Genomics and domestication of field pennycress (*Thlaspi arvense*)

A Dissertation

SUBMITTED TO THE FACULTY OF

UNIVERSITY OF MINNESOTA

BY

Kevin Michael Dorn

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisors

M. David Marks and Donald Wyse

May 2015

© Kevin M. Dorn, 2015

Acknowledgements

This body of work most certainly would not exist without the incredible, ongoing support of a host of others. In temporal order, I first have to thank my parents who have served as my first and loudest supporters and source of inspiration. Their continued message of 'find a career that you love' fortunately brought me to the University of Minnesota, and to the next in line for my sincerest thanks. I first met David Marks in 2004, when I was a freshman his freshman seminar class about 'Life on Mars'. Over 10 years later, I will have learned more about science (and myself) from him that I would have thought possible. Somewhere along that path, I transformed from a clueless pre-med undergrad into a plant scientist, and much of this can be attributed to David. His immense passion for 'the small stuff' in the life of a biologist is an attribute a younger version of myself found confusing (my favorite example - the excitement of looking at a gel), but I'm forever in gratitude to David for instilling this excitement in me. As my mentor and advisor, David's gentle, trusted guidance has shaped my skillsets into his original goals for me at the beginning of my graduate career – to be prepared for a scientific landscape that doesn't yet exist – to be adaptable, cross-disciplinary, and most valued by me, humble throughout it all. David – you have my deepest and sincerest thanks.

A special thanks is also in order for my co-advisor, Don Wyse, from whom I've learned how to frame the basic biological research described in this thesis within the context of global agricultural and environmental issues, as well as the necessary 'soft skills' to successfully raise interested and understanding of our research programs. The day he came over to first talk with David about joining forces to work on pennycress,

Don peered into my office, I’m sure thinking, “Where did David find THIS guy!” I’m forever indebted to both Don and David for their immense support and trust to embark on this line of work, and for propping me up to gain the confidence and leadership skills that will serve me well in my future.

The rest of my PhD advisory committee, Neil Olszewski, Fumi Katagiri, and Russ Gesch also deserve special acknowledgement. Despite my first few confusing and ever-changing first few years of graduate school, their continued support and guidance has helped shape the work described herein and in my development as a scientist. I also am indebted to Jay Hatch, who first took a chance on me as a young teaching assistant when I was an undergrad. From Jay I’ve learned am immense amount about who I am as a teacher, which has already impacted several hundred students, for better or worse. Also deserving of my thanks are my previous research advisors Cynthia Weinig (University of Wyoming), Rob McClung (Dartmouth College), and Keiji Nakajima (Nara Institute of Science and Technology, Japan).

Almost all of the work described in this thesis precipitated out of conversations with my late friend and colleague, Johnathon Fankhauser. Johnathon’s previous experience and excitement about sequencing technologies convinced me that this work was possible, and his assistance and enthusiasm made these projects possible. Our almost daily conversations in line at Subway about next steps, plans for analyses, and potential future projects helped me keep the faith throughout. His untimely death left a deep hole in the hearts of many, but his memory has served as motivation for me to carry on his infectious passion for science and learning in his absence.

The final and biggest thanks goes to my wife and best friend, Katie. She’s literally been by my side since I arrived in Minneapolis to begin college and has been, for over 10 years, my foundation. Without her support, my success in graduate school most certainly would not have happened. Her love and encouragement is what has, and what will continue to be, my motivation for continuing to pursue the passion of science cultivated by the many advisors, mentors, and friends mentioned above.

Dedication

To Katie, my love.

Thesis Abstract

Thlaspi arvense (field pennycress) is a cold tolerant oilseed species that is being domesticated as a new rapid cycling, winter annual cover crop and feedstock for biodiesel production. Pennycress is related to *Arabidopsis thaliana*, a model species that has provided an in-depth understanding of many basic developmental and physiological plant processes, which will provide vital information for the rapid domestication of a wild species into a new crop. By targeting key pennycress traits for improvement, such as reducing seed dormancy, increasing rates of spring flowering and maturity, increasing yield, and modifying seed oil composition, we are poised to develop a new winter cash crop that can fit within the corn/soybean rotation. To enable a mutation breeding approach that utilizes the massive amount of Arabidopsis-based knowledge, genomic resources are needed to identify target genes believed to influence key traits. In this dissertation, the first comprehensive annotated transcriptome assembly and comparative analyses are presented, along with the first draft genome sequence for pennycress. In these analyses, target assembled transcripts and corresponding DNA sequences are identified and compared to Arabidopsis homologs and enable the forward and reverse genetic screening of large-scale mutant populations. An analysis of winter and spring annual pennycress accessions is also presented, which identified several wild alleles of the pennycress FLOWERING LOCUS C homolog which was found to be responsible for differentiating between spring and winter annual phenotypes. The resources presented herein will provide an unprecedented set of tools to enable the rapid domestication of a new crop species.

Table of Contents

	Page
ACKNOWLEDGEMENTS	i
DEDICATION	iv
ABSTRACT	v
LIST OF TABLES	xi
LIST OF FIGURES	xii

CHAPTER 1: On the genomics-based domestication of field pennycress (*Thlaspi arvense* L.) as a new winter oilseed cover crop

Part I: Sustainable intensification and domestication of new crops as a framework meeting the grand challenges of agriculture and the environment	1
Part II: Pennycress – a new winter oilseed cover crop and biodiesel feedstock	5
Part III: Review of environmental and genetic factors affecting flowering time, a key target trait for pennycress domestication	14
Part IV: Conclusions	23
References	24

Chapter 2: De novo assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock

Summary 48

Introduction 49

Methods

- Plant growth conditions and RNA extraction 52
- High throughput RNA sequencing and de novo assembly 54

Results

- Generation of RNA-seq reads and de novo assembly 56
- Annotation and functional characterization of pennycress transcripts 58
- Comparative transcriptomics of pennycress versus other Brassicaceae species 60
- Identification of candidate pennycress genes controlling flowering time and glucocinolate levels 63

Discussion

- Comparative transcriptomics of pennycress and Arabidopsis 67
- Characterization of pennycress glucosinolate metabolism and translocation 68
- Genetics of flowering time in winter annual pennycress 70
- Considerations regarding de novo transcriptome assembly 72
- Future perspectives 74

References 91

CHAPTER 3: A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop

Summary 106

Introduction 107

Methods

Genomic DNA isolation and DNA sequencing 111

Genome assembly, scaffolding, and annotation 114

Comparative genomics 118

Read mapping and variant detection for cleaved amplified polymorphic sequence marker design 120

Comparative analysis of genes involved in glucosinolate metabolism and transport 121

Results and Discussion

Genome sequencing and assembly 122

Genome annotation with MAKER 124

Comparative genomics of the pennycress genome assembly 126

Evaluation of zygosity in the sequenced population 128

Analysis of genes involved in glucosinolate metabolism 131

Identification of predicted orthologs of Arabidopsis genes that confer desirable phenotypes when mutated 136

The future of genomics-based pennycress improvement 138

References 159

CHAPTER 4: Genomic analysis of spring and winter annual pennycress accessions reveals the spring annual growth habit has arisen multiple independent times

Summary 171

Introduction 172

Methods

Thlaspi arvense accession MN108-SA, MN111, and MN111 x MN108-SA F2 population 177

DNA isolation and Illumina genomic DNA sequencing 177

Sequencing data quality control, read mapping, and de novo assembly 178

Sequence comparison of flowering time gene orthologs 180

Sanger sequencing analysis of c.6_7insG FLC mutation in MN111 x MN108-SA F2 population and global spring varieties 180

PCR and sequencing of TaFLC Amplicon 2, Amplicon 3, and c.456bp_del amplicon 181

Results

Flowering time phenotypes of winter and spring annual pennycress 184

Whole Genome Resequencing of MN111 and MN108-SA Individuals 184

Examination of flowering time orthologs in winter and spring annual pennycress 186

Genetic analysis of co-segregation of c.6_7insG mutation with spring annual phenotype in an F2 population 188

Identification of a novel FLC allele from an EMS mutagenized population that confers the spring annual phenotype 190

Geographic distribution of FLC mutations in global spring annual pennycress accessions 192

Global spring annual varieties of unknown genetic origin 194

Discussion

Use of NGS to quickly identify mutations of interest 195

Towards an understanding of flowering time control in pennycress 195

Future directions for pennycress flowering time research 197

References 220

Comprehensive Bibliography 229

List of Tables

Table 2-1
Illumina RNAseq reads and de novo assembly statistics 75

Table 3-1
Pennycress Genome Sequencing Read Data Deposition 140

Table 3-2
Component list of computer used for pennycress genome assembly 141

Table 3-3
Component list of computer used for pennycress genome annotation 143

Table 3-4
Genome sequencing, assembly, and annotation statistics 147

Table 3-5
Complete genome assembly and scaffolding statistics 148

Table 3-6
Repeat Identification Statistics from RepeatMasker 149

Table 3-7
CAPS primer sequences and CAPS locations 150

Table 3-8
Putative orthologs controlling important agronomic traits in pennycress 151

Table 4-1
Summary of Whole Genome Sequencing Results and Variant Detection Analysis in MN108 and MN111 200

Table 4-2
Phenotypes of 50 F2 MN111 x MN108 progeny 202

Table 4-3
Oligonucleotide sequences used for amplification and sequencing of various TaFLC amplicons 203

Table 4-4
Analysis of FLC Mutations in MN108 x MN111 F2 population 207

Table 4-5
Sequence of oligonucleotides used for amplification and sequencing of TaFRI 217

List of Figures

Figure 2-1
Contig length distribution of assembled transcripts 76

Figure 2-2
Taxonomic distribution of top BLAST2GO hits 77

Figure 2-3
Representation of Brassica family phylogenetic tree 78

Figure 2-4
Comparative BLASTx analysis of pennycress to five Brassicaceae species 79

Figure 2-5
Comparative tBLASTn analysis of pennycress to five Brassicaceae species 80

Figure 2-6
Global comparison of the Arabidopsis and pennycress transcriptomes 81

Figure 2-7
Similarity of pennycress transcripts to Arabidopsis genes 83

Figure 2-8
Coverage of pennycress transcripts to Arabidopsis genes 85

Figure 2-9
Similarity and Coverage of pennycress transcripts versus Arabidopsis genes 87

Figure 2-10
Reconstruction of the flowering time pathway in pennycress 89

Figure 3-1
Distribution and accumulated length of genomic scaffolds 145

Figure 3-2
Comparative genomics of pennycress and other Brassicaceae species 153

Figure 3-3
CAPS Analysis of *Thlaspi arve*nse Line MN106 155

Figure 3-4
Analysis of genes involved in glucosinolate metabolism and transport 157

Figure 4-1
MN108 and MN111 accessions used in WGS and to develop F2 population found to segregate for the spring annual phenotype 199

Figure 4-2
Phylogenetic analysis of FLOWERING LOCUS C-like predicted peptides in pennycress and Arabidopsis 204

Figure 4-3
Overview of TaFLC amplicons and Variants 205

Figure 4-4
Identification of a novel, EMS-induced FLC allele conferring the spring annual phenotype. 209

Figure 4-5
Alignments of 5’ end of Arabidopsis and various pennycress FLC alleles and predicted peptides 211

Figure 4-6
Screening and identification of new allele (2385_2841del) conferring the spring annual phenotype 213

Figure 4-7
Geographic distribution of winter and spring annual lines analyzed in this study and confirmed FLC alleles. 215

Figure 4-8
Pennycress FRIGIDA gene structure and cloning strategy to identify c.2085A>T variant 218

Chapter 1: On the genomics-based domestication of field pennycress (*Thlaspi arvense* L.) as a new winter oilseed cover crop

Part I: Sustainable intensification and domestication of new crops as a framework meeting the grand challenges of agriculture and the environment

There is an increasing demand on agriculture to produce both sufficient foodstuffs for a growing population and provide novel sources of renewable biofuels and bioproducts (Garnett *et al.*, 2013, Jordan *et al.*, 2007). In light of a projected global population of 9 billion people by the year 2050, global food security will rely upon the increased productivity of agricultural systems (Godfray *et al.*, 2010). The goal of 'Sustainable Intensification' is to attain the increased agricultural production needs of the growing world population while also aiming to reduce the environmental impacts of agriculture, without impacting our future ability to produce both food and fuel (Tilman *et al.*, 2011). Temporal intensification is one of the proposed methods to attain sustainable intensification (Heaton *et al.*, 2013) through the increase of crop harvest frequency, or the number of harvests per unit area and time (Ray and Foley, 2013). One method to achieve temporal intensification of current major agricultural systems is the integration of fallow season crops into annual cropping systems or the replacement of annual crops with perennials. These new crops epitomize a potential solution for the call of sustainable intensification and meeting food and biofuel feedstock production needs. New fallow season crops hold great promise to increase the resilience of food and biofuel production systems in light of serious environmental issues facing agriculture like deteriorating water quality attributed to agriculture and climate change.

The focus of the research presented in this dissertation is on the development and integration of a new fallow-season crop, field pennycress, into the corn/soybean rotation of the Midwestern United States. Cover crops in major summer annual cropping rotations such as the corn/soybean system have been used to provide ecosystem services to these systems such as limiting soil erosion during the fallow-season, reducing nutrient loss to waterways, and reducing spring weed growth (Snapp *et al.*, 2005, Dabney *et al.*, 2001). However, with little economic benefit other than increasing long-term sustainability of the system, wide scale cover crop adoption has been limited (Singer *et al.*, 2007). The development of new ‘cash cover crops’ that can provide both the necessary ecosystem services that allow for their integration into the current systems (the ‘intensification’) to actually be sustainable, as well as a direct economic return for producers.

Unfortunately, there is a surprising lack of viable cash cover crop species that can immediately be deployed throughout the major agricultural acreage of the United States. The idea of domesticating wild plant species for novel uses by humans stretches throughout the past 10,000 years, accounting for the current forms of most major food crops used today (Doebley *et al.*, 2006). The process of crop domestication over this time period changed the genetic makeup of these species, even in the absence of any knowledge of DNA or genetics. With the modern genetic and genomic tools developed around these major domesticated crops, and the long history of basic research on model plant species, we are now poised to rapidly domesticate novel crop species to meet the demands of sustainable intensification (Varshney *et al.*, 2009). Specifically, the

domestication of fallow season crops will allow for the temporal intensification of the corn/soybean systems currently accounting for over 40 million acres in the United States alone.

Part II: Pennycress – a new winter oilseed cover crop and biodiesel feedstock

Biofuels and bioproducts derived from plants have great potential to reduce anthropogenic carbon emissions contributing to global climate change (Hill *et al.*, 2006). While there has been large investment in developing cellulosic biomass crops for bioethanol production over the past several decades, little progress has been made in bringing these solutions to market (Biello, 2011). As there exists an immediate need for these new crop options and bioproducts, relying on existing technologies and infrastructure systems to harvest and process these products is likely to allow for the most rapid adoption of these novel crops and cropping systems that allow for the sustainable intensification of agricultural systems.

Plant oils are one of the most energy-rich biomolecules and has widely been used by humans for both food, fuel, and industrial uses for centuries (Vollmann and Rajcan, 2009). Of note, seed oil from both soybean (*Glycine max*) and rapeseed (*Brassica napus*) have served as major sources oil for such purposes, and are extensively cultivated worldwide (USDA). As both soybean and oilseed rape are predominantly grown as summer annual crops, and are both produced food-quality oils and protein-rich seedmeal, displacing these crops for biofuel production is not ideal in light of the growing world population and growing demand for food production. One method to supplement the current major corn/soybean rotations is the use of rapid-cycling winter annual crops that

can be planted and harvested when the ground is normally left barren, which can account for up to 6 months of the year.

In this dissertation, the genomics-based domestication and utilization of *Thlaspi arvense* L. (common names: field pennycress, fanweed, stinkweed, frenchweed, mithridate mustard, wild garlic, bastard cress (Best and McIntyre, 1975) - referred to as pennycress herein) as a winter annual oilseed and cover crop is discussed. Pennycress is a member of the Brassicaceae family extended lineage 2 and relative of the model species *Thellungiella halophila, Arabidopsis thaliana*, *Capsella rubella* (both lineage 1), and, as well as the lineage 2 oilseed and vegetable crops *Brassica napus*, *Brassica rapa*, and *Brassica oleraceae* (Franzke *et al.*, 2011). Although native to Eurasia, the introduction and naturalization of pennycress throughout North America has occurred over the past ~150 years (Best and McIntyre, 1975), and is now found throughout the subarctic regions of the northern hemisphere, as well as Australia, New Zealand, and Argentina (Warwick, 2002). Although pennycress can exist as either a winter or spring annual (McIntyre and Best, 1978 , Best and McIntyre, 1976), winter annual pennycress lines already possess many key characteristics of an ideal winter grown oilseed crop. First, it is extremely cold hardy, shown to readily survive winters with temperatures as low as -30° C and has been used a model to study freezing tolerance (Best and McIntyre, 1975, Klebesadel, 1969, Sharma *et al.*, 2007, Zhou *et al.*, 2007). Without the consistent ability to overwinter in northern climates, the viability of wide scale successful adoption would low.

The trait best supporting the financial viability of pennycress adoption lies with it's naturally prolific seed production, with nearly 15,000 seeds per plant (Hume, 1990) and over 1,000 kg/ha of seed production (Best and McIntyre, 1975, Johnson *et al.*, 2015). Seed from pennycress plants have a high total oil content compared to other oilseed crops (20-36% by weight) (Moser *et al.*, 2009b, Moser *et al.*, 2009a), which can be extracted (Evangelista *et al.*, 2012) and utilized as a feedstock for biodiesel (Moser *et al.*, 2009a) or jet fuel (Boateng *et al.*, 2010). After pressing for oil extraction, the remaining seed meal/press cake is high in protein that could have high value uses, such as industrial lubricants, foaming agents, and emulsifiers (Selling *et al.*, 2013, Hojilla-Evangelista *et al.*, 2013). Seed meal has also been examined as a biofumigant that can be applied to fields to inhibit weed germination (Vaughn *et al.*, 2005, Vaughn *et al.*, 2004).

The integration and sustainability of winter annual pennycress into the target corn/soybean rotation has been recently evaluated (Phippen and Phippen, 2012, Fan *et al.*, 2013, Johnson *et al.*, 2015). Even in light of recent extreme variation of soybean and corn prices, major changes to cropping systems, including the integration of a new fallow-season species that could temporally displace the major summer annual crops for even a short time period, could hamper wide scale adoption. Additionally, any significant alteration of yield of seed quality on subsequent crops could equally prevent adoption. Phippen and Phippen (2012) demonstrated that the use of pennycress as a winter crop did not have a major impact on the quality of subsequent soybean crops over a two-year study, and in fact increased soybean dry weight yield compared to a fallow control plot.

Johnson et al. (2015) evaluated the effects of pennycress planting data and seeding rate on the soybean production, as well as examining the role of the winter pennycress crop in weed management. In these series of experiments, it was found that in all cases, the total oilseed yield (soybean plus pennycress) was higher compared to soybean grown without a winter pennycress crop. Additionally, a pennycress cover crop was shown to dramatically reduce total weed biomass, regardless of pennycress seeding rate or the use of a companion cover crop (either oat – *Avena sativa* or forage radish – *Raphanus sativus*), which the authors suggest could be attributed to allelopathic chemicals produced by pennycress (Johnson *et al.*, 2015), as pennycress seed meal has been previously found to inhibit the germination of other plant species (Vaughn *et al.*, 2009).

While pennycress naturally possessed numerous desirable traits of a winter oilseed cover crop, it still retains many wild traits that could hamper wide scale adoption. For example, current pennycress accessions exhibit widely variable rates of seed dormancy, a common trait among weedy species (Baskin and Baskin, 1989, Hazebroek and Metzger, 1990a, Hume, 1990, Hazebroek and Metzger, 1990b, Hume, 1994). As the success of a winter crop like pennycress can hinge upon strong stand establishment in the fall, low rates of fall germination impact both pennycress seed yield and the potential ecosystem services a fall-established crop could provide. Of the winter annual varieties in our collection, spring flowering time and maturity rates are also variable. As the goal remains to integrate pennycress into the corn/soybean rotation without displacing the subsequent soybean planting, early flowering and maturity of the pennycress stands is

essential, especially in northern climates with short and highly variable spring growing seasons. With end-use of pennycress products being a key financial driver behind adoption, optimizing target traits like oil composition for either biodiesel or jet fuel production, reducing glucosinolate levels in seeds to allow for the addition to animal feed, or modifying seed meal protein composition for industrial uses are also focal areas for improvement.

The availability of genomic resources opens vast possibilities for accelerated improvement of crop species while providing a foundation for functional and comparative genomics (Varshney *et al.*, 2009). The recent developments in DNA sequencing technologies has revolutionized biology, mainly the wide scale adoption of 2nd- generation technologies like Illumina sequencing (commonly referred to as next generation sequencing, or NGS), which is characterized by the massively parallel generation of hundreds of millions of short sequencing reads (Delseny *et al.*, 2010, Koboldt *et al.*, 2013, Feuillet *et al.*, 2011, Caccamo and Grotewold, 2013). In stark contrast with first generation technologies, which generally relied on single reactions in single tubes, required massive investments of both money and manpower to generate even a portion of the data now commonplace in just a single run of Illumina sequencing. First generation genome projects mainly relied on Sanger sequencing of large insert artificial clones, such as Bacterial Artificial Chromosomes (BACs), including the human genome (Lander *et al.*, 2001, Venter *et al.*, 2001). The human genome project, for example, required hundreds of scientists working in parallel over a decade to produce

these first draft genome sequences. The public human genome project (Lander et al., 2001) cost $2.7 billion (Fiscal Year 2001 dollars) in United States federal funding alone (www.genome.gov), whereas the cost of re-sequencing a single human genome today is nearing the $1,000 mark with the adoption of the latest Illumina HiSeq X Ten system (Hayden, 2014). Similarly, the first flowering plant species to have its genome sequenced was *Arabidopsis thaliana* (Kaul *et al.*, 2000), which likely cost tens to hundreds of millions of dollars throughout the 1990s, re-sequencing the Arabidopsis genome at over 200X coverage using the latest technology would cost under $3,000 today.

The development of genomic resources to support crop improvement has previously been limited to major, staple crop species such as corn (Schnable *et al.*, 2009), soybean (Schmutz *et al.*, 2010), and wheat (Mayer *et al.*, 2014), and while this limitation has predominantly been due to funding restraints holding back the development of these resources for minor crop species, the development and wide scale adoption of NGS has allowed for the characterization and comparison of a growing list of new plant species (Wang *et al.*, 2012, Varshney *et al.*, 2013, Wang *et al.*, 2014, Kagale *et al.*, 2014, Liu *et al.*, 2014, Slotte *et al.*, 2013, Wu *et al.*, 2012, Varshney *et al.*, 2012). Along with the sheer technical and financial ability to sequence the genome of a new crop species comes the critical mass of basic plant biology research that can be harnessed to both understand the biology of a new species and use that knowledge to make rapid trait improvements. Perhaps one of the most attractive traits lies within its evolutionary history and relation to the model plant species, Arabidopsis. Fortunate for the genomics-based improvement of pennycress, many of the genetic mechanisms controlling key agronomic traits have been

elucidated in Arabidopsis, and often in Brassica species relatives as well, such as flowering time and seed oil biosynthesis. With the development of equivalent genomic resources for pennycress, the application of knowledge gleaned from the billions of public dollars invested in Arabidopsis research is now possible.

Traditional breeding methods are currently being used to improve pennycress, however, these programs are still quite young (Sedbrook *et al.*, 2014). However, these programs rely on sufficient natural variation for key traits, and its currently unclear whether these programs will be capable of making the necessary improvements within a reasonable time period. As there remains an immediate need for new crop options that provide both economic value to producers as well as ecosystem services, and the rapid development of such species will enable society to immediately reap the potential benefits. Two additional strategies are being used in parallel with the traditional pennycress breeding programs: mutagenesis based breeding and genome editing. Both of these additional approaches rely heavily on the genomic resources presented in Chapters 2 and 3 of this dissertation.

The foundation of the mutation breeding approach is several large-scale mutagenized pennycress populations. Currently, ethyl methanesulfonate (EMS), fast neutron (FN), and gamma ray mutagenized populations have been advanced to the M2 generation within the University of Minnesota program. Within these populations, forward genetic screens have already allowed for the identification of novel phenotypic

variation. A mutation breeding approach in a species closely related to a model species as widely characterized as Arabidopsis allows for the prediction of homologous phenotypes of Arabidopsis mutants that could be achieved using mutagenesis. For example, mutations in key Arabidopsis regulatory genes controlling flowering time, maturity, oil synthesis, and seed dormancy can confer what could be considered desirable phenotypes. Forward genetic screens have already been used in these populations to identify underlying mutated genes conferring phenotypes similar to those found in homologues Arabidopsis mutants (unpublished data).

The inherent power of a reverse genetic screen of a large-scale mutant population is the ability to locate similar subtle phenotypes, such as altered chemical composition (seed oil composition, glucosinolate content), which are otherwise difficult to phenotype. Targeting Induced Local Lesions In Genomes (TILLING) is a common approach that has been used to identify mutations in specific genes, and can be used in the case of pennycress, in light of a now sequenced genome and identification of genes of interest, as is presented in Chapters 2 and 3 of this dissertation. A massively parallel version of TILLING called TILLING-by-sequencing has also been utilized to identify target mutations by using multidimensional pooling of PCR products from mutant individuals and using NGS sequencing to identify single individuals within a large population (Tsai *et al.*, 2011). Deletion TILLING (or DeTILLING) can also be used to identify large deletions in target regions caused by FN mutagenesis (Rogers *et al.*, 2009).

The use of genome editing technologies in pennycress is also on the horizon, as pennycress has been shown to be easily transformed using a modified *Agrobacterium* mediated floral dip method (Sedbrook *et al.*, 2014). A host of potential methods exist, including Transcription Activator-Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 (Cermak *et al.*, 2011, Miao *et al.*, 2013). In both of these methods, precise mutations in target DNA sequences can be induced. While the use of these technologies in an emerging crop like pennycress remain experimental, and at least in the near future, will only serve as a resource for basic research on pennycress biology, as public perception and regulation of genome editing technologies could prevent the release of a pennycress variety in which such technologies were used.

Part III: Review of environmental and genetic factors affecting flowering time, a key target trait for pennycress domestication

As previously mentioned, optimization of spring flowering time and maturity is essential for the domestication of pennycress. Being able to harvest a winter pennycress cover as early as possible will hasten adoption, especially in northern climates, where there can be significant yield tradeoffs of planting later maturity data soybean. As such, a major focus of the pennycress improvement program is developing winter annual varieties that flower and mature several weeks earlier than wild accessions. A series of studies by Best and McIntyre throughout the 1970s presented the first information regarding environmental and genetic factors affecting flowering in pennycress (Best and McIntyre, 1972, McIntyre and Best, 1975, Best and Mc Intyre, 1976, McIntyre and Best, 1978). The first of these studies revealed the existence of genetically distinct early flowering and late flowering types, and that flowering in pennycress is dramatically reduced by a lengthened photoperiod, indicating that pennycresss is a quantitative long day species (Best and McIntyre, 1972). In later field and growth chamber studies, the same authors investigated the effects of planting date on flowering time. Both early and late flowering types (Best and McIntyre, 1972) were transplanted into the field over a three month period (April – June of 1972 in Saskatchewan, Canada), resulting in the finding that the early flowering type showed a reduction in days to flower with later planting date, whereas late flowering types showed an increase in days to flower (Mc Intyre and Best, 1975). Subsequent results in the same study revealed an effect of temperature on days to flower in each type,

with increasing temperature accelerating flowering in the early flowering strain and delaying flowering in the late flowering strain. Finally, a study on late flowering strains showed a 2 week cold treatment of 2°C was sufficient to dramatically reduce days to flower, and increasing the length of vernalization further reduced days to flower (Mc Intyre and Best, 1975). The third paper in this series presented a more comprehensive study of the effects of vernalization on pennycress flowering, which showed that vernalization in the field and growth chamber hastened flowering in the late flowering phenotypes, and that cotyledon-stage seedlings are able to respond to vernalization and reduce days to flower (Best and McIntyre, 1976). The final paper in this series presented a genetic evaluation of the early and late flowering strains, in which an F2 cross between the two types was used to determine the genetic difference between the two was caused by a single gene with complete dominance of the late flowering allele (McIntyre and Best, 1978). This finding was a main motivating factor driving the study described in Chapter 4 of this dissertation.

While these previous studies have provided important information regarding flowering time and vernalization in pennycress, significant strides are needed to unravel the undoubtedly complex network of genes and environmental factors to meet the goal of producing rapid flowering winter annual accessions. Mentioned above, the ability to utilize information from Arabidopsis and related Brassicaceae species provides working models for likely gene homologs and genetic networks controlling not just flowering, but a host of highly conserved developmental and physiological traits (Sedbrook *et al.*,

2014). As such, a review of flowering time and vernalization in Arabidopsis is warranted and thus follows.

As sessile organisms, plants must sense and respond to environmental cues to survive and reproduce. These environmental stimuli include photoperiod, ambient temperature changes, seasonal temperature changes (like the prolonged cold of winter), all of which serve as major signals integrated by plants to regulate timing of flowering. Internal molecular pathways also influence the transition to flowering, including age, the autonomous pathway, and gibberellin. Each of these pathways will be discussed below, reviewing the discovery and characterization of key genes and pathways controlling each, mainly in the model species *Arabidopsis thaliana*, unless noted otherwise.

A critical environmental cue perceived by plants is day length. Photoperiodic flowering, or flowering in response to quantitative changes in day length, was first used to describe the short-day flowering habit of tobacco (*Nicotiana tabacum* var Maryland Mammoth), which flowers only under short day conditions (Garner and Allard, 1922). Plants can also be classified as either long day (flower only under long day lengths) or day neutral (Redington, 1929). The photoperiodic signal was thought to be sensed in the leaves, and after the inductive day length was perceived, a flowering hormone termed ‘florigen’ was transported to the shoot apical meristem (SAM) that initiated flowering (Chailakhyan, 1936). Later grafting experiments demonstrated that florigen was indeed a mobile signal perceived in the leaves and transmitted to the SAM (Zeevaart, 1976).

Several decades later, it is now known the major component of the 'florigen hormone' is actually a peptide, encoded by the gene FLOWERING LOCUS T (Tamaki *et al.*, 2007, Corbesier *et al.*, 2007). A host of FT-interacting genes are responsible for the photoperiodic regulation of FT expression, including CONSTANS (CO). In Arabidopsis, CO encodes a zinc finger transcription factor which activates FT transcription by directly interacting with the promoter of FT (Tiwari *et al.*, 2010). The photoperiodic regulation of CONSTANS is mediated via both transcriptional and post-translational means to allow for FT expression in long days only. The transcription of CO is mediated by its interactions with GIGANTEA (GI), FLAVIN KELCH F BOX 1 (FKF1), and CYCLING DOF FACTOR proteins (CDF1, CDF2, CDF3, and CDF5) which are also central components of the circadian clock (Sawa *et al.*, 2007, Song *et al.*, 2012b, Imaizumi *et al.*, 2005, Fornara *et al.*, 2009).

The convergence of the photoperiodic pathway on FT is not unique to the pathway, as each of the major pathways influencing flowering time are integrated by FT, along with SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) (Helliwell *et al.*, 2006, Moon *et al.*, 2003, Samach *et al.*, 2000, Liu *et al.*, 2008). The integration of the environmental and endogenous pathways by these 'floral integrator genes' coordinates the developmental mechanisms in the transition to flowering, including the transition of the shoot apical meristem from a vegetative meristem to an inflorescence meristem. This transition is controlled by the activation of floral meristem identity genes that induce flowering, including APETALA1 (AP1), APETALA2 (AP2),

APETALA3 (AP3)/PISTILLATA (PI), LEAFY (LFY), FRUITFUL (FUL), CAULIFLOWER (CAL), and AGAMOUS (AG) (Blázquez *et al.*, 1998, Blázquez *et al.*, 1997, Weigel *et al.*, 1992, Schultz and Haughn, 1991, Huala and Sussex, 1992, Mandel *et al.*, 1992, Bowman *et al.*, 1993, Ferrándiz *et al.*, 2000, Wagner *et al.*, 1999, Riechmann *et al.*, 1996, Liljegren *et al.*, 1999).

An additional environmental cue perceived by plants that influences flowering is temperature (Song *et al.*, 2013). In Arabidopsis, flowering under long day conditions is accelerated by growth in 23°C versus 16°C (Blázquez *et al.*, 2003). Increased temperatures have been shown to promote flowering via increased FT expression (Balasubramanian *et al.*, 2006). PHYTOCHROME INTERACTING FACTOR 4 (PIF4) activates expression of FT at high temperatures (27° C) by binding to the promoter of FT (Kumar *et al.*, 2012, Proveniers and van Zanten, 2013). SHORT VEGETATIVE PHASE (SVP) also appears to regulate the ambient temperature pathway, as *svp* mutants flower early regardless of temperature changes and represses the expression of FT at high temperatures (Lee *et al.*, 2007). SVP also negatively regulates the expression of microRNA 172 (miR172), a key signal in both the temperature and age-dependent flowering pathways.

