identified in the promoters of CBF2, as well as the transcription factor ZAT12, conferring both negative and positive regulation. One of these binding sites (CAMTA), was shown to be a positive regulator of CBF2 expression, with mutant plants impaired in freezing tolerance. CAMTA proteins may play a role in cold acclimation by

linking Ca+ and calmodulin signaling with expression of COR genes (Doherty *et al.*, 2009). Both ICE1 and CAMTA binding sites are found in the promoter of CBF2, potentially directly linking Ca+ signaling to cold response.

Low temperature induction of the *Arabidopsis* CBFs is also gated by the circadian clock (Fowler *et al.*, 2005) with the highest and lowest levels of cold-induced CBF1-3 transcript occurring at 4 and 16 h after subjective dawn, respectively. Other transcription factors induced by cold in parallel with CBF1-3 are also gated by the circadian clock; however cycle in the opposite phase. This suggests nonidentical, though potentially overlapping, signaling pathways. Similar results in wheat (Badawi *et al.*, 2007) and tomato (Pennycooke *et al.*, 2008) suggest circadian regulation under homeostatic conditions concurring with dawn and dusk periods, maybe overlapping with stomatal aperture changes.

Light is also implicated in regulating the CBF pathway, for instance a low red to far-red ratio of light is sufficient to increase CBF gene expression and confer freezing tolerance at temperatures higher than those required for cold acclimation (Franklin *et al.*, 2007), providing evidence for a second temperature-regulated step in this pathway. Phytochrome-Interacting Factor7 (PIF7) functions as a transcriptional repressor for DREB1C (CBF2) expression and its activity is regulated by components of the red light photoreceptor, and circadian oscillator (Kidokoro *et al.*, 2009). DREB1/CBF expression may be important for avoiding plant growth retardation by the accumulation of DREB1/CBF proteins under unstressed conditions (Kidokoro *et al.*, 2009). Downregulation occurs through a complex network of transcription factors, such as ZAT12 downregulating CBF2 (Vogel *et al.*, 2005), and MYB15 interacting with ICE1, subsequently binding to the MYB recognition sequences in the CBF promoters (Agarwal *et al.*, 2006)

The CBF pathway is not only present in dicots such as *Arabidopsis*, but is widespread through monocots and multiple plant genera, including those native to warm regions and not inherently cold tolerant, with variation in CBF gene copy numbers (Qin *et al.*, 2004, Skinner *et al.*, 2006, Badawi *et al.*, 2007, Stockinger *et al.*, 2007, Tamura *et al.*, 2007, , Pennycooke *et al.*, 2008, Knox *et al.*, 2010). A recent paper by one of the pioneers of the field presents a detailed overview of the status of CBF research today (Thomashow, 2010), and readers wishing for further detail are directed to this review

Alternative cold tolerance pathways also initiate transcription of cold-responsive genes, for example *Arabidopsis* SFR2 encodes a novel β-glycosidase, contributing to freezing tolerance and distinct from the CBF pathway (Thorlby *et al.*, 2004). The null mutant (sfr2-1) causes freezing sensitivity in *Arabidopsis* possibly due to electrolyte leakage. Homologous genes are present and expressed in many terrestrial plants, including those unable to tolerate freezing. Each of these homologues however, has the ability to complement the freezing sensitivity of the *Arabidopsis* sfr2 mutant (Fourrier *et al.*, 2008). In *Arabidopsis* the SFR2 protein is localized to the chloroplast outer envelope membrane, with the chloroplasts of the sfr2 mutant displaying rapid damage post-freezing.

MYBS3 is a single DNA-binding repeat MYB transcription factor previously shown to mediate sugar signaling in rice and also indicated to play a novel role in cold adaptation (Su *et al.*, 2010). Transgenic rice constitutively overexpressing MYBS3 displayed no yield penalty in normal field conditions while tolerating temperatures of 4°C for at least 1 week. Su *et al.* demonstrated repression of CBF-dependent signaling by MYBS3 at the transcriptional level, with distinct pathways likely acting in parallel for short-

and long-term cold stress in rice. This previously undiscovered cold adaptation pathway adds another layer to the complex web of plant responses to cold stress.

RNA processing and nuclear export / stabilization are critical mechanisms in a plants response to cold stress. Recent research has shown cold shock proteins (CSPSs) play roles in promoting cold tolerance, however much remains to be discovered in order to determine the mechanism in plants to promote cold tolerance. A protein from wheat with homology to an E. coli cold shock protein has been linked to the regulation of translation under low temperature; potentially by acting as a RNA chaperone to destabilize secondary structure (Nakaminami *et al.*, 2006). Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis* have been shown to promote cold adaptation process E. coli (Kim *et al.*, 2006). Excitingly, 2 novel cold shock domain proteins were cloned and characterized from rice, a plant unable to cold acclimatize. Nonetheless, in vivo functional analysis confirmed these OsCSPs complement a cold-sensitive bacterial strain that lacks four endogenous cold shock proteins (Chaikam *et al.*, 2008). Transcripts were also shown to be upregulated during temperature decreases. Two structurally differed CSPs of *Arabidopsis* perform different functions in seed germination and growth under stress conditions, even rescuing cold tolerance from an RNA-binding protein null mutant (Park *et al.*, 2009). Other RNA-binding proteins such as known splicing factors alter expression under cold stress, for example the serine-arginine (SR) rich splicing factor SRP34 (At1g02840), as recently reviewed (Filichkin *et al.*, 2010), displays exon skipping under both drought and cold conditions, with several novel introns predicted through alternative splicing. Alternative splicing of another SR protein, SR1, was reported (Iida *et al.*, 2004) under cold stress, as well as in response to hormones (Palusa *et al.*, 2007). Different isoforms of a splicing factor likely alter the binding preference and spicing of a host of downstream

targets, presenting an exciting area for future research. The *Arabidopsis* STABILIZED1 gene encodes a U5 snRNP– associated splicing factor required for both pre-mRNA splicing and transcript turnover. Of interest, it is also upregulated by cold stress, and the sta1-1 mutant plants are defective in the splicing of COR15A (Lee *et al.*, 2006).

Thanks to the genomics revolution, the role of microRNAs (miRNAs) in abiotic stress regulation is being elucidated. Endogenous miRNA levels change as plants are exposed to abiotic stresses (Sunkar *et al.*, 2004), and readers are pointed towards reviews (Jones-Rhoades *et al.*, 2006, Sunkar *et al.*, 2007) providing the backstory of miRNA research in plants with regards to stresses and classes of miRNA, A tiling array (Matsiu *et al.*, 2008) global transcriptome analysis of *Arabidopsis* discovered 7,719 non-AGI (*Arabidopsis* Genome Initiative) transcriptional units (TUs) in the unannotated "intergenic" regions of *Arabidopsis* genome, and most of these are hypothetical non-protein-coding RNAs. Close to 80% of the previously un-annotated TUs belonged to pairs of the fully overlapping sense-antisense transcripts, suggesting stress or ABA induction of antisense TUs in the fully overlapping sense. These non-coding small RNAs exhibit stress-responsive expression patterns; however are also implicated in a very broad web of in planta regulation. For example, in wheat (Triticum aestivum) the trans targets of miRNAs include both transcription factors implicated in development and a plethora of genes involved multiple physiological processes (Yao *et al.*, 2010). Further research will be necessary to pinpoint small RNA targets and the effects of their regulation with regards to cold stress, and other stresses.

Transcriptome profiling using microarrays is allowing for the identification of groups and networks of genes that respond to cold stress. A consensus among microarray studies (Fowler and Thomashow 2002, Rabbini *et al.*, 2003, Lee *et al.* 2006, Oono *et al.* 2006) is that genes induced by abiotic

stress fall into 2 categories: functional proteins such as the aforementioned LEA proteins, proteins playing roles in osmoprotection, and transporters. The second category is the regulatory proteins; the transcription factors, kinases, phosphatases, and other molecules dealing with signaling, directly or indirectly, such as those of the MAP kinase cascade. Such broad profiling allows for glimpses into regulatory networks and offers the potential for further research into specific up or down-regulated genes or gene families.

Initial work into engineering for cold tolerance focused on the CBF transcription factors. It was initially shown that overexpression of AtCBF3 in *Arabidopsis* confers freezing and drought tolerance, however also causes a dwarf phenotype (Liu *et al.*, 1998). In comparison when AtCBF3 is overexpressed in rice, which is unable to cold acclimate, increased tolerance to drought and high salinity stress (and not low-temperature) without stunting growth results (Oh *et al.*, 2005). Readers are directed towards a recent review (Thomashow, 2010) for further insights gained from the CBF pathway in genetic engineering.

Rice is a staple food in much of the world and the seedlings are particularly sensitive to chilling in high-elevation areas. Much research has been conducted into enhancing the cold tolerance of rice to allow for growth in different geographic regions to increase production. Recently (Hu *et al.*, 2008) isolated SNAC2, a nuclear stress-responsive NAC gene from upland rice (*Oryza sativa* L. ssp japonica) characterized for its role in stress tolerance. Transgenic plants overexpressing SNAC2 had higher cell membrane stability than wild type during cold stress with over half of the transgenic plants, and none of the WT plants, surviving after 5 days at 4C (Hu *et al.*, 2008). Another transcription factor, OsMYB3R-2, functions in both stress and developmental processes in rice, with transgenic rice plants overexpressing OsMYB3R-2 shown to exhibit enhanced cold tolerance by

regulating the progress of the cell cycle during chilling stress (Ma *et al.*, 2009), suggesting cell cycle regulation as possible resistance mechanism to stress.

Molecules involved in the Ca+ signaling pathway, acting upstream of transcription factors, can also enhance cold tolerance. Over-expression of a calcium-dependent protein kinase and a calreticulin interacting protein has been shown to enhance cold tolerance in rice plants, emphasizing the importance of signaling components in the response to cold stress in rice (Komatsu *et al.*, 2007). As mentioned in the salt section, overexpression of a single Ca+ dependent protein kinase in rice increases cold tolerance (Saijo *et al.*, 2000), as well as salt tolerance, indicative of the crosstalk between abiotic stress response pathways.

Research into ethylene response factor (ERF) proteins is demonstrating their roles in plant stress responses via interaction with DRE/CBF genes, yet the regulatory mechanism is not well elucidated. Overexpressing TERF2/LeERF2 in tobacco not only activates expression of cold-related genes, but reduces electrolyte leakage (Zhang *et al.*, 2010). The authors demonstrated RNAi knockdown TERF2/LeERF2 transgenic lines to have reduced freezing tolerance, rescued to normal levels with treatment of a precursor to ethylene. Overexpression of OsTERF2 in rice enhanced cold tolerance without affecting growth or agronomic traits (Tian *et al.*, 2010). The transgenic lines displayed increased accumulation of osmotic substances and chlorophyll, as well as reduced ROS and decreased electrolyte leakage. The overexpression of OsTERF2 was shown to initiate expression of downstream cold regulated genes such as OsMyb, OsICE1, and OsCDPK7.

Our understanding of the cold response pathway has thus far allowed for single trait genetic engineering to improve or alter cold tolerance, often with

very promising results however not without the caveats present in most abiotic stress research, such the differences between monocots and dicots. The traditional model system is the dicot *Arabidopsis*, and as evidenced by the OEX CBF growth phenotype differences between *Arabidopsis* and the monocot rice, homologous genes don't always confer conserved responses. The recent development of molecular, genetic, and genomic resources for the grasses *Brachypodium distachyon*, *Setaria italica*, and *Setaria viridis* provide model platforms for future studies of cold, and other abiotic stress research in general, in monocot systems.

Severe desiccation and water deficit: heat and drought stress. Plants face additive and interacting responses to drought and heat stress such as water loss through the evapotranspiration resulting from the opening of stomata for heat dissipation, and detrimental alterations to photosynthesis. There are subsets of genes that are induced by a combination of heat and drought stress that are not induced by each stress independently, as in a laboratory growth chamber. Under field growing conditions resulting in limited water supply, crop plants would be exposed to both stresses simultaneously. Therefore heat and drought stresses will be profiled in the same overarching section, and will have an additional portion at the end focusing on the overlap and further insights into the crosstalk present between these networks.

Heat stress. A transient elevation in temperature, 10-15C above ambient, is typically defined as heat stress (Wahid *et al.*, 2007); however the effects vary with the duration and amount of temperature increase. Plants differ in their abilities to cope with rising temperatures; corn and rice are more thermotolerant than wheat, for example. As with all stresses, the onset of heat immediately changes the cellular state, alters membrane fluidity and lipid composition, and initiates the signaling cascades that ultimately lead to transcript accumulation for genes encoding protective and chaperone

activities. The following section will profile the cellular changes that occur post-heat stress, with emphasis on how genetic engineering is utilizing these response mechanisms to both elucidate the stress response network and improve heat tolerance in agriculturally important crop species. For a detailed overview of plant heat tolerance, readers are directed to the aforementioned review by Wahid *et al.* (2007).

Gradual nonlethal heat treatment confers a phenomenon known as thermotolerance; an increase in heat resistance over non-acclimated plants similar in principle to the cold acclimation detailed in the previous section, however mechanistically different and fully elucidated. What is understood however, is genetic manipulations of some aspects of the response pathway are able to confer single trait heat tolerance to the resulting transgenic plants. Such advances and insights thus far are profiled in the following section. Engineering for tolerance encompasses many facets of the cells natural defense mechanisms to stress, such as reducing the damaging effects of oxidative stress and subsequent buildup of ROS during heat (and other abiotic stresses). Transgenic potato plants were generated containing both the superoxide dismutase (SOD) and ascorbate peroxidase (APX) genes encoding two key chloroplast enzymes for ROS detoxification under the control of the chloroplast SWPA2 oxidative stress-inducible promoter (Tang *et al.*, 2006). Under high temperature treatment, the transgenic plants displayed a photosynthetic activity decrease of only 6%, whereas wild-type plants displayed a 29% decrease. These results suggest that manipulation of the antioxidative mechanism is likely a valuable tool for the creation of heat tolerant crop plants.

The osmolyte glycinebetaine (mentioned earlier in the section on salt) has been implicated in heat tolerance, although the exact mechanism though which tolerance is gained remains unknown. Tobacco (*Nicotiana tabacum*)

lines transgenically accumulating glycinebetaine display higher thermotolerance than WT plants, especially when heat stress occurs under light, suggesting that the accumulation of glycinebetaine leads to increased tolerance to heat-enhanced photoinhibition. This tolerance is likely achieved by accelerating repair of photosystem II (PSII), possibly due to the reduced accumulation of ROS in the transgenic plants with elevated levels of glycinebetaine (Yang *et al.*, 2007). Isoprene is a volatile compound emitted from leaves of many plant species, and has also has been implemented in heat tolerance. Recently the *Populus alba* isoprene synthase gene was introduced into *Arabidopsis* and shown to confer elevated heat tolerance in the transgenic lines over wild type (Sasaki *et al.*, 2007).

As with chilling stress, it is becoming evident that heat stress promotes fatty acid unsaturation in membrane lipids, altering the ratio between membrane fatty acids and proteins and resulting in membrane fluidity changes. Protein transfer across membranes is mediated by protein machinery embedded in the membrane, with different lipid classes within a membrane is known to influence the efficiency of some protein translocation processes (Ma *et al.*, 2006). To this end, membrane associated proteins involved in lipid metabolism have been successfully utilized to increase thermotolerance in both model and crop plants. Fatty acid omega-3 desaturase (FAD) is the key enzyme catalyzing the formation of trienoic fatty acids, the most common fatty acids in membrane lipids, comprising 70% of the membrane lipids in the chloroplast and implemented with defense response (Yaeno *et al.*, 2004).

By investigating transgenic tobacco plants with reduced trienoic fatty acid content (Murakami *et al.*, 2000) it was revealed that decreased contents of trienoic fatty acids play an important role in high-temperature tolerance. Transgenic rice plants in which the content of dienoic fatty acids was increased were more tolerant to high temperatures than WT, having increases

in both chlorophyll content and growth. The maximum photochemical efficiency of PSII was also higher in transgenic plants plants upon high temperature stress (Sohn *et al.*, 2007). Recently, antisense expression of tomato chloroplast omega-3 fatty acid desaturase gene (LeFAD7) was demonstrated to enhance high-temperature tolerance, again through reductions of trienoic fatty acids and increases of dienoic fatty acids (Liu X. *et al.*, 2010).