Members of the microRNA 156 (miR156), miR172, and SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) network regulate the age-dependent flowering pathway, and are also important for the temperature-dependent regulation of

flowering (Lee *et al.*, 2010), leaf morphogenesis, and root architecture (Sun, 2011). In *Arabidopsis*, miR156a expression is high during the juvenile vegetative phase, while miR172b expression remains low. This trend reverses throughout the transition to flowering (Jung *et al.*, 2011, Wu *et al.*, 2009, Yang *et al.*, 2011). The overexpression of miR156a causes a prolonged juvenile phase, resulting in a larger rosette, while miR172b overexpression promotes early flowering (Huijser and Schmid, 2011). miR156 and miR172 levels are also temperature-dependent. The abundance of miR156 is several fold higher at lower temperatures (16°C vs. 23°C), while the trend reverses in the case of miR172, where miR172 levels are increased at 23°C (Lee *et al.*, 2010). A key target of miR156 is SPL3. The protein product of SPL3 binds directly to GTAC motifs within the promoter of FT (Kim et al., 2012). The reduced expression of miR172 under low temperatures (16°C) is caused by post-transcriptional processing of the primary miR172 transcript, which is mediated by FCA (Cho *et al.*, 2012, Jung *et al.*, 2012). The FCA gene product is an RNA-binding protein that binds to the primary miR172 transcript and positively regulates the processing of mature miR172 transcipt (Jung *et al.*, 2012). Low expression miR172 levels results in the increased expression of the miR172 targets TARGET OF EAT1 1 (TOE1), TOE2, and SCHLAFMUTZE (SMZ) (Lee *et al.*, 2010).

The vernalization response in plants is characterized by the acceleration of flowering in response to cold temperatures provided by winter (Kim *et al.*, 2009). Studies in Arabidopsis have revealed the required roles of two negative regulators of flowering, FLOWERING LOCUS C (FLC) and FRIGIDA (FRI). FLC encodes a MADS domain

transcription factor that represses flowering prior to vernalization (Michaels and Amasino, 1999), whereas FRI (Johanson, 2000) is now known to act as a scaffolding protein for the FRIGIDA Complex (FRI-C) that consists of FRIGIDA LIKE 1 (FRL1), FRIGIDA-ESSENTIAL 1 (FES1), SUPPRESSOR OF FRIGIDA4 (SUF4), and FLC EXPRESSOR (FLX) (Choi *et al.*, 2011). Components of FRI-C were found to have specific roles in the interaction with FLC and FRI-C's role as a transcriptional activation complex for FLC expression (Choi *et al.*, 2011). In Arabidopsis and pennycress, there are wild accessions of both spring and winter annual plants, where winter annual plants require vernalization to promote flowering, similar to the 'late flowering' pennycress strains evaluated by McIntyre and Best (1978). In Arabidopsis, the vernalization requirement is mediated by functional alleles of either FRI or FLC, with loss of function alleles in either gene resulting in a rapid flowering phenotype (Michaels and Amasino, 1999, Johanson *et al.*, 2000, Shindo *et al.*, 2005, Michaels and Amasino, 2001).

The effect of vernalization on the removal of FLC-mediated repression of flowering has been a widely addressed question. Prolonged cold treatment epigenetically silences the expression of FLC through H3K27me3 histone methylation (Bastow *et al.*, 2004). This epigenetic silencing of FLC is mediated by chromatin remodeling processes by VERNALIZATION INSENSITIVE 3 (VIN3) and VIN3-LIKE, which are required for the initial repression of FLC, along with a the activity of a Polycomb Respressive Complex 2 involved in the H3K27 methylation that is necessary to maintain repression over time (Jarillo and Piñeiro, 2011). This long-term repression also requires LIKE

HETEROCHROMATIN PROTEIN 1(LHP1)/TERMINAL FLOWER 2 (TFL2) which has been shown to bind H3K27me3 chromatin mark to maintain stable repression (Mylne *et al.*, 2006, Turck *et al.*, 2007, Sung *et al.*, 2006, Zhang *et al.*, 2007). Parallel to vernalization-based repression of FLC is the autonomous pathway that also regulates FLC levels, consisting of FCA, FY, FPA, FVE, LUMINIDEPENDENS (LD), FLOEWRING LATE KH MOTIF (FLK), and FLOWERING LOCUS D (FLD). Each of these autonomous pathway members repress FLC expression at a variety of levels, including RNA binding and processing, as well as chromatin remodeling (He *et al.*, 2003, Simpson *et al.*, 2003, Liu *et al.*, 2007, Marquardt *et al.*, 2006).

Gibberellins (GA) are an endogenous growth regulator that also promotes flowering (Lang, 1957). The synthesis of the bioactive GA4 is catalyzed by GA 20-oxidase (GA20ox) and GA 3-oxidase (GA3ox). In Arabidopsis leaves, ASYMMETRIC LEAVES 1 (AS1), a transcription factor, positively regulates the expression of GA20ox1, as well as interacting with CONSTANS to regulate the expression of FT (Song *et al.*, 2012a). Mutant analysis in Arabidopsis also showed that disruptions in either the biosynthesis of GA or GA signaling pathways can affect flowering time (Jacobsen and Olszewski, 1993, Wilson *et al.*, 1992). An EMS mutant screen in pennycress additionally identified a GA biosynthesis mutant with several developmental phenotypes, including delayed flowering that could be reversed with the application of exogenous GA (Metzger and Hassebrock, 1990). The promotion of flowering by GA is also in part due to its role to increase expression of LFY (Blázquez *et al.*, 1998). Integration of the GA-dependent

pathway has also been shown to converge upon the floral integrator gene SOC1 (Moon *et al.*, 2003). Gibberellins also regulate the transition to flowering through DELLA proteins, which are negative regulators in GA signaling. Expression of SPL3, a positive regulator of FT expression, is repressed by DELLA proteins in leaves, along with SPL3, SPL4, and SPL5 in the apex (Galvão *et al.*, 2012). DELLAs also influence flowering through the reduction of leaf miR172 levels in long day conditions through the regulation of SPL activity (Galvão *et al.*, 2012, Yu *et al.*, 2012).

Part IV: Conclusions

As previously discussed, significant progress has been made in understanding the basic physiological and developmental mechanisms controlling important plant processes that can serve as useful guides for the *de novo* domestication of new crops. There is an unprecedented demand on agricultural systems to provide more food and fuel than has ever been produced, and dramatic changes in these systems will be needed to meet these goals sustainably. While incremental improvements have been made through the refinement of germplasm of the major annual crops to meet these demands, more significant changes to the systems and biodiversity of these systems will likely be needed. With the development of new species that can either replace or supplement the current major systems will come options to meet these great needs.

In this dissertation, the domestication of one such new crop, field pennycress, is discussed. Through the development of novel genomic resources, including a transcriptome (Chapter 2), and genome sequence (Chapter 3) for pennycress has allowed for the identification of key genetic targets for improvement, which is based on knowledge from the massive public investment in basic *Arabidopsis thaliana* research. Chapter 4 presents an analysis of genes responsible for flowering time variation in pennycress, which is one of the key targets of pennycress domestication goals. Using the resources and analyses presented in this dissertation, the domestication of field pennycress as a new winter cover crop and biodiesel feedstock should be possible within a matter of years, not centuries.

References

Sureshkumar Balasubramanian, Sridevi Sureshkumar, Janne Lempe & Detlef Weigel 2006. Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. *PLoS Genet,* 2**,** e106.

Jm Baskin & Cc Baskin 1989. Role of temperature in regulating timing of germination in soil seed reserves of Thlaspi arvense L. *Weed Research,* 29**,** 317-326.

Ruth Bastow, Joshua S Mylne, Clare Lister, Zachary Lippman, Robert A Martienssen & Caroline Dean 2004. Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature,* 427**,** 164-167.

Kf Best & Gi Mcintyre 1975. The Biology of Canadian Weeds 9. Thlaspi arvense L. . *Canadian Journal of Plant Science,* 55**,** 279-292.

Kf Best & Gi Mc Intyre 1976. Studies on the flowering of Thlaspi arvense L. III. The influence of vernalization under natural and controlled conditions. *Botanical Gazette***,** 121-127.

Kf Best & Gi Mcintyre 1972. Studies on the Flowering Time of Thlaspi arvense L. I. The Influence of some Environmental and Genetic Factors. *Botanical Gazette,* 133**,** 454 - 459.

David Biello 2011. The false promise of biofuels. *Scientific American,* 305**,** 58-65.

Miguel A Blázquez, Ji Hoon Ahn & Detlef Weigel 2003. A thermosensory pathway controlling flowering time in Arabidopsis thaliana. *Nat Genet,* 33**,** 168-171.

Miguel A Blázquez, Roland Green, Ove Nilsson, Michael R Sussman & Detlef Weigel 1998. Gibberellins promote flowering of Arabidopsis by activating the LEAFY promoter. *The Plant Cell Online,* 10**,** 791-800.

Miguel A Blázquez, Lara N Soowal, Ilha Lee & Detlef Weigel 1997. LEAFY expression and flower initiation in Arabidopsis. *Development,* 124**,** 3835-3844.

A. A. Boateng, C. A. Mullen & N. M. Goldberg 2010. Producing Stable Pyrolysis Liquids from the Oil-Seed Presscakes of Mustard Family Plants: Pennycress (Thlaspi arvense L.) and Camelina (Camelina sativa)†. *Energy & Fuels,* 24**,** 6624-6632.

John L Bowman, John Alvarez, Detlef Weigel, Elliot M Meyerowitz & David R Smyth 1993. Control of flower development in Arabidopsis thaliana by APETALA 1 and interacting genes. *DEVELOPMENT-CAMBRIDGE-,* 119**,** 721-721.

Mario Caccamo & Erich Grotewold 2013. Turning over a new leaf in plant genomics. *Genome Biol,* 14**,** 403.

T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove & D. F. Voytas 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res,* 39**,** e82.

M Chailakhyan 1936. New facts in support of the hormonal theory of plant development *Proceedings of the USSR Academy of Sciences,* 4**,** 79-83.

Hyun Jung Cho, Jae Joon Kim, Jeong Hwan Lee, Wanhui Kim, Jae-Hoon Jung, Chung-Mo Park & Ji Hoon Ahn 2012. SHORT VEGETATIVE PHASE (SVP) protein negatively regulates miR172 transcription via direct binding to the pri-miR172a promoter in Arabidopsis. *FEBS Lett,* 586**,** 2332-2337.

K. Choi, J. Kim, H. J. Hwang, S. Kim, C. Park, S. Y. Kim & I. Lee 2011. The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *Plant Cell,* 23**,** 289-303.

Laurent Corbesier, Coral Vincent, Seonghoe Jang, Fabio Fornara, Qingzhi Fan, Iain Searle, Antonis Giakountis, Sara Farrona, Lionel Gissot & Colin Turnbull 2007. FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science,* 316**,** 1030-1033.

Sm Dabney, Ja Delgado & Dw Reeves 2001. Using winter cover crops to improve soil and water quality. *Communications in Soil Science and Plant Analysis,* 32**,** 1221-1250.

Michel Delseny, Bin Han & Yue Ie Hsing 2010. High throughput DNA sequencing: The new sequencing revolution. *Plant Science,* 179**,** 407-422.

John F Doebley, Brandon S Gaut & Bruce D Smith 2006. The molecular genetics of crop domestication. *Cell,* 127**,** 1309-1321.

Roque L. Evangelista, Terry A. Isbell & Steven C. Cermak 2012. Extraction of pennycress (Thlaspi arvense L.) seed oil by full pressing. *Industrial Crops and Products,* 37**,** 76-81.

J. Q. Fan, D. R. Shonnard, T. N. Kalnes, P. B. Johnsen & S. Rao 2013. A life cycle assessment of pennycress (Thlaspi aruense L.) -derived jet fuel and diesel. *Biomass & Bioenergy,* 55**,** 87-100.

Cristina Ferrándiz, Qing Gu, Robert Martienssen & Martin F Yanofsky 2000. Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development,* 127**,** 725-734.

Catherine Feuillet, Jan E Leach, Jane Rogers, Patrick S Schnable & Kellye Eversole 2011. Crop genome sequencing: lessons and rationales. *Trends Plant Sci,* 16**,** 77-88.

Fabio Fornara, Kishore Cs Panigrahi, Lionel Gissot, Nicolas Sauerbrunn, Mark Rühl, José A Jarillo & George Coupland 2009. Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev Cell,* 17**,** 75-86.

A. Franzke, M. A. Lysak, I. A. Al-Shehbaz, M. A. Koch & K. Mummenhoff 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci,* 16**,** 108-16.

Vinicius C Galvão, Daniel Horrer, Frank Küttner & Markus Schmid 2012. Spatial control of flowering by DELLA proteins in Arabidopsis thaliana. *Development,* 139**,** 4072-4082.

Ww Garner & Ha Allard 1922. Photoperiodism, the response of the plant to relative length of day and night. *Science***,** 582-583.

Tara Garnett, Mc Appleby, A Balmford, Ij Bateman, Tg Benton, P Bloomer, B Burlingame, M Dawkins, L Dolan & D Fraser 2013. Sustainable intensification in agriculture: premises and policies. *Science,* 341**,** 33-34.

H Charles J Godfray, John R Beddington, Ian R Crute, Lawrence Haddad, David Lawrence, James F Muir, Jules Pretty, Sherman Robinson, Sandy M Thomas & Camilla Toulmin 2010. Food security: the challenge of feeding 9 billion people. *Science,* 327**,** 812-818.

Erika Check Hayden 2014. Technology: the $1,000 genome. *Nature,* 507**,** 294-5.

Jan P Hazebroek & James D Metzger 1990a. Environmental control of seed germination in Thlaspi arvense (Cruciferae). *Am J Bot*, 945-953.

Jan P Hazebroek & James D Metzger 1990b. Seasonal pattern of seedling emergence, survival, and reproductive behavior in Thlaspi arvense (Cruciferae). *Am J Bot*, 954-962.

Yuehui He, Scott D Michaels & Richard M Amasino 2003. Regulation of flowering time by histone acetylation in Arabidopsis. *Science,* 302**,** 1751-1754.

Emily A Heaton, Lisa A Schulte, Marisol Berti, Hans Langeveld, Walter Zegada-Lizarazu, David Parrish & Andrea Monti 2013. Managing a second-generation crop portfolio through sustainable intensification: Examples from the USA and the EU. *Biofuels, Bioproducts and Biorefining,* 7**,** 702-714.

Chris A Helliwell, Craig C Wood, Masumi Robertson, W James Peacock & Elizabeth S Dennis 2006. The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. *The Plant Journal,* 46**,** 183-192.

J. Hill, E. Nelson, D. Tilman, S. Polasky & D. Tiffany 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc Natl Acad Sci U S A,* 103**,** 11206-10.

M. P. Hojilla-Evangelista, R. L. Evangelista, T. A. Isbell & G. W. Selling 2013. Effects of cold-pressing and seed cooking on functional properties of protein in pennycress (Thlaspi arvense L.) seed and press cakes. *Industrial Crops and Products,* 45**,** 223-229.

Eva Huala & Ian M Sussex 1992. LEAFY interacts with floral homeotic genes to regulate Arabidopsis floral development. *The Plant Cell Online,* 4**,** 901-913.

P. Huijser & M. Schmid 2011. The control of developmental phase transitions in plants. *Development,* 138**,** 4117-29.

L Hume 1990. Influence of emergence date and strain on phenology, seed production, and germination of Thlaspi arvense L. *Botanical Gazette***,** 510-515.

Larry Hume 1994. Maternal environment effects on plant growth and germination of two strains of Thlaspi arvense L. *International journal of plant sciences***,** 180-186.

Takato Imaizumi, Thomas F Schultz, Frank G Harmon, Lindsey A Ho & Steve A Kay 2005. FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. *Science,* 309**,** 293-297.

Steven E Jacobsen & Neil E Olszewski 1993. Mutations at the SPINDLY locus of Arabidopsis alter gibberellin signal transduction. *The Plant Cell Online,* 5**,** 887-896.

Jose A Jarillo & Manuel Piñeiro 2011. Timing is everything in plant development. The central role of floral repressors. *Plant Science,* 181**,** 364-378.

U. Johanson 2000. Molecular Analysis of FRIGIDA, a Major Determinant of Natural Variation in Arabidopsis Flowering Time. *Science,* 290**,** 344-347.

U. Johanson, J. West, C. Lister, S. Michaels, R. Amasino & C. Dean 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science,* 290**,** 344-7.

Gregg A Johnson, Michael B Kantar, Kevin J Betts & Donald L Wyse 2015. Field Pennycress Production and Weed Control in a Double Crop System with Soybean in Minnesota. *Agronomy Journal*.

N Jordan, G Boody, W Broussard, Jd Glover, D Keeney, Bh Mccown, G Mcisaac, M Muller, H Murray & J Neal 2007. Sustainable development of the agricultural bio-economy. *SCIENCE-NEW YORK THEN WASHINGTON-,* 316**,** 1570.

J. H. Jung, P. J. Seo, S. K. Kang & C. M. Park 2011. miR172 signals are incorporated into the miR156 signaling pathway at the SPL3/4/5 genes in Arabidopsis developmental transitions. *Plant Mol Biol,* 76**,** 35-45.

Jae-Hoon Jung, Pil Joon Seo, Ji Hoon Ahn & Chung-Mo Park 2012. Arabidopsis RNA-binding protein FCA regulates microRNA172 processing in thermosensory flowering. *Journal of Biological Chemistry,* 287**,** 16007-16016.

S. Kagale, C. S. Koh, J. Nixon, V. Bollina, W. E. Clarke, R. Tuteja, C. Spillane, S. J. Robinson, M. G. Links, C. Clarke, E. E. Higgins, T. Huebert, A. G. Sharpe & I. A. P. Parkin 2014. The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nat Commun,* 5.

S. Kaul, H. L. Koo, J. Jenkins, M. Rizzo, T. Rooney, L. J. Tallon, T. Feldblyum, W. Nierman, M. I. Benito, X. Y. Lin, C. D. Town, J. C. Venter, C. M. Fraser, S. Tabata, Y. Nakamura, T. Kaneko, S. Sato, E. Asamizu, T. Kato, H. Kotani, S. Sasamoto, J. R. Ecker, A. Theologis, N. A. Federspiel, C. J. Palm, B. I. Osborne, P. Shinn, A. B. Conway, V. S. Vysotskaia, K. Dewar, L. Conn, C. A. Lenz, C. J. Kim, N. F. Hansen, S. X. Liu, E. Buehler, H. Altafi, H. Sakano, P. Dunn, B. Lam, P. K. Pham, Q. Chao, M. Nguyen, G. X. Yu, H. M. Chen, A. Southwick, J. M. Lee, M. Miranda, M. J. Toriumi, R. W. Davis, R. Wambutt, G. Murphy, A. Dusterhoft, W. Stiekema, T. Pohl, K. D. Entian, N. Terryn, G. Volckaert, M. Salanoubat, N. Choisne, M. Rieger, W. Ansorge, M. Unseld, B. Fartmann, G. Valle, F. Artiguenave, J. Weissenbach, F. Quetier, R. K. Wilson, M. De La Bastide, M. Sekhon, E. Huang, L. Spiegel, L. Gnoj, K. Pepin, J. Murray, D. Johnson, K. Habermann, N. Dedhia, L. Parnell, R. Preston, L. Hillier, E. Chen, M.

Marra, R. Martienssen, W. R. Mccombie, K. Mayer, O. White, M. Bevan, K. Lemcke, T. H. Creasy, C. Bielke, B. Haas, D. Haase, R. Maiti, S. Rudd, J. Peterson, H. Schoof, D. Frishman, B. Morgenstern, P. Zaccaria, M. Ermolaeva, M. Pertea, J. Quackenbush, N. Volfovsky, D. Y. Wu, T. M. Lowe, S. L. Salzberg, H. W. Mewes, S. Rounsley, D. Bush, S. Subramaniam, I. Levin, S. Norris, R. Schmidt, A. Acarkan, I. Bancroft, F. Quetier, A. Brennicke, J. A. Eisen, T. Bureau, B. A. Legault, Q. H. Le, N. Agrawal, Z. Yu, R. Martienssen, G. P. Copenhaver, S. Luo, C. S. Pikaard, D. Preuss, I. T. Paulsen, M. Sussman, A. B. Britt, D. A. Selinger, R. Pandey, D. W. Mount, V. L. Chandler, R. A. Jorgensen, C. Pikaard, G. Juergens, E. M. Meyerowitz, A. Theologis, J. Dangl, J. D. G. Jones, M. Chen, J. Chory, M. C. Somerville & Ar Gen In 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408**,** 796-815.

D. H. Kim, M. R. Doyle, S. Sung & R. M. Amasino 2009. Vernalization: winter and the timing of flowering in plants. *Annu Rev Cell Dev Biol,* 25**,** 277-99.

J. J. Kim, J. H. Lee, W. Kim, H. S. Jung, P. Huijser & J. H. Ahn 2012. The microRNA156-SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3 module regulates ambient temperature-responsive flowering via FLOWERING LOCUS T in Arabidopsis. *Plant Physiol,* 159**,** 461-78.

Lj Klebesadel 1969. Life cycles of field pennycress in the subarctic as influenced by time of seed germination. *Weed Science***,** 563-566.

Daniel C Koboldt, Karyn Meltz Steinberg, David E Larson, Richard K Wilson & Elaine R Mardis 2013. The next-generation sequencing revolution and its impact on genomics. *Cell,* 155**,** 27-38.

S Vinod Kumar, Doris Lucyshyn, Katja E Jaeger, Enriqueta Alós, Elizabeth Alvey, Nicholas P Harberd & Philip A Wigge 2012. Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature,* 484**,** 242-245.

Es Lander, Linton Lm, Birren B, Nusbaum C, Zody Mc, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, Mcewan P, Mckernan K, Meldrim J, Mesirov Jp, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, Mcmurray A, Matthews L, Mercer S, Milne S, Mullikin Jc, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston Rh, Wilson Rk, Hillier Lw, Mcpherson Jd, Marra Ma, Mardis Er, Fulton La, Chinwalla At, Pepin Kh, Gish Wr, Chissoe Sl, Wendl Mc, Delehaunty Kd, Miner Tl, Delehaunty A, Kramer Jb, Cook Ll, Fulton Rs, Johnson Dl, Minx Pj, Clifton Sw, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T Doggett N, Cheng Jf, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs Ra, Muzny Dm, Scherer Se, Bouck Jb, Sodergren Ej, Worley Kc, Rives Cm, Gorrell Jh, Metzker Ml, Naylor Sl, Kucherlapati Rs, Nelson Dl, Weinstock Gm, Sakaki Y Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith Dr, Doucette-Stamm L, Rubenfield M Weinstock K, Lee Hm, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis Rw, Federspiel Na, Abola Ap, Proctor Mj, Myers Rm, Schmutz J, Dickson M, Grimwood J, Cox Dr, Olson Mv, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans Ga, Athanasiou M, Schultz R, Roe Ba, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, Mccombie Wr, De La Bastide M, Dedhia N,

Blöcker H, Hornischer K, Nordsiek G Agarwala R, Aravind L, Bailey Ja, Bateman A, Batzoglou S, Birney E, Bork P, Brown Dg, Burge Cb, Cerutti L, Chen Hc, Church D, Clamp M, Copley Rr, Doerks T, Eddy Sr, Eichler Ee, Furey Ts, Galagan J, Gilbert Jg, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson Ls, Jones Ta, Kasif S, Kaspryzk A, Kennedy S, Kent Wj, Kitts P, Koonin Ev, Korf I, Kulp D, Lancet D, Lowe Tm, Mclysaght A, Mikkelsen T, Moran Jv, Mulder N, Pollara Vj, Ponting Cp, Schuler G, Schultz J, Slater G, Smit Af, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf Yi, Wolfe Kh, Yang Sp, Yeh Rf, Collins F, Guyer Ms, Peterson J, Felsenfeld A, Wetterstrand Ka, Patrinos A, Morgan Mj, De Jong P Catanese Jj, Osoegawa K, Shizuya H, Choi S & Chen Yj 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Anton Lang 1957. The effect of gibberellin upon flower formation. *Proceedings of the National Academy of Sciences,* 43**,** 709-717.

Hanna Lee, Seong Jeon Yoo, Jeong Hwan Lee, Wanhui Kim, Seung Kwan Yoo, Heather Fitzgerald, James C Carrington & Ji Hoon Ahn 2010. Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis. *Nucleic Acids Res***,** gkp1240.

Jeong Hwan Lee, Seong Jeon Yoo, Soo Hyun Park, Ildoo Hwang, Jong Seob Lee & Ji Hoon Ahn 2007. Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. *Genes & Development,* 21**,** 397-402.

Sarah J Liljegren, Cindy Gustafson-Brown, Anusak Pinyopich, Gary S Ditta & Martin F Yanofsky 1999. Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *The Plant Cell Online,* 11**,** 1007-1018.

Chang Liu, Hongyan Chen, Hong Ling Er, Hui Meng Soo, Prakash P Kumar, Jin-Hua Han, Yih Cherng Liou & Hao Yu 2008. Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis. *Development,* 135**,** 1481-1491.

Fuquan Liu, Victor Quesada, Pedro Crevillén, Isabel Bäurle, Szymon Swiezewski & Caroline Dean 2007. The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. *Mol Cell,* 28**,** 398-407.

S. Liu, Y. Liu, X. Yang, C. Tong, D. Edwards, I. A. Parkin, M. Zhao, J. Ma, J. Yu, S. Huang, X. Wang, J. Wang, K. Lu, Z. Fang, I. Bancroft, T. J. Yang, Q. Hu, X. Wang, Z. Yue, H. Li, L. Yang, J. Wu, Q. Zhou, W. Wang, G. J. King, J. C. Pires, C. Lu, Z. Wu, P. Sampath, Z. Wang, H. Guo, S. Pan, L. Yang, J. Min, D. Zhang, D. Jin, W. Li, H. Belcram, J. Tu, M. Guan, C. Qi, D. Du, J. Li, L. Jiang, J. Batley, A. G. Sharpe, B. S. Park, P. Ruperao, F. Cheng, N. E. Waminal, Y. Huang, C. Dong, L. Wang, J. Li, Z. Hu, M. Zhuang, Y. Huang, J. Huang, J. Shi, D. Mei, J. Liu, T. H. Lee, J. Wang, H. Jin, Z. Li, X. Li, J. Zhang, L. Xiao, Y. Zhou, Z. Liu, X. Liu, R. Qin, X. Tang, W. Liu, Y. Wang, Y. Zhang, J. Lee, H. H. Kim, F. Denoeud, X. Xu, X. Liang, W. Hua, X. Wang, J. Wang, B. Chalhoub & A. H. Paterson 2014. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun,* 5**,** 3930.

M Alejandra Mandel, Cindy Gustafson-Brown, Beth Savidge & Martin F Yanofsky 1992. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1.

S Marquardt, Pk Boss, J Hadfield & C Dean 2006. Additional targets of the Arabidopsis autonomous pathway members, FCA and FY. *J Exp Bot,* 57**,** 3379-3386.

Klaus Fx Mayer, Jane Rogers, Jaroslav Doležel, Curtis Pozniak, Kellye Eversole, Catherine Feuillet, Bikram Gill, Bernd Friebe, Adam J Lukaszewski & Pierre

Sourdille 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science,* 345**,** 1251788.

Gi Mc Intyre & Kf Best 1975. Studies on the flowering of Thlaspi arvense L. II. A comparative study of early-and late-flowering strains. *Botanical Gazette***,** 151-158.

Gi Mcintyre & Kf Best 1978 Studies on the Flowering of Thlaspi arvense IV. Genetic and Ecological Differences between Early and Late Flowering Strains *Botanical Gazette,* 139**,** 190-195.

J. D. Metzger & A. T. Hassebrock 1990. Selection and Characterization of a Gibberellin-Deficient Mutant of Thlaspi arvense L. *Plant Physiol,* 94**,** 1655-62.

Jin Miao, Dongshu Guo, Jinzhe Zhang, Qingpei Huang, Genji Qin, Xin Zhang, Jianmin Wan, Hongya Gu & Li-Jia Qu 2013. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res,* 23**,** 1233-1236.

S. D. Michaels & R. M. Amasino 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell,* 11**,** 949-56.

S. D. Michaels & R. M. Amasino 2001. Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell,* 13**,** 935-41.

Jihyun Moon, Sung-Suk Suh, Horim Lee, Kyu-Ri Choi, Choo Bong Hong, Nam-Chon Paek, Sang-Gu Kim & Ilha Lee 2003. The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. *The Plant Journal,* 35**,** 613-623.

B. R. Moser, G. Knothe, S. F. Vaughn & T. A. Isbell 2009a. Production and Evaluation of Biodiesel from Field Pennycress (Thlaspi arvense L.) Oil. *Energy & Fuels,* 23**,** 4149-4155.

B. R. Moser, S. N. Shah, J. K. Winkler-Moser, S. F. Vaughn & R. L. Evangelista 2009b. Composition and physical properties of cress (Lepidium sativum L.) and field pennycress (Thlaspi arvense L.) oils. *Industrial Crops and Products,* 30**,** 199-205.

Joshua S Mylne, Lynne Barrett, Federico Tessadori, Stéphane Mesnage, Lianna Johnson, Yana V Bernatavichute, Steven E Jacobsen, Paul Fransz & Caroline Dean 2006. LHP1, the Arabidopsis homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of FLC. *Proc Natl Acad Sci U S A,* 103**,** 5012-5017.

Winthrop B. Phippen & Mary E. Phippen 2012. Soybean Seed Yield and Quality as a Response to Field Pennycress Residue. *Crop Science,* 52**,** 2767.

Marcel Cg Proveniers & Martijn Van Zanten 2013. High temperature acclimation through PIF4 signaling. *Trends Plant Sci,* 18**,** 59-64.

Deepak K Ray & Jonathan A Foley 2013. Increasing global crop harvest frequency: recent trends and future directions. *Environmental Research Letters,* 8**,** 044041.

George Redington 1929. THE EFFECT OF THE DURATION OF LIGHT UPON THE GORWTH AND DEVELOPMENT OF THE PLANT. *Biological Reviews,* 4**,** 180-208.

Jose Luis Riechmann, Beth Allyn Krizek & Elliot M Meyerowitz 1996. Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proceedings of the National Academy of Sciences,* 93**,** 4793-4798.

C. Rogers, J. Wen, R. Chen & G. Oldroyd 2009. Deletion-based reverse genetics in Medicago truncatula. *Plant Physiol,* 151**,** 1077-86.

Alon Samach, Hitoshi Onouchi, Scott E Gold, Gary S Ditta, Zsuzsanna Schwarz-Sommer, Martin F Yanofsky & George Coupland 2000. Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science,* 288**,** 1613-1616.

Mariko Sawa, Dmitri A Nusinow, Steve A Kay & Takato Imaizumi 2007. FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. *Science,* 318**,** 261-265.

Jeremy Schmutz, Steven B Cannon, Jessica Schlueter, Jianxin Ma, Therese Mitros, William Nelson, David L Hyten, Qijian Song, Jay J Thelen, Jianlin Cheng, Dong Xu, Uffe Hellsten, Gregory D May, Yeisoo Yu, Tetsuya Sakurai, Taishi Umezawa, Madan K Bhattacharyya, Devinder Sandhu, Babu Valliyodan, Erika Lindquist, Myron Peto, David Grant, Shengqiang Shu, David Goodstein, Kerrie Barry, Montona Futrell-Griggs, Brian Abernathy, Jianchang Du, Zhixi Tian, Liucun Zhu, Navdeep Gill, Trupti Joshi, Marc Libault, Anand Sethuraman, Xue-Cheng Zhang, Kazuo Shinozaki, Henry T Nguyen, Rod A Wing, Perry Cregan, James Specht, Jane Grimwood, Dan Rokhsar, Gary Stacey, Randy C Shoemaker & Scott A. Jackson 2010. Genome sequence of the palaeopolyploid soybean. *Nature,* 463**,** 178-183.

Ps Schnable, D Ware, Rs Fulton, Jc Stein, F Wei, S Pasternak, C Liang, J Zhang, L Fulton, Ta Graves, P Minx, Ad Reily, L Courtney, Ss Kruchowski, C Tomlinson, C Strong, K Delehaunty, C Fronick, B Courtney, Sm Rock, E Belter, F Du, K Kim, Rm Abbott, M Cotton, A Levy, P Marchetto, K Ochoa, Sm Jackson, B Gillam, W Chen, L Yan, J Higginbotham, M Cardenas, J Waligorski, E Applebaum, L Phelps, J Falcone, K Kanchi, T Thane, A Scimone, N Thane, J

Henke, T Wang, J Ruppert, N Shah, K Rotter, J Hodges, E Ingenthron, M Cordes, S Kohlberg, J Sgro, B Delgado, K Mead, A Chinwalla, S Leonard, K Crouse, K Collura, D Kudrna, J Currie, R He, A Angelova, S Rajasekar, T Mueller, R Lomeli, G Scara, A Ko, K Delaney, M Wissotski, G Lopez, D Campos, M Braidotti, E Ashley, W Golser, H Kim, S Lee, J Lin, Z Dujmic, W Kim, J Talag, A Zuccolo, C Fan, A Sebastian, M Kramer, L Spiegel, L Nascimento, T Zutavern, B Miller, C Ambroise, S Muller, W Spooner, A Narechania, L Ren, S Wei, S Kumari, B Faga, Mj Levy, L Mcmahan, P Van Buren, Mw Vaughn, K Ying, Ct Yeh, Sj Emrich, Y Jia, A Kalyanaraman, Ap Hsia, Wb Barbazuk, Rs Baucom, Tp Brutnell, Nc Carpita, C Chaparro, Jm Chia, Jm Deragon, Jc Estill, Y Fu, Ja Jeddeloh, Y Han, H Lee, P Li, Dr Lisch, S Liu, Z Liu, Dh Nagel, Mc Mccann, P Sanmiguel, Am Myers, D Nettleton, J Nguyen, Bw Penning, L Ponnala, Kl Schneider, Dc Schwartz, A Sharma, C Soderlund, Nm Springer, H Sun Q Wang, M Waterman, R Westerman, Tk Wolfgruber, L Yang, Y Yu, L Zhang, S Zhou, Q Zhu, Jl Bennetzen, Rk Dawe, J Jiang, N Jiang, Gg Presting, Sr Wessler, S Aluru, Ra Martienssen, Sw Clifton, Wr Mccombie, Ra Wing & Rk Wilson 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science,* 326**,** 1112-1115.

Elizabeth A Schultz & George W Haughn 1991. LEAFY, a homeotic gene that regulates inflorescence development in Arabidopsis. *The Plant Cell Online,* 3**,** 771-781.

J. C. Sedbrook, W. B. Phippen & M. D. Marks 2014. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (Thlaspi arvense L.). *Plant Sci,* 227**,** 122-32.

G. W. Selling, M. P. Hojilla-Evangelista, R. L. Evangelista, T. Isbell, N. Price & K. M. Doll 2013. Extraction of proteins from pennycress seeds and press cake. *Industrial Crops and Products,* 41**,** 113-119.

N. Sharma, D. Cram, T. Huebert, N. Zhou & I. A. Parkin 2007. Exploiting the wild crucifer Thlaspi arvense to identify conserved and novel genes expressed during a plant's response to cold stress. *Plant Mol Biol,* 63**,** 171-84.

C. Shindo, M. J. Aranzana, C. Lister, C. Baxter, C. Nicholls, M. Nordborg & C. Dean 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant Physiol,* 138**,** 1163-73.

Gordon G Simpson, Paul P Dijkwel, Victor Quesada, Ian Henderson & Caroline Dean 2003. FY is an RNA 3′ end-processing factor that interacts with FCA to control the Arabidopsis floral transition. *Cell,* 113**,** 777-787.

Jw Singer, Sm Nusser & Cj Alf 2007. Are cover crops being used in the US corn belt? *Journal of Soil and Water Conservation,* 62**,** 353-358.

T. Slotte, K. M. Hazzouri, J. A. Agren, D. Koenig, F. Maumus, Y. L. Guo, K. Steige, A. E. Platts, J. S. Escobar, L. K. Newman, W. Wang, T. Mandakova, E. Vello, L. M. Smith, S. R. Henz, J. Steffen, S. Takuno, Y. Brandvain, G. Coop, P. Andolfatto, T. T. Hu, M. Blanchette, R. M. Clark, H. Quesneville, M. Nordborg, B. S. Gaut, M. A. Lysak, J. Jenkins, J. Grimwood, J. Chapman, S. Prochnik, S. Shu, D. Rokhsar, J. Schmutz, D. Weigel & S. I. Wright 2013. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nat Genet,* 45**,** 831-5.

Ss Snapp, Sm Swinton, R Labarta, D Mutch, Jr Black, R Leep, J Nyiraneza & K O'neil 2005. Evaluating cover crops for benefits, costs and performance within cropping system niches. *Agronomy Journal,* 97**,** 322-332.

Y. H. Song, I. Lee, S. Y. Lee, T. Imaizumi & J. C. Hong 2012a. CONSTANS and ASYMMETRIC LEAVES 1 complex is involved in the induction of

FLOWERING LOCUS T in photoperiodic flowering in Arabidopsis. *Plant J,* 69**,** 332-42.

Young Hun Song, Shogo Ito & Takato Imaizumi 2013. Flowering time regulation: photoperiod-and temperature-sensing in leaves. *Trends Plant Sci,* 18**,** 575-583.

Young Hun Song, Robert W Smith, Benjamin J To, Andrew J Millar & Takato Imaizumi 2012b. FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. *Science,* 336**,** 1045-1049.

G. Sun 2011. MicroRNAs and their diverse functions in plants. *Plant Mol Biol*.

Sibum Sung, Yuehui He, Tifani W Eshoo, Yosuke Tamada, Lianna Johnson, Kenji Nakahigashi, Koji Goto, Steve E Jacobsen & Richard M Amasino 2006. Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1. *Nat Genet,* 38**,** 706-710.

Shojiro Tamaki, Shoichi Matsuo, Hann Ling Wong, Shuji Yokoi & Ko Shimamoto 2007. Hd3a protein is a mobile flowering signal in rice. *Science,* 316**,** 1033-1036.

David Tilman, Christian Balzer, Jason Hill & Belinda L Befort 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences,* 108**,** 20260-20264.

Shiv B Tiwari, Yu Shen, Han-Chang Chang, Yanli Hou, Amanda Harris, Siu Fong Ma, Megan Mcpartland, Graham J Hymus, Luc Adam & Colleen Marion 2010. The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. *New Phytologist,* 187**,** 57-66.

Helen Tsai, Tyson Howell, Rebecca Nitcher, Victor Missirian, Brian Watson, Kathie J Ngo, Meric Lieberman, Joseph Fass, Cristobal Uauy & Robert K Tran 2011.

Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiol,* 156**,** 1257-1268.

Franziska Turck, François Roudier, Sara Farrona, Marie-Laure Martin-Magniette, Elodie Guillaume, Nicolas Buisine, Séverine Gagnot, Robert A Martienssen, George Coupland & Vincent Colot 2007. Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet,* 3**,** e86.

Usda. *Global Crop Production Analysis* [Online]. United States Department of Agriculture Foreign Agricultural Service. Available: http://www.pecad.fas.usda.gov/ [Accessed 4/30/15 2015].

R. K. Varshney, S. N. Nayak, G. D. May & S. A. Jackson 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol,* 27**,** 522-30.

R. K. Varshney, C. Song, R. K. Saxena, S. Azam, S. Yu, A. G. Sharpe, S. Cannon, J. Baek, B. D. Rosen, B. Tar'an, T. Millan, X. Zhang, L. D. Ramsay, A. Iwata, Y. Wang, W. Nelson, A. D. Farmer, P. M. Gaur, C. Soderlund, R. V. Penmetsa, C. Xu, A. K. Bharti, W. He, P. Winter, S. Zhao, J. K. Hane, N. Carrasquilla-Garcia, J. A. Condie, H. D. Upadhyaya, M. C. Luo, M. Thudi, C. L. Gowda, N. P. Singh, J. Lichtenzveig, K. K. Gali, J. Rubio, N. Nadarajan, J. Dolezel, K. C. Bansal, X. Xu, D. Edwards, G. Zhang, G. Kahl, J. Gil, K. B. Singh, S. K. Datta, S. A. Jackson, J. Wang & D. R. Cook 2013. Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nat Biotechnol,* 31**,** 240-6.

Rajeev K Varshney, Wenbin Chen, Yupeng Li, Arvind K Bharti, Rachit K Saxena, Jessica A Schlueter, Mark Ta Donoghue, Sarwar Azam, Guangyi Fan & Adam M Whaley 2012. Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. *Nat Biotechnol,* 30**,** 83-89.

S. F. Vaughn, T. A. Isbell, D. Weisleder & M. A. Berhow 2005. Biofumigant compounds released by field pennycress (Thlaspi arvense) seedmeal. *J Chem Ecol,* 31**,** 167-77.

S. Vaughn, T. Isbell, D. Weisleder & M. Berhow 2004. Biofumigation Potential of Field Pennycress (Thlaspi arvense) Seedmeal. *Hortscience,* 39**,** 745-745.

Steven F Vaughn, Debra E Palmquist, Sandra M Duval & Mark A Berhow 2009. Herbicidal activity of glucosinolate-containing seedmeals.

J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. Mckusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K.

Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. Mccawley, T. Mcintosh, I. Mcmullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. Mcdaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh & X. Zhu 2001. The sequence of the human genome. *Science,* 291**,** 1304-51.

Johann Vollmann & Istvan Rajcan 2009. *Oil Crops*, Springer Science & Business Media.

Doris Wagner, Robert Wm Sablowski & Elliot M Meyerowitz 1999. Transcriptional activation of APETALA1 by LEAFY. *Science,* 285**,** 582-584.

K. Wang, Z. Wang, F. Li, W. Ye, J. Wang, G. Song, Z. Yue, L. Cong, H. Shang, S. Zhu, C. Zou, Q. Li, Y. Yuan, C. Lu, H. Wei, C. Gou, Z. Zheng, Y. Yin, X. Zhang, K.

Liu, B. Wang, C. Song, N. Shi, R. J. Kohel, R. G. Percy, J. Z. Yu, Y. X. Zhu, J. Wang & S. Yu 2012. The draft genome of a diploid cotton Gossypium raimondii. *Nat Genet,* 44**,** 1098-103.

L. Wang, S. Yu, C. Tong, Y. Zhao, Y. Liu, C. Song, Y. Zhang, X. Zhang, Y. Wang, W. Hua, D. Li, D. Li, F. Li, J. Yu, C. Xu, X. Han, S. Huang, S. Tai, J. Wang, X. Xu, Y. Li, S. Liu, R. K. Varshney, J. Wang & X. Zhang 2014. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol,* 15**,** R39.

S.I. Warwick, Francis, A., Susko, D.J. 2002 The biology of Canadian weeds. 9. Thlaspi arvense L. (updated) *Canadian Journal of Plant Science,* 82**,** 803-823.

Detlef Weigel, John Alvarez, David R Smyth, Martin F Yanofsky & Elliot M Meyerowitz 1992. LEAFY controls floral meristem identity in Arabidopsis. *Cell,* 69**,** 843-859.

Ruth N Wilson, John W Heckman & Chris R Somerville 1992. Gibberellin is required for flowering in Arabidopsis thaliana under short days. *Plant Physiol,* 100**,** 403-408.

G. Wu, M. Y. Park, S. R. Conway, J. W. Wang, D. Weigel & R. S. Poethig 2009. The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell,* 138**,** 750-9.

H. J. Wu, Z. Zhang, J. Y. Wang, D. H. Oh, M. Dassanayake, B. Liu, Q. Huang, H. X. Sun, R. Xia, Y. Wu, Y. N. Wang, Z. Yang, Y. Liu, W. Zhang, H. Zhang, J. Chu, C. Yan, S. Fang, J. Zhang, Y. Wang, F. Zhang, G. Wang, S. Y. Lee, J. M. Cheeseman, B. Yang, B. Li, J. Min, L. Yang, J. Wang, C. Chu, S. Y. Chen, H. J. Bohnert, J. K. Zhu, X. J. Wang & Q. Xie 2012. Insights into salt tolerance from the genome of Thellungiella salsuginea. *Proc Natl Acad Sci U S A,* 109**,** 12219-24.

L. Yang, S. R. Conway & R. S. Poethig 2011. Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. *Development,* 138**,** 245-9.

Sha Yu, Vinicius C Galvão, Yan-Chun Zhang, Daniel Horrer, Tian-Qi Zhang, Yan-Hong Hao, Yu-Qi Feng, Shui Wang, Markus Schmid & Jia-Wei Wang 2012. Gibberellin regulates the Arabidopsis floral transition through miR156-targeted SQUAMOSA PROMOTER BINDING–LIKE transcription factors. *The Plant Cell Online,* 24**,** 3320-3332.

Jan Ad Zeevaart 1976. Physiology of flower formation. *Annual Review of Plant Physiology,* 27**,** 321-348.

Xiaoyu Zhang, Sophie Germann, Bartlomiej J Blus, Sepideh Khorasanizadeh, Valerie Gaudin & Steven E Jacobsen 2007. The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat Struct Mol Biol,* 14**,** 869-871.

N. Zhou, S. J. Robinson, T. Huebert, N. J. Bate & I. A. Parkin 2007. Comparative genome organization reveals a single copy of CBF in the freezing tolerant crucifer Thlaspi arvense. *Plant Mol Biol,* 65**,** 693-705.

CHAPTER 2: *De novo* assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock

STATEMENT ON PREVIOUSLY PUBLISHED MATERIAL

The work described in this chapter was published in The Plant Journal (2013, Volume 75, Issue 6, pages 1028-1038, DOI: 10.1111/tpj.12267). This article is available under the terms of the Creative Commons Attribution Non-Commercial License and permits the non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited. Permission is not required for non-commercial reuse. Included in this chapter are references to large electronic supplementary datasets and data files included as a part of this publication that will be referenced as Dorn et al. (2013).

SUMMARY

Field pennycress (*Thlaspi arvense* L.) has potential as an oilseed crop that can be grown during fall and winter months and harvested in the early spring as a biodiesel feedstock in the Midwestern United States. There has been little agronomic improvement in pennycress through traditional breeding. The recent advances in genomic technologies allow for the development of genomic tools to enable rapid improvements to be made through genomic assisted breeding. Here we report the first annotated transcriptome for pennycress. RNA was isolated from representative plant tissues, and 203 million unique Illumina RNAseq reads were produced and used in the transcriptome assembly. The draft transcriptome assembly consists of 33,873 contigs with an average length of 1,242 base pairs. A global comparison of homology between the pennycress and *Arabidopsis* transcriptomes, along with four other Brassicaceae species, revealed a high level of global sequence conservation within the family. The final assembly was functionally annotated, which allowed for the identification of putative genes controlling important agronomic traits such as flowering and glucosinolate metabolism. The identification of these genes leads to testable hypotheses concerning their conserved function and to rational strategies to improve agronomic properties in pennycress. Future work to characterize isoform variation between diverse pennycress lines and develop a draft genome sequence for pennycress will further direct trait improvement.

INTRODUCTION

Plant-derived biofuels have the potential to reduce carbon emissions and provide a renewable source of energy (Hill *et al.*, 2006). The replacement of fossil fuels with those derived from plant biomass or oilseeds holds promise to slow global climate change due to anthropologic release of greenhouse gases.

While initial genomics efforts, such as the human genome and the *Arabidopsis thaliana* genome projects, required massive teams and a several billion-dollar investment, the increased access and affordability of next generation sequencing (NGS) resources (i.e. genomics and transcriptomics) provides new approaches for rapidly studying new species (Lander *et al.*, 2001, Venter *et al.*, 2001, Hamilton and Buell, 2012, S. Kaul, 2000). Many different plant species are being considered not only as new sources of biofuel, but also as components of the landscape that improve the environment. Some of these species have only recently been removed from the wild and are in need of significant improvements to remove and replace weedy traits with those that improve their agronomic properties. The application of NGS resources in the development of candidate species should allow for rapid advancement and improvement in these species (Varshney *et al.*).

Biofuel crop species that do not displace land for food production or encourage the destruction of natural lands are especially attractive as alternatives to the biofuel standard, corn derived ethanol (Tilman *et al.*, 2009, Fargione *et al.*, 2008). In addition, new species that provide ecosystem services to reduce the effects of large-scale intensive farming are essential to assure food security. This is especially important in the

Midwestern United States where large portions of the land dedicated to agriculture are left barren nearly half the year, from the time of harvest until the next crop establishment. Planting winter annual crops following the fall harvest has been shown to alleviate soil degradation, topsoil loss through erosion and nutrient runoff, help prevent water pollution by scavenging excess nitrogen from the soil, and limit spring weed growth (Dabney *et al.*, 2001, Snapp *et al.*, 2005).

Pennycress is especially attractive because it can provide a winter cover that uses excess nitrogen and slows soil erosion, provides a spring cover that suppresses weeds, and yields a harvestable oilseed. The combination of these traits makes pennycress one of the best biofuel candidate plant species. Pennycress can be harvested in the spring using conventional machinery and yield up to 1345 kg seed/ha (Mitich, 1996, Best and Mcintyre, 1975). Pennycress seeds are high in oils that can be easily converted into biodiesel (Boateng *et al.*, 2010, Hojilla-Evangelista, 2013, Isbell and Cermak, 2012, Moser *et al.*, 2009a, Moser *et al.*, 2009b). A recent study showed that pennycress can be planted as a winter cover crop after corn in the fall and harvested in the spring without impeding subsequent soybean cultivation, or dramatically impacting soybean: yield, protein content, oil quantity and oil quality (Phippen and Phippen, 2012). Thus, implementation of pennycress will not require any new land or displace traditional food crops. A recent life cycle assessment indicates pennycress-derived fuels could qualify as advanced biofuels by the EPA Renewable Fuels Standard (Fan, 2013).

While the inherent agronomic properties of pennycress are tractable, efforts are needed to maximize oilseed yield, content and composition, while reducing seed

dormancy and glucosinolates. Previous studies compared various genetic aspects of pennycress to its close relative *Thlaspi caerulescens*, which hyperaccumulates zinc and cadmium (Hammond *et al.*, 2006, Milner and Kochian, 2008). The analysis of over 600 pennycress ESTs revealed a close relationship between pennycress and *Arabidopsis* (Sharma *et al.*, 2007). The limited genetic divergence between *Arabidopsis* and its wild relatives, such as pennycress, will facilitate the translation of the basic knowledge gleaned from years of *Arabidopsis* research.

Here we report the sequencing, de novo assembly, and annotation of the transcriptome of several pennycress tissues, including roots, leaves, shoots, flowers, and seed pods. The draft transcriptome consists of 33,873 transcripts. Comparative analyses with other Brassicaceae species showed a high degree of conservation, which serves as a validation of the assembly. The comparative analysis to *Arabidopsis thaliana* allowed us to identify many pennycress orthologs likely responsible for controlling flowering time and glucosinolate metabolism. The pennycress dataset, along with further development of genomic tools and germplasm resources should provide unprecedented tools for beginning a breeding program.

Materials and Methods

Plant growth conditions and RNA extraction

Seed from a small natural population of *Thlaspi arvense* L. was collected near Coates, MN. Seeds were planted in moist Berger BM2 germination mix (Berger Inc., www.berger.ca) stratified for 7 days at 4º C, and then placed in a 21º C growth chamber. Individual seedlings were transferred to 4-inch pots after 2 weeks and were grown under banks of 6400K T5 fluorescent lights with a 16h/8h day/night cycle. To initiate flowering, 6-week old plants with established rosettes were covered and transferred to a 4º C cold room for 14-29 days in the dark. After vernalization, plants were transferred back to the growth chamber and grown under 400W metal halide bulbs with a 16h/8h day/night cycle. Roots, hypocotyls, cotyledons, and young leaves were obtained by planting sterilized seed on 1X Murashige and Skoog medium with 0.8% agar. Seed was stratified at 4º C for 3 days, and then grown for 7 days in constant light under standard T12 fluorescent bulbs.

RNA was extracted from 1.) roots from 12 seedlings grown on MS plates 2.) hypocotyls, cotyledons, young meristems, and first leaves from 12 seedlings grown on MS plates 3.) 4 new leaves from each of two 120 day old unvernalized plants 4.) aerial leaves and stems from 128 day old flowering plants and 5.) flowers and seed pods from 128 day old flowering plants. RNA was purified using the RNeasy Plant Mini Kit (Qiagen, http://www.qiagen.com) following the manufacturer's recommendations. Following the initial total RNA extraction, samples were treated with Ambion TURBO DNase (Life Technologies, http://www.lifetechnologies.com) following the

manufacturer's recommended protocol, then immediately followed by the RNA cleanup procedure from the Qiagen RNeasy kit.

High Throughput RNA sequencing and de novo assembly

A pooled sample containing equal amounts of purified total RNA from each of the 5 tissue samples were submitted to the University of Minnesota Biomedical Genomics Center for sequencing. RNA was subjected to quality control using the Invitrogen RiboGreen RNA assay (Life Technologies) and RNA integrity was analyzed using capillary electrophoresis on the Agilent BioAnalyzer 2100 (Agilent Technologies, http://www.agilent.com). Polyadentylated RNA was selected using oligo dT purification and reverse transcribed to cDNA. cDNA was fragmented, blunt ended, and ligated to the Illumina TruSeq Adaptor Index 3 (Illumina, Inc., http://www.illumina.com). The library was size selected for an insert size of 200 bp and quantified using the Invitrogen PicoGreen dsDNA assay (Life Technologies). The pooled RNA sample was sequenced using the Illumina HiSeq 2000 platform using 100 cycle, paired end reads, producing 374M reads above Q30. Read pairs had an average insert size of 200 bp. Duplicate reads were removed and the first 10 nucleotides were trimmed from the 5' end of each read using the corresponding tools in CLC Genomics Workbench 5.5 (CLC Bio, http://www.clcbio.com). The additional trimming parameters were: removal of low quality sequence-limit = 0.05; removal of ambiguous nucleotides: maximal 2 nucleotides allowed; removal of terminal nucleotides: 10 nucleotides from the 5' end; removal of adaptor sequence: Illumina TruSeq Indexed Adapter 3 and Universal Adapter.

Reads were de novo assembled into contigs using the CLC Genomics Workbench 5.5 de novo assembly tool. A series of independent assemblies were conducted to analyze the effects of varying the de novo assembly parameters. Assemblies were conducted with

varying word size (18, 24, 30, 36, 40, 46, 52, 58, 64), and with length fractions of 0.7 and 0.95 for each. An additional 23 assemblies were conducted varying outside of these parameters, with a total of 41 assemblies conducted. The remaining assembly parameters were: auto bubble size: yes; minimum contig length: 300 bp; perform scaffolding: yes; mismatch cost: 3; insertion cost: 3; deletion cost: 3; update contigs: yes. Functional annotations and gene ontologies were assigned to each assembled contig from the final assembly using Blast2GO using the following parameters: BLASTx against NCBI non-redundant protein database, minimum E-value=1.0E-3, and reporting the top 20 hits. Comparative BLAST searches against *Arabidopsis* were conducted using CLC Genomics Workbench BLAST function, using sequences obtained from the TAIR10 release (www.arabidopsis.org) of the *Arabidopsis* transcriptome and proteome (Lamesch *et al.*, 2012). Sequences for *Arabidopsis lyrata* (Hu *et al.*, 2011), *Capsella rubella* (Slotte, 2013), *Brassica rapa* (Wang *et al.*, 2011b), and *Thellungiella halophila* were obtained from Phytozome v9.1 (www.phytozome.net). Further statistical analysis and figures were prepared using R (Team, 2008). The final assembly described here has been submitted to DDBJ/EMBL/GenBank under the accession GAKE01000000. The complete, annotated FASTA file is available at http://www.cbs.umn.edu/lab/marks/pennycress/transcriptome.

Results

Generation of RNAseq reads and de novo assembly

RNA was isolated from five different pennycress tissue types (see Experimental Procedures) and was sequenced on a single lane of the Illumina HiSeq 2000 platform (2x100 bp, insert size 200bp). A total of 374,725,460 reads with an average quality score >Q30 were obtained. After removing duplicate reads, trimming adaptors, and filtering for low quality sequences, a total of 203,003,444 unique, clean reads were obtained with an average length of 87.6 bp. The full, unfiltered short read dataset was deposited to the NCBI Short Read Archive under accession SRR802670.

The filtered reads were de novo assembled using the CLC Genomics Workbench software package. The effect of varying de novo assembly parameters was examined by performing 41 separate assemblies. Word size (k-mer), match length (the percent length of a read needed to match the initial contig build), and match percent (the percent sequence identity needed to match a read to the initial contig build) were varied and the effect on various assembly statistics was examined. Regardless of match length and percent, smaller word sizes produced assemblies with smaller average contig lengths, but assembled a few contigs that were significantly larger (16-18 kb) compared to assemblies with word sizes of 52 or larger (15 kb). These large contigs are likely misassembled because each contained sequences similar to multiple *Arabidopsis* genes. The assemblies created with 95% match length and 95% match percent parameters were chosen for further comparison of how word size affected the relative assembly quality. Increasing the word size caused the percent of reads used in the final assembly and average contig

length to increase, while decreasing the number of contigs assembled. The assemblies with larger word sizes also had a higher percentage of contigs built with BLASTx hits to at least one *Arabidopsis* peptide. The statistics regarding the assembly optimization and BLAST results for the assembly with word size 64, 95% match length, and 95% match percent are shown in Supporting Table S1 of Dorn et al. (2013).

The assembly with a word size of 64, 95% match length, and 95% match percent was chosen for further analysis and annotation due to the high assembly statistics, while yielding a high proportion of assembled transcripts with significant matches to *Arabidopsis* genes. A summary of sequencing reads and assembly statistics is shown in Table 2-1. The sequences from this assembly have been submitted to NCBI under Bioproject ID PRJNA183631. A total of 33,874 contigs were built using these parameters. This includes a spiked phiX174 genome sequence that serves as a sequencing control, which was subsequently removed from the final assembly and total assembly length. The mean contig length was 1,242 bp, with a minimum and maximum contig length of 215 bp and 15,516 bp, respectively. The size distribution of contig lengths is shown in Figure 2-1. The N50 statistic was 1,729 bp, meaning all contigs this size or larger encompassed at 50% of the total 42,069,800 bp assembly length.

Annotation and functional characterization of pennycress transcripts

The pennycress transcriptome sequences were annotated using Blast2GO Pro (Conesa *et al.*, 2005). The database used in this analysis only contains well-characterized sequences and does not include sequences from resources such as newly assembled draft genomes. The taxonomic distribution from this analysis was examined (Figure 2-2). Over 20,000 transcripts had top hits to an *Arabidopsis* species, among them, 11,936 transcripts had a top hit to *Arabidopsis thaliana*, while 11,364 transcripts had top hits to *Arabidopsis lyrata.* Nearly 75% of the pennycress transcripts had top BLAST hits within the Brassicaceae family. The sister genus, Brassica, had a large proportion of these top hits: *Brassica rapa* (283), *Brassica napus* (233), and *Brassica oleraceae* (164). 713 transcripts had top matches to other plant sequences outside the Brassicaceae. 24% of the transcripts had top BLAST hits to non-plant sequences or no significantly similar sequence in the public database, with ~4% and ~19%, respectively. The complete dataset from the final assembly including annotations and associated GO terms from this analysis are available in Supporting Table S2 of Dorn et al. (2013). The complete file containing the annotated pennycress transcript sequences is available at pennycress.umn.edu.

Annotations and associated cellular component (CC), molecular function (MF), and biological process (BP) gene ontology (GO) terms were produced for each pennycress transcript. 27,456 transcripts had a significant hit in the public databases (BLAST E-value << 0.01). 26,797 transcripts received at least one GO annotation. The most highly represented BP GO terms were oxidation-reduction process (1403 transcripts) and DNA-dependent regulation of transcription (1255 transcripts). GO terms

associated with response to cold (727 transcripts), the vegetative to reproductive phase transition of the meristem (462 transcripts), and the regulation of flower development (411) were also highly represented. The 50 most represented GO-terms are shown in Supporting Figure S1 of Dorn et al. (2013).

Comparative transcriptomics of the pennycress and to other Brassicaceae species

Previous molecular analyses of the Brassicaceae have divided the family into three basic lineages, recently reviewed in (Franzke *et al.*, 2011). *Thlaspi arvense* is a member of the expanded lineage 2, and is more closely related to *Thellungiella halophila* and other *Eutrema*/*Thellungiella* species than the *Brassica* species in lineage 2 (Figure 2-3). *Arabidopsis thaliana*, *Arabidopsis lyrata*, and *Capsella rubella* are members of lineage 1. To explore the relationship between pennycress and other Brassicaceae at the transcriptome level, we compared the assembled translated pennycress transcriptome to a peptide database derived from the sequenced genomes of *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Capsella rubella*, *Brassica rapa*, and *Thellungiella halophila*. A BLASTx comparison of the pennycress transcriptome to this peptide database showed that 16,298 of the 33,873 pennycress contigs had significant ($e \leq 0.05$) top hits to *T. halophila* (Figure 2-4). *B. rapa* had the next highest number of top hits (4,972), and the Lineage 1 species having approximately 3,000 each. A BLASTx comparison of the remaining sequences without significant hits to one of the five Brassicaceae species revealed that 3,386 sequences had no significant hit in the NCBI non-redundant (nr) peptide database. This BLAST search returned 779 pennycress contigs with significant hits in the nr peptide database, including 424 fungi. Many of these fungal hits (273) were fungal plant pathogens, including *Fusarium*, *Pyrenophora*, *Phaeosphaeria*, *Leptosphaeria*, and *Bipolaris* (Supporting Table S3 of Dorn et al. (2013). These fungal transcripts were left in the assembly as the association between pennycress and these fungi may be informative in future analyses.

To examine the degree of conservation between pennycress and other sequenced Brassicaceae species, five pairwise tBLASTn comparisons were made between pennycress and each of the five Brassicaceae species (Figure 2-5). *T. halophila* had the highest number of sequences with significant hits to the pennycress database (e ≤ 0.05 and ≥70% positive match percent), along with the greatest proportion of peptides with significant matches (24,411/29,284). All of these five species had at least 72% of their proteins significantly represented in the pennycress database. All five Brassicaceae genomes share 14,677 of the pennycress transcripts (e ≤ 0.05 and ≥70% positive percent). An additional 4,547 sequences were shared between pennycress and at least one of the other Brassicaceae species. The tBLASTn results from this analysis are available in Supporting Table S4 of Dorn et al. (2013). A global view of the top pennycress transcript and the similarity to each *Arabidopsis thaliana* peptide (primary transcripts only) is shown in Figure 2-6. 14,186 of the 27,416 *Arabidopsis* loci had transcripts with >70% similarity and >70% coverage in the pennycress transcriptome.

To more closely examine the level of global sequence conservation between pennycress and *Arabidopsis thaliana*, we further examined a BLASTx comparison of the pennycress transcriptome assembly to the *Arabidopsis* TAIR10 peptide database (primary transcripts only). The relative homology of each predicted peptide to the most similar *Arabidopsis* protein was measured by the percent of positive sequence similarity (Figure 2-7) and percent coverage (Figure 2-8). A smooth scatter plot representing the percent similarity and percent coverage for each pennycress sequence compared to the closest *Arabidopsis* peptide sequence is shown in Figure 2-9. A large proportion (>85%) of

transcripts possess at least 70% similarity to an *Arabidopsis* protein. A total of 16,556 pennycress transcripts had at least one match to an *Arabidopsis* gene with >70% similarity / >70% coverage (Figure 2-9, boxes). 4,846 pennycress transcripts were ≥95% similarity and coverage, 9,685 transcripts were between 80-95% similarity and coverage, while 2,025 transcripts were between 70-80% similarity and coverage. 17,317 transcripts were <70% similarity and coverage. 4,783 transcripts lacked a significant BLASTx hit (e ≤ 0.05) to an *Arabidopsis* peptide.

Identification of candidate pennycress genes controlling flowering time and glucosinolate levels

The close evolutionary relationship between pennycress and *Arabidopsis* enabled the identification of pennycress orthologs likely responsible for controlling important agronomic traits like time to flower and glucosinolate metabolism. For each pennycress transcript, the top 20 BLASTx hit against the *Arabidopsis* peptide database was mined for hits to *Arabidopsis* genes known to control these traits. For these transcripts, the longest theoretical translation was obtained to explore protein sequence conservation. The nucleotide sequences and predicted peptides for each sequence, along with the amino acid alignment to their respective *Arabidopsis* homolog are shown in Supporting Data S1 of Dorn et al. (2013).

To investigate the conservation of the flowering time pathway in pennycress, we attempted to reconstruct the flowering pathway in *Arabidopsis* using predicted peptides from the transcriptome assembly (Jung and Müller, 2009). Full length (Met to Stop) predicted peptides with high homology to their respective *Arabidopsis* peptides for VRN1, VRN2, VIN3, LHP1, FLC, SVP, TSF, AGL19, SOC1, FT, LHY, TOC1, PRR7, PRR9, FKF1, and GI. Partial or incomplete matches were found for MSI1, FD, TFL1, AP1, FUL, CCA1, and CO. None of the pennycress transcripts had a top hit to the *Arabidopsis* FRI locus (AT4G00650), however, we found a 613aa predicted peptide similar to *B. napus* FRI.a [GenBank: AFA43306.1], previously shown to be a major determinant of flowering in rapeseed (Wang *et al.*, 2011a). A pennycress ortholog for LEAFY (AT5G61850) was also not found in the final assembly. However, truncated

transcripts similar to the *Arabidopsis* LEAFY sequence were detected in the assemblies created using word size 24, 30, and 46 (95% match length and percent). Putative orthologs of the FRIGIDA protein complex were also found (Choi *et al.*, 2011). Full length predicted peptides with high sequence similarity were found for EFS (AT1g77300), SUF4 (AT1G30970), FLX (AT2G30120), FES1 (AT2G33835), FRL1 (AT5G16320), and YAF9 (AT5G45600). No unique matches were found for TAF14 (AT2G18000). A reconstruction of the flowering time pathway using pennycress transcripts is shown in Figure 2-10.