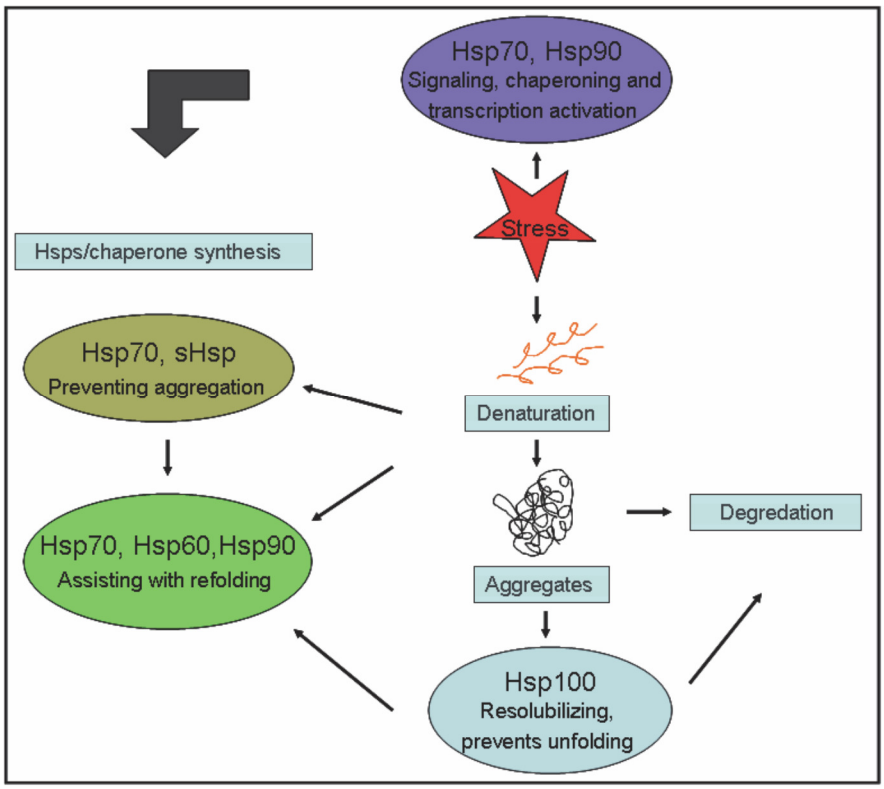
Photosynthesis, the light driven carbon dioxide assimilation process and the primary means of energy production in plants, is extremely sensitive to elevated temperatures. Heat stress inhibits photosynthesis in part by reducing the activation of Rubisco, due initially to the denaturation of Rubisco activase (Salvucci *et al.*, 2001). Loss of activase activity during heat stress is caused by exceptional sensitivity of the protein to thermal denaturation and is responsible in part for deactivation of Rubisco itself. The effects of heat stress on *Arabidopsis* plants in which Rubisco activase or chloroplast thylakoid membrane fluidity had been altered demonstrated that a) plants having less polyunsaturation of thylakoid lipids display lower net photosynthetic rates than the WT and b) the rate of Rubisco deactivation affects the temperature dependence of photosynthesis (Kim *et al.*, 2005). To test the hypothesis that a non-degraded Rubisco activase can improve photosynthesis under elevated temperatures, several thermostable *Arabidopsis* isoforms of Rubisco activase were introduced into a Rubisco activase null mutant line. The transgenics displayed higher photosynthetic rates, with increased biomass and increased seed yields, compared to wild-type activase, providing evidence for Rubisco activase as a limiting factor in photosynthesis elevated temperatures (Kurek *et al.*, 2007). Rubisco activase is a potential target for future genetic manipulation in improving crop plants productivity under heat stress (Kurek *et al.*, 2007). In addition, down-regulation of photosynthesis in temperature stressed plants is caused by

reduced post-translational import into the chloroplast of plastidic proteins required for the replacement of impaired proteins coded by the nuclear genome (Dutta *et al.*, 2010).

Heat stress also inhibits synthesis and promotes degradation of cytokinins, important hormones for regulation of growth and development processes, such as cell division, leaf senescence, and root growth (Xu *et al.*, 2010), however the underlying mechanisms are poorly understood. Xu *et al.* used transgenic *Agrostis stolonifera*, a C3 perennial grass species, to survey protein changes in response to elevated temperatures. The gene controlling cytokinin synthesis was used to create 2 transgenic lines, each with different inducible promoters, and a null mutant line. Protein content changes in leaf and root tissue were found to primarily regulate energy metabolism, protein destination and storage. In the transgenic lines, 6 leaf proteins and 9 root proteins were found to be elevated or remain at steady state comparable WT levels, and among these was the small subunit of Rubisco, Hsp90, and glycolate oxidase, a suggesting a definite regulatory role for cytokinins in metabolic pathway regulation associated with heat tolerance in C3 perennial grass species (Xu *et al.*, 2010).

Much research has been conducted on heat shock proteins (HSPs), the molecular chaperones regulating proper protein folding, localization, degradation, and stabilization of under homeostatic and stress conditions (Feder *et al.*, 1999). There are several families of HSPs present in both plants and animals, named based on their respective molecular weights. There are 5 classes of HSPs in plants (for comprehensive reviews see Baniwal *et al.*, 2004, Wang *et al.*, 2004, Kotak *et al.*, 2007); the Hsp70 class which prevents protein aggregation and assists with transcriptional activation and import, the Hsp60 chaperonin class which assists with folding and re-folding, the Hsp90 class which plays a role in assisting other signaling molecules, the Hsp100

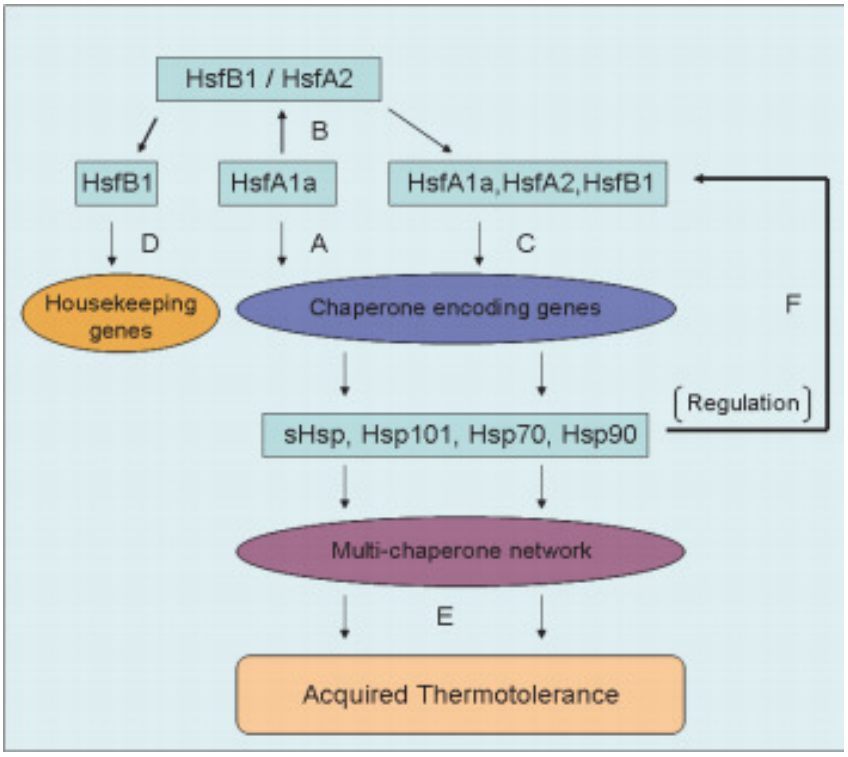
class preventing unfolding, and finally the sHSP class which act to stabilize non-native proteins (**Figure 4.2**).



Much of our current knowledge regarding HSPs contribution towards plant survival under heat stress is based off homology with other eukaryotes and extrapolation based molecular chaperoning activity and in vitro, with little specific in vivo information (Kotak *et al.*, 2007).

In the absence of heat shock, cytosolic HSP90 appears to negatively regulate heat-inducible genes by actively suppressing Hsp function, however in is transiently inactivated following heat shock, leading to Hsf activation (Yamada *et al.*, 2007). AtHsp101, when constitutively expressed in rice, enhances thermotolerance (Katiyar-Agarwal *et al.*, 2003).

It is the heat stress transcription factor (Hsf) family of more than 20 members, which are the central regulation proteins of heat stress response and defense (Baniwal *et al.*, 2004). These modular classes (A, B, and C) share motifs for DNA binding and transcriptional activation, and are defined by differences in the hydrophobic amino acid residues required for oligomerization. The B and C class Hsfs are believed to function in conjunction with class A Hsfs to amplify or regulate signals, rather than function on their own. The majority of our knowledge into the mechanisms into plant Hsfs has come from studies with 2 dicots: *Arabidopsis* and tomato (**Figure 4.3**).



Tomato has 17 members of the Hsf family, however despite this complexity; HsfA1 has a unique function as being the "master regulator" for induced thermotolerance and cannot be replaced with any of the other Hsf members (Mishra *et al.*, 2002). This is not the case in *Arabidopsis* however, where

sequencing of *Arabidopsis* genome revealed unique complexity of the Hsf family. Hsfs comprising 21 members were assigned to 3 classes and 14 groups based structural and phylogenetic comparison to homologues in other eukaryotes and plants (Nover et al 2001). No master regulator has been yet identified in *Arabidopsis*, where even double knockouts only affect a small subset of genes. While offering a beginning for homology comparisons, there is not complete overlap between the Hsfs in tomato and *Arabidopsis* however, for example while HsfA1a and HsfA1b are highly conserved between species. In *Arabidopsis* (unlike tomato) they have the capacity to functionally replace each other. Recent work implicates HsfA1a/1b in cooperation at a number of target gene promoters also regulated by HsfA2, possibly indicating a recruitment of HsfA2 and replacement of HsfA1a/A1b at the same target gene promoters (Li *et al.*, 2010).

o date only a few of the Hsfs have been studied in depth, and the following brief summary of the Hsf network comes from an excellent review by von Koskull-Doring *et al.*, from 2007, which readers are encouraged to read for a detailed summary of Hsf structure and function. HsfA2 is the dominant Hsf present in thermotolerant cells in both *Arabidopsis* and tomato, and may initiate transcription of a core subset of heat stress induced genes. Recent work has also implicated AtHsfA2 in anoxia tolerance in *Arabidopsis* (Banti *et al.*, 2010), further demonstrating the overlapping and redundancy of this complex network. AtHsfA2 also plays an important role in linking heat shock with oxidative stress signals (Li *et al.*, 2005). A recent study (Cohen-Peer *et al.*, 2010) has demonstrated AtSUMO1 of AtHsfA2 to be involved with the plant's regulatory response to heat stress and acquired thermotolerance. Post-translational modification of target proteins by SUMO proteins (see cold section for background) regulates many cellular processes, and adds a further layer to this complex network In a recent study to identify potential regulatory components involved in thermotolerance, a reverse genetics

approach was used by screening *Arabidopsis* T-DNA insertion mutants for lines displaying phenotypic decreased thermotolerance. The Hsf AtHsfA2 fell out as the only mutant line more sensitive to severe heat stress than WT following long recovery periods, and able to be complemented by the introduction of WT AtHsfA2. This depicts HsfA2 as a heat-inducible transactivator, sustaining expression of Hsp genes and extending the duration of acquired thermotolerance in *Arabidopsis* (Charng *et al.*, 2007), as well as being an attractive candidate for continued research in the orthologous genes of crop plants under field conditions.

LeHsfB1 interacts with HsfA1a in a synergistic fashion to form an "enhanceosome" complex to possibly regulate the expression of housekeeping genes during periods of heat stress. Both tomato and *Arabidopsis* HsfA5 acts as an inhibitor of the activator HsfA4, by initiating the formation of hetero-oligomer complexes. HsfA9 plays a role in seed development and maturation, likely working in conjunction with other networks during heat stress, and is shown to induce expression of small heat shock proteins (sHsps) and Hsp101 in *Arabidopsis* leaves under non-stressed conditions (Koskukll-Doring *et al.*, 2007, and all references contained therein). HsfA3 is implicated is a crosstalk network with drought stress, with transcription in fact being induced DREB2A, in a cascade resulting in the transcription of genes encoding protective Hsps (Schramm *et al.*, 2008, Yoshida *et al.*, 2008, Chen *et al.*, 2010).

The Hsp/Hsf network in plants response to heat stress is quite complex, and still being fully elucidated. Heat responses in monocots may increase the complexity of the network yet again. In contrast to tomato and *Arabidopsis* containing only one HsfA2, rice has five HsfA2 genes (von Koskull-Döring *et al.*, 2007). Expression profiles of 12 class A OsHsfAs suggest different regulatory networks between heat and non-heat stress (Liu *et al.*, 2009). A

population of *Arabidopsis* was transformed with a full-length rice cDNA library in order to isolate the rice genes responsible for high-temperature stress tolerance (Yokotani *et al.*, 2008),. A thermotolerant line encoding the rice the heat stress transcription factor OsHsfA2 fell out of the analysis as highly expressing several classes of heat-shock proteins (Yokotani *et al.*, 2008), and also displaying tolerance to high-salinity stress. A genome-wide analysis of rice, including *Oryza sativa* L. ssp japonica and *Oryza sativa* L. ssp indica (Wang *et al.*, 2009), identified 25 rice Hsf genes. Promoter analysis identified a number of stress-related cis-elements in the promoter regions, however no correlation was found between heat-shock gene responses and their cis-elements. This study sets the foundation for future research into OsHsf function and tolerance. A recent study in the dicot grape (*Vitis vinifera*) identified four genes strongly upregulated by heat stress, whose overexpression resulted in the acquisition of thermotolerance in *Arabidopsis* (Kobayashi *et al.*, 2010). Further in vivo studies in grape are underway to elucidate chaperone mechanisms, localization, and functions under stress conditions (Wang *et al.*, 2009).

The complete mechanisms of Hsp mediated thermotolerance remains to be fully elucidated in plants; however work with transgenics has shown that altered levels of Hsps and Hsfs have dramatic effects on plants resistance to elevated temperatures. This offers a promising outlook for future research, utilizing non-model organisms and trials under realistic field conditions. Indeed, based on transcriptional response profiling, *Arabidopsis* Hsf and Hsp expression has been shown to be strongly induced by heat, cold, salt, (stresses sharing osmotic components), and upon wounding, suggesting an interaction point between multiple stress response pathways, warranting functional analysis under conditions apart from heat shock treatments, presenting another area for future research (Swindell *et al.*, 2007).

Drought. A water deficit, along with freezing and increased Na+ concentration, all disturb the water content of the cell, thus altering membrane fluidity, protein stability, and water potential gradients. This osmotic stress leads to wilting associated with loss of turgor pressure, and ultimately complete desiccation. Cellular sensors initially perceive and respond to the drought induced signaling, triggering gene expression changes to synthesize additional signals such as ABA. Further signaling cascades are then initiated to signal new gene expression patterns that are proposed to play a role in cellular adaptation to water-deficit stress (Bray 2002). Drought stress shares many of the same response pathways as the other osmotic stress pathways profiled earlier on the sections on salt and cold.

As with these other osmotic stressors, there are two classes of proteins synthesized as a result of cellular perception of drought stress. First the regulatory proteins; kinases, components of signaling transduction and amplification pathways such as ABA, and the transcription factors activating genes encoding protective proteins. The second class of proteins in the functional proteins, those serving protecting and chaperoning roles such as LEA proteins, osmoprotectants such as proline, proteins regulating water channels for turgor pressure, and proteases. It is these classes of proteins that have been the focus of genetic engineering for drought tolerance, as well as the focus of the following section.

A plethora of recent reviews exist on the physiological responses of the plant to drought stress (Yordanov *et al.*, 2000, Wang *et al.*, 2003, Bartels *et al.*, 2005, Umezawa *et al.*, 2006, Barnabas *et al.*, 2008), and methods to engineer tolerance to water deficit. Readers interested in detailed background on drought stress are encouraged to read these reviews; this section offers only a small summary of some of the most recent discoveries and genetic engineering. Dehydrins are members of the LEA family of proteins and as

mentioned previously aid in stabilizing proteins and other molecules during stresses, likely by replacing water to maintain homeostasis. Other hypotheses for roles played by LEA proteins are: compensating for the increasing ionic concentration by binding ions in dehydrated cells, and interaction with carbohydrates to prevent cellular damage during dehydration (Bartels 2005). Dehydrins are also present in fungi as well as plants; in the white truffle (Tuber borchii), a novel dehydrin-like coding gene displays increases transcript abundance during cellular dehydration (Abba *et al.*, 2006). A promoter region from a dehydrin in coffee (*Coffea canephora*) has been cloned and implicated in osmotic stress-specific gene expression (Hinniger *et al.*, 2005), and will be useful for studying control of gene expression during osmotic stress in coffee, an important crop.

Xerophytes are plants that are adapted to life in a low water environment, typically by employing altered root function. Watermelon (*Citrullus lanatus*) is one of these plants, and a recent study (Yoshimura *et al.*, 2008) provides insights into the molecular mechanisms behind their adapted root system. In the early stages of drought stress watermelon displays enhanced root development, a drought avoidance mechanism for absorbing water deeper beneath the surface layer of soil. Proteome analysis revealed proteins induced in the early stage of drought stress to be involved in root morphogenesis and carbon/nitrogen metabolism, likely promoting rapid root development and growth. In later stages of drought stress however, the protein ratios shifted to lignin synthesis-related proteins and molecular chaperones, enhancing desiccation tolerance and protein stability. Developed root systems are not the only method plants use to survive in arid environments. Succulent xenophytes also show a greater abundance of free proline, up to 16 times greater than plants native to non-arid environments, as well as larger accumulations of Na+ rather than K+ for osmotic adjustment (Wang *et al.*, 2004), perhaps acting as an effective strategy for their adaptation to arid

environments. The tonoplast Na+/H+ antiporter (NHX) is involved in the compartmentalization of cytosolic Na+ into vacuoles. Zygophyllum xanthoxylum is a succulent xerophyte with a recently characterized ZxNHX antiporter demonstrated to be most active in the leaves. The transcript abundance of ZxNHX under salt stress up to 8.4 times greater than unstressed plants and up to 4.4 times greater under drought conditions than unstressed controls (Wu *et al.*, 2011), and may prove useful for future studies with crop species to predispose tolerance to both Na+ and drought.

Desiccation tolerance is an adaptation to extreme environmental conditions, perhaps leading to abundant expression of hydrophilic proteins as a survival mechanism, such as the "resurrection plant", *Craterostigma plantagineu* (Bartels 2005). C. plantagineum has limited genomic information available, yet it is becoming evident (and not surprisingly) that desiccation tolerance is a complex trait, and probably linked to the recognition of cis-regulatory sequences (Bartels 2005). However, it has been possible, based on homology inferences, to define four broad categories of *C. plantagineum* genes induced by dehydration: protective proteins including such as hydrophilins, regulatory proteins and RNA, carbohydrate metabolism enzymes, and proteins involved in water transport (Bartels 2005). Other insights have come from naturally occurring desiccation tolerant plants, recently a 31-kDa putative dehydrin polypeptide was discovered in the desiccation-tolerant fern Polypodium polypodioide, found to be localized at the cell walls and present only during drying (Layton *et al.* 2010). The protein rapidly dissipated upon tissue rehydration, along with changing the hydrophilicity of leaf surfaces and enabling reversible cell wall deformation. This suggests this protein potentially plays a role in avoiding mechanical failure during drought, and another angle to pursue with genetic engineering.

Transcriptome analysis of sub-lethal drought stress conditions in *Arabidopsis* identified three distinct stages of plant responses: initially an early "priming and preconditioning" stage, with early accumulation of ABA and associated signaling genes, which with a decrease in stomatal conductance, an intermediate stage preparatory for acclimation, and a late stage of new homeostasis with reduced growth. This is accompanied by a peak in expression of genes involved in cell wall expansion, likely as a preparatory step toward drought acclimation by the adjustment of the cell wall (Harb *et al.*, 2010).

A recent microarray investigating genes responsible for drought tolerance between genotypes of barley; two drought-insensitive and one drought-sensitive, identified 17 genes that may play a role in enhancing tolerance (Guo *et al.*, 2009). These genes are likely constitutively expressed in the two drought-insensitive genotypes, with their encoded proteins playing a role in their tolerance. These genes include those controlling stomatal closure via carbon metabolism (NADP malic enzyme, NADP-ME), those synthesizing the osmoprotectant glycine-betaine, those generating protectants against ROS scavenging, and those stabilizing membranes and proteins. Also found were genes enhancing Ca+ signaling and molecular chaperoning. These findings allow a basis for selecting single genes conferring drought tolerance on cereals by transgenic means, and engineering for drought avoidance has taken advantage of the mechanisms for stomatal closure. For example, expression of the NADP-ME gene from maize has resulted in altered stomatal behavior and water relations when introduced into tobacco. The majority of water lost from plants occurs through stomata. When stomata are open, ions accumulate in order to increase the turgor pressure of the guard cells, which results in increased pore size (Laporte *et al.*, 2002). Guard cells are present in pairs on the underside of leaves and surround the stomatal pores. These control both the CO2 influx required for photosynthesis and the

loss of water to the atmosphere due to transpiration. Drought induced ABA synthesis signals stomatal closing, thus reducing stomatal aperture and ultimately reducing water loss (Schroeder *et al.*, 2001). The transgenic NADP-ME tobacco displayed reduced stomatal conductance, yet importantly remained similar to the wild type in their growth and rate of development (Laporte *et al.*, 2002).

MAPK cascades are implicated various signaling pathways involved in plant development and stress responses. A tobacco MAPKKK (NPK1) was constitutively expressed in maize, improving drought tolerance by maintaining higher photosynthesis rates than WT without affecting yield, producing a kernel yield comparable to well-watered WT plants (Shou *et al.*, 2004). A novel MAPKKK gene, DSM1, has recently been characterized in rice, which functions as an early signaling component in regulating responses to drought stress by regulating scavenging of ROS (Ning *et al.*, 2010). Overexpression of DSM1 in rice increases resistance to dehydration stress at the seedling stage.

In a study exposing nodulated alfalfa plants to drought conditions, it was shown much of the loss of alfalfa performance was due to reduced photosynthesis (in the leaves) and nitrogenase activity (in the nodules). Proteomic profiling showed a marked increase in proline levels, likely as a result of the intracellular increase in ROS (Aranjuelo *et al.*, 2010). High levels of the osmoprotectant proline also resulted from the introduction of AtDREB1A/CBF3, driven by the inducible rd29A promoter, into tall fescue (*Festuca arundinacea*) (Zhao *et al.*, 2007), which displayed increased resistance to drought.

The NAC family of plant-specific transcription factors plays roles in plant organ development, division, and resistance to pathogen attack (Hu *et al.*,

2006) and all references therin). Some members of the NAC family are stress responsive, for example SNAC1 (STRESS-RESPONSIVE NAC 1) in induced in guard cells by drought stress, and when overexpressed in rice confers enhanced drought and salt resistance, without phenotypic changes or yield penalty under field conditions (Hu *et al.*, 2006). Another novel rice NAC gene, ONAC045, is induced by drought, salt, cold, and ABA treatment (Zheng *et al.*, 2009). ONAC045 was shown to function as a transcriptional activator, and rice plants overexpressing ONAC045 displayed enhanced tolerance to both drought and salt. Of the 140 OsNAC genes predicted in rice, 18 have been identified as being induced by stress conditions (Jeong *et al.*, 2010). Of these, a recent functional genomics approach identified a rice NAC-domain gene, OsNAC10, which when under the control root-specific promoter (rather that the constitutive promoter GOS2) RCc3 displayed improved the drought tolerance and yield of transgenic rice plants grown under field drought conditions (Jeong *et al.*, 2010). This represents another example of how spatial or temporal expressions of transgenes affect the growth habit and yield of the plant. Future research would benefit from utilizing promoters other that those offering constitutive expression when creating transgenics to both resist stress and maintain reproduction and yield.

As introduced in the salt section, promising results have been demonstrated by overexpressing vacuolar membrane H+ pumps (Gaxiola *et al.*, 2001, Park *et al.*, 2005), allowing for an increase in vacuolar solute content, allowing for enhanced osmotic adjustment capacity. Recently AtAVP1 was introduced into cotton (Pasapula *et al.*, 2010), with the transgenic phenotype displaying an increased vacuolar proton gradient, resulting in solute accumulation and water retention. The AVP1-expressing cotton plants also displayed a 20% increase in fiber yield when grown under field conditions. These results suggest a promising role for AVP1 in both drought and salt tolerance,

perhaps offering the ability to reclaim farmland in arid regions (Pasapula *et al.*, 2010).

The C4 grass foxtail millet (*Setaria italica*) not only harbors novel genes for increased salt tolerance (Puranik *et al.*, 2011), but is also resistant to dehydration stress. Comparative transcriptome analysis under early and late drought stress identified the major upregulated transcripts to be involved in metabolism, signaling, transcriptional regulation, and proteolysis (Lata *et al.*, 2010). Five cultivars of varying drought sensitivity were also screened for their dehydration tolerance, with differentially expressed transcripts identified between them. Selected examples of upregulated transcripts include: DREB2 (with a 5 fold increase after 6 hours. and an 11-fold increase after 24 hours), Ca+ dependent kinases (likely due to enhanced Ca+ signaling), a member of the aquaporin superfamily, and the osmoprotectant thionin. Also shown to be upregulated in the more drought tolerant cultivars was the U2-snRNP, one of the 5 small ribonucleoprotein particles that make up the spliceosome, the regulator of both constitutive and alternative splicing in eukaryotes. This suggests an altered network of alternative splicing and gene regulation in foxtail millet under drought stress, offering an avenue for future research to pursue the gene targets and transcripts that undergo alternative splicing.

Combinations of abiotic stressors: Profiles of heat and drought. The combined physiological and molecular effects of heat and drought stress are quite complex, and it remains extremely difficult, if not impossible, to deduce these effects from observing the responses from one stress alone. For example, high leaf temperatures are a result of the combined effect because plants lose the ability for transpirational cooling when water availability is limited. When faced with high temperatures, plants will open their stomata in an effort to cool, however when drought is also introduced plants reduce their

stomatal aperature in an effort to reduce water loss, which in turn increases temperatures within the leaf. This increase greatly perturbs cellular homeostasis and the activities of enzymes, membranes, and cellular homeostasis. A recent study in the perennial grass *Leymus chinensis* indicates high temperatures, combined with drought stress, reduces the function of PSII, weakens nitrogen anabolism, increases protein degradation, and provokes the peroxidation of lipids (Xu and Zhou, 2006).

The knowledge of the molecular effects of this combination in cereals remains rather limited, however a recent review (Barnabus *et al.*, 2008) offers insights into the current physiological knowledge, and readers are directed to this detailed overview for more information. From initial development, to fertilization, to the development of reproductive organs and successful seed set, high light and drought stress put severe pressures on cereals. Of agricultural importance is the combination of these stresses on grain filling, the final stage of growth in cereals where fertilized ovaries develop into caryopses. This process is dependent on the remobilization of carbon from vegetative tissues to developing grain. Water stress during the grain-filling period induces early senescence, reduced photosynthesis, and shortens the grain-filling period; it increases the remobilization of nonstructural carbohydrates from the vegetative tissues to the grain (for a review see Yang *et al.*, 2005). Under a combination of drought and heat stress, the amount of starch accumulation is greatly reduced, along with the activity of the enzymes responsible for starch synthesis, reducing grain weight. Around 65% of the dry weight of cereals can be accounted for by starch (Barnabus *et al.*, 2008). ADP-glucose pyrophosphorylase (AGPase) is considered the rate limiting step in starch synthesis, and differs in thermostability between plants, for example the AGPases of cereal endosperms are heat labile, while those in potato (Solanum tuberosum) tubers, are heat stable (Linebarger *et al.*, 2005). Recent research identified an

N-terminal motif unique to heat-stable AGPases, and when inserted into corn (*Zea mays*) was shown to increase heat stability more than 300-fold. This thermostability
stems from a cysteine residue within the motif, giving rise to small subunit homodimers not found in the wild-type maize enzyme (Linebarger *et al.*, 2005).

Trancriptome analysis is shedding light on the extent of the crosstalk network that exists between these abiotic stresses, which often result in the transcription of both overlapping and unique gene sets. This is not surprising given that a) several abiotic stresses share an osmotic (water loss) component and b) outside of a laboratory setting plants are exposed to combinations of stresses simultaneously, such as drought/heat or salt/drought, with the combined effects initiating unique transcriptome responses. *Arabidopsis* plants subjected to drought and heat stress display partial overlap of the two stress defense pathways individually, as well as 454 transcripts found to be specifically expressed during a combination of drought and heat stress (Rizhsky *et al.*, 2004). These transcripts are characterized by enhanced respiration, suppressed photosynthesis, a complex expression pattern of defense and metabolic transcripts, and the accumulation of sucrose and other sugars. Interestingly heat stress was found to ameliorate the toxicity of proline to cells, suggesting that during a combination of drought and heat stress sucrose replaces proline in plants as the major osmoprotectant (Rizhsky *et al.*, 2004). Microarray analysis of transgenic overexpressing DREB2A discovered upregulation of known drought and salt responsive genes but also heat shock related genes (Sakuma *et al.*, 2006), implicating its function in an ABA independent regulon (Nakashima and Yamaguchi-Shinozaki *et al.*, 2005).

Significant research is required to tease apart the molecular basis for this additive effect, yet given the current state of the field and the advent of high-throughput sequencing technologies combined with molecular cloning and characterization, the future remains bright for engineering plants with one or more additive genes conferring tolerance to heat and drought stress.

High light. Increases in light intensity over and above that which a plant can utilize in photosynthetic reactions is considered high light stress, and is extremely detrimental to the plant. This is in part due to accumulation of ROS, as well as the severely detrimental effects on photosynthesis and carbon fixation, all of which lead to cellular perturbations and ultimately result in crop loss or yield reduction. This high light induced photoinhibition causes reductions in the photosystem II (PSII) complex, and reduced photosynthetic CO2 fixation (Krause *et al.*, 2005). The light driven PSII is found in the thylakoid membrane of chloroplasts, and also cyanobacteria, with the D1 and D2 protein complexes at its core. These sub-units act as the reaction center, binding chlorophyll, phenophytin, and plastoquinine co-factors involved in transmembrane induced charge separation (Nixon *et al.*, 2010), with the redox state of the plastoquinine pool affecting signaling and chlorophyll fluorescence (Hohmann-Marriott *et al.*, 2010). High light also synthesizes chloroplast antioxidant enzymes, with plastoquinol shown to be the main lipid-soluble antioxidant synthesized in *Arabidopsis* during the acclimation process (Szymańska *et al.*, 2009).

Arabidopsis leaves respond to high light conditions by a gradual loss of chlorophyll; decreases to 79%, 78%, and 66% of the initial value after 24, 48, and 72 h of high light acclimation, respectively, have been observed (Zelisko *et al.*, 2005). The 2010 review by Nixon *et al.*, summarizes the past 30 years of research into the assembly and repair of PSII, and readers are directed to this manuscript in depth discussion and mechanisms. The D1 and D2 proteins

are subject to photodamage under high light. In mature chloroplasts, expression of the genes encoding D1 and D2 are transcriptionally upregulated in response to light, maintaining high rates of synthesis of the reaction centers, and therefore PSII activity under high intensity light conditions (Onda *et al.*, 2008 and references therein).

The light harvesting complex (LHC II) of PSII is located in the thylakoid membrane of the chloroplast, collecting energy from sunlight and transferring it to the PSII reaction centers (Zelisko *et al.*, 2005), and although the main function of LHC II is energy collection and transfer, it also is involved in the distribution of excitation energy between PS II and PS I. LHC II also plays a role in preventing damage to photosynthetic machinery when there is an excess of light, necessary for high light acclimatization. Recent work into elucidating the regulatory network in of proteases responsible discovered a chloroplast-targeted protease, AtFtsH6, identified as being responsible for the degradation of LHC II in *Arabidopsis*, with an ortholog in Populus trichocarpa. It is likely that FtsH6 is a general LHC II protease and that FtsH6-dependent LHC II proteolysis is a feature of all higher plants (Zelisko *et al.*, 2005), and may play a role in the high light acclimatization process.

Leaf anatomy changes during photosynthetic light acclimation, for example leaves under shade display a reduction in the mesophyll cell palisade layer, allowing a wider area for light harvesting tissues, while chloroplasts under sunlight display more active carbon fixation carriers (such as Rubisco) and reaction centers (Weston *et al.*, 2000), with lower amounts of thylakoids per chloroplast area. *Arabidopsis* leaves have been shown to develop elongated palisade mesophyll cells and increase leaf thickness under exposure to increased fluence rates (Weston *et al.*, 2000). In addition to their role as photosynthetic centers, the chloroplasts also produce fatty acids and amino

acids that act as secondary messengers and the building blocks of protein synthesis (López-Juez 2007).

Microarray studies have offered insights into global gene expression changes in response to high light stress in *Arabidopsis*. Enzymes of the phenylpropanoid pathway, specifically those involved in lignin and anthocyanin synthesis, were shown to accumulate under long exposures to high light (Kimura *et al.*, 2003), perhaps as acting cellular protection mechanisms. Upregulation of stress-specific sigma factors (the Sig family) are evident as well. AtSig5 (AiSigE) is upregulated in response to light stress, suggesting a regulatory role in chloroplast gene expression under high light, and has also been shown to be induced under blue-light (470 nm) illumination (Onda *et al.*, 2008). AtSIG5 likely protects plants from stresses by assisting and increasing repair of the PSII reaction center (Nagashima *et al.*, 2004), as well as play a crucial role in plant reproduction (Yao *et al.*, 2003). There appears to be at least some homology between dicot and monocot systems; recently the nuclear genes OsSIG5 and OsSIG6 were identified and demonstrated to encode chloroplast localized sigma factors in rice, as the first example of Sig5 in crop plants (Kubota *et al.*, 2007).

Interestingly, but not entirely surprising, was the fact the array demonstrated high light stress induces genes associated with other abiotic stresses such as LEA14, COR15a, KIN1, and RD29a, as well as fibrillins (suggesting a role for the lipid protein plastoglobulin in chloroplast protection) and lipid transfer proteins. This is not surprising given the amounts of crosstalk in the abiotic stresses, as these genes are responsible for the encoding of proteins involved in protection of chaperoning, membrane protection, and other cellular components. The transcription factor DREB2A again fell out as overlapping with drought and high light specifically, likely induced by increasing ROS levels in chloroplasts under high light conditions (Kimura *et*

al., 2003). One of the genes demonstrated to be upregulated more than 3-fold by Kimura *et al.*, was a member of the early light-inducible protein (ELIP) family, ELIP2. Photoinhibition by high light also induces ELIP transcription in thylakoid membranes, corresponding to the degree of photoinhibition (Adamska *et al.*, 1992), with chloroplast ELIP levels paralleling the decrease in the amount of D1 protein subunit of PSII. The expression pattern of ELIPs suggests a role in protection of the photosynthetic apparatus against photooxidative damage. Since *Arabidopsis* carries two ELIP genes (ELIP1/2), a double null mutant was created, of which the sensibility to photoinhibition and ability to recover from light stress was not different from WT (Rossini *et al.*, 2006), raising questions about the photoprotective function of these proteins. Constitutive expression of AtELIP2 in *Arabidopsis* leaves decreased chloroplast chlorophyll content and caused a decrease in all photosynthetic pigments, however did not alter the composition, organization, or functionality of the photosystems. This indicates ELIPs are likely not directly involved in the synthesis and assembly of specific photosynthetic complexes, but rather affect the biogenesis of all chlorophyll-binding complexes (Tzvetkova-Chevolleau *et al.*, 2007). Continued study will be necessary to fully elucidate the photoprotective role of the ELIP family, perhaps suggesting they may not be the best candidates for genetic engineering to increase high light tolerance.

There has been little research using genetic engineering to increase the photosynthetic capability of agriculturally important crop species, yet given the insights we have gained from global transcriptome studies, and traditional genetic approaches have characterized genes enhancing cellular protection, altering photosynthesis, and involved with various aspects of high light acclimatization. The coming years now have the benefit of a wealth of genomic information, and identification of factors participating in signaling between the nucleus and chloroplast; allowing for directed studies into

increased photochemical quenching, dissipation of excess light energy, and reduction of ROS.

Future of abiotic stress research: Incorporating the genomics revolution. The next decade of research into abiotic stress tolerance promises to be both an exiting and fruitful one. It has the advantage of an existing bank of knowledge in the form of public gene expression data from microarray and HTS experiments, new emerging monocot model systems closely related to the cereals, and the coupling of traditional breeding with genetic engineering. New insights into the gene regulatory networks regulating stress-relevant pathways are continuing to emerge, and natural variation between cultivars or accessions, when coupled with high-throughput sequencing and quantitative phenotyping for improved stress tolerance, can pinpoint candidate genes for future study.