Through comparative transcriptomics, we have identified orthologs of both myrosinases and specifier proteins likely responsible for controlling the break down of glucosinolates in pennycress. We performed a BLASTx comparison of the pennycress transcriptome assembly against the *Arabidopsis* proteome for the main myrosinases (TGG1-6 and PEN2) and specifier proteins (ESP/ESR and NSP1-5) responsible for glucosinolate breakdown in *Arabidopsis*. An *Arabidopsis* TGG1-ortholog was found whose longest open reading frame produced a predicted peptide with high sequence conservation compared to *Arabidopsis*. The top BLASTp hit against the non redundant protein database was a *Eutrema wasabi* myrosinase [GenBank:BAE16356]. The predicted peptide from another pennycress transcript was found to be highly similar to *Arabidopsis* TGG4. This predicted peptide had a top BLASTp hit to a myrosinase from *Amoracia rusticana* (horseradish) [AEZ01595.1]. An ortholog for the Arabidiopsis atypical myrosinase PEN2 was also found. The pennycress PEN2 predicted peptide has

95% sequence identity conservation to an unnamed protein product from *T. halophila* [GenBank:BAJ34425.1].

The conservation of specifier proteins was also examined. Three pennycress transcripts were found to have high homology to three *Arabidopsis* NSP genes. A full length predicted peptide similar to the AtNSP1 peptide, but was most similar to the nitrile-specifier protein from another member of the Brassicaceae, *Schouwia purpurea* [GenBank:AFP47629.1]. Another transcript was found which encoded for a 1073 aa predicted peptide with high sequence similarity to the C-terminal Kelch domain containg region of AtNSP4. The N-terminus of this peptide has high similarity to other AtNSPs. A third transcript was found to encode for a predicted peptide with high similarity to AtNSP5. We also identified orthologs to the glucosinolate transporters GTR1 and GTR2 in the pennycress transcriptome. These predicted peptides have significant homology to the *Arabidopsis* GTR1 (AT3G47960) and GTR2 (AT5G62680) peptides.

This comparison of assembly coverage is at least qualitatively indicative of expression level differences in the total RNA library. Albeit, directly comparing the non-normalized statistic of average coverage across transcripts for quantification is inappropriate, we observed many high coverage transcripts related to glucosinolate metabolism. Interestingly, we observed that among the 100 transcripts with the highest average coverage, 6 were similar to beta-glucosidase (2 transcripts), myrosinases (3 transcripts), and myrosinase-binding protein (1 transcript). The remaining 94 transcripts in this group could be considered ‘house keeping’ genes. Predictably, most of these transcripts are involved in photosynthetic processes. It remains unknown whether the

high levels of glucosinolates and glucosinolate byproducts in pennycress is simply due to high expression of these myrosinates and/or specifier proteins, unique activity, unique hormonal regulation of activity or expression, or some combination of these.

Discussion

Comparative transcriptomics of pennycress and *Arabidopsis*

We have sequenced, assembled, and annotated the pennycress transcriptome. The draft transcriptome consists of 33,873 unique sequences, of which 27,442 were annotated with the Blast2GO pipeline. Of these transcripts, 35% were most similar to an *Arabidopsis thaliana* gene, and 74% had top hits in the Brassicaceae, indicating a high level of sequence conservation across the family. BLAST comparisons between pennycress and five other sequenced Brassicaceae species showed our pennycress transcriptome has good coverage of homologous sequences. These analyses are consistent with previous phylogenetic findings that pennycress is more closely related to *T. halophila* than to *Brassica* species (Franzke *et al.*, 2011).

The total transcriptome assembly length was over 42 Mbp. The pennycress genome (2n=14) is approximately 539 Mbp (Johnston, 2005, Hume *et al.*, 1995). Comparatively, the *Arabidopsis* genome (2n=10) is estimated to be 125 Mbp (Kaul *et al.*, 2000), with the latest genome annotation release (TAIR10) containing 33,602 genomic features, including 27,416 protein-coding genes. *Brassica rapa* has 41,174 protein coding genes with an average transcript/coding lengths of 2015/1172 bp (Wang *et al.*, 2011b). The number of genes identified here in pennycress, along with the estimated genome size, matches similar observations on total gene number in the *Arabidopsis* and *B. rapa* genomes.

Characterization of pennycress glucosinolate metabolism and translocation

Many plants in the order Brassicales produce high levels of glucosinolates and glucosinolate hydrolysis products, which are thought to serve a defensive function (Bones and Rossiter, 1996). Glucosinolates are one of the most highly characterized secondary metabolites in *Arabidopsis* (Wittstock and Burow, 2010). Myrosinases, also known as thioglucoside glucohydrolases, hydrolyze the glucosinolate, forming an intermediate aglycone. The aglycone is either spontaneously rearranged to form isothiocyanates, or converted to a simple nitrile, epithionitrile, or thiocyanate by specifier proteins. The levels of glucosinolates and glucosinolate byproducts have been characterized in pennycress (Warwick *et al.*, 2002, Kuchernig *et al.*, 2011), motivated by the characteristic 'garlic-like' odor of the species. This has led to another common name of this species: 'stinkweed'. A single thiocyanate-forming protein (TaTFP) has previously been identified and characterized in pennycress (Kuchernig *et al.*, 2011). Pennycress seed has also been investigated for its biofumigant properties-likely due to the high levels of glucosinolates in the seeds (Vaughn *et al.*, 2005). After oil is pressed from pennycress seed, the remaining presscake still has economic potential. Pennycress seedcake has high levels of protein (25%), which has the potential to serve as an animal feed supplement or use in industrial products (Selling *et al.*, 2013). However, the high levels glucosinolates, which in high levels can be toxic to animals, would prohibit this use (Vaughn *et al.*, 2005, Best and Mcintyre, 1975, Warwick *et al.*, 2002). Previous work in *Arabidopsis* identified key glucosinolate transporters responsible for translocating glucosinolates (Nour-Eldin *et al.*, 2012). The *Arabidopsis* double mutant *gtr1 gtr2* showed significantly reduced levels

of glucosinolates in seed. We predict loss of function mutations in the pennycress GTR-like genes identified here would cause a reduction in seed glucosinolate levels.

Genetics of flowering time in winter annual pennycress

The genetic mechanisms controlling the transition from vegetative to reproductive growth has been widely studied in *Arabidopsis* and other plant species (Simpson and Dean, 2002, Amasino, 2005, Kim *et al.*, 2009). In many species adapted to winter climates, a period of cold provided by overwintering is required to render plants competent to flower, a process known as vernalization. In many crucifer species there is natural variation in populations adapted to different climates. Much of this variation is attributed to the complex interaction of FRIGIDA (FRI), the FRIGIDA protein complex, and FLOWERING LOCUS C (FLC), which serve as the main response to vernalization (Choi *et al.*, 2011). The period of vernalization provided by winter epigenetically represses FLC (Michaels and Amasino, 2001, Sheldon *et al.*, 2000). This lifts transcriptional repression of FLC on FLOWERING LOCUS T (FT), a main integrator of environmental cues promoting flowering. 'Fast-cycling' lines of *Arabidopsis* contain a loss of function mutation in FRI (Gazzani, 2003, Johanson *et al.*, 2000).

Variation of FRI and FLC orthologs in *B. rapa* (Schranz *et al.*, 2002, Yuan *et al.*, 2009), *B.oleracea* (Irwin *et al.*, 2012), and *B. napus* (Wang *et al.*, 2011a, Tadege *et al.*, 2001) are associated with vernalization and flowering. Both 'early' and 'late' flowering lines of pennycress have been reported (Best and Mcintyre, 1976). Much like the fast cycling lines of *Arabidopsis*, the 'early' pennycress lines will flower without a period of vernalization, exhibiting a spring annual habit. The late flowering lines instead will grow for a period of time in the fall as a vegetative rosette, but not flower until the spring. The genetic differences between winter and spring annual pennycress lines was determined to

be caused by a single dominant allele (Mcintyre and Best, 1978). We predict that natural variation between spring and winter lines are due to mutations in FRI or FLC-like genes. In order for pennycress to be easily integrated as a winter cover crop throughout different climates, precise control of spring flowering time is needed. Perturbations of the flowering time pathway in cultivated species has served as an important tool for controlling flowering time through breeding and genetic modification (Jung and Müller, 2009). Our identification of the orthologous genes likely responsible for controlling flowering time will be a useful tool for making rapid improvements in the pennycress germplasm.

Considerations regarding de novo transcriptome assembly

Varying de novo assembly parameters using short read data has been shown to assemble unique transcripts corresponding to real genes (Zhao *et al.*, 2011). In this study, we chose a single assembly due to the assembly statistics and high number of transcripts with significant similarity to *Arabidopsis* peptides. The finding of a high number of potentially orthologous sequences in the pennycress and its relatives provides one validation of the pennycress assembly. However, different assembly programs and parameters can affect the assembly of both highly and lowly expressed transcripts (Gongora-Castillo and Buell, 2013, Zhao *et al.*, 2011). For example, large word sizes poorly assemble lowly expressed genes (Gruenheit *et al.*, 2012). Thus, it is not expected that any one assembly will truly represent the complete biological transcriptome. This was highlighted in the current analysis between pennycress and *Arabidopsis*. We predicted that a LEAFY-like ortholog should be represented in our RNA pools, but it was not assembled in the final assembly. Using smaller word sizes (24, 30, 46) did result in the assembly of LEAFY-like transcripts - see transcript sequences in Supporting Data S1 of Dorn et al. (2013). These transcripts had low coverage (7x average) with few mapped reads. Combined with the high number of reads used to create our final assembly (over 200 million), this indicates the pennycress LEAFY ortholog was expressed at low levels in our sample and likely not assembled in the final assembly due to the larger word size. In our optimization, smaller word sizes also resulted in the assembly of some obviously misassembled transcripts where multiple transcripts from unlinked genes were joined

together. These results further support the need for a full characterization of the potential changes caused by various de novo assembly parameters.

Future Perspectives

We have identified pennycress homologs likely responsible for controlling key agronomic traits like seed glucosinolate levels and flowering time, which are primary targets for future research in order to improve the pennycress germplasm. It should be straightforward to make improvements using reverse genetic approaches to identify inactive or altered alleles by using well established TILLING protocols (Kurowska *et al.*, 2011, McCallum *et al.*, 2000). Our on going sequencing of the pennycress genome will enable rapid screening of TILLING populations through NGS. In addition, the ability to make improvements using transgenic approaches to modify gene expression by overexpression or knockout down of endogenous gene expression or via expression of novel genes lies on the immediate horizon as we have found that pennycress is relatively easy to regenerate (unpublished observation). Pennycress has tremendous agronomic potential as a winter cover and new source of oilseeds. A recent report by the MIT Joint Program on Science and Policy of Global Change indicates the pennycress could be grown on over 40 million acres each year, yielding up to 6 billion gallons of oil that can be converted to biodiesel (Moser *et al.*, 2009a, Winchester, 2013). This represents approximately 15% of the 40 billion gallons of diesel consumed annually in the US. The recent advances in ‘omics-based’ technologies will allow the resources developed here to make rapid improvements to the pennycress germplasm.

Number of raw, unfiltered reads	374,725,460
Total length of reads, pre-filtering (bp)	37,472,546,000
Total length of reads, post-filtering (bp)	17,799,652,172
Number of trimmed, unique reads	203,003,444
Number of contigs	33,873
Average contig length (bp)	1,242
Minimum/maximum contig length (bp)	215/15,516
N50 (bp)	1,729
Total assembly length (bp)	42,069,800

Table 2-1 Illumina RNAseq reads and de novo assembly statistics

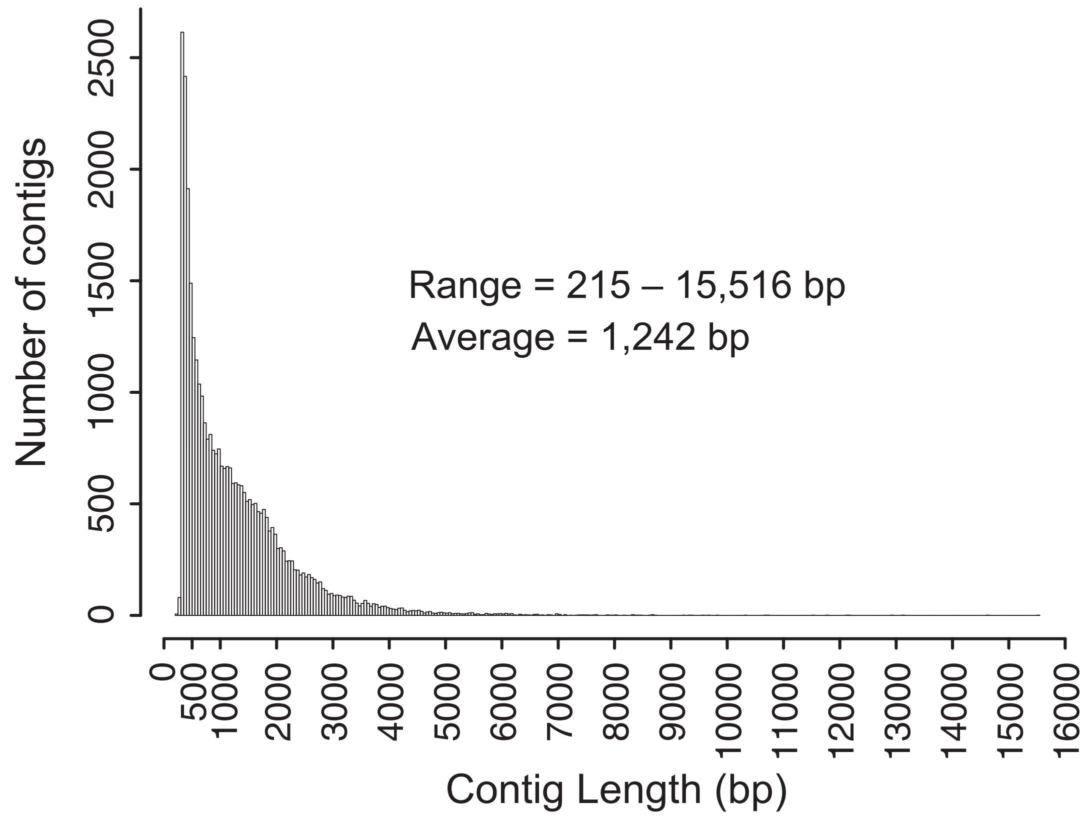


Figure 2-1 - Contig length distribution of assembled transcripts

Histogram of the length distribution of assembled contigs.

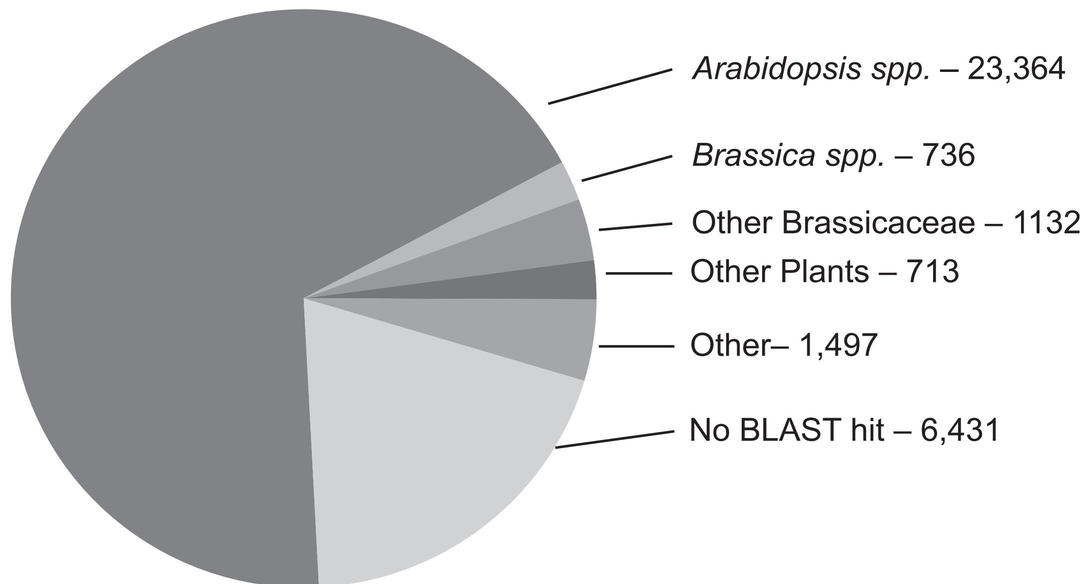


Figure 2-2 – Taxonomic distribution of top BLAST2GO hits

The taxonomic distribution of the top BLAST hits for each transcript in the de novo transcriptome assembly from Blast2GO. Only taxonomic data for the top BLAST result of each transcript is shown.

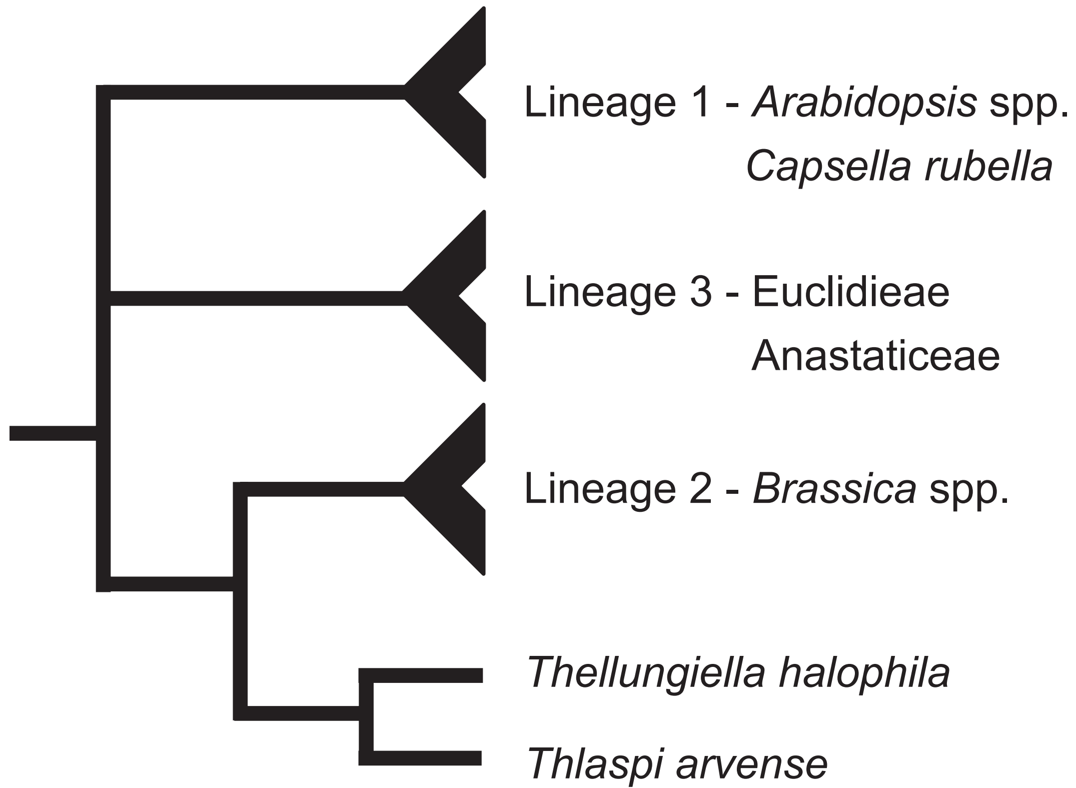


Figure 2-3 – Representation of Brassica family phylogenetic tree

Brassicaceae phylogeny, adapted from Beilstein et al. (2010) and Franzke et al. (2011).

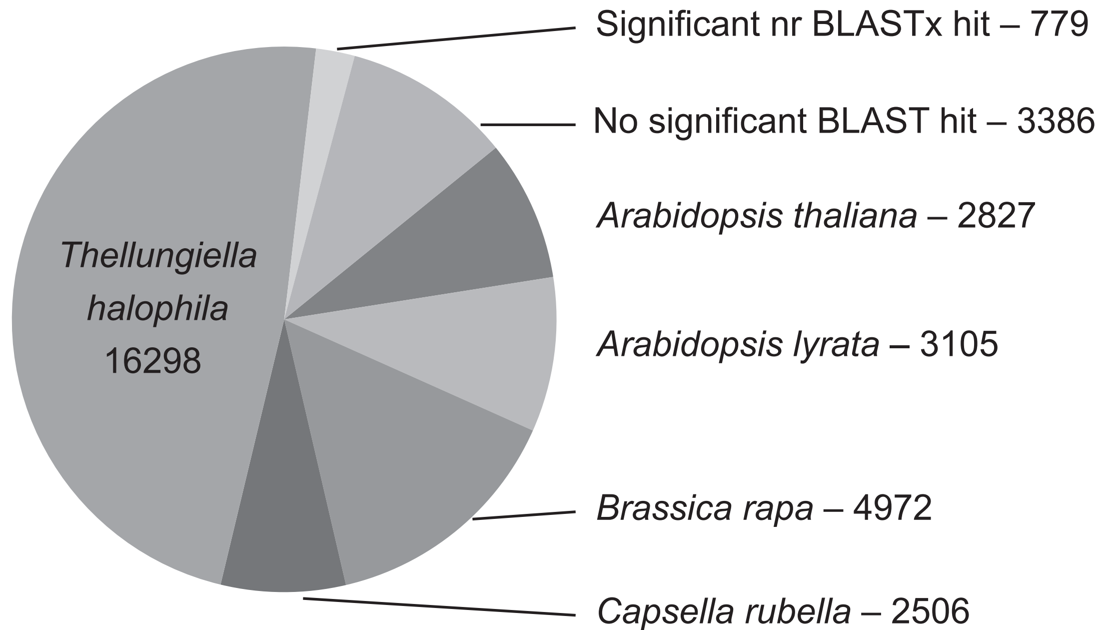


Figure 2-4 - Comparative BLASTx analysis of pennycress to five Brassicaceae species

BLASTx comparison of the pennycress transcriptome assembly to *A. thaliana*, *A. lyrata*, *B. rapa*, *C. rubella*, and *T. halophila*. The top BLAST hit (e ≤ 0.05) for each pennycress transcript to the five species is shown. Contigs without significant hits were then compared to the NCBI peptide non-redundant (nr) database.

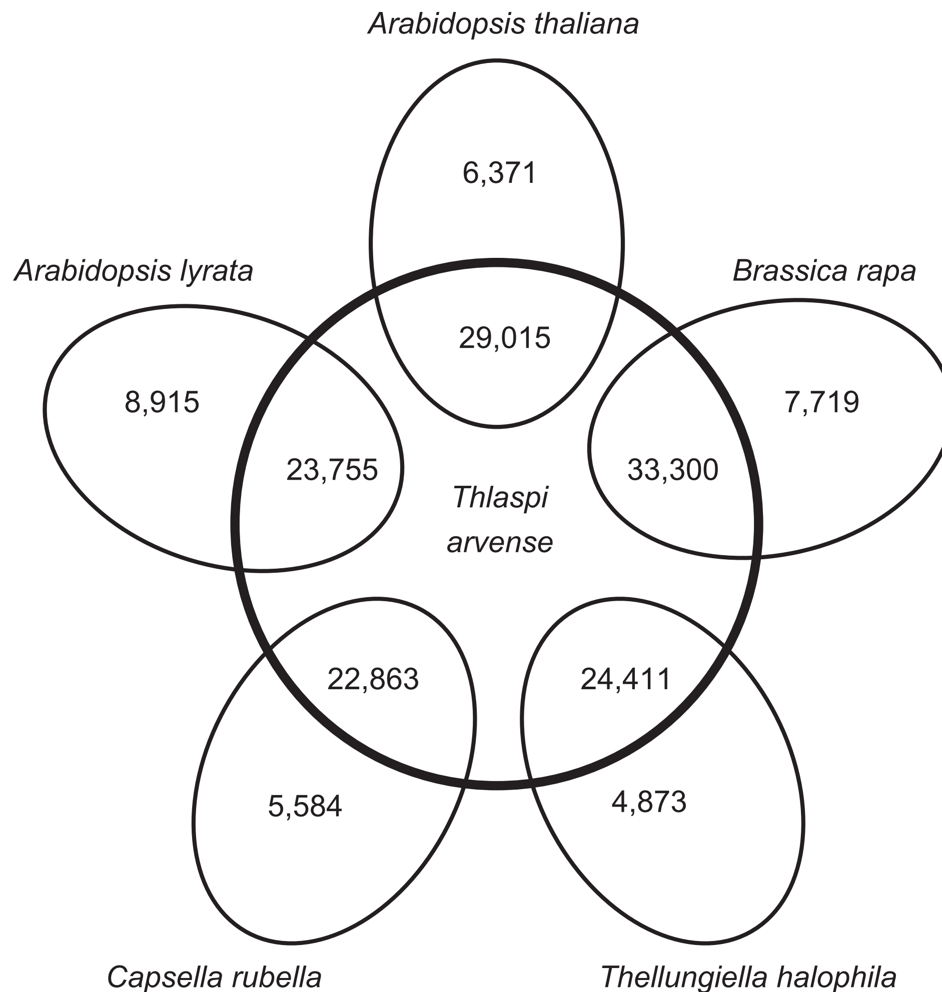
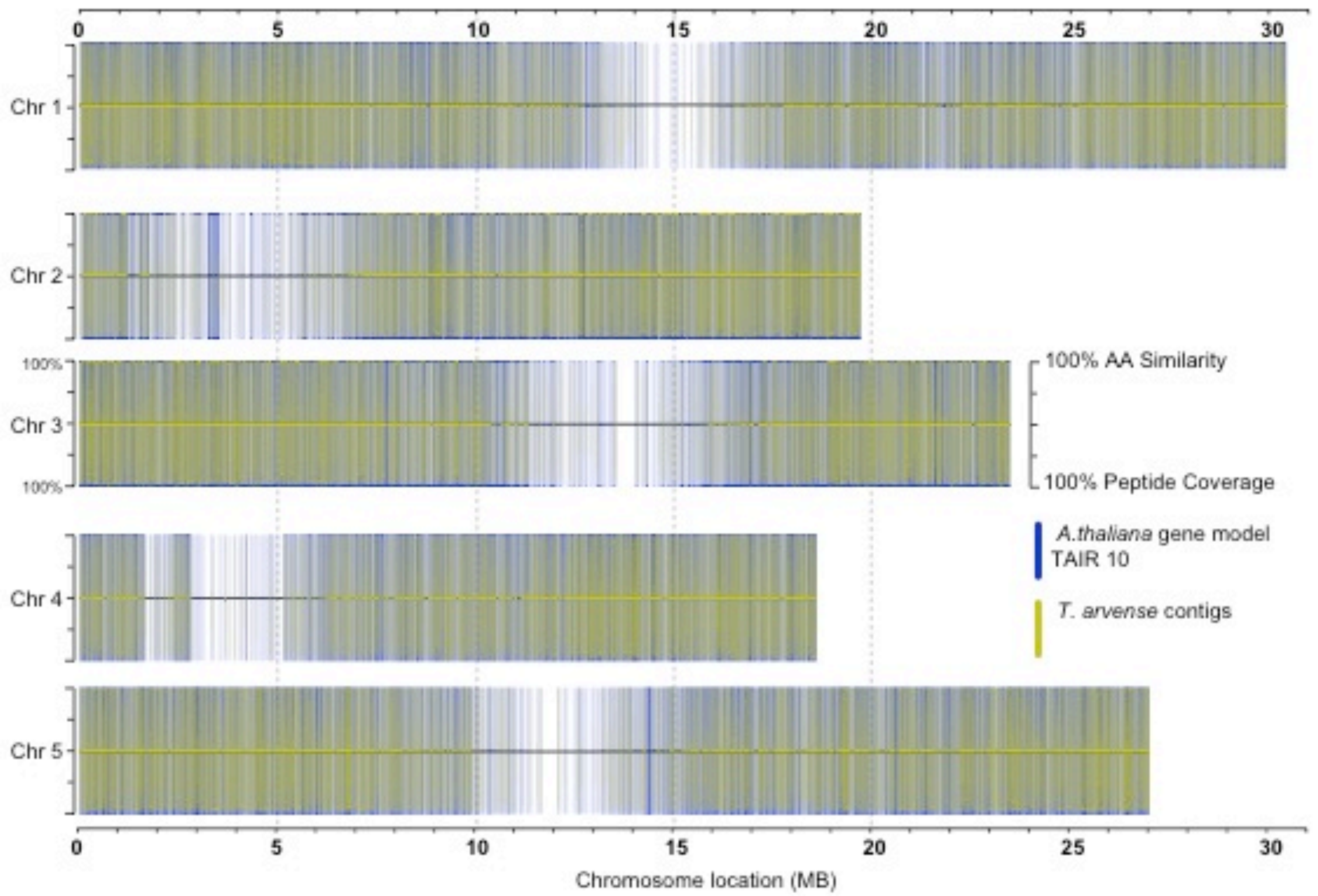


Figure 2-5 - Comparative tBLASTn analysis of pennycress to five Brassicaceae species

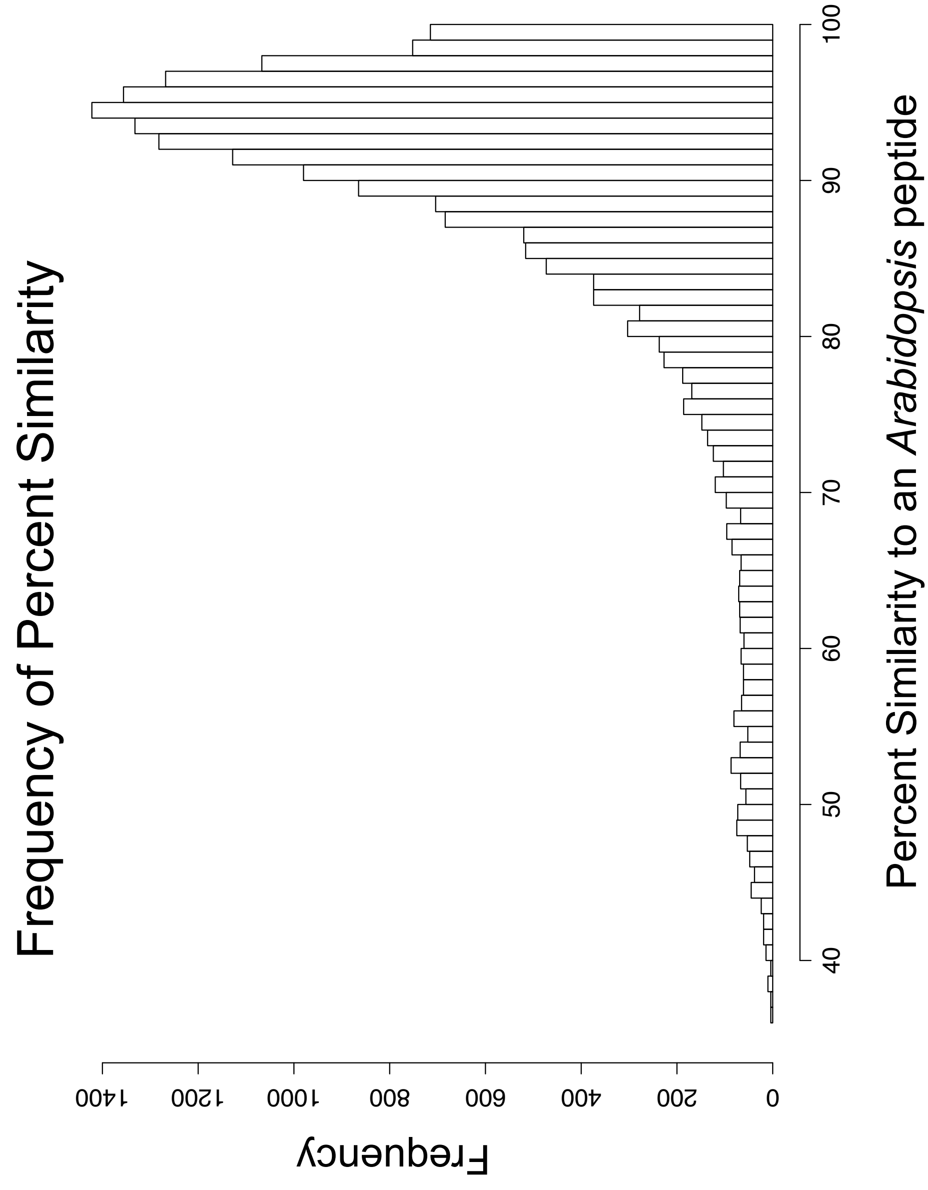
Five pairwise tBLASTn comparisons of Brassicaceae species to the pennycress transcriptome assembly. Sequences with significant homology ($e \leq 0.05$ and positive percent $\leq 70\%$) shared between the five Brassicaceae species and pennycress (*Thlaspi arvense*) are shown on the inner circle.



Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
100%
100%
0
5
10
15
20
25
30
Chromosome location (MB)
100% AA Similarity
100% Peptide Coverage
A.thaliana gene model TAIR 10
T. arvense contigs

Figure 2-6 Global comparison of the Arabidopsis and pennycress transcriptomes

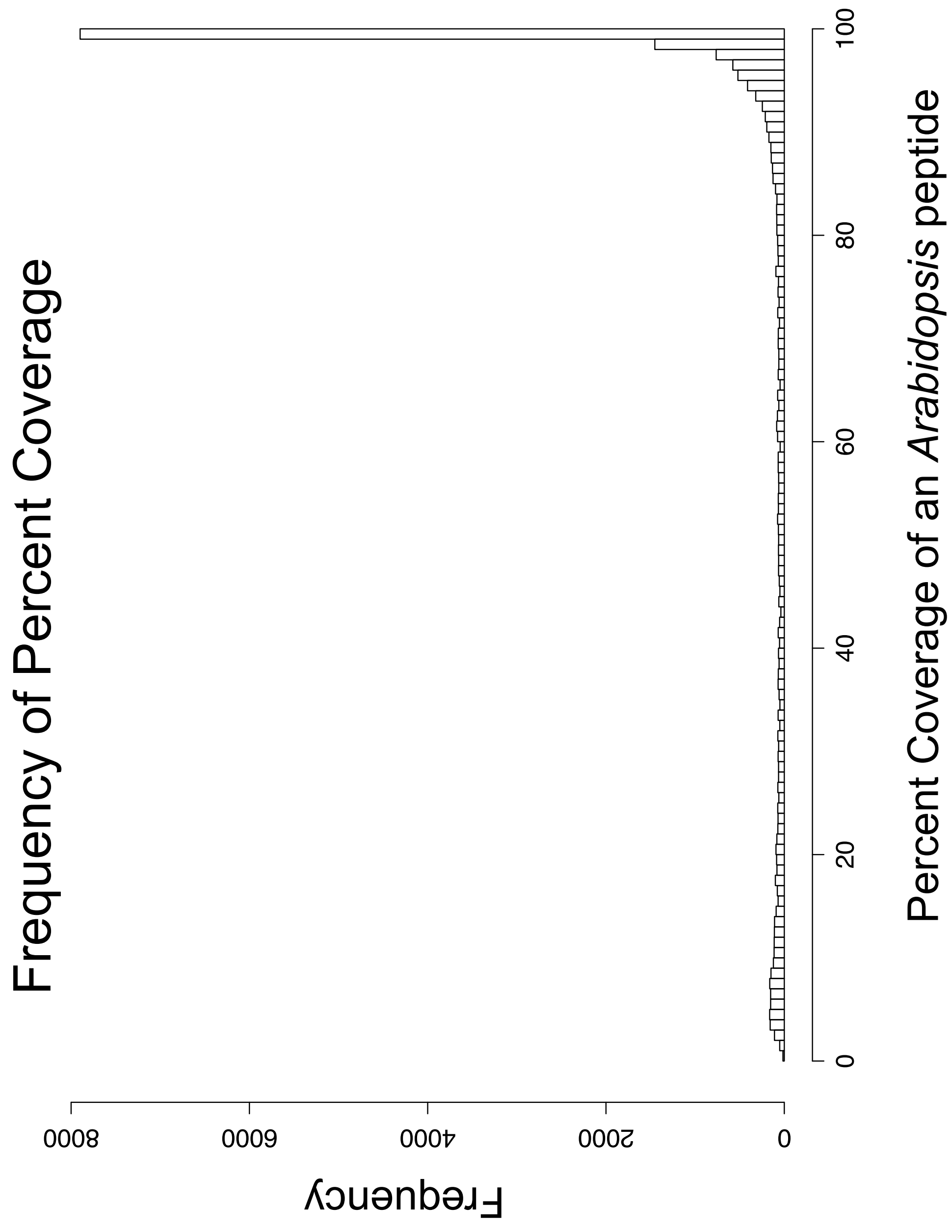
Global representation of the Arabidopsis transcriptome reconstructed using pennycress transcripts. Each of the five Arabidopsis chromosomes (Chr 1-5) is shown with each gene model relative to chromosomal position with dark bars. The percent of sequence similarity is shown in the positive y-axis, and percent of the Arabidopsis peptide covered with this similarity in the negative y-axis (light bars). White regions represent non-genic regions. The similarity and coverage of the most similar pennycress transcript is shown as a light bar for each Arabidopsis gene.



Frequency of Percent Similarity
Frequency
0
200
400
600
800
1000
1200
1400
40
50
60
70
80
90
100
Percent Similarity to an Arabidopsis peptide

Figure 2-7 - Similarity of pennycress transcripts to Arabidopsis genes

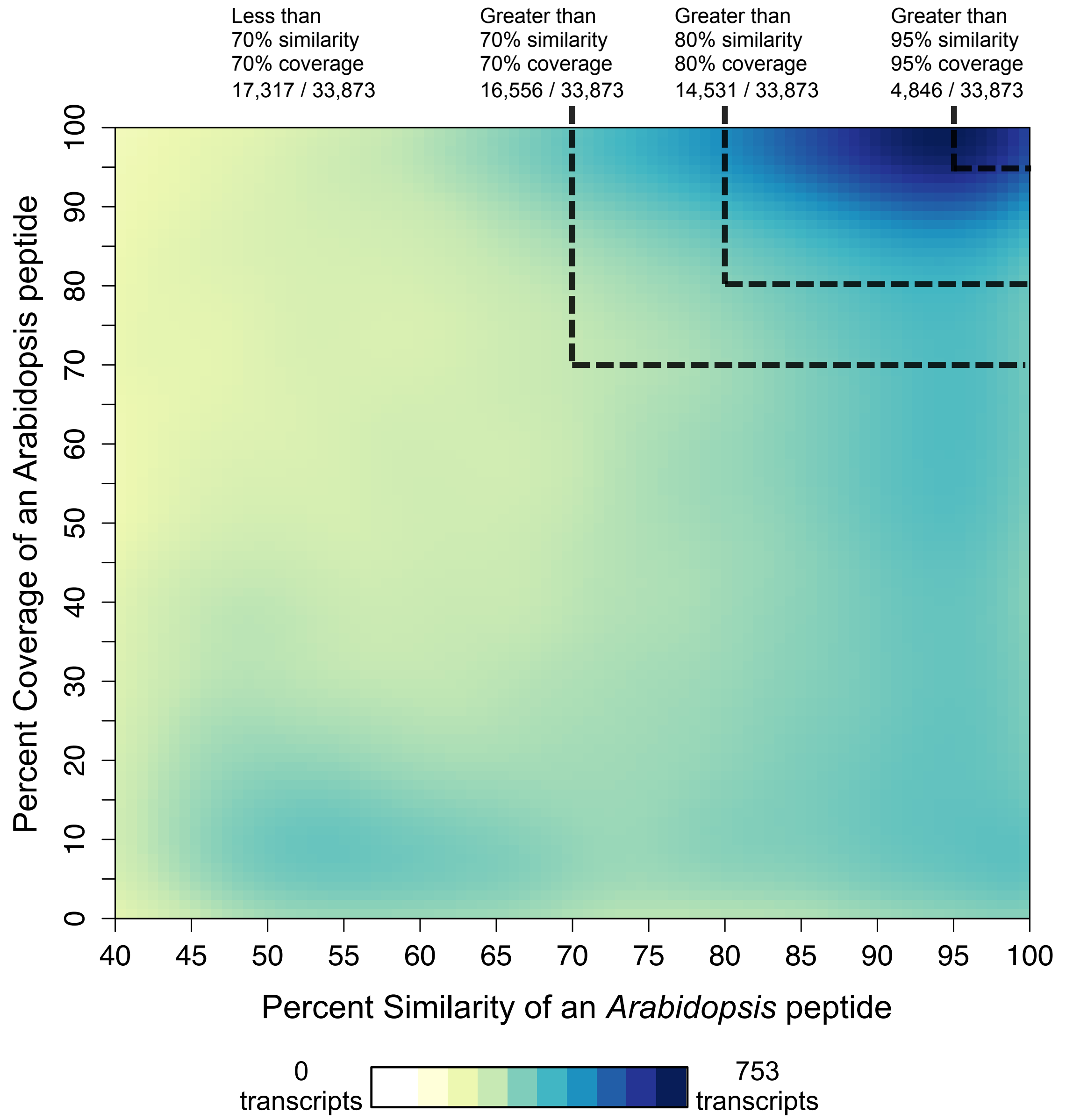
Histogram showing frequency vs. percent similarity (positive amino acid identity) of pennycress contigs to an *Arabidopsis* peptide from a global BLASTx comparison of the pennycress transcriptome the Arabidopsis predicted protein set.



Frequency of Percent Coverage
Frequency
0
2000
4000
6000
8000
0
20
40
60
80
100
Percent Coverage of an *Arabidopsis* peptide

Figure 2-8 - Coverage of pennycress transcripts to Arabidopsis genes

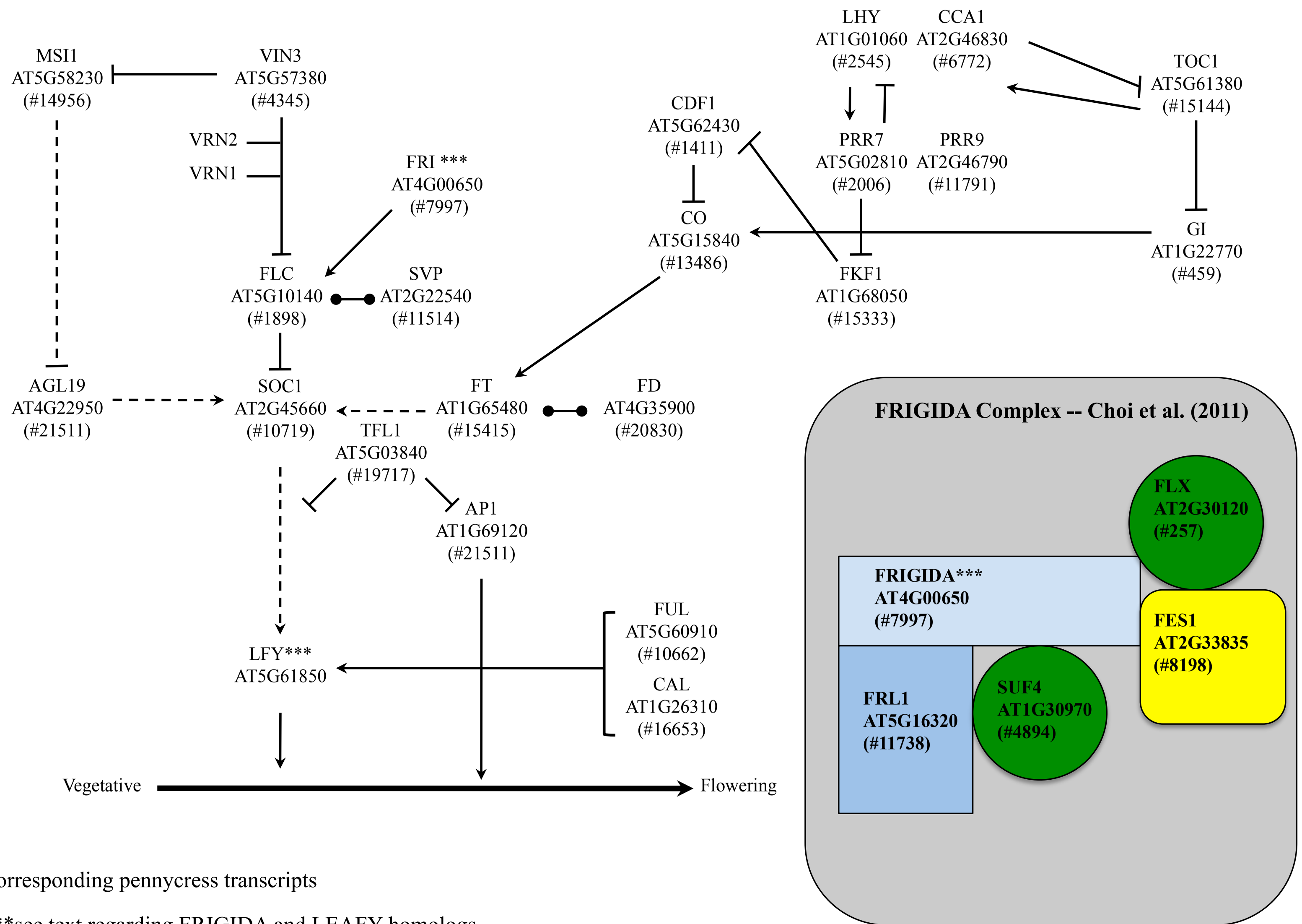
Histogram showing Frequency vs. percent coverage (longest positive hit/peptide length) of pennycress contigs to an *Arabidopsis* peptide from a global BLASTx comparison of the pennycress transcriptome the Arabidopsis predicted protein set. (note: most assembled pennycress transcripts have high coverage which greatly skews the histogram to the right)



Less than
70% similarity
70% coverage
17,317 / 33,873
Greater than
70% similarity
70% coverage
16,556 / 33,873
Greater than
80% similarity
80% coverage
14,531 / 33,873
Greater than
95% similarity
95% coverage
4,846 / 33,873
Percent Coverage of an Arabidopsis peptide
0
10
20
30
40
50
60
70
80
90
100
40
45
50
55
60
65
70
75
80
85
90
95
100
Percent Similarity of an *Arabidopsis* peptide
0
transcripts
753
transcripts

Figure 2-9 Similarity and Coverage of pennycress transcripts versus Arabidopsis genes

Smoothed color density representation of the percent similarity (x-axis) of each pennycress transcript plotted against the percent coverage of the *Arabidopsis* protein similarity (y-axis). Plot produced using the ‘smoothScatter’ function in R (R Team, 2008), which produces a smoothed density representation of the scatterplot using a kernel density estimate (nbin=100). Darker color indicates a higher density of transcripts in a given position, with the darkest ‘bin’ containing over 700 transcripts. Boxes encompassing transcripts encoding peptides with 70%, 80%, and 95% sequence similarity and coverage are shown in the upper right corner. Raw similarity and coverage data is available in Supporting Table S2 of Dorn et al. (2013).



Corresponding pennycress transcripts

***see text regarding FRIGIDA and LEAFY homologs

Figure 2-10 – Reconstruction of the flowering time pathway in pennycress

A model for the pathway controlling flowering time in pennycress adapted from Jung and Müller (2009). Arabidopsis gene identifiers and pennycress transcriptome contig numbers are shown under each gene name.

References

Richard M. Amasino 2005. Vernalization and flowering time. *Current Opinion in Biotechnology,* 16**,** 154-158.

K. F. Best & G. I. Mcintyre 1975. Biology of Canadian Weeds, 9. Thlaspi arvense L. . *Canadian Journal of Plant Science,* 55**,** 279-292.

K. F. Best & G. I. Mcintyre 1976. Studies on Flowering of Thlaspi-Arvense L .3. Influence of Vernalization under Natural and Controlled Conditions. *Botanical Gazette,* 137**,** 121-127.

A. A. Boateng, C. A. Mullen & N. M. Goldberg 2010. Producing Stable Pyrolysis Liquids from the Oil-Seed Presscakes of Mustard Family Plants: Pennycress (Thlaspi arvense L.) and Camelina (Camelina sativa). *Energy & Fuels,* 24**,** 6624-6632.

A. M. Bones & J. T. Rossiter 1996. The myrosinase-glucosinolate system, its organisation and biochemistry. *Physiologia Plantarum,* 97**,** 194-208.

K. Choi, J. Kim, H. J. Hwang, S. Kim, C. Park, S. Y. Kim & I. Lee 2011. The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *Plant Cell,* 23**,** 289-303.

A. Conesa, S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon & M. Robles 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics,* 21**,** 3674-3676.

S. M. Dabney, J. A. Delgado & D. W. Reeves 2001. Using Winter Cover Crops to Improve Soil and Water Quality. *Communications in Soil Science and Plant Analysis,* 32**,** 1221-1250.

J. Fan, Shonnard, D., Kalnes, T., Johnsen, P., Rao, S. 2013. A life cycle assessment of pennycress (Thlaspi arvense L.) -derived jet fuel and diesel. *Biomass and Bioenergy,* In Press.

J. Fargione, J. Hill, D. Tilman, S. Polasky & P. Hawthorne 2008. Land clearing and the biofuel carbon debt. *Science,* 319**,** 1235-1238.

A. Franzke, M. A. Lysak, I. A. Al-Shehbaz, M. A. Koch & K. Mummenhoff 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci,* 16**,** 108-116.

S. Gazzani 2003. Analysis of the Molecular Basis of Flowering Time Variation in Arabidopsis Accessions. *Plant Physiology,* 132**,** 1107-1114.

E. Gongora-Castillo & C. R. Buell 2013. Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Nat Prod Rep*.

N. Gruenheit, O. Deusch, C. Esser, M. Becker, C. Voelckel & P. Lockhart 2012. Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants. *BMC Genomics,* 13.

J. P. Hamilton & C. R. Buell 2012. Advances in plant genome sequencing. *Plant J,* 70**,** 177-90.

J. P. Hammond, H. C. Bowen, P. J. White, V. Mills, K. A. Pyke, A. J. M. Baker, S. N. Whiting, S. T. May & M. R. Broadley 2006. A comparison of the Thlaspi caerulescens and Thlaspi arvense shoot transcriptomes. *New Phytologist,* 170**,** 239-260.

J. Hill, E. Nelson, D. Tilman, S. Polasky & D. Tiffany 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 11206-11210.

Mila P. Hojilla-Evangelista, Roque L. Evangelista, Terry A. Isbell, and Gordon W. Selling 2013. Effects of cold-pressing and seed cooking on functional properties of protein in pennycress (< i> Thlaspi arvense</i> L.) seed and press cakes. *Industrial Crops and Products,* 45**,** 223-229.

T. T. Hu, P. Pattyn, E. G. Bakker, J. Cao, J. F. Cheng, R. M. Clark, N. Fahlgren, J. A. Fawcett, J. Grimwood, H. Gundlach, G. Haberer, J. D. Hollister, S. Ossowski, R. P. Ottilar, A. A. Salamov, K. Schneeberger, M. Spannagl, X. Wang, L. Yang, M. E. Nasrallah, J. Bergelson, J. C. Carrington, B. S. Gaut, J. Schmutz, K. F. X. Mayer, Y. V. De Peer, I. V. Grigoriev, M. Nordborg, D. Weigel & Y. L. Guo 2011. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat Genet,* 43**,** 476-+.

L. Hume, M. D. Devine & S. Shirriff 1995. The Influence of Temperature Upon Physiological Processes in Early-Flowering and Late-Flowering Strains of Thlaspi-Arvense L. *International Journal of Plant Sciences,* 156**,** 445-449.

J. A. Irwin, C. Lister, E. Soumpourou, Y. W. Zhang, E. C. Howell, G. Teakle & C. Dean 2012. Functional alleles of the flowering time regulator FRIGIDA in the Brassica oleracea genome. *BMC Plant Biology,* 12.

Terry A. Isbell & Steven C. Cermak 2012. Extraction of pennycress (Thlaspi arvense L.) seed oil by full pressing. *Industrial Crops & Products,* 37**,** 6p.

U. Johanson, J. West, C. Lister, S. Michaels, R. Amasino & C. Dean 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science,* 290**,** 344-7.

J. S. Johnston 2005. Evolution of Genome Size in Brassicaceae. *Annals of Botany,* 95**,** 229-235.

Christian Jung & Andreas E. Müller 2009. Flowering time control and applications in plant breeding. *Trends in Plant Science,* 14**,** 563-573.

S. Kaul, H. L. Koo, J. Jenkins, M. Rizzo, T. Rooney, L. J. Tallon, T. Feldblyum, W. Nierman, M. I. Benito, X. Y. Lin, C. D. Town, J. C. Venter, C. M. Fraser, S. Tabata, Y. Nakamura, T. Kaneko, S. Sato, E. Asamizu, T. Kato, H. Kotani, S. Sasamoto, J. R. Ecker, A. Theologis, N. A. Federspiel, C. J. Palm, B. I. Osborne, P. Shinn, A. B. Conway, V. S. Vysotskaia, K. Dewar, L. Conn, C. A. Lenz, C. J. Kim, N. F. Hansen, S. X. Liu, E. Buehler, H. Altafi, H. Sakano, P. Dunn, B. Lam, P. K. Pham, Q. Chao, M. Nguyen, G. X. Yu, H. M. Chen, A. Southwick, J. M. Lee, M. Miranda, M. J. Toriumi, R. W. Davis, R. Wambutt, G. Murphy, A. Dusterhoft, W. Stiekema, T. Pohl, K. D. Entian, N. Terryn, G. Volckaert, M. Salanoubat, N. Choisne, M. Rieger, W. Ansorge, M. Unseld, B. Fartmann, G. Valle, F. Artiguenave, J. Weissenbach, F. Quetier, R. K. Wilson, M. De La Bastide, M. Sekhon, E. Huang, L. Spiegel, L. Gnoj, K. Pepin, J. Murray, D. Johnson, K. Habermann, N. Dedhia, L. Parnell, R. Preston, L. Hillier, E. Chen, M. Marra, R. Martienssen, W. R. Mccombie, K. Mayer, O. White, M. Bevan, K. Lemcke, T. H. Creasy, C. Bielke, B. Haas, D. Haase, R. Maiti, S. Rudd, J. Peterson, H. Schoof, D. Frishman, B. Morgenstern, P. Zaccaria, M. Ermolaeva, M. Pertea, J. Quackenbush, N. Volfovsky, D. Y. Wu, T. M. Lowe, S. L. Salzberg,

H. W. Mewes, S. Rounsley, D. Bush, S. Subramaniam, I. Levin, S. Norris, R. Schmidt, A. Acarkan, I. Bancroft, F. Quetier, A. Brennicke, J. A. Eisen, T. Bureau, B. A. Legault, Q. H. Le, N. Agrawal, Z. Yu, R. Martienssen, G. P. Copenhaver, S. Luo, C. S. Pikaard, D. Preuss, I. T. Paulsen, M. Sussman, A. B. Britt, D. A. Selinger, R. Pandey, D. W. Mount, V. L. Chandler, R. A. Jorgensen, C. Pikaard, G. Juergens, E. M. Meyerowitz, A. Theologis, J. Dangl, J. D. G. Jones, M. Chen, J. Chory, M. C. Somerville & Ar Gen In 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408**,** 796-815.

Dong-Hwan Kim, Mark R. Doyle, Sibum Sung & Richard M. Amasino 2009. Vernalization: Winter and the Timing of Flowering in Plants. *Annual Review of Cell and Developmental Biology,* 25**,** 277-299.

J. C. Kuchernig, A. Backenkohler, M. Lubbecke, M. Burow & U. Wittstock 2011. A thiocyanate-forming protein generates multiple products upon allylglucosinolate breakdown in Thlaspi arvense. *Phytochemistry,* 72**,** 1699-1709.

M. Kurowska, A. Daszkowska-Golec, D. Gruszka, M. Marzec, M. Szurman, I. Szarejko & M. Maluszynski 2011. TILLING - a shortcut in functional genomics. *Journal of Applied Genetics,* 52**,** 371-390.

P. Lamesch, T. Z. Berardini, D. H. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander, M. Garcia-Hernandez, A. S. Karthikeyan, C. H. Lee, W. D. Nelson, L. Ploetz, S. Singh, A. Wensel & E. Huala 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research,* 40**,** D1202-D1210.

E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. Fitzhugh, R. Funke, D. Gage, K. Harris, A. Heaford, J.

Howland, L. Kann, J. Lehoczky, R. Levine, P. Mcewan, K. Mckernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. Mcmurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. Mcpherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. Mccombie, M. De La Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E.

Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. Mclysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. De Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen & Consortium International Human Genome Sequencing 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

C. M. Mccallum, L. Comai, E. A. Greene & S. Henikoff 2000. Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiology,* 123**,** 439-442.

G. I. Mcintyre & K. F. Best 1978. Studies on Flowering of Thlaspi-Arvense L .4. Genetic and Ecological Differences between Early-Flowering and Late-Flowering Strains. *Botanical Gazette,* 139**,** 190-195.

S. D. Michaels & R. M. Amasino 2001. Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell,* 13**,** 935-941.

M. J. Milner & L. V. Kochian 2008. Investigating heavy-metal hyperaccumulation using Thlaspi caerulescens as a model system. *Ann Bot,* 102**,** 3-13.

L. W. Mitich 1996. Field pennycress (Thlaspi arvense L) - The stinkweed. *Weed Technology,* 10**,** 675-678.

B. R. Moser, G. Knothe, S. F. Vaughn & T. A. Isbell 2009a. Production and Evaluation of Biodiesel from Field Pennycress (Thlaspi arvense L.) Oil. *Energy & Fuels,* 23**,** 4149-4155.

B. R. Moser, Shailesh N. Shah, Jill K. Winkler-Moser, Steven F. Vaughn & Roque L. Evangelista 2009b. Composition and physical properties of cress (Lepidium sativum L.) and field pennycress (Thlaspi arvense L.) oils. *Industrial Crops & Products,* 30**,** 7p.

H. H. Nour-Eldin, T. G. Andersen, M. Burow, S. R. Madsen, M. E. Jorgensen, C. E. Olsen, I. Dreyer, R. Hedrich, D. Geiger & B. A. Halkier 2012. NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature,* 488**,** 531-4.

Winthrop B. Phippen & Mary E. Phippen 2012. Soybean Seed Yield and Quality as a Response to Field Pennycress Residue. *Crop Science,* 52**,** 7p.

H.L. Koo S. Kaul, J. Jenkins, M. Rizzo, T. Rooney, L.J. Tallon, T. Feldblyum, W. Nierman, M.-I. Benito, M.-I.X. Lin, C.D. Town, J.C. Venter, C.M. Fraser, S.Tabata, Y. Nakamura, T. Kaneko, S. Sato, E. Asamizu, T. Kato, H. Kotani, S. Sasamoto, J.R. Ecker, A. Theologis, N.A. Federspiel, C.J. Palm, B.I. Osborne, P. Shinn, A.B. Conway, V.S. Vysotskaia, K. Dewar, L. Conn, C.A. Lenz, C.J. Kim, N.F. Hansen, S.X. Liu, E. Buehler, H. Altafi, H. Sakano, P. Dunn, B. Lam, P.K. Pham, Q. Chao, M. Nguyen, G. Yu, H. Chen, A. Southwick, J.M. Lee, M. Miranda, M.J. Toriumi, R.W. Davis, R. Wambutt, G. Murphy, A. Düsterhöft, W. Stiekema, T. Pohl, K.-D. Entian, N. Terryn, G. Volckaert, M. Salanoubat, N. Choisne, M. Rieger, W. Ansorge, M. Unseld, B. Fartmann, G. Valle, F. Artiguenave, J. Weissenbach, F. Quetier, R.K. Wilson, M. De La Bastide, M. Sekhon, E. Huang, L. Spiegel, L. Gnoj, K. Pepin, J. Murray, D. Johnson, K. Habermann, N. Dedhia, L. Parnell, R. Preston, L. Hillier, E. Chen, M. Marra, R.

Martienssen, W.R. Mccombie, K. Mayer, O. White, M. Bevan, K. Lemcke, T.H. Creasy, C. Bielke, B. Haas, D. Haase, R. Maiti, S. Rudd, J. Peterson, H. Schoof, D. Frishman, B. Morgenstern, P. Zaccaria, M. Ermolaeva, M. Pertea, J. Quackenbush, N. Volfovsky, D. Wu, T.M. Lowe, S.L. Salzberg, H.-W. Mewes, S. Rounsley, D. Bush, S. Subramaniam, I. Levin, S. Norris, R. Schmidt, A. Acarkan, I. Bancroft, F. Quetier, A. Brennicke, J.A. Eisen, T. Bureau, B.-A. Legault, Q.-H. Le, N. Agrawal, Z. Yu, R. Martienssen, G.P. Copenhaver, S. Luo, C.S. Pikaard, D. Preuss, I.T. Paulsen, M. Sussman, A.B. Britt, J.A. Eisen. D.A. Selinger, R. Pandey, D.W. Mount, V.L. Chandler, R.A. Jorgensen, C. Pikaard, G. Juergens, E.M. Meyerowitz, J.R. Ecker, A. Theologis, J. Dangl, J.D.G. Jones, M. Chen, J. Chory, C. Somerville. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408**,** 796-815.

M. E. Schranz, P. Quijada, S. B. Sung, L. Lukens, R. Amasino & T. C. Osborn 2002. Characterization and effects of the replicated flowering time gene FLC in Brassica rapa. *Genetics,* 162**,** 1457-1468.

G. W. Selling, M. P. Hojilla-Evangelista, R. L. Evangelista, T. Isbell, N. Price & K. M. Doll 2013. Extraction of proteins from pennycress seeds and press cake. *Industrial Crops and Products,* 41**,** 113-119.

N. Sharma, D. Cram, T. Huebert, N. Zhou & I. A. P. Parkin 2007. Exploiting the wild crucifer Thlaspi arvense to identify conserved and novel genes expressed during a plant's response to cold stress. *Plant Mol Biol,* 63**,** 171-184.

C. C. Sheldon, D. T. Rouse, E. J. Finnegan, W. J. Peacock & E. S. Dennis 2000. The molecular basis of vernalization: The central role of FLOWERING LOCUS C (FLC). *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 3753-3758.

G. G. Simpson & C. Dean 2002. Arabidopsis, the Rosetta stone of flowering time? *Science,* 296**,** 285-9.

T. Slotte, Hazzouri, K. M., Ågren, J. A., Koenig, D., Maumus, F., Guo, Y.-L., Steige, K., Platts, A. E., Escobar, J. S., Newman, L., Wang, W., Mandáková, T., Vello, E., Smith, L. M., Henz, S. R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., Andolfatto, P., Hu, T. T., Blanchette, M., Clark, R. M., Quesneville, H., Nordborg, M., Gaut, B. S., Lysak, M. A., Jenkins, J., Grimwood, J., Chapman, J., Prochnick, S., Shu, S., Rokhsar, D., Schmutz, J., Weigel, D., and Wright, S. I. 2013. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nature Genetics,* accepted for publication.

S. S. Snapp, S. M. Swinton, R. Labarta, D. Mutch, J. R. Black, R. Leep, J. Nyiraneza & K. O'neil 2005. Evaluating cover crops for benefits, costs and performance within cropping system niches. *Agronomy Journal,* 97**,** 322-332.

M. Tadege, C. C. Sheldon, C. A. Helliwell, P. Stoutjesdijk, E. S. Dennis & W. J. Peacock 2001. Control of flowering time by FLC orthologues in Brassica napus. *Plant Journal,* 28**,** 545-553.

R Development Core Team 2008. *R: A language and environment for statistical computing,* Vienna, Austria R Foundation for Statistical Computing.

D. Tilman, R. Socolow, J. A. Foley, J. Hill, E. Larson, L. Lynd, S. Pacala, J. Reilly, T. Searchinger, C. Somerville & R. Williams 2009. Beneficial Biofuels-The Food, Energy, and Environment Trilemma. *Science,* 325**,** 270-271.

Rajeev K. Varshney, Spurthi N. Nayak, Gregory D. May & Scott A. Jackson 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology,* 27**,** 522-530.

S. F. Vaughn, T. A. Isbell, D. Weisleder & M. A. Berhow 2005. Biofumigant compounds released by field pennycress (Thlaspi arvense) seedmeal. *Journal of Chemical Ecology,* 31**,** 167-177.

J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. Mckusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. Mccawley, T. Mcintosh, I. Mcmullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R.

Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. Mcdaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh & X. Zhu 2001. The sequence of the human genome. *Science,* 291**,** 1304-51.

N. A. Wang, W. Qian, I. Suppanz, L. J. Wei, B. Z. Mao, Y. Long, J. L. Meng, A. E. Muller & C. Jung 2011a. Flowering time variation in oilseed rape (Brassica napus L.) is associated with allelic variation in the FRIGIDA homologue BnaA.FRI.a. *Journal of Experimental Botany,* 62**,** 5641-5658.

X. W. Wang, H. Z. Wang, J. Wang, R. F. Sun, J. Wu, S. Y. Liu, Y. Q. Bai, J. H. Mun, I. Bancroft, F. Cheng, S. W. Huang, X. X. Li, W. Hua, J. Y. Wang, X. Y. Wang, M. Freeling, J. C. Pires, A. H. Paterson, B. Chalhoub, B. Wang, A. Hayward, A. G. Sharpe, B. S. Park, B. Weisshaar, B. H. Liu, B. Li, B. Liu, C. B. Tong, C. Song, C. Duran, C. F. Peng, C. Y. Geng, C. S. Koh, C. Y. Lin, D. Edwards, D. S. Mu, D. Shen, E. Soumpourou, F. Li, F. Fraser, G. Conant, G. Lassalle, G. J. King, G. Bonnema, H. B. Tang, H. P. Wang, H. Belcram, H. L. Zhou, H. Hirakawa, H. Abe, H. Guo, H. Wang, H. Z. Jin, I. A. P. Parkin, J. Batley, J. S. Kim, J. Just, J.

W. Li, J. H. Xu, J. Deng, J. A. Kim, J. P. Li, J. Y. Yu, J. L. Meng, J. P. Wang, J. M. Min, J. Poulain, J. Wang, K. Hatakeyama, K. Wu, L. Wang, L. Fang, M. Trick, M. G. Links, M. X. Zhao, M. N. Jin, N. Ramchiary, N. Drou, P. J. Berkman, Q. L. Cai, Q. F. Huang, R. Q. Li, S. Tabata, S. F. Cheng, S. Zhang, S. J. Zhang, S. M. Huang, S. Sato, S. L. Sun, S. J. Kwon, S. R. Choi, T. H. Lee, W. Fan, X. Zhao, X. Tan, X. Xu, Y. Wang, Y. Qiu, Y. Yin, Y. R. Li, Y. C. Du, Y. C. Liao, Y. Lim, Y. Narusaka, Y. P. Wang, Z. Y. Wang, Z. Y. Li, Z. W. Wang, Z. Y. Xiong & Z. H. Zhang 2011b. The genome of the mesopolyploid crop species Brassica rapa. *Nature Genetics,* 43**,** 1035-U157.

S. I. Warwick, A. Francis & D. J. Susko 2002. The biology of Canadian weeds. 9. Thlaspi arvense L. (updated). *Canadian Journal of Plant Science,* 82**,** 803-823.

N. Winchester, Mcconnachie, D., Wollersheim, C, Waitz, I. 2013. Market cost of renewable jet fuel adopiton in the United States Massachusetts Institute of Technology

U. Wittstock & M. Burow 2010. Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. *Arabidopsis Book,* 8**,** e0134.

Y. X. Yuan, J. Wu, R. F. Sun, X. W. Zhang, D. H. Xu, G. Bonnema & X. W. Wang 2009. A naturally occurring splicing site mutation in the Brassica rapa FLC1 gene is associated with variation in flowering time. *Journal of Experimental Botany,* 60**,** 1299-1308.

Q. Y. Zhao, Y. Wang, Y. M. Kong, D. Luo, X. Li & P. Hao 2011. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *Bmc Bioinformatics,* 12.

CHAPTER 3: A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop

STATEMENT ON PREVIOUSLY PUBLISHED MATERIAL

The work described in this chapter was published in DNA Research (2015, Volume 22, Issue 2, pages 121-131, DOI: 10.1093/dnares/dsu045, an Oxford University Press journal. This article was published as 'Open Access', distributed under the terms of the Creative Commons Attribution License, which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. Included in this chapter are references to large electronic supplementary datasets and data files included as a part of this publication that will be referenced as Dorn et al. (2015).

SUMMARY

Field pennycress (*Thlaspi arvense* L.) is being domesticated as a new winter cover crop and biofuel species for the Midwestern United States that can be double-cropped between corn and soybeans. A genome sequence will enable the use of new technologies to make improvements in pennycress. To generate a draft genome a hybrid sequencing approach was used to generate 47 Gb of DNA sequencing reads from both the Illumina and PacBio platforms. These reads were used to assemble 6,780 genomic scaffolds. The draft genome was annotated using the MAKER pipeline, which identified 27,390 predicted protein-coding genes, with almost all of these predicted peptides having significant sequence similarity to Arabidopsis proteins. A comprehensive analysis of pennycress gene homologs involved in glucosinolate biosynthesis, metabolism, and transport pathways revealed high sequence conservation compared to other Brassicaceae species, and helps validate the assembly of the pennycress gene space in this draft genome. Additional comparative genomic analyses indicate that the knowledge gained from years of basic Brassicaceae research will serve as a powerful tool for identifying gene targets whose manipulation can be predicted to result in improvements for pennycress.

Introduction

Next generation sequencing (NGS) has enabled the characterization and comparison of whole genomes for a growing list of plant species (Yang *et al.*, 2013, Wang *et al.*, 2014, Kagale *et al.*, 2014, Liu *et al.*, 2014, Slotte *et al.*, 2013). This same technology is being used to speed up and facilitate the breeding of crop plants (Kagale *et al.*, 2014, Varshney *et al.*, 2013, Varshney *et al.*, 2014, Ray and Satya, 2014). NGS also has the potential to enable new approaches to domesticate crops (Sedbrook *et al.*, 2014). One species targeted for domestication is *Thlaspi arvense* (field pennycress, pennycress herein). Pennycress is a member of the Brassicaceae in the tribe Thlaspideae native to Eastern Europe. Pennycress plants are diploid, propagate largely via self-fertilization and have a 1C DNA content of 539 Mb (Johnston *et al.*, 2005). Naturalized populations are distributed worldwide, including North America, South America and Australia. Previous molecular studies showed that the original *Thlaspi* genus was polyphylogenic and several members of the *Thlaspi* genus were kept in a monophylogenic Thlaspideae tribe (Mummenhoff *et al.*, 1997). *Thlaspi caerulescens*, which is a well-studied heavy metal accumulator (Milner *et al.*, 2014, Pence *et al.*, 2000), was moved to the genus *Noccaea* during this reorganization (Mummenhoff *et al.*, 1997, Koch and German, 2013). Two members of the original *Thlaspi* genus, *T. ceratocarpum* and *T. alliaceum* (roadside pennycress), are in the Thlaspideae along with *Alliaria petiolata* (garlic mustard) and *Peltaria alliacea* (garlic cress) and several other species. The Brassicaceae are divided into three lineages and the members of the Thlaspideae are in lineage II, which includes the Brassica genus, but not Arabidopsis (lineage I) (Franzke *et al.*, 2011). Similar to *B.*

napus and *B. rapa*, pennycress produces seeds with high oil content (30% by weight) (Moser *et al.*, 2009b).

Field pennycress is being developed as a new winter cover crop and biodiesel feedstock for the Midwestern United States that can be seeded into standing corn (Dorn *et al.*, 2013, Fan *et al.*, 2013, Moser *et al.*, 2009a, Sedbrook *et al.*, 2014). As a winter annual, pennycress germinates in the fall and overwinters as a vegetative rosette. It exhibits extreme winter hardiness, surviving the harsh winters that are common to the Canadian Plains and Midwestern United States (Best and McIntyre, 1975). Pennycress flowers and matures early in the spring, producing up to 1,300 kilograms/hectare of seed that can be harvested in time for planting an additional summer annual crop of soybeans (Sedbrook *et al.*, 2014, Phippen and Phippen, 2012). The oils found in pennycress seed are suitable for biodiesel production (Moser *et al.*, 2009a) and the remaining seedmeal is high in protein that could serve as an additional revenue source for farmers (Warwick, 2002). As a winter cover, pennycress would provide important ecosystem services. In much of the Midwest, the ground lies barren from late fall until early summer, which can increase nutrient runoff and soil erosion (Tilman *et al.*, 2002). The addition of a fall-planted pennycress cover crop should reduce nutrient leaching and erosion, which would help sustain current farming practices. In the United States alone, over 16 million hectares currently used in the corn to soybean rotation could be used for pennycress production without displacing current food production systems (Fan *et al.*, 2013, Winchester, 2013).

As a weedy species, there are many challenges that will need to be addressed in order to convert pennycress into a new crop species (Sedbrook, 2014). Pennycress already has many natural attributes such as the high seed yield and oil content described above. However, earlier maturing varieties are needed to ensure that pennycress can be harvested early, allowing for full-season soybean production. Seed dormancy also is an issue (Karimmojeni, 2014, Saini *et al.*, 1987), as is common in many weed species; we have observed highly variable germination rates in preliminary field trials. Pennycress oil and protein meal are not currently suitable for animal or human consumption (Evangelista *et al.*, 2012). Making pennycress seeds edible would add value to the crop. In addition, the fatty acids contained in the oil are adequate for conversion to biodiesel, but can be further optimized (Moser *et al.*, 2009b, Moser *et al.*, 2009a). It is unclear if traditional breeding programs can address these concerns, as there may be insufficient natural variation in wild populations.

Arabidopsis thaliana has been used as a key organism to address questions concerning plant development for the past thirty years. Arabidopsis research has resulted in an understanding of many plant developmental processes, such as the control of time to flower, of seed maturation, and of seed oil synthesis (Baud and Lepiniec, 2009, Holdsworth *et al.*, 2008, Jung and Muller, 2009). During the course of these studies, many mutations have been identified in Arabidopsis that confer what can be considered agronomically desirable phenotypes (Sedbrook, 2014). The function of pennycress genes can be predicted on the basis of their similarity to genes with known function in

Arabidopsis (Sedbrook, 2014). With this information, it should be possible to identify mutations that confer desirable traits in pennycress using mutation-based breeding tools such as TILLING (McCallum *et al.*, 2000), DeTILLING (Rogers *et al.*, 2009), and CRISPR-Cas (Belhaj *et al.*, 2013). To enable the use of these technologies and to take advantage of information derived from research on Arabidopsis and other Brassica species, a genomic sequence is needed for pennycress.

To generate a draft genome for pennycress, a natural population of plants was isolated from a roadside near Coates, Minnesota called MN106. Little is known about natural pennycress populations at the genome sequence level. In this report we have examined this population at the genome level using several different next generation sequencing strategies to *de novo* assemble a draft genome. This assembly captures the vast majority of pennycress gene space as predicted by comparison with other Brassica species. The draft genome detailing the pennycress gene space has allowed the function of pennycress genes to be predicted on the basis of their similarity to genes with known function in Arabidopsis. In addition, the analysis revealed that even in mixed natural populations, the selfing mode of reproduction maintains individual plants in a highly homozygous state.

Materials and Methods

Genomic DNA isolation and DNA sequencing

Thlaspi arvense line MN106 has been previously described and originates from Coates, Minnesota (Dorn *et al.*, 2013). Seed from a MN106 bulk planting was germinated on moist Berger BM2 germination mix (Berger Inc, www.berger.ca), stratified at 4°C for 7 days, and grown in climate-controlled growth chambers at the University of Minnesota (21°C, 16 h/8 h day/night cycles at 100 μmol/m^2/s PAR). Individual plants were transplanted to 4 inch pots two weeks after germination. Six week old plants with established rosettes were vernalized at 4°C for 30 days in the dark. After vernalization, plants were returned to growth chambers. In all, nine plants were selected for DNA extraction. After tissue harvest, the same nine plants were maintained and seeds were individually collected from each plant. To enrich the total amount of nuclear DNA sequenced, nuclei were purified from young leaf tissue using a series of density gradient centrifugation steps using an adapted protocol (Olszewski *et al.*, 1988). All centrifugation steps were performed at 4°C. A total of 15 grams of young leaves were washed in cold, sterile ddH_2O and homogenized in 300ml of 4°C nuclear isolation buffer (1 M sucrose, 10 mM Tris HCl (pH 7.2), 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol) in a Brinkmann Polytron homogenizer. Homogenized tissue was filtered through four layers of cheesecloth and one layer of Miracloth (EMD Millipore, www.emdmillipore.com) two times. Filtered lysate was centrifuged for 15 minutes at 9,500 RPM in a Beckman JA14 rotor at 4°C. The supernatant was discarded and the resulting pellet was resuspended in

50 ml of nuclear isolation buffer plus 250 μl of Triton X-100. The resuspended pellet was then twice centrifuged for 10 minutes at 9,500 RPM in a Beckman JA20 rotor, resuspending the pellet as above. The final pellet was resuspended in 50 ml of nuclear isolation buffer and placed onto a Percoll density gradient consisting of 60% and 35% layers (1x Percoll Buffer - 1M sucrose, 10 μM Tris-HCl, 10μl $MgCl_2$, 5 μM 2-mercaptoethanol in Percoll, gradient layers were diluted with nuclear isolation buffer). The gradient was centrifuged at 7,755 RPM for 20 minutes in a Cervall HB-4 rotor, and nuclei were harvested from the interface between the two density layers. Purified nuclei were diluted in five volumes of nuclear isolation buffer and centrifuged for 10 minutes in the JA20 rotor at 7,300 RPM. The supernatant was discarded, and the pellet was resuspended in 200 μl of 1X Tris-EDTA. The DNA was purified from the nuclei using the DNeasy Plant Mini Kit (Qiagen, www.qiagen.com).

Illumina sequencing (100bp paired end library run on Illumina HiSeq 2000, 250bp paired end library run on Illumina MiSeq, 50bp long-jump mate pair libraries library run on Illumina HiSeq 2000) was completed at the University of Minnesota Genomics Center. Pacific Biosciences (PacBio RSII) sequencing was completed at the Mayo Clinic Molecular Biology Core (Rochester, Minnesota).

Paired-end libraries were sequenced on full lanes of both the Illumina HiSeq (100 bp paired-end, insert size = 280 bp, Illumina TruSeq Adaptor #18 - CTCCGC) and the Illumina MiSeq (250 bp paired-end, insert size = 460 bp). The MiSeq sequencing sample

was sequenced without using an index read as this was a uniplex run for the purposes of barcoding in the Illumina index read. Three Illumina Nextera mate pair libraries were also sequenced using the Illumina HiSeq 2000 platform (50 bp paired-end) with 2 kb (Adaptor #2 - CGATGT), 3.5 kb (Adaptor #7 - CAGATC), and 7 kb (Adaptor #19 - GTGAAA) inserts. Pacific Biosciences (PacBio RSII) Sequencing was completed at the Mayo Clinic Molecular Biology Core (Rochester, Minnesota). A genomic DNA library with insert target size of 10 kb was prepared using the Pacific Biosciences low input 10 kb library preparation protocol using the MagBead Station. This library was run on four SMRT cells with 1x120 minute movies each. All raw sequencing files have been submitted to NCBI Sequence Read Archive under accession number SRP033211 (Table 1).

Genome assembly, scaffolding, and annotation

FASTQ files from all sequencing runs were imported into CLC Genomics Workbench Version 6.5 (CLC Bio, www.clcbio.com). Illumina reads were subjected to quality control using the Sequencing QC Report tool. 100 bp paired-end Illumina reads were trimmed and filtered using the following parameters: Ambiguous Trim = Yes, Ambiguous Limit = 5, Quality Limit = 0.05, Use Colorspace = No, Search Reverse Sequence = Yes, Remove 5' terminal nucleotides = Yes, Number of 5' Terminal Nucleotides to Remove = 5, Remove 3' terminal nucleotides = No, Discard Long Reads = No, Save Broken Pairs = No. 250 bp paired-end Illumina reads were trimmed and filtered using the following parameters: Ambiguous Trim = Yes, Ambiguous Limit = 10, Quality Limit = 0.05, Use Colorspace = No, Search Reverse Sequence = No, Remove 5' terminal nucleotides = Yes, Number of 5' Terminal Nucleotides to Remove = 10, Remove 3' terminal nucleotides = No, Discard Long Reads = No, Save Broken Pairs = No. After trimming and filtering, the 250 bp paired-end reads were merged using the Merged Overlapping Pairs tool in CLC Genomics Workbench (Mismatch cost = 4, Minimum Score = 8, Gap Cost = 4, Max Unaligned = 0). Unmerged reads were saved and also used in the initial assembly. 50 bp mate-pair Illumina reads were trimmed and filtered using the following parameters: Ambiguous Trim = Yes, Ambiguous Limit = 2, Quality Limit = 0.05, Use Colorspace = No, Search Reverse Sequence = Yes, Remove 5' terminal nucleotides = Yes, Number of 5' Terminal Nucleotides to Remove = 2, Remove 3' terminal nucleotides = No, Discard Long Reads = No, Save Broken Pairs = No. The mate-pair libraries were exported as FASTQ files for scaffolding as described below.

Genome sequencing reads were trimmed and de novo assembled in CLC Genomics Workbench Version 6.5 (CLC Bio, www.clcbio.com). The CLC assembler has previously been used in assembling complex plant genomes, such as the barley (*Hordeum vulgare*) (International Barley Genome Sequencing *et al.*, 2012) and rubber tree (*Hevea brasillensis*) (Rahman *et al.*, 2013) genome projects. Additionally, the CLC assembler has a low Random Access Memory (RAM) requirement, opposed to other popular *de Bruijn* graph assemblers that can require hundreds of gigabytes of RAM.

The pennycress draft genome was assembled and annotated using two desktop computers built specifically for this purpose (Tables 2 and 3). Trimmed and filtered 100 bp paired-end and 250 bp paired-end (merged and unmerged) reads were *de novo* assembled in CLC Genomics Workbench using the De Novo Assembly tool with the following parameters: Mapping Mode = Map reads to contigs (slow), Automatic Bubble Size = Yes (50), Minimum Contig Length = 200 bp, Automatic Word Size = No, Word Size = 64, Perform Scaffolding = Yes, Automatically Detect Paired Distribution = Yes, Mismatch Cost = 3, Insertion Cost = 3, Deletion Cost = 3, Length Fraction = 0.95, Similarity Fraction = 0.9. PacBio reads were used as guidance only in producing the initial assembly. Scaffolds (≥ 1,000 bp) from the initial assembly were further scaffolded using SSPACE (Boetzer *et al.*, 2011) (parameters: -x = 1, -z = 0, -k = 3, -a = 0.7, -n = 10, -T = 8, -p = 1, -o = 15, -t = 0, -m = 32, -r = 0.9) with the Illumina mate-pair and 100 bp paired-end libraries (Dorn *et al.*, 2013). Remaining gaps in the assembly after the mate-pair

scaffolding step were filled using the Illumina 100 bp paired-end reads using GAPFILLER (Boetzer and Pirovano, 2012) (parameters: -o = 2, -m = 29, -r = 0.7, -n = 10, -g = 1, -d = 50, -t = 10, -i =10). Assembly statistics of both contigs and scaffolds are listed in Table S1. Both SSPACE and GAPFILLER were run through the corresponding PERL scripts using Strawberry Perl (http://strawberryperl.com) on Windows. The final scaffolds were exported for annotation using the MAKER pipeline (Cantarel *et al.*, 2008). Twelve scaffolds were subsequently removed from the assembly following the NCBI contamination screen.

An additional computer was assembled using the components listed in Table 3 to conduct genome annotation. The BioLinux 7 operating system (http://nebc.nerc.ac.uk/tools/bio-linux) was installed on the computer used for genome annotation. The MAKER annotation pipeline and dependent programs were installed on this computer for in-lab genome annotation, including repeat identification, model training, and gene prediction. Two rounds of annotation using the MAKER pipeline were completed. The first round of annotation was used to create a training set for the SNAP gene prediction software (Korf, 2004) using our previously published *Thlaspi arvense* transcriptome (est2genome=1) and *Eutrema salsungineum* (formerly *Thellungiella halophila*) protein set (Thalophila_173_protein) available from www.phytozome.net (protein2genome=1) on *Thlaspi arvense* genome scaffolds over 300 kb long. Repeat masking for both the training and final annotations were identified and masked using RepeatMasker (*Arabidopsis thaliana* repeat library athrep.fasta, available from www.giriinst.org) and RepeatRunner

(repeat library ‘te_proteins.fasta’ available in RepeatRunner - http://www.yandell-lab.org/software/repeatrunner.html). A SNAP HMM training file was generated for the 1,000 best gene models using the programs Fathom and Forge (part of the SNAP package). The final round of annotation was completed using this HMM training file, and resulting GFF3 files for all scaffolds were compiled using the PERL script ‘gff3merge.pl’ included in MAKER. The predicted peptide sequences from the pennycress genome annotation were extracted and compared to the Arabidopsis TAIR10 peptide database (TAIR10_pep_20110103_representitive_model_updated) using BLASTp (Expectation value = 10.0, word size = 7, mask lower case = no, filter low complexity = yes, max number of hits = 10, protein matrix and gap costs: BLOSUM62, Existence = 11, Extension = 1). RNAseq reads from the previously published de novo transcriptome assembly (Dorn et al., 2013) were used to further validate gene models. RNAseq reads from this experiment (NCBI Short Read Archive accession number SRR802670) were trimmed (Trim Adaptor = Illumina TruSeq Indexed Adaptor 3, Ambiguous Trim = Yes, Ambiguous Limit = 2, Quality Trim = Yes, Quality Limit = 0.05, Remove 5’ terminal nucleotides = Yes, Number of 5’ terminal nucleotides to remove = 10, Save broken pairs = Yes). Trimmed paired and orphan reads were mapped to the genome assembly in CLC Genomics Workbench (Mapping type = Map to gene regions only, Maximum number of hits for a read = 10, Strand specific = Both, Count paired reads as two = Yes, Reference type = Genome annotated with genes and transcripts, Global alignment = no, Automatically detect paired distances = Yes, Similarity fraction = 0.8, Length fraction = 0.8, Mismatch cost = 2, Insertion cost = 3, Deletion cost = 3).

Comparative genomics

Comparative studies on the structural arrangement and synteny of the pennycress genome assembly were performed using SynMap (www.genomeevolution.org/CoGE/SynMap.pl). To examine the synteny between the *Eutrema salsugineum* and *Thlaspi arvense* genomes, seven pseudo-chromosome sequences were constructed by concatenating the corresponding genomic scaffolds(Yang *et al.*, 2013). *T. arvense* genomic scaffolds greater than 75 kb in length were compared to the seven *E. salsugineum* pseudo-chromosomes using a Syntenic Path Assembly in SynMap (DAGChainer – Relative Gene order, -D=20, -A=5, skip random/unknown chromosomes). BLASTp comparisons of the 27,390 predicted pennycress peptides were performed in CLC Genomics Workbench using the predicted peptide databases for *Arabidopsis thaliana* (Arabidopsis Genome, 2000), *Arabidopsis lyrata* (Hu *et al.*, 2011), *Brassica rapa* (Wang *et al.*, 2011), *Capsella rubella* (Slotte *et al.*, 2013), and *Eutrema salsugineum* (Yang *et al.*, 2013) using the following BLAST parameters: Expectation Value = 10, Word Size = 7, Filter Low Complexity = Yes, Protein Matrix = BLOSUM62, Gap Existence = 11, Gap Extension = 1. Peptide sequences for *A. thaliana* were obtained from ftp://ftp.arabidopsis.org/home/tair/Proteins/TAIR10_protein_lists/TAIR10_pep_20101214. Sequences for the remaining species were obtained from Phytozome (www.phytozome.net). BLASTn analyses of the previously published pennycress transcriptome (Dorn et al. 2013) against the genome assembly were performed using the following parameters: Match cost = 2, Mismatch cost = 3, Gap Existence = 2, Gap

Extension = 2, Expectation value = 10, Filter Low Complexity = Yes, Maximum number of hits = 15.

Read mapping and variant detection for CAPS marker design

Trimmed and filtered sequencing reads from the Illumina HiSeq 2000 100 bp paired-end experiment were re-mapped to the genome assembly to identify potential heterozygosity or variation in the MN106 accession sequenced. Reads were mapped to the assembly in CLC Genomics Workbench using the 'Map Reads to Reference' tool (masking mode = no masking, mismatch cost = 3, insertion cost = 3, deletion cost = 3, length fraction = 0.95, similarity fraction = 0.95, global alignment = no, auto detect paired distances = yes, non-specific match handling = map randomly) and variants called using the Probabilistic Variant Detection tool (ignore non-specific matches = yes, ignore broken pairs = yes, minimum coverage = 10, variant probability = 90, require both forward and reverse reads = no, maximum expected variants = 2, ignore quality scores = no). Sites declared as SNPs were manually examined to identify CAPS markers in which one variant was a member of a six base recognition site for a DNA restriction endonuclease and the other variant resulted in the loss of the restriction site. Four sites were identified and primers were designed to amplify these regions for CAPS analysis (Supplementary Table S4 of of (Dorn *et al.*, 2015)). As a control, the primers were designed to flank both a CAPS site and a conserved restriction site that would be cleaved in both variants. DNA was isolated using the Mag-Bind EZ Plant DNA Kit (Omega BioTek, www.omegabiotek.com). Polymerase chain reactions were performed using Q5 High Fidelity DNA Polymerase (New England Biolabs, www.neb.com) with these DNA extracts and the corresponding CAPS primers to confirm the potential variants.

Comparative analysis of genes involved in glucosinolate metabolism and transport

Arabidopsis thaliana, *Brassica rapa*, and *Brassica olereaceae* genes previously identified to be involved in glucosinolate biosynthesis, breakdown, and transport were derived from a previous study (Liu *et al.*, 2014). Alignments of predicted peptide sequences for each gene were performed in CLC Genomics Workbench using the Create Alignment tool (Gap open cost = 10, Gap extension cost = 1, End gap cost = Free, Alignment mode = Very accurate). Neighbor Joining trees were created in CLC Genomics Workbench (Protein distance measure = Jukes-Cantor, Perform bootstrap analysis = Yes, Replicates = 100). To examine expression levels of predicted gene models, Illumina RNAseq reads from the previously published transcriptome assembly(Dorn *et al.*, 2013) were trimmed and filtered (Illumina TruSeq Trim Adaptor 3, Ambiguous Trim = Yes, Ambiguous Limit = 2, Quality Trim = Yes, Quality Limit = 0.05, Also search on reversed sequence = Yes, Remove 5' terminal nucleotides = Yes, Number of 5' terminal nucleotides to remove = 10, Remove 3' terminal nucleotides = no, Save broken pairs = Yes) and mapped to the annotated draft genome (Mapping Type = Map to gene regions only, Maximum number of hits for a read = 10, Strand Specific = Both, Count paired read as two = Yes, Expression Value = RPKM, Reference type = Genome annotated with genes and transcripts, Global alignment = no, Auto detect paired distances = Yes, Similarity fraction = 0.8, Length Fraction = 0.8, Mismatch cost = 2, Insertion cost = 3, Deletion cost = 3).

Results and Discussion

Genome Sequencing and Assembly

Pennycress is a diploid species with a haploid number of seven chromosomes and a genome size of 539 Mbp (Johnston *et al.*, 2005). In order to isolate a sufficient quantity of nuclear DNA for the various sequencing libraries used, nuclear DNA was isolated from nine plants derived from a MN106 isolate that had been maintained in the lab for several generations. This DNA was sequenced using both the Illumina and PacBio platforms, generating over 47 Gbp of sequencing data representing over 87x coverage of the predicted genome size (Table 4). Illumina HiSeq 2000 and MiSeq reads were *de novo* assembled using the de Bruijn graph-based CLC Genomics Workbench assembler. PacBio reads were not used to create the *de Bruijn* graph, but instead were only used to resolve ambiguities during the graph building stage.

The initial CLC assembly resulted in 206,726 initial scaffolds ≥200bp, encompassing 392,190,998 bp. Of this initial assembly, initial scaffolds over 1,000 bp (n=50,064, 322,949,692 bp total length) were further joined using long-insert Illumina mate pair reads in SSPACE and remaining gaps were filled using GAPFILLER. This analysis resulted in the formation of 6,768 final scaffolds that encompass over 343 Mbp with an average scaffold length of 50,681 bp and N50 value of 140,815 bp (Table 4). Over 60% of the assembled scaffolds are over 10,000 bp long, with 902 scaffolds over 100,000 bp, and 9 scaffolds over 1,000,000 bp (Table 5). The longest 3,000 scaffolds represented over 85% of the assembly length (Figure 1). The 156,662 small scaffolds

(<1000 bp) from the initial CLC assembly that were excluded from the final scaffolding and gapfilling manipulations encompassed 69,241,306 bp (Table 5). These small initial scaffolds were excluded from further analyses as they likely represent sequences from repetitive regions of the genome, which are difficult to assemble and were unlikely to contribute to gene identification efforts. These small initial scaffolds represent a significant portion (12.8%) of the predicted genome size. With the addition of the initial small scaffolds, the total assembled length of the draft genome presented here is 412,253,695 bp, 76.5% of the predicted pennycress genome size of 539Mb.

As the draft genome is incomplete (<80% of the predicted genome size) and fragmented, developing a more complete and contiguous assembly will be important for both plant improvement efforts and answering basic questions about the genomics and evolution of pennycress. However, the reported contig N50 for (21 kb) is well within the norm of other recently reported genome assemblies based on Illumina reads such those for *Aquilaria agacha* - 14.6 kb (Chen *et al.*, 2014), *Sesamum indium* - 52.2 kb (Wang *et al.*, 2014), *Citrullus lantus* - 26.4 kb (Guo *et al.*, 2013), and *Cicer arietinum* - 23.54 kb (Varshney *et al.*, 2013). The sequence of contigs directly reflects the sequenced gene space which is especially important for a species like pennycress where this information enables one to identify genes of interest whose manipulation via either over expression or knockdown can be predicted to confer agronomically desirable phenotypes.

Genome Annotation with MAKER

Genome annotation using the MAKER pipeline (Cantarel *et al.*, 2008) annotated 27,390 predicted protein coding genes with an average total length of 2,195 bp, average coding sequence length of 1,238 bp, and an average of 5.541 exons/gene (Supplementary Dataset S2 of Dorn et al. 2015). More than 89% of the predicted gene models are supported by at least 5 RNA sequencing (RNAseq) reads from the previously published *de novo* transcriptome assembly, while 1452 gene models (5.3%) lack RNAseq read support (Supplementary Dataset S2 of (Dorn *et al.*, 2015)). Over 85% of the predicted peptides (23,538) have at least one highly significant BLASTp hit ($e<1x10^{-5}$ and >70% positive percent) to an *A. thaliana* (TAIR10) predicted peptide (Table 4). Another 1,876 predicted peptides (6.8%) show a significant hit ($e<1x10^{-5}$), but at a lower positive percent value (>60%). Only 173 predicted peptides lacking an *A.thaliana* BLASTp hit were found (Expectation value >10) (Supplementary Dataset S2 of (Dorn *et al.*, 2015)). BLASTn analyses of the 33,873 *de novo* assembled contigs from the pennycress transcriptome (Dorn *et al.*, 2013) against the draft genome indicate a high level of completeness of the predicted gene space. Over 88% of transcripts (30,053) had BLASTn hits (≥95% identity and $e\leq1x10^{-5}$) in the genome, while 95.8% of assembled transcripts (32,458) had a significant hit ($e\leq1x10^{-5}$) but at a lower percent identity threshold (>75%) (Supplementary Dataset S3 of (Dorn *et al.*, 2015)). Only 167 transcriptome contigs lacked a BLASTn hit in the genome. (Expectation value >10) A separate BLASTn analysis of these 167 transcripts against the genomes of *Arabidopsis thaliana* (Arabidopsis Genome, 2000), *Arabidopsis lyrata* (Hu *et al.*, 2011), *Brassica rapa* (Wang

et al., 2011), *Capsella rubella* (Slotte *et al.*, 2013), and *Eutrema salsugineum (Yang et al., 2013)* revealed that 40 transcripts lacked any match to these five species, while the remaining 127 consisted of low quality and short hits, likely indicative of misassembled sequences from the de novo assembled transcriptome (Supplementary Dataset S4 of (Dorn *et al.*, 2015)). The repetitive DNA content of the final pennycress genome scaffolds was assessed by RepeatMasker (Smit, 1996-2010). It was found that repetitive elements constituted 24% of the draft genome. This analysis identified 78,812 retroelements encompassing >71 Mb of the assembly, consisting mainly of Gypsy/DIRS1 type long terminal repeats (60.3 Mb). 12,382 DNA transposons representing >5 Mb of the draft genome were also found, including hobo-Activator (0.8 Mb) and Tourist/Harbinger (0.9 Mb) type elements (Table 6). Both the raw sequencing reads and assembled sequences have been submitted to NCBI, which has been supplemented by a pennycress genome database containing a JBrowse genome browser (Westesson *et al.*, 2013), BLAST database, and data repository available at pennycress.umn.edu.

Comparative genomics of the pennycress genome assembly

Of the sequenced Brassicaceae genomes, pennycress is most closely related to *E. salsugineum*, which possesses a much smaller genome (241 Mb), but the same karyotype (n=7) (Yang *et al.*, 2013, Wu *et al.*, 2012). To evaluate the relative completeness of the genome assembly, we used a syntenic path assembly comparison of the pennycress assembly to the *E. salsugineum* reference genome(Yang *et al.*, 2013). Pennycress genomic scaffolds greater than 75 kb long, representing 241 Mb (>70%) of the assembly were compared to the 241 Mb reference genome of *E. salsugineum*. Large portions of the seven *E. salsugineum* pseudochromosomes possess a high degree of synteny with the pennycress assembly, indicative of the close evolutionary relationship between these two species, as well as a high level of completeness of conserved regions in the pennycress genome (Figure 2A).

To evaluate the quality of the genome assembly and predicted gene models, a combination of comparative analyses were used to compare the draft pennycress genome to the gene models of *A. thaliana*, *A. lyrata*, *B. rapa*, *C. rubella*, and *E. salsugineum*. BLASTp analyses of the 27,390 predicted pennycress peptides against five separate databases containing these five Brassicaceae species revealed over 23,000 of the pennycress peptides had highly similar hits in all five species, with *E. salsugineum* possessing the highest proportion (>89%) of highly similar predicted peptides (Figure 2B and Supplementary Dataset S5 of (Dorn *et al.*, 2015)). Similarly, in a BLASTp analysis of the pennycress predicted peptides against a single database containing all five

Brassicaceae species, 15,414 predicted pennycress peptides had highly similar (e $\leq 1x10^{-5}$ and $\geq$70% hit length) hits to *E. salsungieneum* (Supplementary Dataset S6). An additional 8,965 predicted peptides had highly similar hits to *B. rapa*, *A. lyrata*, *A. thaliana*, and *C. rubella,* while 2,903 BLASTp hits fell outside of these parameters, and 108 of the predicted pennycress peptides lacked a BLASTp hit (Figure 2C).