Since the advent of genome-wide surveys of expression patterns and differential regulation under various conditions, huge datasets of stress-specific genes have begun to amass. These datasets are incorporated into public databases, and freely searchable by the research community. Now that we have begun to identify subsets of genes and gene families induced under stresses, it is time to utilize this knowledge towards high-throughput screens of transgenic plants expressing genes under stress induced or tissue specific promoters. It is through such large scale functional genomic approaches that genes or gene combinations will be identified that are capable of conferring tolerance to the abiotic stress of interest without detrimental effects to reproduction or yield. To date most such transgenic studies have relied most often on a candidate gene first identified in the dicot *Arabidopsis*, fused to a constitutive promoter, and grown under laboratory conditions for a short duration. While often resulting in a plant demonstrating improved stress tolerance, such studies are of limited value unless conducted in a crop species

and under realistic field conditions, and can be likened to chipping away at an iceberg. Given the genome scale datasets available now it is plausible to directly identify novel genes or groups of genes in a crop itself or a closely related model system suited for laboratory study. Two relevant and recently emerging models are the grasses *Brachypodium* and *Setaria*. *Brachypodium* is member of the Pooideae subfamily of grasses and a well suited model system due to its relatively small fully sequenced and annotated genome, growing mutant collection, transcriptome sequence data, and other genetic resources (The International *Brachypodium* Initiative *et al.*, 2010). *Setaria* is a C4 grass (Brutnell *et al.*, 2010), as are corn, sugarcane, and sorghum, and therefore lends itself as a model for these agriculturally important crops. C4 plants have the ability to withstand higher light intensities and temperatures than C3 plants (wheat, barley, etc.) and information derived from *Setaria* may allow for improved viability of other crop species in new geographic regions.

One crucial issue that has only been touch upon briefly in this chapter, due to space constraints, is traditional breeding for increased stress tolerance. Abiotic stress tolerance is a complex trait, and it remains difficult to breed for tolerance without effecting yield or viability. There are many previous reviews (Bruce *et al.*, 2001, Price *et al.*, 2002, Withcombe *et al.*, 2008, Ashraf *et al.*, 2010) focusing on QTL and breeding cereals for stress tolerance, as well as genetic engineering coupled with breeding; and readers are directed to these reviews for further information. Future efforts will likely combine breeding and genetic engineering to maximize the benefits to both tolerance and yield. For example, a QTL involved in stress tolerance may bring undesirable closely linked traits, which may in turn be compensated by with complementary transgenes. The technology for such approaches is available now, and the challenge will be translating the laboratory discoveries into field studies and vice versa.

Research into improving stress tolerance has historically focused primarily on transcription factors. Transcription factors are master regulators of the response network, directly controlling either a single gene of multiple gene products. In addition, post-transcriptional regulation is mediated by splicing factors, specifically, by members the SR family of splicing factors, that are themselves alternatively spliced under abiotic stresses (Palusa *et al.*, 2007, Filichkin *et al.*, 2010). This layer of regulation of gene expression likely alters the splicing of a host of downstream genes in response to abiotic stresses, including transcription factors, and may simultaneously target multiple response mechanisms. Future research towards understanding the regulatory web of transcription factors, splicing factors, and their targets will be necessary in order to elucidate the foundations of abiotic stress tolerance in plants.

The next decade of abiotic stress research in plants has the potential to take great strides towards fully understanding stress response gene networks and translating this combined knowledge into increased crop yields. The knowledge gained from high-throughput and genome-scale technologies, coupled with the work of breeders, may allow us to meet the world's ever increasing demand for food, despite our growing population.

Literature cited:

Abba, S., Ghignone, S., Bonfante, P. 2006. A dehydration-inducible gene in the truffle Tuber borchii identifies a novel group of dehydrins. BMC Genomics, Vol.7, No.39, (March 2006), ISSN 1471-2164

Abraham, E., Rigo, G., Szekely, G., Nagy, R., Koncz, C., Szabados, L. 2003. Light-dependent induction of proline biosynthesis by abscisic acid and

salt stress is inhibited by brassinosteroid in *Arabidopsis*. Plant Molecular Biology, Vol.51, No.3, (Febuary 2003), pp. 363-372, ISSN 1677-0420

Adamska, I. 1997. ELIPs - Light-induced stress proteins. Physiologia Plantarum , Vol.100, No.4, (August 1997), pp. 794-805, ISSN 0031-9317

Agarwal, M., Hao, Y., Kapoor, A., Dong, C.-H., Fujii, H., Zheng, X., *et al.* 2006. A R2R3 Type MYB Transcription Factor Is Involved in the Cold Regulation of CBF Genes and in Acquired Freezing Tolerance. Journal of Biological Chemistry ,Vol. 281, No.49, (October 2006), pp. 37636-37645, ISSN 0021-9258

Apse, M. P., Aharon, G. S., Snedden, W. A., Blumwald, E. 1999. Salt Tolerance Conferred by Overexpression of a Vacuolar Na+/H+ Antiport in *Arabidopsis*. Science,Vol.285, No.5431, (August 1999), pp. 1256-1258, ISSN 0028-0836

Aranjuelo, I., Molero, G., Erice, G., Avice, J. C., Noques, S. 2010. Plant physiology and proteomics reveals the leaf response to drought in alfalfa (*Medicago sativa* L.). Journal of Experimental Botany, Vol.62, No.1, (August 2010), pp.1-13, ISSN 0022-0957

Ashraf, M. 2010. Inducing drought tolerance in plants: Recent advances. Biotechnology Advances, No. 28,Vol.1, (January 2010), pp. 169-183, ISSN 0734-9750

Badawi, M., Reddy, Y. V., Agharbaoui, Z., Tominaga, Y., Danyluk, J., Sarhan, F., Houde, M. 2008. Structure and functional analysis of wheat

ICE (Inducer of CBF Expression) genes. Plant and Cell Physiology, Vol.48, No.9, (July 2008), pp. 1237-1249, ISSN 0032-0781

Baniwal, S., Bharti, K., Chan, K., Fauth, M., Ganguli, A., Kotak, S., Mishra, S., Nover, L., Port, M., Scharf, K.-D., Tripp, J., Weber, C., Zielinski, D., von Koskull-Doring, P. 2004. Heat stress response in plants, a complex game with chaperones and more than twenty heat stress transcription factors. Journal of Biosciences, Vol.29, No.4, (December 2004), pp. 471-487, ISSN 0250-5991

Banti, V., Mafessoni, F., Loreti, E., Alpi, A., Perata, P. 2010. The Heat-Inducible Transcription Factor HsfA2 Enhances Anoxia Tolerance in *Arabidopsis*. C, Vol.152, No.3, (March 2010), pp. 1471-1483, ISSN 0032-0889

Barnabás, B., Jäger, K., Fehér, A. 2008. The effect of drought and heat stress on reproductive processes in cereals. Plant cell environment, Vol.31, No.1, (January 2008), pp. 11-38, ISSN 0140-7791

Bartels, D. 2005. Desiccation Tolerance Studied in the Resurrection Plant Craterostigma plantagineum. Integrative and Comparative Biology, Vol.45, No.5, (December 2001), pp. 696-701, ISSN 1540-7063

Biamonti, G. and Caceres, J. F. 2009. Cellular stress and RNA splicing. Trends in Biochemical Sciences, Vol.34, No.3, (March 2009), pp.146-153, ISSN 0968-0004

Bohnert, H., J., Nelson, D. E. & Jensen, R. G. 1995 Adaptations to Environmental Stresses. The Plant Cell Online, Vol.7, No.7, (July 1005), pp. 1099-1111, ISSN 1040-4651

Bray, E. A. 2002. Abscisic acid regulation of gene expression during water-deficit stress in the era of the *Arabidopsis* genome. Plant, Cell Environment, Vol.25, No.2, (Febuary 2002), pp. 153-161, ISSN 0140-7791

Brinker, M., Brosch, M., Vinocur, B., Abo-Ogiala, A., Fayyaz, P., Janz, D. 2010. Linking the Salt Transcriptome with Physiological Responses of a Salt-Resistant Populus Species as a Strategy to Identify Genes Important for Stress Acclimation. Plant Physiology, Vol.154, No.4, (December 2010), pp. 1697-1709, ISSN 0032-0889

Bruce, W. B., Edmeades, G. O., Barker, T. C. 2002. Molecular and physiological approaches to maize improvement for drought tolerance. Journal of Experimental Botany, Vol.53, No.366, (Jamuary 2002), pp. 13-25, ISSN 0022-0957

Brutnell, T. P., Wang, L., Swartwood, K., Goldschmidt, A., Jackson, D., Zhu, X.-G., Kellogg, E. & Van Eck, J. 2010. *Setaria viridis*, A Model for C4 Photosynthesis.The Plant Cell Online, Vo.22, No.8, (August 2010), pp. 2537-2544, ISSN 1532-298X

Carden, D. E., Walker, D. J., Flowers, T. J., Miller, A. J. 2003. Single-Cell Measurements of the Contributions of Cytosolic Na+ and K+ to Salt Tolerance. Plant Physiology, Vol.131, No.2, (Febuary 2003), pp. 676-683, ISSN 0032-0889

Chaikam, V., Karlson, D. 2008. Functional characterization of two cold shock domain proteins from Oryza sativa. Plant, Cell Environment,Vol.31, No.7, (July 2008), pp. 995-1006, ISSN 0140-7791

Charng, Y., Liu, H., Liu, N., Chi, W., Wang, C., Chang, S. 2007. A Heat-Inducible Transcription Factor, HsfA2, Is Required for Extension of Acquired Thermotolerance in *Arabidopsis*. Plant Physiology,Vol.143, No.1, (January 2007), pp. 251-262, ISSN 0032-0889

Chaves, M. M., Flexas, J., Pinheiro, C. 2009. Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. Annals of Botany, Vol.103, No.4, (Febuary 2009), pp. 551-560, ISSN 0305-7364

Chen, H., Hwang, J. E., Lim, C. J., Kim, D. Y., Lee, S. Y., Lim, C. O. 2010. *Arabidopsis* DREB2C functions as a transcriptional activator of HsfA3 during the heat stress response. Biochemical and Biophysical Research Communications,Vol. 401, No.2, (October 2010), pp.238-244, ISSN 1090-2104

Chen, W., J., Zhu, T. 2004. Networks of transcription factors with roles in environmental stress response. Trends in Plant Science, Vol.9, no.12, (December 2004), pp. 591-596, ISSN 1360-1385

Chinnusamy, V., Zhu, J., Zhu, J.-K. 2007. Cold stress regulation of gene expression in plants. Trends in Plant Science, Vol. 12, No.10, (October 2007), pp. 444-451, ISSN 1360-1385

Cohen-Peer, R., Schuster, S., Meiri, D., Breiman, A., Avni, A. 2010. Sumoylation of *Arabidopsis* heat shock factor A2 (HsfA2) modifies its activity during acquired thermotholerance. Plant Molecular Biology, Vol.74, No.1-2, (September 2010), pp. 33-45, ISSN 0167-4412

Craterostigma plantagineum. Integrative and Comparative Biology, Vol.45, No.5, (November 2005), pp. 696-701, ISSN 1540-7063

Cuevas, J. C., Lopez-Cobollo, R., Alcazar, R., Zarza, X., Koncz, C., Altabella, T., Salinas, J., Tiburcio, A. F., Ferrando, A. 2008. Putrescine is involved in *Arabidopsis* freezing tolerance and cold acclimation by regulating ABA levels in response to low temperature. Plant Physiology, Vol.148, No.2, (October 2008), pp. 1094–1105, ISSN 0032-0889

Doherty, C. J., Van, H. A., Myers, S. J., Thomashow, M. F. 2009. Roles for *Arabidopsis* CAMTA Transcription Factors in Cold-Regulated Gene Expression and Freezing Tolerance. The Plant Cell, Vol.21, No.3, (March 2009), pp. 972-984, ISSN 1040-4651

Dong, C., Danyluk, J., Wilson, K. E., Pocock, T., Huner, N. P., Sarhan, F. 2002. Cold-Regulated Cereal Chloroplast Late Embryogenesis Abundant-Like Proteins. Molecular Characterization and Functional Analyses. Plant Physiology, Vol.129, No.3, (July 2002), pp.1368-1381, ISSN 0032-0889

Dong, C.-H., Agarwal, M., Zhang, Y., Xie, Q., Zhu, J.-K. 2006. The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1. Proceedings of the National Academy of Sciences, Vol.103, No.21, (May 2006), pp. 8281-8286, ISSN 0027-8424

Egawa, C., Kobayashi, F., Ishibashi, M., Nakamura, T., Nakamura, C., Takumi, S. 2006. Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions

in common wheat. Genes Genetic Systems, Vol.81, No.2, (April 2006), pp. 77-91. ISSN 1341-7568

Feder, M. E., Hofmann, G. E. 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annual Review of Physiology,Vol.61, No.1, (March 1999), pp. 243-282, ISSN 0066-4278

Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., Wong, W. K., Mockler, T. C. 2009. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Research, Vol.20, No.1, (January 2010), pp. 45-58, ISSN 1088-9051

Flowers, T., Garcia, A., Koyama, M. & Yeo, A. 1997. Breeding for salt tolerance in crop plants — the role of molecular biology Acta Physiologiae Plantarum, Vol.10, No.4, (1997), pp. 427-433, ISSN 1861-1664

Forment, J., Naranjo, M. Á., Roldán, M., Serrano, R., Vicente, O. 2002. Expression of *Arabidopsis* SR-like splicing proteins confers salt tolerance to yeast and transgenic plants. The Plant Journal, Vol.30, No.5, (June 2002), pp. 511-519, ISSN 0960-7412

Fourrier, N., Bédard, J., Lopez-Juez, E., Barbrook, A., Bowyer, J., Jarvis, P., Warren, G., Thorlby, G. 2008 A role for SENSITIVE TO FREEZING2 in protecting chloroplasts against freeze-induced damage in *Arabidopsis*. The Plant Journal, Vol.55, No.5, (September 2008), pp.734-745, ISSN 0960-7412

Fowler, S. G., Cook, D., Thomashow, M. F. 2005. Low Temperature Induction of *Arabidopsis* CBF1, 2, and 3 Is Gated by the Circadian Clock. Plant Physiology, Vol.137, No.3, (March 2005), pp. 961-968, ISSN 0032-0889

Franklin, K., Whitelam, G. 2007. Light-quality regulation of freezing tolerance in *Arabidopsis* thaliana. Nature Genetics, Vol.39, No.5, (November 2007), pp. 1410-1413, ISSN 1061-4036

Fukuda, A., Nakamura, A., Tagiri, A., Tanaka, H., Miyao, A., Hirochika, H., Tanaka, Y. (2004). Function, Intracellular Localization and the Importance in Salt Tolerance of a Vacuolar Na+/H+ Antiporter from Rice. Plant and Cell Physiology, Vol.45, No.2, (January 2004), pp. 146-159, ISSN 0032-0781

Fursova, O. V., Pogorelko, G. V., Tarasov, V. A. 2009. Identification of ICE2, a gene involved in cold acclimation which determines freezing tolerance in *Arabidopsis* thaliana. Gene, Vol.429, No.1-2, (January 2009), pp. 98-103, ISSN 0378-1119

Garciadeblás, B., Haro, R., Benito, B. 2007. Cloning of two SOS1 transporters from the seagrass *Cymodocea nodosa*, SOS1 transporters from Cymodocea and *Arabidopsis*mediate potassium uptake in bacteria. Plant Molecular Biology, Vol.63, No.4, (March 2007), pp. 479-490, ISSN 0167-4412

Gaxiola, R. A., Li, J., Undurraga, S., Dang, L. M., Allen, G. J., Alper, S. L., Fink, G. R. 2001. Drought- and salt-tolerant plants result from overexpression of the AVP1 H+-pump. Proceedings of the National

Academy of Sciences, Vol.98, No.20, (September 2001), pp. 11444-11449, ISSN 0027-8424

Ghoulam, C., Foursy, A., Fares, K. 2002. Effects of salt stress on growth, inorganic ions and proline accumulation in relation to osmotic adjustment in five sugar beet cultivars. Environmental and Experimental Botany,Vol.47, No.1, (January 2002), pp. 39-50, ISSN 0098-8472

Gilmour, S. J., Zarka, D. G., Stockinger, E. J., Salazar, M. P., Houghton, J. M., Thomashow, M. F. 1998. Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. The Plant Journal, Vol.16. No.4, (November 1998), pp. 433-442, ISSN 1365-313X

Gray, W. M., Ostin, A., Sandberg, G., Romano, C. P., Estelle, M. 1998. High temperature promotes auxin-mediated hypocotyl elongation in *Arabidopsis*. Proceedings of the National Academy of Sciences, Vol.95, No.12, (June 1998), pp. 7197-7202, ISSN 0027-8424

Guo, P., Baum, M., Grando, S., Ceccarelli, S., Bai, G., Li, R. 2009. Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. Journal of Experimental Botany, Vol.60, No.12, (June 2009) , pp. 3531-3544, ISSN 0022-0957

Guo, Y.-Q., Tian, Z.-Y., Qin, G.-Y., Yan, D.-L., Zhang, J., Zhou, W.-Z, Qin, P. 2009. Gene expression of halophyte *Kosteletzkya virginica* seedlings under salt stress at early stage. Genetica, Vol.137, No.2, (November 2009), pp. 189-199, ISSN 0016-6707

Haake, V., Cook, D., Riechmann, J., Pineda, O., Thomashow, M. F., Zhang, J. Z. 2002. Transcription Factor CBF4 Is a Regulator of Drought Adaptation in *Arabidopsis*. Plant Physiology, Vol.130, No.2, (October 2002), pp. 639-648, ISSN 0032-0889

Halfter, U., Ishitani, M., Zhu, J.-K. 2000. The *Arabidopsis* SOS2 protein kinase physically interacts with and is activated by the calcium-binding protein SOS3. Proceedings of the National Academy of Sciences, Vol.97, No.1, (March 2000), pp. 3735-3740, ISSN 0027-8424

Harb, A., Krishnan, A., Ambavaram, M. M., Pereira, A. 2010. Molecular and Physiological Analysis of Drought Stress in *Arabidopsis* Reveals Early Responses Leading to Acclimation in Plant Growth. Plant Physiology, Vol.154, No.3, (August 2010), pp.1254-1271, ISSN 0032-0889

Hare, P., Cress, W. & van Staden, J. 1999. Proline synthesis and degradation, a model system for elucidating stress-related signal transduction. Journal of Experimental Botany, Vol.50, No.333, (January 1999), pp.413-434, ISSN 0022-0957

Hinniger, C., Caillet, V., Michoux, F., BenAmor, M., Tanksley, S., Lin, C., Mccarthy, J. 2006. Isolation and Characterization of cDNA Encoding Three Dehydrins Expressed During *Coffea canephora* (Robusta) Grain Development. Annals of Botany, Vol.97, No.5, (May 2006), pp. 755-765, ISSN 0305-7364

Hohmann-Marriott, M. F., Takizawa, K., Eaton-Rye, J. J., Mets, L., Minagawa, J. 2010. The redox state of the plastoquinone pool directly modulates minimum chlorophyll fluorescence yield in *Chlamydomonas*

reinhardtii. FEBS Letters, Vol.584, No.5, (March 2010), pp. 1021-1026, ISSN 0014-5793

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q., Xiong, L. 2006. Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. Proceedings of the National Academy of Sciences, Vol.103, No.35, (June 2006), pp. 12987-12992, ISSN 0027-8424

Hu, H., You, J., Fang, Y., Zhu, X., Qi, Z., Xiong, L. 2008. Characterization of transcription factor gene SNAC2 conferring cold and salt tolerance in rice. Plant Molecular Biology,Vol.67, No.102, (May 2008), pp. 169-181, ISSN 0167-4412

Huang, J., Hirji, R., Adam, L., Rozwadowski, K. L., Hammerlindl, J. K., Keller, W. A., Selvaraj, G. 2000. Genetic Engineering of Glycinebetaine Production toward Enhancing Stress Tolerance in Plants: Metabolic Limitations. Plant Physiology, Vol.122, No.3, (March 2000), pp. 747-756, ISSN 0032-0889

Hundertmark, M., Hincha, D. 2008. LEA proteins and their encoding genes in *Arabidopsis* thaliana. BMC Genomics, Vol.9, No.1, (March 2008), pp. 118, ISSN 1471-2164

Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., Konagaya, A., Shinozaki, K. 2004. Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. Nucleic Acids Research,Vol. 32, No.17, (September 2004), pp. 5096-5103, ISSN 0305-1048

International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature, Vol.463, No.7282, (February 2010), pp. 763-768, ISSN 0028-0836

Ishitani, M., Liu, J., Halfter, U., Kim, C., Shi, W., Zhu, J. 2000. SOS3 function in plant salt tolerance requires N-myristoylation and calcium binding. Plant Cell, Vol.12, No.9, (September 2000), pp. 1667-78, ISSN 1040-4651

Ishitani, M., Xiong, L., Stevenson, B., Zhu, J. K. 1997. Genetic Analysis of Osmotic and Cold Stress Signal Transduction in *Arabidopsis*: Interactions and Convergence of Abscisic Acid-Dependent and Abscisic Acid-Independent Pathways. The Plant Cell Online , 9 (11), 1935-1949.

Jaspers, P., & Kangasjärvi, J. 2010 Reactive oxygen species in abiotic stress signaling. Physiologia Plantarum, Vol.138, No.4, (April 2010), pp. 405-413, ISSN 1399-3054

Jeong, J. S., Kim, Y. S., Baek, K. H., Jung, H., Ha, S.-H., Choi, Y. D., *et al.* 2010. Root-Specific Expression of OsNAC10 Improves Drought Tolerance and Grain Yield in Rice under Field Drought Conditions. Plant Physiology, Vol.153, No.1, (May 2010), pp. 185-197, ISSN 0032-0889

Johnson-Flanagan, A. M., Huiwen, Z., Thiagarajah, M. R., Saini, H. S. 1991. Role of Abscisic Acid in the Induction of Freezing Tolerance in Brassica napus Suspension-Cultured Cells. Plant Physiology, Vol.95, No.4, (April 1991), pp. 1044-1048, ISSN 0032-0889

Jones-Rhoades, M. W., Bartel, D. P., Bartel, B. 2006. MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology, Vol.57, (June 2006) pp. 19-53, ISSN 1543-5008

Kanaoka, M. M., Pillitteri, L. J., Fujii, H., Yoshida, Y., Bogenschutz, N. L., Takabayashi, J., Zhu, J.-K., Torii, K. U. 2008. SCREAM/ICE1 and SCREAM2 Specify Three Cell-State Transitional Steps Leading to *Arabidopsis* Stomatal Differentiation. The Plant Cell, Vol.20, No.7, (July 2008), pp.1775-1785, ISSN 1040-4651

Katiyar-Agarwal, S., Agarwal, M., Grover, A. 2003. Heat-tolerant basmati rice engineered by over-expression of hsp101. Plant Molecular Biology, Vol.51, No.5, (March 2003), pp. 677-686, ISSN 0167-4412

Katiyar-Agarwal, S., Zhu, J., Kim, K., Agarwal, M., Fu, X., Huang, A., Zhu, J. 2006. The plasma membrane Na+/H+ antiporter SOS1 interacts with RCD1 and functions in oxidative stress tolerance in *Arabidopsis*. Proceedings of the National Academy of Sciences, Vol. 103, No.49, (December 2006) pp. 18816-18821, ISSN 0027-8424

Kern, A., J., and Dyer, W. E. 2004. Glycine Betaine Biosynthesis Is Induced by Salt Stress but Repressed by Auxinic Herbicides in Kochia scoparia. Journal of Plant Growth Regulation, Vol.23, No.1, (March 2004), pp. 9-19, ISSN 0721-7595

Kidokoro, S., Maruyama, K., Nakashima, K., Imura, Y., Narusaka, Y., Shinwari, Z. K., Osakabe, Y., Fujita, Y., Mizoi, J., Shinozaki, K., Yamaguchi-Shinozaki, K. 2009. The Phytochrome-Interacting Factor PIF7 Negatively Regulates DREB1 Expression under Circadian Control

in *Arabidopsis*. Plant Physiology, Vol.151, No.4, (December 2009), pp.2046-2057, ISSN 0032-0889

Kim, D.-Y., Jin, J.-Y., Alejandro, S., Martinoia, E., Lee, Y. 2010. Overexpression of AtABCG36 improves drought and salt stress resistance in *Arabidopsis*. Physiologia Plantarum, Vol.139, No.2, (June 2010), pp. 170-180, ISSN 0031-9317

Kim, J. S., Park, S. J., Kwak, K. J., Kim, Y. O., Kim, J. Y., Song, J., Boseung Jang, B., Jung, C., Kang, H. 2006. Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in Escherichia coli. Nucleic Acids Research, Vol.35, No.2, (December 2006), pp. 506-516, ISSN 0305-1048

Kim, K., Portis, A. R. 2005. Temperature Dependence of Photosynthesis in *Arabidopsis* Plants with Modifications in Rubisco Activase and Membrane Fluidity. Plant and Cell Physiology, Vol.46, No.3, (Febuary 2005), pp. 522-530, ISSN 0032-0781

Kimura, M., Yamamoto, Y. Y., Seki, M., Sakurai, T., Sato, M., Abe, T., Yoshida, S., Manabe, K., Shinozaki, K., Matsui, M. 2003. Identification of *Arabidopsis* Genes Regulated by High Light–Stress Using cDNA Microarray. Photochemistry and Photobiology, Vol.77, No.2, (February 2003), pp.226-233, ISSN 0031-8655

Kishitani, S., Watanabe, K., Yasuada, S., Arakawa, K., Takabe, T. 1994. Accumulation of glycinebetaine during cold acclimation and freezing tolerance in leaves of winter and spring barley plants. Plant, Cell & Environment, Vol.17, No.1, (January 1994), pp. 89-95, ISSN 0140-7791

Knight, H., Zarka, D. G., Okamoto, H., Thomashow, M. F., Knight, M. R. 2004. Abscisic Acid Induces CBF Gene Transcription and Subsequent Induction of Cold-Regulated Genes via the CRT Promoter Element. Plant Physiology, Vol.135, No.3, (July 2004), pp. 1710-1717, ISSN 0032-0889

Knox, A., Dhillon, T., Cheng, H., Tondelli, A., Pecchioni, N., Stockinger, E. 2010. CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. Theoretical and Applied Genetics, Vol.121, No.1, (June 2010), pp. 21-35, ISSN 0040-5752

Komatsu, S., Yang, G., Khan, M., Onodera, H., Toki, S., Yamaguchi, M. 2007. Over-expression of calcium-dependent protein kinase 13 and calreticulin interacting protein 1 confers cold tolerance on rice plants. Molecular Genetics and Genomics, Vol.277, No.6, (June 2007), pp. 713-723, ISSN 1617-4615

Kotak, S., Vierling, E., Bäumlein, H., von Koskull-Döring, P. 2007. A Novel Transcriptional Cascade Regulating Expression of Heat Stress Proteins during Seed Development of *Arabidopsis*. The Plant Cell, Vol.19, No.1, (January 2007), pp. 182-195, ISSN 1040-4651

Kovacs, D., Agoston, B., Tompa, P. 2008.Disordered plant LEA proteins as molecular chaperones. Plant Signal Behavior, Vol.3, No.9, (September 2008), pp.710-713, ISSN 1559-2316

Krause, G. H., Gallé, A., Virgo, A., García, M., Bucic, P., Jahns, P., Winter, K. 2006. High-Light Stress does not Impair Biomass Accumulation of

Sun-Acclimated Tropical Tree Seedlings (*Calophyllum longifolium* Willd. and *Tectona grandis* L. f.). Plant Biology, Vol.8, No.1, (January 2006), pp. 31-41, ISSN 1435-8603

Kreps, J. A., Wu, Y., Chang, H.-S., Zhu, T., Wang, X. & Harper, J. F. 2002. Transcriptome Changes for *Arabidopsis* in Response to Salt, Osmotic, and Cold Stress. Plant Physiology, Vol.130, No.4, (December 2002), pp. 2129-2141, ISSN 1532-2548

Kubota, Y., Miyao, A., Hirochika, H., Tozawa, Y., Yasuda, H., Tsunoyama, Y., Niwa, Y., Imamura, S., Shirai, M., Asayama, M. 2007. Two Novel Nuclear Genes, OsSIG5 and OsSIG6, Encoding Potential Plastid Sigma Factors of RNA Polymerase in Rice: Tissue-Specific and Light-Responsive Gene Expression. Plant and Cell Physiology, Vol.48, No.1, (January 2007), pp. 186-192, ISSN 1471-9053

Kurek, I., Chang, T. K., Bertain, S. M., Madrigal, A., Liu, L., Lassner, M. W., Zhu, G. 2007. Enhanced Thermostability of *Arabidopsis* Rubisco Activase Improves Photosynthesis and Growth Rates under Moderate Heat Stress. The Plant Cell, Vol.19, No.10, (October 2007), pp. 3230-3241, ISSN 1040-4651

Laporte, M. M., Shen, B., Tarczynski, M. C. 2002. Engineering for drought avoidance: expression of maize NADP malic enzyme in tobacco results in altered stomatal function. Journal of Experimental Botany, Vol.53, No.369, (April 2002), pp. 699-705, ISSN 1460-2431

Lata, C., Sahu, P. P., Prasad, M. 2010. Comparative transcriptome analysis of differentially expressed genes in foxtail millet (*Setaria italica* L.) during dehydration stress. Biochemical and Biophysical Research

Communications, Vol.393, No.4, (March 2010), pp. 720-727, ISSN 1090-2104

Layton, B. E., Boyd, M. B., Tripepi, M. S., Bitonti, B. M., Dollahon, M. N. R., Balsamo, R. A. 2010. Dehydration-induced expression of a 31-kDa dehydrin in *Polypodium polypodioides* (Polypodiaceae) may enable large, reversible deformation of cell walls. American Journal of Botany, Vol.97, No.4, (March 2010), pp. 535-544, ISSN 0002-9122

Lee, B., Kapoor, A., Zhu, J., Zhu, J.-K. 2006. STABILIZED1, a Stress-Upregulated Nuclear Protein, Is Required for Pre-mRNA Splicing, mRNA Turnover, and Stress Tolerance in *Arabidopsis*. The Plant Cell, Vol.18, No.7, (July 2006), pp. 1736-1749, ISSN 1040-4651

Lee, B.-h., Lee, H., Xiong, L., Zhu, J.-K. 2002. A Mitochondrial Complex I Defect Impairs Cold-Regulated Nuclear Gene Expression. The Plant Cell, Vol.14, No.6, (June 2002), pp. 1235-1251. ISSN 1040-4651

Li, C., Chen, Q., Gao, X., Qi, B., Chen, N., Xu, S., Chen, J., Wang, X. 2005. AtHsfA2 modulates expression of stress responsive genes and enhances tolerance to heat and oxidative stress in *Arabidopsis*. Science in China Series C: Life Sciences, Vol.46, No.8, (December 2005), pp.540-550, ISSN 1006-9305

Li, M., Berendzen, K., Schöffl, F. 2010. Promoter specificity and interactions between early and late *Arabidopsis* heat shock factors. Plant Molecular Biology, Vol.73, No.4-5 (July 2010), pp. 559-567, ISSN 0167-4412

Li, P., Brutnell, T. P. 2011. *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. Journal of Experimental Botany, Epub ahead of print, (May 2011), ISSN 0022-0957

Li, T., Zhang, Y., Liu, H., Wu, Y., Li, W., Zhang, H. 2010. Stable expression of *Arabidopsis* vacuolar Na+/H+ antiporter gene AtNHX1 and salt tolerance in transgenic soybean for over six generations. Chinese Science Bulletin,Vol. 55, No.12, (April 2010), pp. 1127-1134, ISSN 1001-6538

Li, W., Wang, D., Jin, T., Chang, Q., Yin, D., Xu, S., Liu, B.,Liu, L. 2010. The Vacuolar Na+/H+ Antiporter Gene SsNHX1 from the Halophyte *Salsola soda* Confers Salt Tolerance in Transgenic Alfalfa (*Medicago sativa* L.). Plant Molecular Biology Reporter, Vol.29, No.2, (July 2010), pp. 278-290, ISSN 1572-9818

Linebarger, C. R. L., Boehlein, S. K., Sewell, A. K., Shaw, J., Hannah, L. C. 2005. Heat Stability of Maize Endosperm ADP-Glucose Pyrophosphorylase Is Enhanced by Insertion of a Cysteine in the N Terminus of the Small Subunit. Plant Physiology, Vol.139, No.4, (December 2005), pp. 625-1634, ISSN 0032-0889

Liu, A.-L., Zou, J., Zhang, X.-W., Zhou, X.-Y., Wang, W.-F., Xiong, X.-Y., Chen, L.-Y., Chen, X.-B. 2010. Expression Profiles of Class A Rice Heat Shock Transcription Factor Genes Under Abiotic Stresses. Journal of Plant Biology, Vol.53, No.2, (January 2010), pp. 142-149, ISSN 1226-9239

Liu, J., Zhu, J.-K. 1998. A Calcium Sensor Homolog Required for Plant Salt Tolerance. Science, Vol.280, No.5371, (June 1998), pp. 1943-1945, ISSN 0036-8075

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., Shinozaki, K. 1998. Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought-and Low-Temperature-Responsive Gene Expression, Respectively, in *Arabidopsis*. The Plant Cell, Vol. 10, No.8, (August 1998), pp. 1391-1406, ISSN 1040-4651

Liu, X., Yang, J., Li, B., Yang, X., Meng, Q. 2010 Antisense expression of tomato chloroplast omega-3 fatty acid desaturase gene (LeFAD7) enhances the tomato high-temperature tolerance through reductions of trienoic fatty acids and alterations of physiological parameters. Photosynthetica, Vol.48, No.1, (January 2010) pp. 59-66, ISSN 0300-3604

Liu, Y., Xiong, Y., Bassham, D. C. 2009. Autophagy is required for tolerance of drought and salt stress in plants. Autophagy,Vol. 5, No.7, (October 2009), pp. 954-963, ISSN 1554-8627

Llorente, F., Oliveros, J. C., Martínez-Zapater, J. M., Salinas, J. 2000. A freezing-sensitive mutant of *Arabidopsis* frs1 is a new aba3 allele. Planta, Vol.211, No.5, (October 2000), pp. 648-655, ISSN 0032-0935

Lopez-Juez, E. 2007. Plastid biogenesis, between light and shadows. Journal of Experimental Botany,Vol.58, No.1, (January 2007), pp. 11-26, ISSN 0022-0957

Ma, S., Gong, Q., Bohnert, H. J. 2006. Dissecting salt stress pathways. Journal of Experimental Botany, Vol.57, No.5, (March 2006), pp. 1097-1107, ISSN 0022-0957

Mahajan, S., Tuteja, N. 2005. Cold, salinity and drought stresses: An overview. Archives of Biochemistry and Biophysics, Vol.444, No.2, (December 2005), pp. 139-158, ISSN 0003-9861

Martanez, J.-P., Lutts, S., Schanck, A., Bajji, M., Kinet, J.-M. 2004. Is osmotic adjustment required for water stress resistance in the Mediterranean shrub *Atriplex halimus* L. Journal of Plant Physiology, Vol.161, No.9, (September 2004), pp. 1041-1051, ISSN 0032-0889

Martínez-Atienza, J., Jiang, X., Garciadeblas, B., Mendoza, I., Zhu, J.-K., Pardo, J., M., Quintero, F., J. 2007. Conservation of the Salt Overly Sensitive Pathway in Rice. Plant Physiology, Vol.143, No.2, (February 2007), pp. 1001-1012, ISSN 0032-0889

Matsui, A., Ishida, J., Morosawa, T., Mochizuki, Y., Kaminuma, E., Endo, T. A., Okamoto., M., Nambara., E., Nakajima., M., Kawashima., M., Satou., M., Kim., J., Kobayashi., N., Toyoda., T., Shinozaki., K., Seki., M. 2008. *Arabidopsis* Transcriptome Analysis under Drought, Cold, High-Salinity and ABA Treatment Conditions using a Tiling Array. Plant and Cell Physiology, Vol.49, No.8, (July 2008), pp. 1135-1149, ISSN 0032-0781