Evaluation of zygosity in the sequenced population

The draft genome was constructed using DNA isolated from nine plants that were several generations removed from a single population first identified in Coates, Minnesota. A variant detection analysis was performed to assess the degree of homozygosity among the individual plants used for sequencing (see Materials and Methods for parameters used in this analysis). We detected 131,906 single nucleotide polymorphisms (SNPs), which is approximately 1 SNP per 2.6 Kbp. Given the high level of sequencing coverage and stringent quality control, it is unlikely that the predicted SNPs were solely due to sequencing errors. We developed three hypotheses to explain the SNPs. First, these SNPs represent inappropriately assembled duplicated regions of the pennycress genome. In this analysis, if the first hypothesis were true and the declared SNPs were artificial, then we would not expect any differences between the nine different plants. Second, the SNPs may represent evidence of heterozygosity throughout the genome that would occur if the plants were prone to a high degree of outcrossing. If the second hypothesis were true, then we would expect that the nine plants would segregate at approximately 1:1:2 for the homozygosity of the presence or absence of a SNP or would be heterozygous with one chromatid containing the site and the other lacking the site, respectively. Third, distinct highly homozygous populations made up the original collection of plants used to isolate DNA that was sequenced. In this case, the genomes of any individual in the population would be expected to be largely homozygous at any particular locus.

Cleaved amplified polymorphic sequence (CAPS) (Konieczny and Ausubel, 1993) analysis using DNA isolated from progeny of the individual plants that were used to generate the draft genome was performed to distinguish between these three hypotheses. Primer sequences used to amplify regions used in the CAPS analysis shown in Figure 3A are listed in Table 7. Individuals were shown to either distinctly contain or lack the variant at four restriction enzyme sites, which eliminated the first hypothesis that these SNPs represented divergence in paralogous genes or misassembly of duplicated regions. Furthermore, none of the samples showed evidence of heterozygosity. Plants 3, 5, and 7 lacked the cut sites at the polymorphic regions and plants 1, 2, 4, 6, 8, and 9 were homozygous for the cut sites (Figure 2B). This supports the third hypothesis that the original MN106 population contained at least two distinct, highly homogenous populations. The fact that three individuals lacked all the cut sites and six individuals contained all the cut sites is likely due to the fact that at every CAPS locus, one prominent variant was detected in the variant detection analysis. Loci with the prominent variant that contained the six base restriction site were chosen for the CAPS analysis.

Based on previous analyses, a high degree of homozygosity among wild pennycress individuals was not unexpected, as the cleistogamous development of pennycress flowers (pollination occurring inside a closed flower) minimizes outcrossing (Figure 3C) (Warwick, 2002). These findings will aid pennycress breeding programs, as progeny from crosses will be expected to readily self-pollinate with minimal need for isolation. In addition, the analysis provides and extra level of validation by showing that a

subset of the identified SNPs are real and not created by sequencing or assembly errors. These SNPs have the potential to be used in breeding and geographical studies and are available in Supplementary Dataset S7. These results suggest that the original MN106 population consists of at least two distinct haplotypes on the basis of the CAPS analysis. This suggests that even in mixed populations, the selfing nature of pennycress reduces hetereozyosity. This is an important finding, as it will facilitate the isolation and generation of highly inbred populations.

Analysis of genes involved in glucosinolate metabolism

To explore the functionality of our assembly we determined how well the predicted gene space accounted for genes involved in glucosinolate biosynthesis. Glucosinolates (GSLs) are a diverse class of secondary metabolites common throughout the Brassicaceae that are important for plant/pathogen and plant/herbivore interactions (Halkier and Gershenzon, 2006). The underlying genetic mechanisms controlling the biosynthesis, transport, and breakdown of GSLs within the Brassicaceae has largely been dissected, with many of the genes responsible for this process having been characterized (Grubb and Abel, 2006). Generally, upon biosynthesis of the core glucosinolate structure (derived from one of several amino acids), GSLs generally lack bioactivity until plant tissue is damaged, leading to the hydrolysis of the GSL structure to one of several byproducts. This breakdown process is mediated by enzymes called myrosinases, resulting in an unstable aglycone (Grubb and Abel, 2006). The diversification of the final glucosinolate breakdown byproduct is mediated by several classes of specifier proteins (Kuchernig *et al.*, 2012).

Historically, pennycress has been characterized by its unique ‘garlic-like’ aroma that has been attributed to high levels of allylthiocyanate (Grmelin, 1959), which is a major hydrolysis byproduct of the major GSL in above-ground pennycress tissues: allylglucosinolate (Kuchernig *et al.*, 2011). The high level of GSLs in pennycress gives rise to another common name of the species, stinkweed (Warwick, 2002). The high level of GSLs and GSL byproducts in pennycress is of significant agronomic and economic

interest. Animal feed containing pennycress seed has traditionally been considered undesirable due to the high levels of GSLs in pennycress (Warwick, 2002). However, after seed is pressed for oil, the remaining seed meal remains high in protein (Hojilla-Evangelista *et al.*, 2013, Selling *et al.*, 2013) and presents a potential new source of billions of kilograms per year of high protein meal. If varieties with low GSL-levels in seed can be developed, there is a potential for using defatted pennycress meal as an animal feed supplement.

BLASTp analyses were used to identify putative orthologs to known glucosinolate genes in Arabidopsis (Dataset S2 of Dorn et al. 2015). Genes involved in the GSL core biosynthesis and breakdown pathway were derived from *A. thaliana*, *B. rapa*, and *B. oleraceae* (Liu *et al.*, 2014). Putative orthologs were identified for GSL biosynthesis from methionine and tryptophan (Figure 3A), along with several putative myrosinases (thioglucoside glucohydrolase – TGGs and atypical myrosinases – PEN2 and PEN3). Potential orthologs to several interesting specifier proteins were also identified (Figure 4A – yellow box). To obtain a semi-quantitative estimate of these predicted genes, RNAseq reads from the previously described transcriptome were used to obtain rough expression values for each gene model (Supplementary Dataset S1 of Dorn et al. 2015). As these RNAseq reads represent a global library representing various tissues, this analysis provides an initial probe into the pennycress genes potentially responsible for the unique glucosinolate composition of pennycress. For example, two putative myrosinases (Ta16900 and Ta16899) represent the 120th and 129th most highly

expressed gene models, respectively (Figure 4A, Figure 4B, and Supplementary Dataset S1). A MODIFIED VACUOLE PHENOTYPE1 (MVP1)-like pennycress gene was also identified (Ta16960- Figure 4A). MVP1 in Arabidopsis interacts with the myrosinase TGG2 to modulate myrosinase activity (Agee *et al.*, 2010).

Potential orthologs to many important specifier proteins were also identified in this analysis and help explain the unique GSL activity in pennycress. For example, EPITHIOSPECIFIER MODIFIER1 (ESM1) in Arabidopsis represses the formation of nitriles and favors isothiocyanate production (Zhang *et al.*, 2006). An ESM1-like pennycress gene, Ta16274 (Figure 4A), was among the top 200 expressed genes models, and could explain the high levels of isothiocyanate in pennycress (Kuchernig *et al.*, 2011). Additionally, another potential hydrolysis product of GSLs, epithionitriles, can only be formed from GSLs possessing terminal double bonds in their side chain (allylglucosinolate, for example). As allylglucosinolate represents over 90% of GSLs in above ground tissue (Kuchernig *et al.*, 2011), this ESM1-like pennycress gene could represent a key GSL specifier protein also responsible for epithionitrile production.

The evolution of the identified specifier proteins was also examined. It has been previously reported that GSL-related genes commonly exist in tandemly-duplicated regions of the genome within the Brassicaceae (Hofberger *et al.*, 2013, Liu *et al.*, 2014). The predicted peptides of *A. thaliana* TGG1 (Figure 4B), MVP1 (Figure 4C), ESP (Figure 4D), and ESM1 (Figure 4E) and corresponding orthologs in pennycress, *B. rapa*,

and *B. oleraceae* were compared to determine sequence similarity and retention of tandemly duplicated regions throughout the genome. We identified three putative pennycress myrosinases with high sequence similarity to the Arabidopsis myrosinases TGG1 (AT5G26000) (Figure 4B), with an additional gene (Ta16899) being more similar to *B. oleraceae* TGG2-like genes (Bol9025706 and Bol017328). Only a single MVP1-like pennycress gene was identified in the draft genome (Figure 4C). Three orthologous predicted peptides for AtESP (AT1G54040) were identified (Figure 4D). The previously described TaTFP (Kuchernig *et al.*, 2011), which was cloned from a cDNA library, had a top BLASTp hit to Ta16961. However, the Ta16961 predicted peptide is nearly twice as long as the TaTFP predicted peptide (NCBI Accession JN244735), and appears to indicate either the TaTFP cDNA represented a unique splice variant or an erroneous splice site prediction from the annotation pipeline. An adjacent gene model, Ta16962, also had high similarity to TaTFP (Figure 4D). Two tandem ESM1-like genes were also identified, similar to the tandemly duplicated orthologs in *B. rapa* (Bra027358 and Bra027359) and *B. oleraceae* (Bol005066 and Bol005067) (Figure 4E).

Several orthologs to important GSL transport genes were also identified. The GSL transporters GLUCOSINOLATE TRANSPORTER (GTR) 1 and 2 have been shown to serve as major transporters responsible for the loading of GSLs into developing seeds in Arabidopsis, and loss of function mutations in these genes result in significantly reduced levels of GSLs in seeds (Nour-Eldin *et al.*, 2012, Andersen *et al.*, 2013). The

identification of putative GTR1 and GTR2 orthologs (Figure 4F) represent important targets for improvement in our ongoing breeding program.

Identification of predicted orthologs of Arabidopsis genes that confer desirable phenotypes when mutated

There are several important traits that will need to be addressed during the domestication process to make pennycress a viable crop, such as reducing seed dormancy and increasing rates of flowering. In Arabidopsis, complete or partial loss of function mutations in a number of genes can confer agronomically desirable traits such as increased seed size, improved seed oil composition, earlier flowering and reduced seed dormancy, seed glucosinolate content and seed pod shatter (Table 8). Here we show that predicted orthologs of these genes are present in our draft assembly (Table 8). While this is not an exhaustive list of potential gene targets, this represents a number of well-characterized genes responsible for controlling these phenotypes. Mutations in these genes in Arabidopsis can confer what would be considered agronomically desirable phenotypes. For example, pennycress GTR1 and GTR2 predicted orthologs are highly conserved at the predicted peptide level (88.5% and 94.8% positive match, respectively). Additionally, identifying natural or induced mutations in key regulatory genes like DELAY OF GERMINATION 1 (Bentsink *et al.*, 2006, Nakabayashi *et al.*, 2012) could lead to reduced levels of seed dormancy in pennycress. Similarly, developing winter annual lines that flower and mature rapidly in the spring should be possible through targeting genes involved in the vernalization and photoperiodic flowering time pathways, such as FLOWERING LOCUS C and FRIGIDA(Dorn *et al.*, 2013) (Table 8 and Supplementary Dataset S8 of Dorn et al. 2015) . The development of rapidly maturing lines will be particularly important in the upper Midwestern United States, as the shorter

spring growing seasons at northern latitudes could limit the widespread adaptation of pennycress as a winter cover crop. By targeting specific important agronomic traits, and the underlying genetic mechanisms controlling these traits through the approaches mentioned above, we are poised to rapidly convert a wild weed species into a new highly valuable and sustainable winter oilseed crop.

The future of genomics-based pennycress improvement

While the annotated draft genome for pennycress presented in this report will provide new tools for the domestication of a new winter oilseed crop, there is still much work to be done to develop a broad base of genomic resources for pennycress. For example, the creation of a complete, anchored reference genome is needed. The increasing length and quality of DNA sequencing reads from third generation (single molecule – PacBio or Oxford Nanopore) sequencing technologies and new library preparation techniques such as Illumina TruSeq Synthetic Long-Reads (LRseq, formerly Moleculo) (Voskoboynik *et al.*, 2013) will allow for the *de novo* assembly of highly contiguous yet complex genomes. Combined with our ongoing development of a high-density genetic map using restriction site-associated DNA sequencing (RADseq) (Poland and Rife, 2012, Peterson *et al.*, 2012), the anchoring and ordering of a highly contiguous genome assembly to the genetic linkage map should be on the immediate horizon to produce a complete reference genome for pennycress. With a completed pennycress genome, the phylogenetic resolution of the Brassicaceae expanded lineage 2 can be improved, including providing tools for the phylogenetic resolution of *Thlaspi* species (Koch, 2004), and answering questions on pennycress genome structure. The draft pennycress genome presented here will enable the development of a genomics-based breeding program. In addition, the identification of the gene space will allow the use of techniques such as mutation breeding, TILLING of ethyl methanesulfonate-mutagenized populations, DeTILLING of fast neutron-mutagenized populations, and genomic selection. The use of precise genome editing techniques such as CRISPR/Cas9 and TALENs should also be on the horizon, as

pennycress can be transformed using the *Agrobacterium*-mediated vacuum infiltration floral dip method (John Sedbrook, personal communication). Pennycress has the potential to be planted on over 16 million hectares in the United States alone and produce over 22 billion liters of oil suitable as a biodiesel feedstock (Winchester, 2013, Fan *et al.*, 2013, Moser *et al.*, 2009a). The successful domestication of pennycress has the potential to benefit farmers and the environment, improve energy security, and reduce greenhouse gases associated with global climate change. The genomics-based domestication of pennycress represents an exciting example of the development of a new crop species.

	Accession	Run
Thlaspi arvense line MN106 – Illumina 100 bp paired-end	SRX380646	SRR1034657
Thlaspi arvense line MN106 – Illumina 250 bp paired-end	SRX380649	SRR1034659
Thlaspi arvense line MN106 – Illumina 50 bp mate pair (2 kb insert)	SRX381531	SRR1035703
Thlaspi arvense line MN106 – Illumina 50 bp mate pair (3.5 kb insert)	SRX381541	SRR1035705
Thlaspi arvense line MN106 – Illumina 50 bp mate pair (7 kb insert)	SRX381551	SRR1035715
Thlaspi arvense line MN106 – PacBio RSII 10 kb insert reads	SRX380881	SRR1035588

Table 3-1: Pennycress Genome Sequencing Read Data Deposition

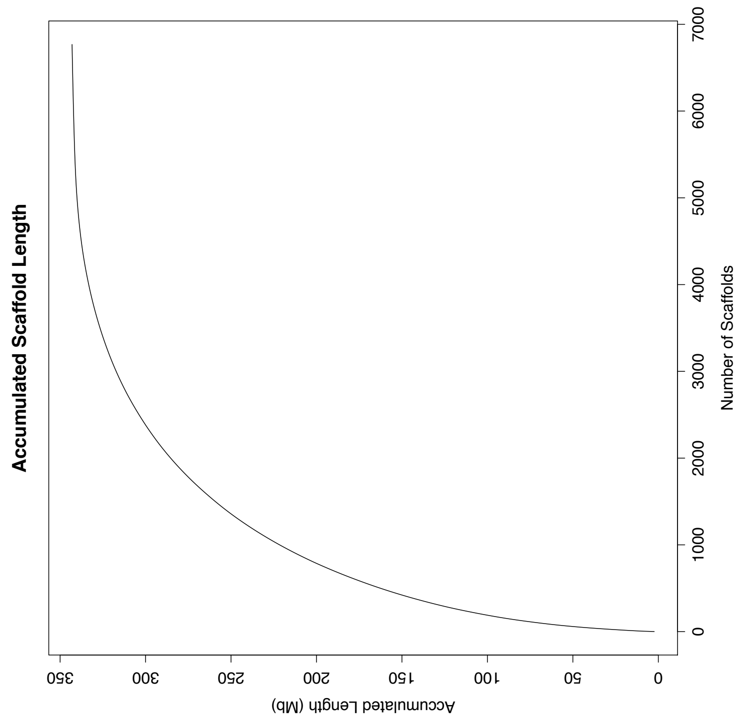
The Illumina and Pacific Biosciences sequencing reads for *Thlaspi arvense* line MN106 are available in the NCBI Sequence Read Archive under accession number SRP033211. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AZNP00000000. The version described in this paper is version AZNP01000000. The genome assembly and related annotation files can also be downloaded from pennycress.umn.edu.

Component	Model and Manufacturer
CPU	Intel Core i7-3820 (Intel)
Motherboard	Asus Sabertooth X79 motherboard (Asustek Computer Inc.)
RAM	(2) G.SKILL Ripjaws Z Series 32 GB (4x8GB) Memory – Model F3-12800CL10Q-32GBZL (G.SKILL)
CPU Cooler	Cooler Master Hyper 212 EVO RR-212E-20PK-R2 CPU Cooler (Cooler Master)
Boot Hard Drive	256 GB Crucial M4 CT256M4SSD2 Solid State Hard Drive (Crucial)
Main Hard Drive	2 TB Western Digital WD Black WD2002FAEX Hard Drive (Western Digital)
SSD Caching	128 GB Samsung 840 Pro Series MZ-7PD128BW Solid State Hard Drive (Samsung)
Power Supply	Corsair Enthusiast Series TX650 V2 650W Power Supply
Video Card	EVGA 02G-P4-3653-KR GeForce GTX 650 Ti 2GB PCIe3.0 Video Card (EVGA)
Optical Drive	LITE-ON iHDS118-04 DVD/CD-ROM
Case	Corsair Carbide Series 500R Computer Case

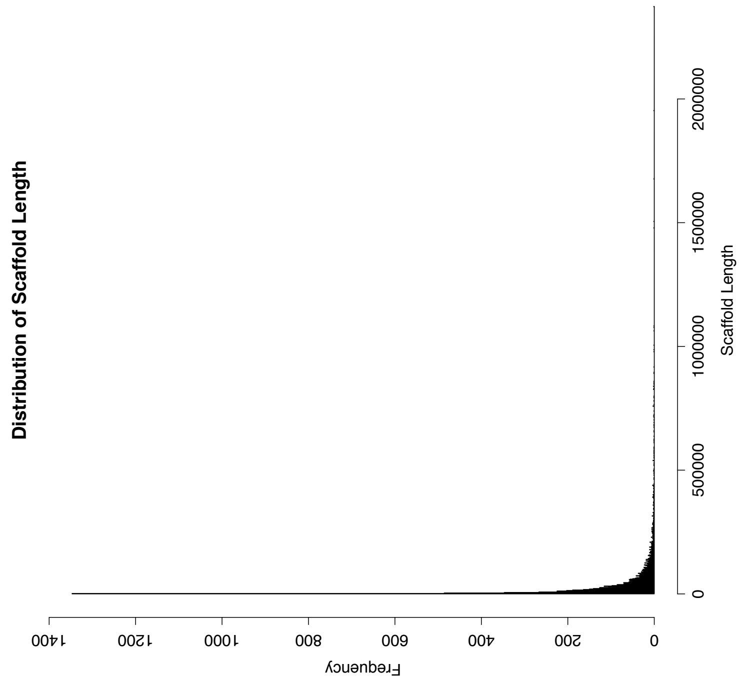
Table 3-2: Component list of computer used for pennycress genome assembly

Component	Model and Manufacturer
CPU	(2) Opteron 6344, 12-core 2.6ghz (AMD)
Motherboard	KGPE-D16 SSI EEB Server Motherboard (Asus)
RAM	(4 x 16GB) ValueRAM ECC Registered KVR16R11D4K4 (Kingston)
CPU Cooler	4-Dual Heat Pipe SSO Bearing CPU Cooler - NH-U9DO (Noctua)
Boot Hard Drive	500 GB - 840 Series Solid State Hard Drive - MZ-7TD500BW (Samsung)
Main Hard Drive	Barracuda 2 TB Hard Drive - ST2000DM001 (Seagate)

Table 3-3: Component list of computer used for pennycress genome annotation



Accumulated Scaffold Length
Accumulated Length (Mb)
Number of Scaffolds
0
50
100
150
200
250
300
350
0
1000
2000
3000
4000
5000
6000
7000



Distribution of Scaffold Length
Frequency
Scaffold Length
0
200
400
600
800
1000
1200
1400
0
500000
1000000
1500000
2000000

Figure 3-1: Distribution and accumulated length of genomic scaffolds.

A. Distribution of de novo assembled *Thlaspi arvense* genomic scaffolds.

B.) Accumulated length of genomic scaffolds. Over 85% of the total assembly length (>343MB) consists of the 3,000 longest scaffolds.

T. arvense Genome Sequencing

Library Type	Number of Reads	# Nucleotides after QC (Mbp)
Illumina HiSeq 2000 (2x100bp Paired-end)	352,394,426	33,190.50
Illumina MiSeq (2x250bp Paired-end)	6,291,688 (merged) 8,548,686 (unmerged)	2,589.33 1,904.90
Illumina HiSeq 2000 (2x50bp Mate-Pair) 2, 3, 5 kb inserts	209,815,249	9,550.66
PacBio SMRT Cell (4 cells)	110,751	214.78

T. arvense Genome Assembly

	Scaffolds	Contigs in Scaffolds
Number	6,768	44,109
Mean length (bp)	50,681	7,375
N50 (bp)	140,815	21,096
Total assembly length (bp)	343,012,389	325,295,785

T. arvense Genome Annotation

Number of Gene Models	27,390
- Gene Models highly similar to TAIR10 peptide ($e<1x10^{-5}$ and >70% positive percent)	85.94%
GC Content	37.99%
Repetitive DNA sequence	24.38%
-Retroelements (20.94%)	78,812
-DNA transposons (1.6%)	12,382

Table 3-4 – Genome sequencing, assembly, and annotation statistics

Summary statistics for DNA sequencing libraries, genome assembly and scaffolding, and genome annotation

Complete genome assembly statistics	
v1.0 pennycress genome assembly scaffolds	
Scaffold Statistics	Value
Number of Scaffolds	6,768
Cumulative Scaffold Length	343,012,389
Number of Scaffolds >1000 bp	6,731
Number of Scaffolds >10,000 bp	4,099
Number of Scaffolds >100,000 bp	902
Number of Scaffolds >1,000,000	9
Mean Scaffold Size	50,681
Median Scaffold Size	17,246
N50 Scaffold Size	140,815
L50 Scaffold Count	561
Average number of contigs per scaffold	4
Average length of breaks (>25 Ns) between contigs in scaffolds	807
%N's in scaffolds	5.17%
Scaffolded contig statistics	
Number of Contigs in final scaffolds	44,109
Cumulative Contig Length (bp)	25,295,785
Number of Contigs >1000 bp	38,572
Number of Contigs >10,000 bp	7,953
Number of Contigs >100,000 bp	238
Mean Contig Size	7,375
Median Contig Size	2,458
N50 Contig Size	21,096
L50 Contig Count	3,513
Small initial scaffold statistics (scaffolds <1000bp from initial CLC assembly)	
Number of scaffolds	156662
Cummulative Scaffold Length	69241306
Longest scaffold	999
Shortest Scaffold	200
Mean Scaffold Size	442
Median Scaffold Size	386
N50 Scaffold Size	505
%N's	0.43%

Table 3-5: Complete genome assembly and scaffolding statistics

Type	Number of elements
Retroelements	78,812
-SINEs	459
-LINEs	3,902
-L1/CIN4	3847
-LTR elements	74,451
-Ty1/Copia	6,214
-Gypsy/DIRS1	66,376
DNA transposons	12,382
-hobo-Activator	2,447
-Tc1-IS630-Pogo	1,651
-En-Spm	140
-Tourist/Harbinger	1,914
Unclassified	2,882
Small RNA	482
Satellites	135
Simple repeats	84,548
Low complexity	23,616

Table 3-6: Repeat Identification Statistics from RepeatMasker.

Location of CAPS marker	Forward Primer Sequence	Reverse Primer Sequence	Restriction Enzyme
Scaffold 99	CATCAACTGTCATTGTAGAGC	TGTGAGCGATTCAGGACTCTTAA	PstI
Scaffold 109	GGGTTTGATGATTAGGTCAAG	CAGACTAGGACAATATCTCTC	BamHI
Scaffold 236	GTACGCTGACCATAATAGATAG	CTGTAAAGTGTAATTTTCTCTGCTG	BclI
Scaffold 92	TATGTCGAACAGGGAGAG	AAACTGTGGTTGCTCTCG	SacI

Table 3-7: CAPS primer sequences and CAPS locations

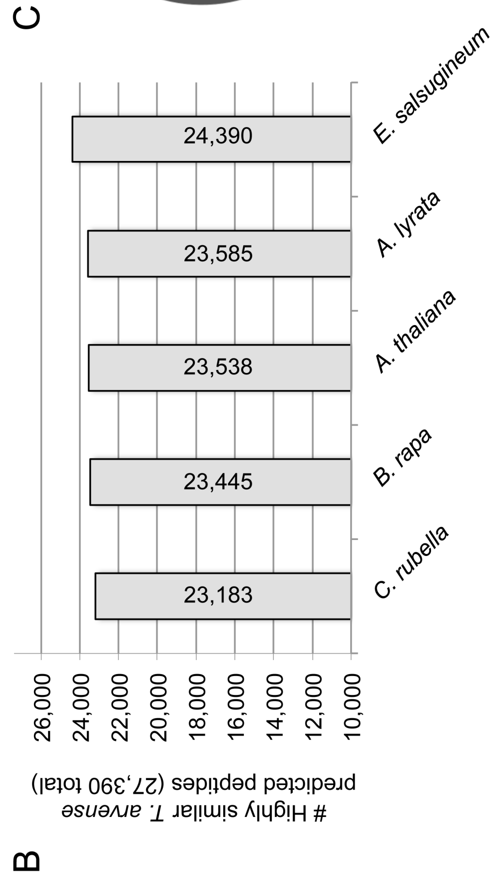
Trait of Interest	Arabidopsis Gene	AGI #	Corresponding Pennycress Predicted Peptide ID	Pennycress Genome Scaffold #	Arabidopsis Peptide Length	Pennycress Peptide Length	% Identity	% Positive
Seed Dormancy	DOG1	AT5G45830	Ta16411	141	291	284	72.9	80.9
	ABI3	AT3G24650	Ta24539	630	720	728	86.8	90.6
	ABI4	AT2G40220	Ta07356	85	328	395	59.3	66.3
	ABI5	AT2G36270	Ta25088	828	442	434	85.8	89
Seed Size	WRI1	AT3G54320	Ta07949	20	438	441	80.2	83.4
	DGAT1	AT3G51520	Ta05453	12	314	566	67.7	70.1
	IKU2	AT3G19700	Ta01151	3	991	1016	79.7	86.8
	KLU	AT1G13710	Ta14711	213	517	488	84.8	88.2
	GL2	AT1G79840	Ta08884	25	776	798	82.6	86
	MUM4	AT1G53500	Ta23332	457	667	597	83.6	85.8
Seed GSL Content	HAG1	AT3G54610	Ta07928	20	568	592	85.8	88.3
	GTR1	AT3G47960	Ta20037	206	636	651	81	88.5
	GTR2	AT5G62680	Ta01982	4	616	613	90.3	94.8
Erucic Acid Content	FAE1	AT4G34520	Ta11742*	45	506	506	86.4	91.5
Linoleic/Linolienic Acid Content	FAD2	AT3G12120	Ta12495	59	383	404	84.3	88.2
Seed Pod Shatter	SHP1	AT3G58780	Ta15094	22	273	248	86.8	89.4
	SHP2	AT2G42830	Ta08438	22	248	290	76.6	80.7
	IND	AT4G00120	Ta25465	1003	198	172	61.5	69
	ALC	AT5G67110	Ta02444	6	210	207	72.8	77.9
	RPL	AT5G02030	Ta15425	92	575	639	80.5	83.3
	FUL	AT5G60910	Ta01807	57	242	208	63.3	68.1
Time To Flower	FLC	AT5G10140	Ta00917	1	196	203	84.3	89.7
	FRI	AT4G00650	Ta26225*	1344	314	367	55	65.8

* MAKER-derived gene model contained errors and were manually corrected to obtain predicted peptide sequence

Table 3-8: Putative orthologs controlling important agronomic traits in pennycress. Identification of pennycress genes with high sequence similarity to Arabidopsis genes controlling key traits of interest, including genomic location and predicted peptide similarity

A

B



C

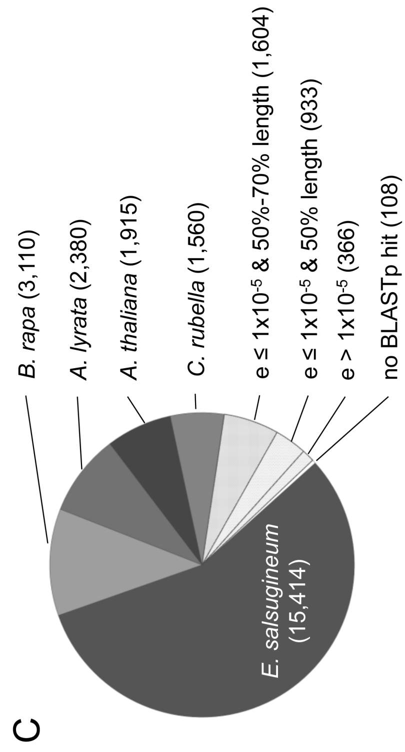
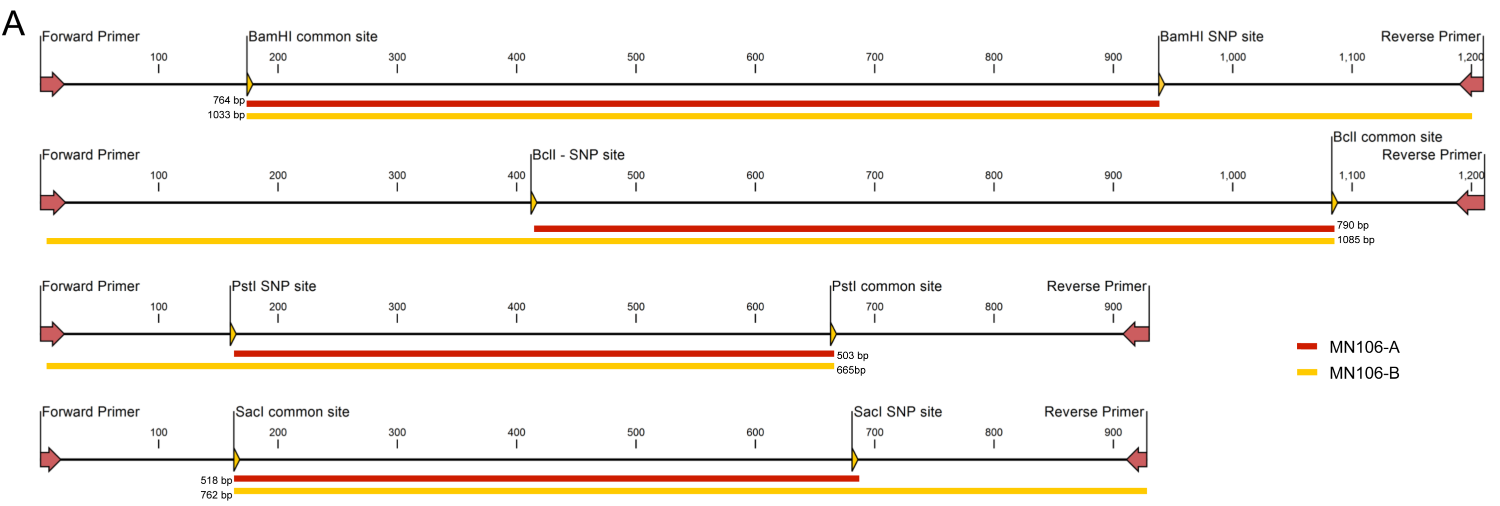


Figure 3-2: Comparative genomics of pennycress and other Brassicaceae species

(A) Syntenic path assembly dotplots comparing pennycress scaffolds >75 kilobases long to the seven *Eutrema salsugineum* pseudo-chromosomes from Yang et al. (2013). **(B)** BLASTp analysis of the 27,390 predicted pennycress peptides *against predicted peptide sets from Capsella rubella (Slotte et al., 2013), Brassica rapa (Wang et al., 2011), Arabidopsis thaliana (Bevan and Initiative, 2000), Arabidopsis lyrata (Hu et al., 2011b),* and *E. salsugineum*. Highly similar is defined as pennycress predicted peptide having at least one BLASTp hit $e < 1x10^{-5}$ and positive sequence similarity >70%. **(C)** BLASTp analysis of predicted pennycress peptides against a protein database containing the predicted peptides of the five Brassicaceae species listed. Predicted peptides with top hits ($e \leq 1x10^{-5}$ and >70% hit length) to a predicted protein from the corresponding species are shown, with pennycress peptides with hits falling below this threshold shown in the lower right half of the piechart.