Maughan, P.,J., Turner, T.,B., Coleman, C.,E., Elzinga, D.,B., Jellen, E.,N., Morales JA, Udall JA, Fairbanks DJ, Bonifacio A. 2009. Characterization of Salt Overly Sensitive 1 (SOS1) gene homoeologs in

quinoa (*Chenopodium quinoa* Willd.). Genome, Vol.52, No.7, (July 2009), pp. 647-657, ISSN 0831-2796

Miller, G., Shulaev, V., Mittler, R. 2008. Reactive oxygen signaling and abiotic stress. Physiologia Plantarum, Vol.133, No.33, (July 2008), pp. 481-489, ISSN 0031-9317

Mishra, S. K., Tripp, J., Winkelhaus, S., Tschiersch, B., Theres, K., Nover, L., Scharf, K.-D. 2002. In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. Genes & Development, Vol.16, No.12, (May 2002), pp. 1555-1567, ISSN 0890-9369

Mittler, R., Kim, Y., Song, L., Coutu, J., Coutu, A., Ciftci-Yilmaz, S., Lee, H., Stevenson, B., Zhu, J. 2006. Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. FEBS Letters, Vol.580, No.28-29, (December 2006), pp. 6537-6542, ISSN 0014-5793

Miura, K., Jin, J. B., Lee, J., Yoo, C. Y., Stirm, V., Miura, T., *et al.* 2007. SIZ1-Mediated Sumoylation of ICE1 Controls CBF3/DREB1A Expression and Freezing Tolerance in *Arabidopsis*. The Plant Cell, Vol.19, No.4, (April 2007), pp. 1403-1414, ISSN 1040-4651

Mizuno, H., Kawahara, Y., Sakai, H., Kanamori, H., Wakimoto, H., Yamagata, H., Oono, Y., Wu, J., Ikawa, H., Itoh, T., Matsumoto, T. 2010. Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa* L.). BMC Genomics, Vol.11, No.1, (December 2010), pp. 683, ISSN 1471-2164

Moellering, E. R., Muthan, B., Benning, C. 2010. Freezing Tolerance in Plants Requires Lipid Remodeling at the Outer Chloroplast Membrane. Science, Vol.330, No.6001, (October 2010), pp. 226-228, ISSN 0036-8075

Mohanty, S. and Tripathy, B. 2010. Early and late plastid development in response to chill stress and heat stress in wheat seedlings. Protoplasma, Epub ahead of print, (November 2010), pp.1-12, ISSN 1615-6102

Monroy, A. F., Dhindsa, R. S. 1995. Low-Temperature Signal Transduction: Induction of Cold Acclimation-Specific Genes of Alfalfa by Calcium at 25C. The Plant Cell, Vol.7, No.3, (March 1995), pp. 321-331, ISSN 1040-4651

Munns, R. 2002. Comparative physiology of salt and water stress. Plant, Cell & Environment, Vol. 25, No.2, (February 2002), pp. 239-250, ISSN 0140-7791

Nagashima, A., Hanaoka, M., Shikanai, T., Fujiwara, M., Kanamaru, K., Takahashi, H., Tanaka, K. 2004. The Multiple-Stress Responsive Plastid Sigma Factor, SIG5, Directs Activation of the psbD Blue Light-Responsive Promoter (BLRP) in *Arabidopsis* thaliana. Plant and Cell Physiology, Vol.55, No.4, (April 2004), pp. 357-368, ISSN 0032-0781

Nakaminami, K., Karlson, D. T., Imai, R. 2006. Functional conservation of cold shock domains in bacteria and higher plants. Proceedings of the National Academy of Sciences, Vol.103, No.26, (June 2006), pp. 10122-10127, ISSN 0027-8424

Nakashima, K., and Yamaguchi-Shinozaki, K. 2005. Molecular Studies on Stress-Responsive Gene Expression in *Arabidopsis* and Improvement of Stress Tolerance in Crop Plants by Regulon Biotechnology. Physiologia Plantarum, Vol.126, No.1, (January 2006), pp. 62–71, ISSN 0021-3551

Nakayama, K., Okawa, K., Kakizaki, T., Inaba, T. 2008. Evaluation of the Protective Activities of a Late Embryogenesis Abundant (LEA) Related Protein, Cor15am, during Various Stresses in Vitro. Bioscience, Biotechnology, and Biochemistry, Vol.72, No.6, (June 2006), pp. 1642-1645, ISSN 0916-8451

Ning, J., Li, X., Hicks, L. M., Xiong, L. A. 2010. Raf-Like MAPKKK Gene DSM1 Mediates Drought Resistance through Reactive Oxygen Species Scavenging in Rice. Plant Physiology, Vol.152, No.2, (February 2010), pp. 876-890, ISSN 0032-0889

Nixon, P. J., Michoux, F., Yu, J., Boehm, M., Komenda, J. 2010. Recent advances in understanding the assembly and repair of photosystem II. Annals of Botany, Vol.106, No.1, (March 2010), pp. 1-16, ISSN 0305-7364

Nover, L., Bharti, K., Döring, P., Mishra, S., K., Ganguli, A., Scharf, K-D. 2001. *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones, Vol.6, No.3, (April 2001), pp.177.189, ISSN 1355-8145

Novillo, F., Alonso, J. M., Ecker, J. R., Salinas, J. 2004. CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in *Arabidopsis*. Proceedings of

the National Academy of Sciences of the United States of America , Vol.101, No.11, (March 2004), pp. 3985-3990, ISSN 0027-8424

Novitskaya, G., V., Suvorova, T. A., Trunova, T. I. 2000. Lipid Composition of Tomato Leaves as Related to Plant Cold Tolerance. Russian Journal of Plant Physiology, Vol. 47, No.6, (January 2000), pp. 728-733, ISSN 1021-4437

Oh, S.-J., Song, S. I., Kim, Y. S., Jang, H.-J., Kim, S. Y., Kim, M., *et al.* 2005. *Arabidopsis* CBF3/DREB1A and ABF3 in Transgenic Rice Increased Tolerance to Abiotic Stress without Stunting Growth. Plant Physiology, Vol.138, No.1, (May 2005), pp. 341-351, ISSN 0032-0889

Ohnishi, N. and Murata, N. 2006. Glycinebetaine Counteracts the Inhibitory Effects of Salt Stress on the Degradation and Synthesis of the D1 Protein during Photoinhibition in *Synechococcus*. Plant Physiology, Vol.141, No.2, (June 2006), pp. 758-765, ISSN 0032-0889

Olías, R., Eljakaoui, Z., Pardo, J. M., Belver, A. 2009. The Na+/H+ exchanger SOS1 controls extrusion and distribution of Na+ in tomato plants under salinity conditions. Plant Signal Behavior, Vol.4, No.10, (October 2009),pp.973–976,ISSN1559-2316

Onda, Y., Yagi, Y., Saito, Y., Takenaka, N., Toyoshima, Y. 2008. Light induction of *Arabidopsis* SIG1 and SIG5 transcripts in mature leaves: differential roles of cryptochrome 1 and cryptochrome 2 and dual function of SIG5 in the recognition of plastid promoters. The Plant Journal, Vol.55, No.6, (September 2008), pp. 968-978. ISSN 0960-7412

Ottow, E. A., Brinker, M., Teichmann, T., Fritz, E., Kaiser, W., Brosché, M., Kangasjärvi, J., Jiang, X., Polle, A. 2005. Populus euphratica Displays Apoplastic Sodium Accumulation, Osmotic Adjustment by Decreases in Calcium and Soluble Carbohydrates, and Develops Leaf Succulence under Salt Stress. Plant Physiology, Vol.139, No.4, (November 2005), pp. 1762-1772, ISSN 0032-0889

Palusa, S. G., Ali, G. S., Reddy, A. S. 2007. Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. The Plant Journal, Vol.49, No.6, (March 2007), pp. 1091-1107, ISSN 0960-7412

Park, S. J., Kwak, K. J., Oh, T. R., Kim, Y. O., Kang, H. 2009. Cold Shock Domain Proteins Affect Seed Germination and Growth of *Arabidopsis* thaliana Under Abiotic Stress Conditions. Plant and Cell Physiology, Vol.50, No.4, (April 2009), pp.869-878, ISSN 0032-0781

Park, S., Li, J., Pittman, J. K., Berkowitz, G. A., Yang, H., Undurraga, S., *et al.* 2005. Up-regulation of a H+-pyrophosphatase (H+-PPase) as a strategy to engineer drought-resistant crop plants. Proceedings of the National Academy of Sciences of the United States of America, Vol.102, No.52, (December 2005), pp. 18830-18835, ISSN 0027-8424

Pasapula, V., Shen, G., Kuppu, S., Paez-Valencia, J., Mendoza, M., Hou, P., Chen, J., Qiu, X., Zhu, L., Zhang, X., Auld, D., Blumwald, E., Zhang, H., Gaxiola, R., Payton, P. 2011. Expression of an *Arabidopsis* vacuolar H+-pyrophosphatase gene (AVP1) in cotton improves drought- and salt tolerance and increases fibre yield in the field conditions. Plant Biotechnology Journal, Vol.9, No.1, (January 2011), pp. 88-99, ISSN 1467-7644

Pennycooke, J., Cheng, H., Roberts, S., Yang, Q., Rhee, S., Stockinger, E. 2008. The low temperature-responsive, *Solanum* CBF1 genes maintain high identity in their upstream regions in a genomic environment undergoing gene duplications, deletions, and rearrangements. Plant Molecular Biology, Vol.67, No.5, (July 2008), pp. 483-497, ISSN 0167-4412

Price, A. H., Cairns, J. E., Horton, P., Jones, H. G., Griffiths, H. 2002. Linking drought-resistance mechanisms to drought avoidance in upland rice using a QTL approach: progress and new opportunities to integrate stomatal and mesophyll responses. Journal of Experimental Botany, Vol.53, No.371, (January 2002), pp. 989-1004, ISSN 0022-0957

Qin, F., Sakuma, Y., Li, J., Liu, Q., Li, Y.-Q., Shinozaki, K., Yamaguchi-Shinozaki, K. 2004. Cloning and Functional Analysis of a Novel DREB1/CBF Transcription Factor Involved in Cold-Responsive Gene Expression in *Zea mays* L. Plant and Cell Physiology, Vol.45, No.8, (May 2004), pp. 1042-1052, ISSN 0032-0781

Rabbani, M. A., Maruyama, K., Abe, H., Khan, M. A., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M., Shinozaki, K., Yamaguchi-Shinozaki., K. 2003. Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses. Plant Physiology, Vol.133, No.4, (December 2003), pp. 1755-1767, ISSN 0032-0781

Reddy, A. S. 2007. Alternative Splicing of Pre-Messenger RNAs in Plants in the Genomic Era. Annual Review of Plant Biology, Vol.58, (June 2007), pp. 267-294, ISSN 1543-5008 Rekarte-Cowie, I., Ebshish, O. S.,

Mohamed, K. S., Pearce, R. S. 2008. Sucrose helps regulate cold acclimation of *Arabidopsis* thaliana. Journal of Experimental Botany, Vol.59, No.15, (November 2008), pp. 4205-4217, ISSN 0022-0957

Rensink, W., Hart, A, Liu, J., Ouyang, S., Zismann, V., Buel,l CR. 2005. Analyzing the potato abiotic stress transcriptome using expressed sequence tags. Genome, Vol.48,, No.4, (August 2005), pp. 598-605, ISSN 0225-7149

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., Mittler, R. 2004. When Defense Pathways Collide. The Response of *Arabidopsis* to a Combination of Drought and Heat Stress. Plant Physiology, Vol.134, No.4, (April 2004), pp.1683-1696, ISSN 0032-0889

Rorat, T. 2006. Plant dehydrins — Tissue location, structure and function. Cellular & Molecular Biology Letters, Vol.11, No.4, (September 2006), pp. 536-556, ISSN 1425-8153

Rossini, S., Casazza, A. P., Engelmann, E. C., Havaux, M., Jennings, R. C., Soave, C. 2006. Suppression of Both ELIP1 and ELIP2 in *Arabidopsis* Does Not Affect Tolerance to Photoinhibition and Photooxidative Stress. Plant Physiology, Vol.141, No.4, (August 2006), pp. 1264-1273, ISSN 0032-0889

Saijo, Y., Hata, S., Kyozuka, J., Shimamoto, K., Izui, K. 2000. Over-expression of a single Ca2+-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. The Plant Journal, Vol.23, No.3, (August 2000), pp. 319-327, ISSN 0960-7412

Sakamoto, A., Murata, N. 2000. Genetic engineering of glycinebetaine synthesis in plants: current status and implications for enhancement of stress tolerance. Journal of Experimental Botany,Vol.51, No.342, (January 2000), pp. 81-88, ISSN 0022-0957

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., Yamaguchi-Shinozaki, K. 2006. Dual function of an *Arabidopsis* transcription factor DREB2A in water and heat-stress-responsive gene expression. Proceedings of the National Academy of Sciences, Vol.103, No.49, (December 2006), pp. 18822-18827, ISSN

Salvucci, M. E., Osteryoung, K. W., Crafts-Brandner, S. J., Vierling, E. 2001. Exceptional Sensitivity of Rubisco Activase to Thermal Denaturation *in vitro* and *in vivo*. Plant Physiology, Vol.127, No.3, (November 2001), pp. 1053-1064, ISSN 0032-0781

Sasaki, K., Saito, T., Lämsä, M., Oksman-Caldentey, K.-M., Suzuki, M., Ohyama, K., Muranaka, T., Ohara, K., Yazaki, K. 2007. Plants Utilize Isoprene Emission as a Thermotolerance Mechanism. Plant and Cell Physiology, Vol.48. No.9, (August 2007), pp. 1254-1262, ISSN 00320781

Schramm, F., Larkindale, J., Kiehlmann, E., Ganguli, A., Englich, G., Vierling, E., Von Koskull-Döring, P. 2008. A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of *Arabidopsis*. The Plant Journal, Vol.53, No.2, (January 2008), pp. 264-274, ISSN 0960-7412

Schroeder, J., Kwak, J., Allen, G. 2001. Guard cell abscisic acid signalling and engineering drought hardiness in plants. Nature, Vol.410, No.6826, (March 2001), pp. 327-30, ISSN 0028-0836

Serrano, R., Mulet, J. M., Rios, G., Marquez, J. A., F., I. i., Leube, M. P., Mendizabal, I., Pascual-Ahuir, A., Proft, M., Ros, R., Montesinos, C. 1999. A glimpse of the mechanisms of ion homeostasis during salt stress. Journal of Experimental Botany, Vol.50, No. Special Issue, (June 1999), pp. 1023-1036, ISSN 0022-0957

Shabala, S. and Cuin, T.A. 2008. Potassium transport and plant salt tolerance. Physiologia Plantarum, Vol.133,No.4,(August 2008), pp. 651-669, ISSN 0031-9317

Shi, H., Kim, Y., Guo, Y., Stevenson, B., Zhu, J.-K. 2002. The *Arabidopsis* SOS5 Locus Encodes a Putative Cell Surface Adhesion Protein and Is Required for Normal Cell Expansion. The Plant Cell Online, Vol.15, No.1, (January 2002), pp. 19-32, ISSN 1532-298X

Shibasaki, K., Uemura, M., Tsurumi, S. & Rahman, A. 2009. Auxin Response in *Arabidopsis* under Cold Stress, Underlying Molecular Mechanisms. The Plant Cell, Vol.21, No.12, (December 2009), pp. 3823–3838, ISSN 1040-4651

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, Vol.3, No.3, (June 2000), pp. 217-223, ISSN

Shinozaki, K., Yamaguchi-Shinozaki, K. 2007. Gene networks involved in drought stress response and tolerance. Journal of Experimental Botany, Vol.58, No.2, (January 2007), pp. 221-227, ISSN 0022-0957

Shou, H., Bordallo, P., Wang, K. 2004. Expression of the Nicotiana protein kinase (NPK1) enhanced drought tolerance in transgenic maize. Journal of Experimental Botany, Vol.75, No.5, (May 2005), pp. 1013-1019, ISSN 0022-0957

Skinner, J., Szucs, P., Zitzewitz, J. v., Marquez-Cedillo, L., Filichkin, T., Stockinger, E., Thomashow, M., F., Chen, T., H., H., Hates, P., M.. 2006. Mapping of barley homologs to genes that regulate low temperature tolerance in *Arabidopsis*. TAG Theoretical and Applied Genetics, Vol.112, No.5, (January 2006), pp.832-842, ISSN 0040-5752

Sohn, S., Back, K. 2007. Transgenic rice tolerant to high temperature with elevated contents of dienoic fatty acids. Biologia Plantarum, Vol.51, No.2, (June 2007), pp. 340-342, ISSN 0006-3134

Steponkus, P., L., Lynch, D., V. 1989. Freeze/thaw-induced destabilization of the plasma membrane and the effects of cold acclimation. Journal of Bioenergetics and Biomembranes, Vol.21, No.1, (February 1989), pp. 21-41, ISSN 0145-479X

Stockinger, E. J., Skinner, J. S., Gardner, K. G., Francia, E., Pecchioni, N. 2007. Expression levels of barley Cbf genes at the Frost resistance-H2 locus are dependent upon alleles at Fr-H1 and Fr-H2. The Plant Journal, Vol.51, No.12, (July 2007), pp.308-321, ISSN 0960-7412

Su, C.-F., Wang, Y.-C., Hsieh, T.-H., Lu, C.-A., Tseng, T.-H., Yu, S.-M. 2010. A Novel MYBS3-Dependent Pathway Confers Cold Tolerance in Rice. Plant Physiology, Vol.153, No.1, (May 2010), pp. 145-158, ISSN 0032-0889

Sung, S., Amasino, R. M. 2004. Vernalization and epigenetics: how plants remember winter. Current Opinion in Plant Biology, Vol.7, No.1, (Feb 2004), pp. 4-10, ISSN 1369-5266

Sunkar, R., Zhu, J.-K. 2004. Novel and Stress-Regulated MicroRNAs and Other Small RNAs from *Arabidopsis*. The Plant Cell Online, Vol.16, No.8, (August 2004), pp. 201-2019, ISSN 1040-4651

Sunkar, R., Chinnusamy, V., Zhu, J., Zhu, J.-K. 2007. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends in Plant Science, Vol.12, No.7, (July 2007), pp. 301-309, ISSN 1360-1385

Sutton, F., Ding, X., Kenefick, D. G. 1992. Group 3 LEA Gene HVA1 Regulation by Cold Acclimation and Deacclimation in Two Barley Cultivars with Varying Freeze Resistance. Plant Physiology, Vol.99, No.1, (May 1992), pp. 338-340, ISSN 0032-0889 Swindell, W., Huebner, M., Weber, A. (2007). Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC Genomics,Vol.8, No.125 (March 2007), ISSN 1471-2164

Szymanska, R. and Kruk, J. 2010. Plastoquinol is the Main Prenyllipid Synthesized During Acclimation to High Light Conditions in *Arabidopsis* and is Converted to Plastochromanol by Tocopherol

Cyclase. Plant and Cell Physiology, Vol.51, No.4, (February 2010), pp. 537-545, ISSN 1471-9053

Takuhara, Y., Kobayashi, M., Suzuki, S. 2011 Low-temperature-induced transcription factors in grapevine enhance cold tolerance in transgenic *Arabidopsis* plants. Journal of Plant Physiology, Vol.168, No.9, (June 2011), pp. 967 – 975, ISSN 0176-1617

Tamura, K., Yamada, T. 2007. A perennial ryegrass CBF gene cluster is located in a region predicted by conserved synteny between *Poaceae* species. TAG Theoretical and Applied Genetics, Vol.114, No.2, (January 2007), pp. 273-283, ISSN 0040-5752

Tanabe, N., Yoshimura, K., Kimura, A., Yabuta, Y., Shigeoka, S. 2007. Differential Expression of Alternatively Spliced mRNAs of *Arabidopsis* SR Protein Homologs, atSR30 and atSR45a, in Response to Environmental Stress. Plant and Cell Physiology, Vol.48, No.7, (July 2007), pp. 1036-1049, ISSN 0032-0781

Tang, L., Kwon, S.-Y., Kim, S.-H., Kim, J.-S., Choi, J., Cho, K., Chang K. Sung, K., C., Kwak, S., Le, H. 2006. Enhanced tolerance of transgenic potato plants expressing both superoxide dismutase and ascorbate peroxidase in chloroplasts against oxidative stress and high temperature. Plant Cell Reports, Vol.25, No. 12, (July 2006), pp. 1380-1386 ISSN 0721-7714

Tester, M. and Davenport, R. 2003. Na+ Tolerance and Na+ Transport in Higher Plants. Annals of Botany, Vol.91, No.5, (April 2003), pp. 503-527, ISSN 0305-7364

Thomashow, M. 1999. Plant cold acclimation, Freezing tolerance genes and regulatory mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology, Vol.50, (June 1999), pp. 571-599, ISSN 1040-2519

Thomashow, M. F. 1998. Role of Cold-Responsive Genes in Plant Freezing Tolerance. Plant Physiology, Vol.118, No.1, (September 1998), pp. 1-8, ISSN 0032-0889

Thomashow, M. F. 2001. So What's New in the Field of Plant Cold Acclimation? Lots! Plant Physiology, Vol.125, No.1, (January 2001), pp. 89-93, ISSN 0032-0889

Thomashow, M., F. 2010. Molecular basis of plant cold acclimation, insights gained from studying the CBF cold response pathway. Plant Physiology, Vol.154, No.2, (October 2010), pp. 571-577, ISSN 0032-0889

Thorlby, G., Fourrier, N., Warren, G. 2004. The SENSITIVE TO FREEZING2 Gene, Required for Freezing Tolerance in *Arabidopsis* thaliana, Encodes a beta-Glucosidase. The Plant Cell, Vol. 16, No.8, (August 2004), pp. 2192-2203, ISSN 1040-4651

Tian, Y., Zhang, H., Pan, X., Chen, X., Zhang, Z., Lu, X., Huang, R. 2010. Overexpression of ethylene response factor TERF2 confers cold tolerance in rice seedlings. Transgenic Research, Epub ahead of print, (December 2010), ISSN 0962-8819

Tzvetkova-Chevolleau, T., Franck, F., Alawady, A. E., Dall'Osto, L., Carrière, F., Bassi, R., Grimm, B., Nussaume, L., Havaux, M. 2007. The light stress-induced protein ELIP2 is a regulator of chlorophyll synthesis

in *Arabidopsis* thaliana. The Plant Journal , Vol. 50, No.5, (June 2007), pp. 795-809, ISSN 0960-7412

Uemura, M., Joseph, R., A., Steponkus, P.,l. 1995. Cold Acclimation of *Arabidopsis* thaliana (Effect on Plasma Membrane Lipid Composition and Freeze-Induced Lesions). Plant Physiology, Vol.109, No.1, (September 1995), pp. 15-30, ISSN 0032-0889

Umezawa, T., Fujita, M., Fujita, Y., Yamaguchi-Shinozaki, K., Shinozaki, K. 2006. Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. Current Opinion in Biotechnology, Vol.17, No.2, (April 2006), pp. 113-122, ISSN 0958-1669

Vera-Estrella, R., Barkla, B. J., García-Ramírez, L. and Pantoja, O. 2005. Salt Stress in *Thellungiella halophila* Activates Na+ Transport Mechanisms Required for Salinity Tolerance. Plant Physiology, Vol.139, No.3, (November 2005), pp. 1507-1517, ISSN 0032-0889

Villalobos, M. A., Bartels, D. & Iturriaga, G. 2004. Stress Tolerance and Glucose Insensitive Phenotypes in *Arabidopsis* Overexpressing the CpMYB10 Transcription Factor Gene. Plant Physiology,Vol.135, No.1, (May 2004), pp. 309-324 , ISSN 1532-2548

Vogel, J. T., Zarka, D. G., Van, H. A., Fowler, S. G., Thomashow, M. F. 2005. Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. The Plant Journal, Vol.41, No.2, (January 2005), pp. 195-211, ISSN 0960-7412

von Koskull-Döring, P., Scharf, K.-D., Nover, L. (2007). The diversity of plant heat stress transcription factors. Trends in Plant Science, Vol.12, No.10, (October 2007), pp. 452– 457, ISSN 1360-1385

Wahid, A., Gelani, S., Ashraf, M., Foolad, M. 2007. Heat tolerance in plants, An overview. Environmental and Experimental Botany, Vol.61, No.3, (December 2007), pp. 199 – 223, ISSN 0098-8472

Wang, C., Zhang, Q., Shou, H.-x. 2009. Identification and expression analysis of OsHsfs, in rice. Journal of Zhejiang University - Science B, Vol.10, No.4, (April 2009), pp. 291-300, ISSN 1673-1581

Wang, S., Wan, C., Wang, Y., Chen, H., Zhou, Z., Fu, H., Sosebee, R., E. 2004. The characteristics of Na+, K+ and free proline distribution in several drought-resistant plants of the Alxa Desert, China. Journal of Arid Environments, Vol.56 No.3, (Febuary 2004), pp. 525-539, ISSN 0140-1963

Wang, W., Vinocur, B., Altman, A. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta, Vol.218, No.1, (November 2003), pp. 1-14, ISSN 0032-0935

Wang, W., Vinocur, B., Shoseyov, O., Altman, A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in Plant Science,Vol. 9, No.5, (May 2004), pp. 244-252, ISSN 1360-1385

Wang, W., Vinocur, B., Shoseyov, O., Altman, A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response.

Trends in Plant Science, Vol.9, No.5, (May 2004), pp. 244 – 252, ISSN 1360-1385

Wang, X., Li, W., Li, M., Welti, R. 2006. Profiling lipid changes in plant response to low temperatures. Physiologia Plantarum, Vol.126, No.1, (January 2006), pp. 90-96, ISSN 0031-9317

Watanabe, S., Kojima, K., Ide, Y., Sasaki, S. 2000. Effects of saline and osmotic stress on proline and sugar accumulation in *Populus euphratica* in vitro. Plant Cell, Tissue and Organ Culture, Vol.63, No. 3, (November 2000), pp. 199-206, ISSN 0167-6857

Webb, M. S., Uemura, M., Steponkus, P. L. 1994. A Comparison of Freezing Injury in Oat and Rye: Two Cereals at the Extremes of Freezing Tolerance. Plant Physiology, Vol.104, No.2, (Febuary 1994), pp. 467-478. ISSN 0032-0889

Webb, M. S., Uemura, M., Steponkus, P. L. 1994. A Comparison of Freezing Injury in Oat and Rye, Two Cereals at the Extremes of Freezing Tolerance. Plant Physiology, Vol.104, No.2, (February 1994), pp. 467–478, ISSN 0032-0889

Weston, E., Thorogood, K., Vinti, G., López-Juez, E. 2000. Light quantity controls leaf-cell and chloroplast development in *Arabidopsis* thaliana, wild type and blue-light-perception mutants. Planta, Vol.211, No.6, (November 2000) pp. 807-815, ISSN 0032-0935

Williams, M. E., Torabinejad, J., Cohick, E., Parker, K., Drake, E. J., Thompson, J. E., Hortter, M., DeWald, D., B. 2005. Mutations in the *Arabidopsis* Phosphoinositide Phosphatase Gene SAC9 Lead to

Overaccumulation of PtdIns(4,5)P2 and Constitutive Expression of the Stress-Response Pathway. Plant Physiology, Vol.138, No.2, (June 2005), pp. 686-700, ISSN 0032-0889

Witcombe, J., Hollington, P., Howarth, C., Reader, S., Steele, K. 2008. Breeding for abiotic stresses for sustainable agriculture. Philosophical Transactions of the Royal Society B: Biological Sciences, Vol.363, No.1492, (Febuary 2008), pp. 703-716, ISSN 1471-2970

Woodward, A. J., Bennett, I. J. 2005. The effect of salt stress and abscisic acid on proline production, chlorophyll content and growth of in vitro propagated shoots of *Eucalyptus camaldulensis*. Plant Cell, Tissue and Organ Culture, Vol.82, No.2, (January 2005), pp. 189-200, ISSN 0167-6857

Wu, G.-Q., Xi, J.-J., Wang, Q., Bao, A.-K., Ma, Q., Zhang, J.-L., Wang, S-M. 2011. The ZxNHX gene encoding tonoplast Na+/H+ antiporter from the xerophyte *Zygophyllum xanthoxylum* plays important roles in response to salt and drought. Journal of Plant Physiology, Vol.168, No.8, (May 2011), pp. 758-767, ISSN 0032-0889

Wu, Y.-Y.,Chen, Q.-J., Chen, M., Chen, J., and Wang, X.-C. 2005. Salt-tolerant transgenic perennial ryegrass (*Lolium perenne* L.) obtained by Agrobacterium tumefaciens-mediated transformation of the vacuolar Na+/H+ antiporter gene. Plant Science, Vol.169, No.1, (July 2005), pp. 65-73, ISSN 0306-4484

Xie, C., G., Lin, H., Deng, H., W., Guo, Y. 2009. Roles of SCaBP8 in salt stress response. Plant Signal Behavior, Vol.4, No.10, (October 2009), pp. 956-958, ISSN1559-2316

Xiong, L. and Yang, Y. 2003. Disease Resistance and Abiotic Stress Tolerance in Rice Are Inversely Modulated by an Abscisic Acid Inducible Mitogen-Activated Protein Kinase. The Plant Cell, Vol.15, No.3, (March 2003), pp.45-759, ISSN 1040-4651

Xu, K., Hong, P., Luo, L., Xia, T. 2009. Overexpression of AtNHX1, a Vacuolar Na+/H+ Antiporter from *Arabidopsis* thalina, in Petunia hybrida Enhances Salt and Drought Tolerance. Journal of Plant Biology, Vol.52, No.5, (August 2009), pp. 453-461, ISSN 1226-9239

Xu, Y., Gianfagna, T., Huang, B. 2010. Proteomic changes associated with expression of a gene (ipt) controlling cytokinin synthesis for improving heat tolerance in a perennial grass species. Journal of Experimental Botany,Vol.61, No.12, (June 2010), ISSN 1460-2431

Xu, Z. and Zhou, G. 2006. Combined effects of water stress and high temperature on photosynthesis, nitrogen metabolism and lipid peroxidation of a perennial grass *Leymus chinensis*. Planta, Vol.224, No.5, (October 2006), pp. 1080-1090, ISSN 0032-0935

Xue, Z.-Y., Zhi, D.-Y., Xue, G.-P., Zhang, H., Zhao, Y.-X., Xia, G.-M. 2004. Enhanced salt tolerance of transgenic wheat (*Tritivum aestivum* L.) expressing a vacuolar Na+/H+ antiporter gene with improved grain yields in saline soils in the field and a reduced level of leaf Na+. Plant Science, Vol.167, No.4, (October 2004), pp. 849 – 859, ISSN 0306-4484

Yaeno, T., Matsuda, O., Iba, K. 2004. Role of chloroplast trienoic fatty acids in plant disease defense responses. The Plant Journal, Vol.40, No.6, (September 2004), pp. 931-941, ISSN 0960-7412

Yamada, K., Fukao, Y., Hayashi, M., Fukazawa, M., Suzuki, I., Nishimura, M. 2007. Cytosolic HSP90 Regulates the Heat Shock Response That Is Responsible for Heat Acclimation in *Arabidopsis* thaliana. Journal of Biological Chemistry, Vol.282, No.52, (December 2007), pp. 37794-37804, ISSN 0021-9258

Yang, J. and Zhang, J. 2006. Grain filling of cereals under soil drying. New Phytologist, Vol.169, No.2, (January 2006), pp. 223-236, ISSN 0028-646X

Yang, X., Wen, X., Gong, H., Lu, Q., Yang, Z., Tang, Y., *et al.* 2007. Genetic engineering of the biosynthesis of glycinebetaine enhances thermotolerance of photosystem II in tobacco plants. Planta, Vol.225, No.3 (September 2007), pp. 719-733, ISSN 0032-0935 Yao, J., Roy-Chowdhury, S., Allison, L. A. (2003). AtSig5 Is an Essential Nucleus-Encoded *Arabidopsis* s-Like Factor. Plant Physiology, Vol.132, No.2, (June2003), pp. 739-747, ISSN 0032-0889

Yao, Y., Ni, Z., Peng, H., Sun, F., Xin, M., Sunkar, R., *et al.* 2010. Non-coding small RNAs responsive to abiotic stress in wheat (*Triticum aestivum*). Functional &; Integrative Genomics,Vol.10, No.2, (May 2010), pp. 187-190, ISSN 1438-7948

Yeo, A. 1998. Predicting the interaction between the effects of salinity and climate change on crop plants. Scientia Horticulturae, Vol.78, No.1-4, (November 1998), pp. 159-174, ISSN 0304-4238

Yokotani, N., Ichikawa, T., Kondou, Y., Matsui, M., Hirochika, H., Iwabuchi, M., *et al.* 2008. Expression of rice heat stress transcription

factor OsHsfA2e enhances tolerance to environmental stresses in transgenic *Arabidopsis*. Planta, Vol.227, No.5, (April 2008), pp. 957-967, ISSN 0032-0935

Yordanov, I., Velikova, V., Tsonev, T. 2000. Plant responses to drought and stress tolerance. Photosynthetica. Vol.30, No.2, (July 2000), pp.187-206, ISSN 03003604

Yoshida, T., Sakuma, Y., Todaka, D., Maruyama, K., Qin, F., Mizoi, J., *et al.* 2008. Functional analysis of an *Arabidopsis* heat-shock transcription factor HsfA3 in the transcriptional cascade downstream of the DREB2A stress-regulatory system. Biochemical and Biophysical Research Communications, Vol.368, No.3, (April 2008), pp.515-521., ISSN 1090-2104

Yoshimura, K., Masuda, A., Kuwano, M., Yokota, A., Akashi, K. 2008. Programmed Proteome Response for Drought Avoidance/Tolerance in the Root of a C3 Xerophyte (Wild Watermelon) Under Water Deficits. Plant and Cell Physiology, Vol. 49. No.2, (January 2008), pp. 226-241, ISSN 0032-0781

Zelisko, A., García-Lorenzo, M., Jackowski, G., Jansson, S., Funk, C. 2005. AtFtsH6 is involved in the degradation of the light-harvesting complex II during high-light acclimation and senescence. Proceedings of the National Academy of Sciences of the United States of America, Vol.102, No.38, (September 2005), pp. 13699-13704, ISSN 0027-8424

Zhang, G.-H., Su, Q., An, L.-J., Wu, S. 2008. Characterization and expression of a vacuolar Na+/H+ antiporter gene from the monocot

halophyte *Aeluropus littoralis*. Plant Physiology and Biochemistry, Vol.46, No.2, (February 2008), pp. 117-126, ISSN 0981-9428

Zhang, H. & Blumwald, E. 2001. Transgenic salt-tolerant tomato plants accumulate salt in foliage but not in fruit. Nature Biotechnology, Vol.19, No.8, (August 2001), pp. 756-758, ISSN 1087-0156

Zhang, H., Irving, L. J., McGill, C., Matthew, C., Zhou, D., Kemp, P. 2010. The effects of salinity and osmotic stress on barley germination rate: sodium as an osmotic regulator. Annals of Botany,Vol.106, No.6, (December 2010), pp. 1027-1035, ISSN 03057364

Zhang, J., Jia, W., Yang, J., Ismail, A. M. 2006. Role of ABA in integrating plant responses to drought and salt stresses. Field Crops Research,Vol.97, No.1, (May 2006), pp. 111-119, ISSN 0378-4290

Zhang, J.-L., Flowers, T., Wang, S.-M. 2010. Mechanisms of sodium uptake by roots of higher plants. Plant and Soil. Vol.326, No.1-2, (July 2010), pp. 45-60, ISSN 1573-5036

Zhang, Z. and Huang, R. 2010. Enhanced tolerance to freezing in tobacco and tomato overexpressing transcription factor TERF2/LeERF is modulated by ethylene biosynthesis. Plant Molecular Biology, Vol.73, No.3, (June 2010), pp. 241-249, ISSN 0167-4412

Zhao, J., Ren, W., Zhi, D., Wang, L., Xia, G. 2007. *Arabidopsis* DREB1A/CBF3 bestowed transgenic tall fescue increased tolerance to drought stress. Plant Cell Reports, Vol.26, No.9, (September 2010), pp. 1521-1528, ISSN 0721-7714

Zheng, X., Chen, B., Lu, G., Han, B. 2009. Overexpression of a NAC transcription factor enhances rice drought and salt tolerance. Biochemical and Biophysical Research Communications, Vol.379, No.4, (Febuary 2009), pp. 985-989, ISSN 0006-291X

Zhou, S., Wei, S., Boone, B., Levy, S. 2007. Microarray analysis of genes affected by salt stress in tomato. African Journal of Environmental Science and Technology, Vol.1, No.2, (September 2007), pp.014-026, ISSN 1996-0786

Zhu, J.-K. 2002. Salt and Drought Stress Signal Transduction In Plants. Annual Review of Plant Biology, Vol.53, No.1, (June 2002), pp. 247-273, ISSN 1 543-5008

Chapter 5:

Discovery of SNP markers in expressed genes of hazelnut

Doug W. Bryant, Sam E. Fox, Erik R. Rowley, Henry D. Priest, Rongkun Shen,

Weng-Keen Wong, and Todd C. Mockler

Acta Horticulturae
ISHS Secretariat
PO Box 500
3001 Leuven 1, Belgium
859, 289-294 (2010)
DOI: 10.17660/ActaHortic.2010.859.33

Abstract:

Polymorphisms associated with gene coding regions are useful tools for use as molecular markers, for example in marker-assisted breeding efforts. In particular, single nucleotide polymorphisms (SNPs) have emerged as a preferred marker for high-throughput genotyping studies. The objective of this study was to develop a set of SNPs for European hazelnut (*Corylus avellana*) by exploiting the exceptional depth and breadth of transcriptome sequencing made possible by high-throughput RNA sequencing (RNA-seq) using the Illumina Genome Analyzer platform. Bioinformatics tools were used to mine SNPs from a database of EST sequences derived from *de novo* assembly of Illumina RNA-seq reads representing hazelnut transcripts from multiple accessions. This resulted in the identification of 5,398 SNPs. This set of SNP markers for hazelnut provides a new resource for genetic mapping of important agronomic traits and will be useful markers for aligning future physical and genetic maps.

Introduction:

Hazelnut, Corylus avellana, is a valuable Oregon (USA) crop and a full 99% of the hazelnut produced in the United States comes from the Willamette Valley, which accounts for 4% of the world crop. In addition to its relevance to Oregon agriculture, hazelnut possesses several characteristics that make it an ideal *Betulaceae* model including a relatively small stature, small genome size of ~350 Mb, relatively short life cycle of ~5 years to first flowering, available genetic linkage map, available BAC library, genetically diverse collection of over 700 Corylus accessions, and amenity to transformation. Prior to this study no large scale single nucleotide polymorphism (SNP) database has yet been made available for this important crop, limiting the use of genotyping with these preferred markers.

Single nucleotide polymorphisms (SNPs) are the most common DNA sequence variants in most organisms and when available they are often the marker system of choice (reviewed in Ganal *et al.*, 2009). SNPs, defined as genetic variation in a DNA sequence that occurs when a single nucleotide is altered, are found in most regions of a genome, including coding regions. Coding region SNPs can be readily mapped, rendering SNPs effective genetic markers (reviewed in Jones *et al.*, 2009). In order to use SNPs as genetic markers they first must be discovered, a task to which high-throughput RNA sequencing (RNA-seq) is aptly suited.

Current high-throughput sequencing technologies (reviewed in Shendure and Ji, 2008; Fox *et al.*, 2009) have made deep interrogation of expressed transcript sequences practical using the RNA-seq approach. While the current generation of high-throughput sequencing platforms represents a dramatic improvement over previous technologies, the massive datasets produced by these new technologies require processing with specialized computer algorithms.

Using data obtained from RNA-seq high-throughput sequencing experiments and

appropriate bioinformatics tools we provide a set of SNP markers for hazelnut as a new genetic mapping resource. This resource will be useful in aligning future physical and genetic maps and will assist in the genetic mapping of important agronomic traits and marker assisted breeding efforts.

Materials and methods:

Plant materials, growth conditions, and tissue collection.

Bark, leaves, whole seedlings and catkins were used as plant materials in this study. Bark and leaves were collected from field-grown examples of the Jefferson accession (OSU 703.007). Catkins were collected from a field-grown example of the Barcelona accession (PI 557037; CCOR 36.001). The whole seedlings, including roots, were progeny of a cross: OSU 954.076 (cutleaf) x OSU 976.091 (contorta).

EST Sources and Database Construction. Total RNA was isolated from hazelnut tissue samples. Full-length enriched (FL) and randomly primed (RP) cDNA libraries were prepared and sequenced using an Illumina GA2 Genome Analyzer essentially as described in Fox *et al.* (2009) and Filichkin *et al.* (2009). The FL cDNA libraries for bark, leaves and catkins were generated using the SMART method (Zhu *et al.*, 2001). The RP libraries for whole seedlings were generated by random hexamer primed reverse transcription of 2X oligo-dT isolated polyA+ mRNA. These cDNA libraries were then subjected to RNA-seq analysis. Illumina RNA-seq was performed by personnel in the Oregon State University Center for Genome Research and Biocomputing core facility.

Raw Illumina reads were obtained after base calling in the Solexa Pipeline. In order to reduce uninformative and spurious assemblies reads containing ambiguous base calls (Ns), homopolymers, low-quality sequence regions, matching primers, adapters, and a surrogate chloroplast genome reference (Morus indica; NC_008359) were removed prior to assembly. A total of ~71 million 36 nt

Illumina reads from the four cDNA libraries were pooled and assembled using Velvet v0.7.55 (Zerbino and Birney, 2008) with a hash length of 31 and a minimum reported contig length of 101 nt.

SNP Discovery. The Illumina RNA-seq reads used for the EST contig assembly described above were aligned back to the contigs generated by Velvet using BLAT (Kent, 2002) (options: -maxGap=0 -maxIntron=0 -minScore=18 -noHead -out=pslx). The resulting BLAT alignments were filtered using a Perl script to retain only those read matches to the Velvet contigs over the entire 36 nt length of the input reads, allowing for up to two mismatches. Matches corresponding to reads hitting more than one location or more that one Velvet assembled EST contig were discarded, leaving ~22 million reads for SNP analysis. The resulting filtered BLAT alignments were used as an input to RGA-SNP (http://rga.cgrb.oregonstate.edu/). RGA-SNP was run using the following options: -e 0.10 -v 20 -n 4 -p 0.50. The resulting RGA output was filtered using a Perl script implementing a set of ad hoc rules in order to further identify only bi-allelic SNP calls.

Results:

De novo *assembly of hazelnut EST contigs from Illumina RNA-seq data*.

Through Illumina sequencing of the total RNA as described above under "EST Sources and Database Construction" a total of 110,648 Velvet-assembled EST contigs were obtained (http://hazelnut-files.oregonstate.edu), representing a total estimated sequence length of 27.9 Mb of expressed sequences (**Table 5.1**).

Table 5.1. RNA-seq data and EST contigs used for SNP detection

	Number	Cumulative size (Mbp)
Starting 36 nt reads	118281029	4258.1
Reads used in assembly	71087739	2559.2
Velvet contigs	110648	26.9
Reads re-mapped to contigs	21946736	790.1
Contigs containing SNPs	4361	

Each different tissue and genotype was sequenced separately in order to retain information on the accessions and tissue of origin.From the stringent filtering process described above under "SNP Discovery" 5,398 SNP calls in 4,361 Velvet-assembled hazelnut EST contigs were made. **Table 5.2** summarizes overall statistics for the Velvet output contigs.

Table 5.2. Summary statistics for Velvet output contigs.

Number of contigs	110648	
Minimum length	101	
Maximum length	4353	
Mean length	239	
Median length	161	
Minimum depth of coverage	1.01	*
Maximum depth of coverage	1086.4	*
Mean depth of coverage	23.85	*
Median depth of coverage	6.43	*
Mean coverage over length	0.97	#
Median coverage over length	0.99	#

*Velvet k-mer coverage.
#RGA coverage based on remapped Illumina reads (BLAT; max 2 mismatches over entire length of 36 nt read).

Out of a total 110,648 contigs 291 output by Velvet, the mean contig length was 239, with a mean coverage over the length of 0.97. Overall mean coverage depth was about 24.

SNP discovery in the hazelnut dataset. The Velvet-assembled hazelnut EST contigs were next processed through a SNP discovery pipeline. The entire set of 110,648 contigs representing an estimated length of 27.9 Mbp was used for SNP detection, as described above. In order to minimize detection of sequencing errors, each SNP was required to be supported by at least four independent RNA-seq reads with both variants being detected at least twice. A representative randomly chosen example is shown in **Figure 5.1.**

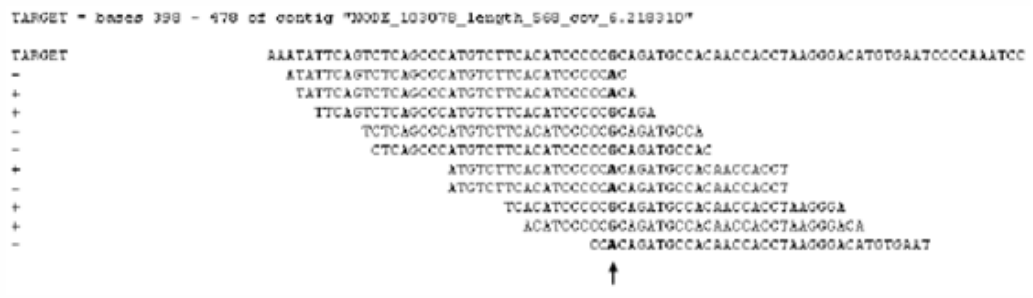


Figure 5.1. Alignment of RNA-seq reads to a portion of a Velvet-assembled hazelnut EST contig and inferred SNP

A total of 5,398 putative bi-allelic SNPs were detected in 4,361 contigs (Supplemental Table 1; http://hazelnut-files.oregonstate.edu), corresponding to an average occurrence of one SNP every 193 bp. A total of 3,194 transitions (1,591 A<>G and 1,603 C<>T) and 2,204 transversions (583 A<>C, 603A<>T, 478 C<>G, 540 G<>T) (Fig. 2) were detected, with A<>T being the most common (603; 27.3%) and C<>G the least common (478; 21.7%) (**Figure 5.2**)

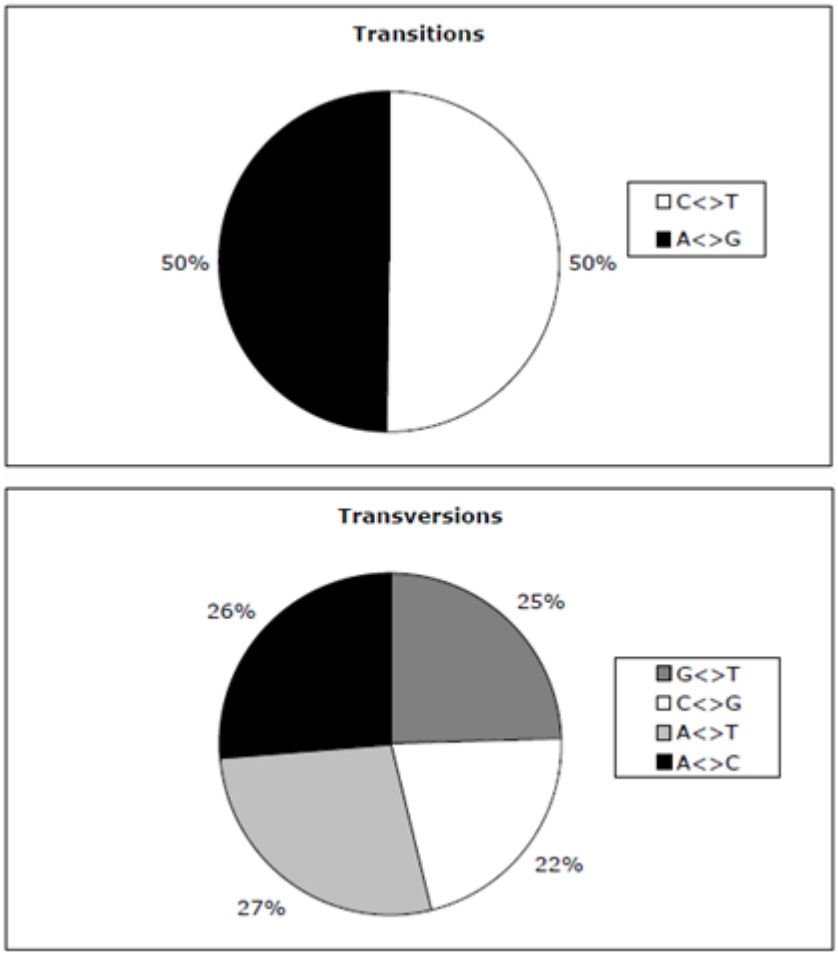


Figure 5.2. Classes of SNP discovered in hazelnut

Discussion:

High-throughput sequencing, specifically RNA-seq transcriptome sequencing experiments, allow for relatively simple, cost effective in silico SNP detection in coding regions. Previous technologies such as Sanger-sequencing platforms were limited to producing ESTs through the analysis of several hundred nucleotides at the terminal ends of cDNAs or cDNA fragments as plasmid inserts. With these older technologies it is impossible to sequence full-length cDNAs without arduous, low-throughput primer walking (Yamada *et al.*, 2003). By contrast, existing next-generation sequencing platforms allow direct shotgun sequencing of cDNAs to efficiently interrogate internal gene structures without the bias toward terminal ends, and at a much lower cost than traditional methods. This capability

is a tremendous benefit when working with horticultural crops and species without a genomic reference, lowering considerably the barrier for SNP discovery and use of SNPs as genetic markers. While exciting, there are several issues that must be kept in mind when using this technology for SNP discovery in horticultural crops.

During contig assembly, different paralogous genes may become incorrectly assembled into a single merged gene. In these types of cases, without post validation, it may be difficult or impossible to differentiate between actual SNPs and false positives arising from misassembly. Moreover, false positive SNP calls can arise from sequencing errors incorporated during the HTS experiment. Differential gene expression for different alleles of the same gene within or between accessions may lead to biased transcript contig assemblies that represent sequence data from one allele more than another, contributing to false negatives. Because the transcript data used in this study represented multiple genotypes another potential source of false-negatives in this study was our filtering for biallelic SNPs. Finally, to associate the detected SNPs with specific genes, more data or a reference genome must be available.

The results generated in this study represent an extremely high quality dataset in terms of accuracy, represented by mean coverage length, and depth of sequencing. However, a logical next step is the separate experimental validation of at least some of the putative SNPs that were detected in this study. Following such validation, genuine SNPs may be reliably used for mapping and marker assisted breeding. Further, the RNA-seq data will also be useful for the future empirical annotation of a hazelnut genome sequence.

Acknowledgements:
We thank Dr. Sergei Filichkin, Mark Dasenko, Chris Sullivan and Scott Givan for assistance with Illumina sequencing. We thank Rongkun Shen for coding the

RGA-SNP program. We also thank Dr. Shawn Mehlenbacher and Rob Hilles for kindly providing the hazelnut tissue samples that were used in this study. This work was supported by Oregon State University startup funds to TCM and partially supported by National Science Foundation Plant Genome Grant DBI 0605240 to TCM. HDP was supported by a Computational and Genome Biology Initiative Fellowship from Oregon State University. SEF and EER collected hazelnut tissues, prepared RNAs and cDNA libraries, and prepared samples for Illumina sequencing. RS coded the RGA-SNP program. DWB, HDP, and TCM conducted the bioinformatics analysis. WKW consulted on bioinformatics issues. DB and TCM wrote the paper.

Literature Cited:

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.-K. and Mockler, T.C. 2009. Genome wide mapping of alternative splicing in *Arabidopsis* thaliana. Genome Res. Published in Advance October 26, 2009, doi:10.1101/gr.093302.109.

Fox, S., Filichkin, S. and Mockler, T. 2009. Applications of ultra high throughput sequencing. Methods Mol. Biol. 553:79-108.

Ganal, M.W., Altmann, T. and Röder M.S. 2009. SNP identification in crop plants. Curr. Opin. Plant Biol. 12:211-7.

Jones, N., Ougham, H., Thomas H. and Pasakinskiene, I. 2009. Markers and mapping revisited: finding your gene. New Phytol. 183:935-66.

Kent, W.J. 2002. BLAT – The BLAST-Like Alignment Tool. Genome Res. 12:656–664.

Shendure, J. and Ji, H. 2008. Next-generation DNA sequencing. Nature

Biotechnol. 26:1135-1145.

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G., Liu, S.X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H.L., Tripp, M., Chang, C.H., Lee, J.M., Toriumi, M., Chan, M.M.H., Tang, C.C., Onodera, C.S., Deng, J.M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A.D., Gurjal, M., Hansen, N.F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V.W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P.X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M,, Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E.K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R. 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302:842-846.

Zerbino, D.R. and Birney, E. 2008. Velvet: algorithms for de novo short read assembly using *de Bruijn* graphs. Genome Res. 18:821-829.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. and Siebert, P.D. 2001. Reverse transcriptase template switching, a SMART approach for full-length cDNA library