A
Forward Primer
BamHI common site
BamHI SNP site
Reverse Primer
764 bp
1033 bp
BclI - SNP site
BclI common site
790 bp
1085 bp
PstI SNP site
PstI common site
503 bp
665bp
SacI common site
SacI SNP site
518 bp
762 bp
MN106-A
MN106-B

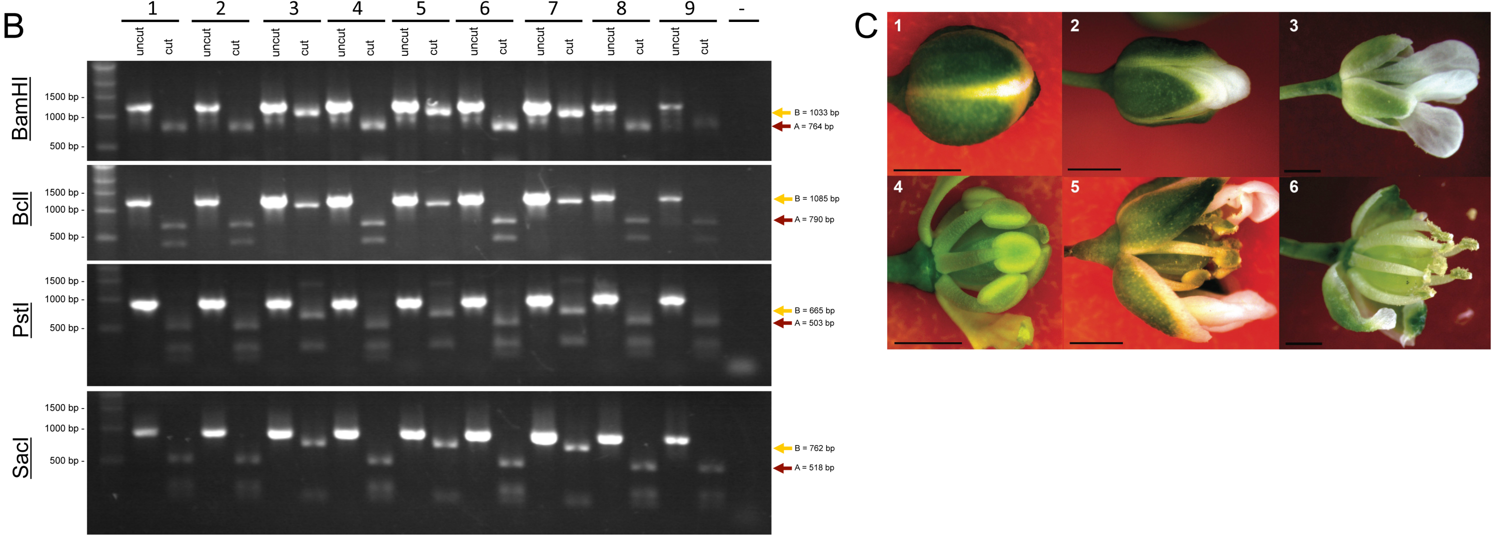

B
BamHI
BclI
PstI
SacI
uncut
cut
1500 bp
1000 bp
500 bp
B = 1033 bp
A = 764 bp
B = 1085 bp
A = 790 bp
B = 665 bp
A = 503 bp
B = 762 bp
A = 518 bp
C

Figure 3-3: CAPS Analysis of *Thlaspi arve*nse Line MN106 (A) Schematic of the four PCR fragments produced by the primer sets listed in Supplementary Table S4. The largest fragments used to distinguish between individuals containing the SNP are shown in maroon (corresponding to the MN106-A genotype) and gold (MN106-B genotype). (B) DNA was isolated from progeny of each of the nine plants used to produce the draft genome assembly and analyzed using four CAPS markers. PCR products for each plant are shown side-by-side undigested (uncut) and post-digestion (cut) with the corresponding restriction endonucleases. In all cases, samples 3, 5, and 7 share restriction digest patterns, corresponding to the MN106-B genotype. A negative control for the PCR reaction is shown in the last lane. (C) Morphology of developing *Thlaspi arvense* flowers. The top panel (1,3,5) shows the morphology of the unaltered flowers, while the bottom panel (2,4,6) shows the same series of flowers with sepals and petals either removed or rearranged to reveal the status of the stamens with regards to filament elongation and the shedding of pollen. 2.) Neither filament elongation nor pollen shedding have commenced in (1). 4.) Filaments have elongated and pollen is being shed inside of the closed flower shown in (3). 5.) Pollen densely covers the stigmatic surface by the time the flower is fully open in (6). All scale bars equal 1 millimeter.

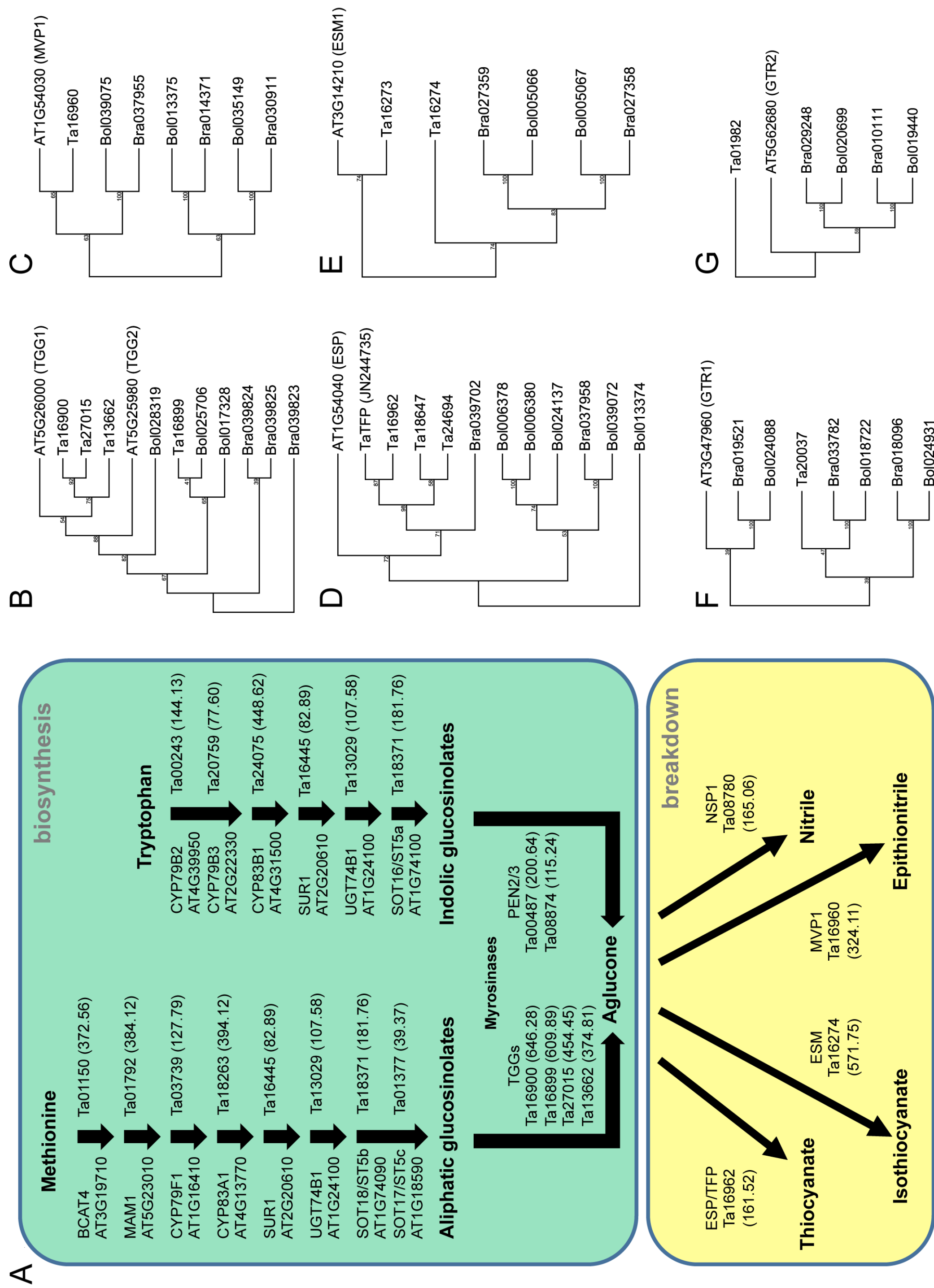

A
biosynthesis
Methionine
BCAT4 AT3G19710
Ta01150 (372.56)
MAM1 AT5G23010
Ta01792 (384.12)
CYP79F1 AT1G16410
Ta03739 (127.79)
CYP83A1 AT4G13770
Ta18263 (394.12)
SUR1 AT2G20610
Ta16445 (82.89)
UGT74B1 AT1G24100
Ta13029 (107.58)
SOT18/ST5b AT1G74090
Ta18371 (181.76)
SOT17/ST5c AT1G18590
Ta01377 (39.37)
Aliphatic glucosinolates
Tryptophan
CYP79B2 AT4G39950
Ta00243 (144.13)
CYP79B3 AT2G22330
Ta20759 (77.60)
CYP83B1 AT4G31500
Ta24075 (448.62)
SUR1 AT2G20610
Ta16445 (82.89)
UGT74B1 AT1G24100
Ta13029 (107.58)
SOT16/ST5a AT1G74100
Ta18371 (181.76)
Indolic glucosinolates
Myrosinases
TGGs
Ta16900 (646.28)
Ta16899 (609.89)
Ta27015 (454.45)
Ta13662 (374.81)
PEN2/3
Ta00487 (200.64)
Ta08874 (115.24)
Aglucone
breakdown
ESP/TFP Ta16962 (161.52)
Thiocyanate
ESM Ta16274 (571.75)
Isothiocyanate
MVP1 Ta16960 (324.11)
Epithionitrile
NSP1 Ta08780 (165.06)
Nitrile
B
AT5G26000 (TGG1)
Ta16900
Ta27015
Ta13662
AT5G25980 (TGG2)
Bol028319
Ta16899
Bol025706
Bol017328
Bra039824
Bra039825
Bra039823
C
AT1G54030 (MVP1)
Ta16960
Bol039075
Bra037955
Bol013375
Bra014371
Bol035149
Bra030911
D
AT1G54040 (ESP)
TaTFP (JN244735)
Ta16962
Ta18647
Ta24694
Bra039702
Bol006378
Bol006380
Bol024137
Bra037958
Bol039072
Bol013374
E
AT3G14210 (ESM1)
Ta16273
Ta16274
Bra027359
Bol005066
Bol005067
Bra027358
F
AT3G47960 (GTR1)
Bra019521
Bol024088
Ta20037
Bra033782
Bol018722
Bra018096
Bol024931
G
Ta01982
AT5G62680 (GTR2)
Bra029248
Bol020699
Bra010111
Bol019440

Figure 3-4: Analysis of genes involved in glucosinolate metabolism and transport

(A) Overview of glucosinolate biosynthesis core structure (green) via methionine and tryptophan and breakdown (yellow) and corresponding orthologs in the pennycress genome - pathway derived from Liu et al., (2014). Expression values (RPKM, in parentheses) are shown for each putative ortholog derived from the global RNAseq reads previously described (Dorn et al., 2013). **(B-G)** Neighbor-joining trees of TGG1/TGG2, MVP1, ESP, ESM1, GTR1, and GTR2-like predicted peptides (100 bootstrap replicates) from pennycress (identified in this study), *B. rapa* and *B. oleraceae* (Liu et al., 2014).

References

A. E. Agee, M. Surpin, E. J. Sohn, T. Girke, A. Rosado, B. W. Kram, C. Carter, A. M. Wentzell, D. J. Kliebenstein, H. C. Jin, O. K. Park, H. Jin, G. R. Hicks & N. V. Raikhel 2010. MODIFIED VACUOLE PHENOTYPE1 is an Arabidopsis myrosinase-associated protein involved in endomembrane protein trafficking. *Plant Physiol,* 152**,** 120-32.

T. G. Andersen, H. H. Nour-Eldin, V. L. Fuller, C. E. Olsen, M. Burow & B. A. Halkier 2013. Integration of biosynthesis and long-distance transport establish organ-specific glucosinolate profiles in vegetative Arabidopsis. *Plant Cell,* 25**,** 3133-45.

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

S. Baud & L. Lepiniec 2009. Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. *Plant Physiol Biochem,* 47**,** 448-55.

K. Belhaj, A. Chaparro-Garcia, S. Kamoun & V. Nekrasov 2013. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods,* 9**,** 39.

L. Bentsink, J. Jowett, C. J. Hanhart & M. Koornneef 2006. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proc Natl Acad Sci U S A,* 103**,** 17042-7.

Kf Best & Gi Mcintyre 1975. The Biology of Canadian Weeds 9. Thlaspi arvense L. . *Canadian Journal of Plant Science,* 55**,** 279-292.

M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler & W. Pirovano 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics,* 27**,** 578-9.

M. Boetzer & W. Pirovano 2012. Toward almost closed genomes with GapFiller. *Genome Biol,* 13**,** R56.

B. L. Cantarel, I. Korf, S. M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sanchez Alvarado & M. Yandell 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res,* 18**,** 188-96.

C. H. Chen, T. C. Kuo, M. H. Yang, T. Y. Chien, M. J. Chu, L. C. Huang, C. Y. Chen, H. F. Lo, S. T. Jeng & L. F. Chen 2014. Identification of cucurbitacins and assembly of a draft genome for Aquilaria agallocha. *BMC Genomics,* 15**,** 578.

K. M. Dorn, J. D. Fankhauser, D. L. Wyse & M. D. Marks 2013. De novo assembly of the pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant J,* 75**,** 1028-38.

K. M. Dorn, J. D. Fankhauser, D. L. Wyse & M. D. Marks 2015. A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *DNA Res,* 22**,** 121-131.

Roque L. Evangelista, Terry A. Isbell & Steven C. Cermak 2012. Extraction of pennycress (Thlaspi arvense L.) seed oil by full pressing. *Industrial Crops and Products,* 37**,** 76-81.

J. Q. Fan, D. R. Shonnard, T. N. Kalnes, P. B. Johnsen & S. Rao 2013. A life cycle assessment of pennycress (Thlaspi aruense L.) -derived jet fuel and diesel. *Biomass & Bioenergy,* 55**,** 87-100.

A. Franzke, M. A. Lysak, I. A. Al-Shehbaz, M. A. Koch & K. Mummenhoff 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci,* 16**,** 108-116.

R and Virtanen Grmelin, A 1959 A new type of enzymatic cleavage of mustard oil glucosides. Formation of allylthiocyanate in Thlasip arvense L and Benzylthiocynate in Lepidium ruderale L and Lepidium sativum L. . *Acta Chem. Scand,* 13.

C. D. Grubb & S. Abel 2006. Glucosinolate metabolism and its control. *Trends Plant Sci,* 11**,** 89-100.

S. Guo, J. Zhang, H. Sun, J. Salse, W. J. Lucas, H. Zhang, Y. Zheng, L. Mao, Y. Ren, Z. Wang, J. Min, X. Guo, F. Murat, B. K. Ham, Z. Zhang, S. Gao, M. Huang, Y. Xu, S. Zhong, A. Bombarely, L. A. Mueller, H. Zhao, H. He, Y. Zhang, Z. Zhang, S. Huang, T. Tan, E. Pang, K. Lin, Q. Hu, H. Kuang, P. Ni, B. Wang, J. Liu, Q. Kou, W. Hou, X. Zou, J. Jiang, G. Gong, K. Klee, H. Schoof, Y. Huang, X. Hu, S. Dong, D. Liang, J. Wang, K. Wu, Y. Xia, X. Zhao, Z. Zheng, M. Xing, X. Liang, B. Huang, T. Lv, J. Wang, Y. Yin, H. Yi, R. Li, M. Wu, A. Levi, X. Zhang, J. J. Giovannoni, J. Wang, Y. Li, Z. Fei & Y. Xu 2013. The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. *Nat Genet,* 45**,** 51-8.

B. A. Halkier & J. Gershenzon 2006. Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol,* 57**,** 303-33.

J. A. Hofberger, E. Lyons, P. P. Edger, J. Chris Pires & M. Eric Schranz 2013. Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family. *Genome Biol Evol,* 5**,** 2155-73.

M. P. Hojilla-Evangelista, R. L. Evangelista, T. A. Isbell & G. W. Selling 2013. Effects of cold-pressing and seed cooking on functional properties of protein in pennycress (Thlaspi arvense L.) seed and press cakes. *Industrial Crops and Products,* 45**,** 223-229.

M. J. Holdsworth, L. Bentsink & W. J. Soppe 2008. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytol,* 179**,** 33-54.

T. T. Hu, P. Pattyn, E. G. Bakker, J. Cao, J. F. Cheng, R. M. Clark, N. Fahlgren, J. A. Fawcett, J. Grimwood, H. Gundlach, G. Haberer, J. D. Hollister, S. Ossowski, R. P. Ottilar, A. A. Salamov, K. Schneeberger, M. Spannagl, X. Wang, L. Yang, M. E. Nasrallah, J. Bergelson, J. C. Carrington, B. S. Gaut, J. Schmutz, K. F. Mayer, Y. Van De Peer, I. V. Grigoriev, M. Nordborg, D. Weigel & Y. L. Guo 2011. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat Genet,* 43**,** 476-81.

Consortium International Barley Genome Sequencing, K. F. Mayer, R. Waugh, J. W. Brown, A. Schulman, P. Langridge, M. Platzer, G. B. Fincher, G. J. Muehlbauer, K. Sato, T. J. Close, R. P. Wise & N. Stein 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature,* 491**,** 711-6.

J. S. Johnston, A. E. Pepper, A. E. Hall, Z. J. Chen, G. Hodnett, J. Drabek, R. Lopez & H. J. Price 2005. Evolution of genome size in Brassicaceae. *Ann Bot,* 95**,** 229-35.

C. Jung & A. E. Muller 2009. Flowering time control and applications in plant breeding. *Trends Plant Sci,* 14**,** 563-73.

S. Kagale, C. S. Koh, J. Nixon, V. Bollina, W. E. Clarke, R. Tuteja, C. Spillane, S. J. Robinson, M. G. Links, C. Clarke, E. E. Higgins, T. Huebert, A. G. Sharpe & I.

A. P. Parkin 2014. The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nat Commun,* 5.

H. ; Taab Karimmojeni, A.; Rashidi, B.; Bazrafshan, Ah. 2014. Dormancy breaking and seed germination of the annual weeds Thlaspi arvense, Descurainia sophia and Malcolmia africana (Brassicaceae). *JOURNAL OF PLANT PROTECTION RESEARCH,* 54**,** 179-187.

M and Al-Shehbaz Koch, I 2004. Taxonomic and Phylogenetic Evaluation of the American "Thlaspi" species: Identity and relationship to the Eurasian genus Noccaea (Brassicaceae). *Systematic Botany,* 29**,** 375-384.

M. A. Koch & D. A. German 2013. Taxonomy and systematics are key to biological information: Arabidopsis, Eutrema (Thellungiella), Noccaea and Schrenkiella (Brassicaceae) as examples. *Front Plant Sci,* 4.

A. Konieczny & F. M. Ausubel 1993. A Procedure for Mapping Arabidopsis Mutations Using Codominant Ecotype-Specific Pcr-Based Markers. *Plant Journal,* 4**,** 403-410.

I. Korf 2004. Gene finding in novel genomes. *BMC Bioinformatics,* 5**,** 59.

J. C. Kuchernig, A. Backenkohler, M. Lubbecke, M. Burow & U. Wittstock 2011. A thiocyanate-forming protein generates multiple products upon allylglucosinolate breakdown in Thlaspi arvense. *Phytochemistry,* 72**,** 1699-709.

J. C. Kuchernig, M. Burow & U. Wittstock 2012. Evolution of specifier proteins in glucosinolate-containing plants. *BMC Evol Biol,* 12**,** 127.

S. Liu, Y. Liu, X. Yang, C. Tong, D. Edwards, I. A. Parkin, M. Zhao, J. Ma, J. Yu, S. Huang, X. Wang, J. Wang, K. Lu, Z. Fang, I. Bancroft, T. J. Yang, Q. Hu, X. Wang, Z. Yue, H. Li, L. Yang, J. Wu, Q. Zhou, W. Wang, G. J. King, J. C. Pires, C. Lu, Z. Wu, P. Sampath, Z. Wang, H. Guo, S. Pan, L. Yang, J. Min, D. Zhang, D. Jin, W. Li, H. Belcram, J. Tu, M. Guan, C. Qi, D. Du, J. Li, L. Jiang, J. Batley, A. G. Sharpe, B. S. Park, P. Ruperao, F. Cheng, N. E. Waminal, Y. Huang, C. Dong, L. Wang, J. Li, Z. Hu, M. Zhuang, Y. Huang, J. Huang, J. Shi, D. Mei, J. Liu, T. H. Lee, J. Wang, H. Jin, Z. Li, X. Li, J. Zhang, L. Xiao, Y. Zhou, Z. Liu, X. Liu, R. Qin, X. Tang, W. Liu, Y. Wang, Y. Zhang, J. Lee, H. H. Kim, F. Denoeud, X. Xu, X. Liang, W. Hua, X. Wang, J. Wang, B. Chalhoub & A. H. Paterson 2014. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun,* 5**,** 3930.

C. M. Mccallum, L. Comai, E. A. Greene & S. Henikoff 2000. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol,* 123**,** 439-42.

M. J. Milner, N. Mitani-Ueno, N. Yamaji, K. Yokosho, E. Craft, Z. J. Fei, S. Ebbs, M. C. Zambrano, J. F. Ma & L. V. Kochian 2014. Root and shoot transcriptome analysis of two ecotypes of Noccaea caerulescens uncovers the role of NcNramp1 in Cd hyperaccumulation. *Plant Journal,* 78**,** 398-410.

B. R. Moser, G. Knothe, S. F. Vaughn & T. A. Isbell 2009a. Production and Evaluation of Biodiesel from Field Pennycress (Thlaspi arvense L.) Oil. *Energy & Fuels,* 23**,** 4149-4155.

Bryan R. Moser, Shailesh N. Shah, Jill K. Winkler-Moser, Steven F. Vaughn & Roque L. Evangelista 2009b. Composition and physical properties of cress (Lepidium sativum L.) and field pennycress (Thlaspi arvense L.) oils. *Industrial Crops and Products,* 30**,** 199-205.

K. Mummenhoff, A. Franzke & M. Koch 1997. Molecular data reveal convergence in fruit characters used in the classification of Thlaspi s. l. (Brassicaceae). *Botanical Journal of the Linnean Society,* 125**,** 183-199.

K. Nakabayashi, M. Bartsch, Y. Xiang, E. Miatton, S. Pellengahr, R. Yano, M. Seo & W. J. J. Soppe 2012. The Time Required for Dormancy Release in Arabidopsis Is Determined by DELAY OF GERMINATION1 Protein Levels in Freshly Harvested Seeds. *Plant Cell,* 24**,** 2826-2838.

H. H. Nour-Eldin, T. G. Andersen, M. Burow, S. R. Madsen, M. E. Jorgensen, C. E. Olsen, I. Dreyer, R. Hedrich, D. Geiger & B. A. Halkier 2012. NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature,* 488**,** 531-4.

N. E. Olszewski, F. B. Martin & F. M. Ausubel 1988. Specialized binary vector for plant transformation: expression of the Arabidopsis thaliana AHAS gene in Nicotiana tabacum. *Nucleic Acids Res,* 16**,** 10765-82.

N. S. Pence, P. B. Larsen, S. D. Ebbs, D. L. D. Letham, M. M. Lasat, D. F. Garvin, D. Eide & L. V. Kochian 2000. The molecular physiology of heavy metal transport in the Zn/Cd hyperaccumulator Thlaspi caerulescens. *Proc Natl Acad Sci U S A,* 97**,** 4956-4960.

B. K. Peterson, J. N. Weber, E. H. Kay, H. S. Fisher & H. E. Hoekstra 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS One,* 7.

W. B. Phippen & M. E. Phippen 2012. Soybean Seed Yield and Quality as a Response to Field Pennycress Residue. *Crop Science,* 52**,** 2767-2773.

J. A. Poland & T. W. Rife 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome,* 5**,** 92-102.

A. Y. Rahman, A. O. Usharraj, B. B. Misra, G. P. Thottathil, K. Jayasekaran, Y. Feng, S. Hou, S. Y. Ong, F. L. Ng, L. S. Lee, H. S. Tan, M. K. Sakaff, B. S. Teh, B. F. Khoo, S. S. Badai, N. A. Aziz, A. Yuryev, B. Knudsen, A. Dionne-Laporte, N. P. Mchunu, Q. Yu, B. J. Langston, T. A. Freitas, A. G. Young, R. Chen, L. Wang, N. Najimudin, J. A. Saito & M. Alam 2013. Draft genome sequence of the rubber tree Hevea brasiliensis. *BMC Genomics,* 14**,** 75.

S. Ray & P. Satya 2014. Next generation sequencing technologies for next generation plant breeding. *Front Plant Sci,* 5.

C. Rogers, J. Wen, R. Chen & G. Oldroyd 2009. Deletion-based reverse genetics in Medicago truncatula. *Plant Physiol,* 151**,** 1077-86.

H. S. Saini, P. K. Bassi, J. S. Goudey & M. S. Spencer 1987. Breakage of Seed Dormancy of Field Pennycress (Thlaspi-Arvense) by Growth-Regulators, Nitrate, and Environmental-Factors. *Weed Science,* 35**,** 802-806.

J. C. Sedbrook, W. B. Phippen & M. D. Marks 2014. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: Example pennycress (Thlaspi arvense L.). *Plant Sci,* 227C**,** 122-132.

Jc Sedbrook, Phippen, Wb, Marks, Md 2014. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (Thlaspi arvense L.). *Plant Sci,* 227**,** 122-132.

G. W. Selling, M. P. Hojilla-Evangelista, R. L. Evangelista, T. Isbell, N. Price & K. M. Doll 2013. Extraction of proteins from pennycress seeds and press cake. *Industrial Crops and Products,* 41**,** 113-119.

T. Slotte, K. M. Hazzouri, J. A. Agren, D. Koenig, F. Maumus, Y. L. Guo, K. Steige, A. E. Platts, J. S. Escobar, L. K. Newman, W. Wang, T. Mandakova, E. Vello, L. M. Smith, S. R. Henz, J. Steffen, S. Takuno, Y. Brandvain, G. Coop, P. Andolfatto, T. T. Hu, M. Blanchette, R. M. Clark, H. Quesneville, M. Nordborg, B. S. Gaut, M. A. Lysak, J. Jenkins, J. Grimwood, J. Chapman, S. Prochnik, S. Shu, D. Rokhsar, J. Schmutz, D. Weigel & S. I. Wright 2013. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nat Genet,* 45**,** 831-5.

Afa Smit, Hubley, R & Green, P. 1996-2010. RepeatMasker Open-3.0

D. Tilman, K. G. Cassman, P. A. Matson, R. Naylor & S. Polasky 2002. Agricultural sustainability and intensive production practices. *Nature,* 418**,** 671-7.

R. K. Varshney, C. Song, R. K. Saxena, S. Azam, S. Yu, A. G. Sharpe, S. Cannon, J. Baek, B. D. Rosen, B. Tar'an, T. Millan, X. Zhang, L. D. Ramsay, A. Iwata, Y. Wang, W. Nelson, A. D. Farmer, P. M. Gaur, C. Soderlund, R. V. Penmetsa, C. Xu, A. K. Bharti, W. He, P. Winter, S. Zhao, J. K. Hane, N. Carrasquilla-Garcia, J. A. Condie, H. D. Upadhyaya, M. C. Luo, M. Thudi, C. L. Gowda, N. P. Singh, J. Lichtenzveig, K. K. Gali, J. Rubio, N. Nadarajan, J. Dolezel, K. C. Bansal, X. Xu, D. Edwards, G. Zhang, G. Kahl, J. Gil, K. B. Singh, S. K. Datta, S. A. Jackson, J. Wang & D. R. Cook 2013. Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nat Biotechnol,* 31**,** 240-6.

R. K. Varshney, R. Terauchi & S. R. Mccouch 2014. Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLoS Biol,* 12.

A. Voskoboynik, N. F. Neff, D. Sahoo, A. M. Newman, D. Pushkarev, W. Koh, B. Passarelli, H. C. Fan, G. L. Mantalas, K. J. Palmeri, K. J. Ishizuka, C. Gissi, F. Griggio, R. Ben-Shlomo, D. M. Corey, L. Penland, R. A. White, I. L. Weissman & S. R. Quake 2013. The genome sequence of the colonial chordate, Botryllus schlosseri. *Elife,* 2.

L. Wang, S. Yu, C. Tong, Y. Zhao, Y. Liu, C. Song, Y. Zhang, X. Zhang, Y. Wang, W. Hua, D. Li, D. Li, F. Li, J. Yu, C. Xu, X. Han, S. Huang, S. Tai, J. Wang, X. Xu, Y. Li, S. Liu, R. K. Varshney, J. Wang & X. Zhang 2014. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol,* 15**,** R39.

X. Wang, H. Wang, J. Wang, R. Sun, J. Wu, S. Liu, Y. Bai, J. H. Mun, I. Bancroft, F. Cheng, S. Huang, X. Li, W. Hua, J. Wang, X. Wang, M. Freeling, J. C. Pires, A. H. Paterson, B. Chalhoub, B. Wang, A. Hayward, A. G. Sharpe, B. S. Park, B. Weisshaar, B. Liu, B. Li, B. Liu, C. Tong, C. Song, C. Duran, C. Peng, C. Geng, C. Koh, C. Lin, D. Edwards, D. Mu, D. Shen, E. Soumpourou, F. Li, F. Fraser, G. Conant, G. Lassalle, G. J. King, G. Bonnema, H. Tang, H. Wang, H. Belcram, H. Zhou, H. Hirakawa, H. Abe, H. Guo, H. Wang, H. Jin, I. A. Parkin, J. Batley, J. S. Kim, J. Just, J. Li, J. Xu, J. Deng, J. A. Kim, J. Li, J. Yu, J. Meng, J. Wang, J. Min, J. Poulain, J. Wang, K. Hatakeyama, K. Wu, L. Wang, L. Fang, M. Trick, M. G. Links, M. Zhao, M. Jin, N. Ramchiary, N. Drou, P. J. Berkman, Q. Cai, Q. Huang, R. Li, S. Tabata, S. Cheng, S. Zhang, S. Zhang, S. Huang, S. Sato, S. Sun, S. J. Kwon, S. R. Choi, T. H. Lee, W. Fan, X. Zhao, X. Tan, X. Xu, Y. Wang, Y. Qiu, Y. Yin, Y. Li, Y. Du, Y. Liao, Y. Lim, Y. Narusaka, Y. Wang, Z. Wang, Z. Li, Z. Wang, Z. Xiong, Z. Zhang & Consortium Brassica Rapa Genome Sequencing Project 2011. The genome of the mesopolyploid crop species Brassica rapa. *Nat Genet,* 43**,** 1035-9.

S.I. Warwick, Francis, A., Susko, D.J. 2002 The biology of Canadian weeds. 9. Thlaspi arvense L. (updated) *Canadian Journal of Plant Science,* 82**,** 803-823.

O. Westesson, M. Skinner & I. Holmes 2013. Visualizing next-generation sequencing data with JBrowse. *Brief Bioinform,* 14**,** 172-7.

N. Winchester, Mcconnachie, D., Wollersheim, C. And Waitz, I. 2013. Market cost of renewable jet fuel adoption in the United States. *MIT Joint Program on the Science and Policy of Global Change*.

H. J. Wu, Z. Zhang, J. Y. Wang, D. H. Oh, M. Dassanayake, B. Liu, Q. Huang, H. X. Sun, R. Xia, Y. Wu, Y. N. Wang, Z. Yang, Y. Liu, W. Zhang, H. Zhang, J. Chu, C. Yan, S. Fang, J. Zhang, Y. Wang, F. Zhang, G. Wang, S. Y. Lee, J. M. Cheeseman, B. Yang, B. Li, J. Min, L. Yang, J. Wang, C. Chu, S. Y. Chen, H. J. Bohnert, J. K. Zhu, X. J. Wang & Q. Xie 2012. Insights into salt tolerance from the genome of Thellungiella salsuginea. *Proc Natl Acad Sci U S A,* 109**,** 12219-24.

R. Yang, D. E. Jarvis, H. Chen, M. A. Beilstein, J. Grimwood, J. Jenkins, S. Shu, S. Prochnik, M. Xin, C. Ma, J. Schmutz, R. A. Wing, T. Mitchell-Olds, K. S. Schumaker & X. Wang 2013. The Reference Genome of the Halophytic Plant Eutrema salsugineum. *Front Plant Sci,* 4**,** 46.

Z. Zhang, J. A. Ober & D. J. Kliebenstein 2006. The gene controlling the quantitative trait locus EPITHIOSPECIFIER MODIFIER1 alters glucosinolate hydrolysis and insect resistance in Arabidopsis. *Plant Cell,* 18**,** 1524-36.

CHAPTER 4: Genomic analysis of spring and winter annual pennycress accessions reveals the spring annual growth habit has arisen multiple independent times

SUMMARY

Field pennycress (*Thlaspi arvense* L.) is currently being developed as a new cold-tolerant oilseed crop. In natural populations, pennycress, like many Brassicaceae relatives, can exhibit either a winter or spring annual phenotype. Pennycress is a diploid relative of *Arabidopsis thaliana,* a model species that has been used to study many adaptive phenotypes, including flowering time and developmental timing. In Arabidopsis and other Brassicaceae species, mutations in negative regulators of flowering, including FLOWERING LOCUS C and FRIGIDA can cause the transition to a spring annual habit. The genetics underlying the difference between spring and winter annual pennycress lines are currently unknown. Here we report the identification of three new alleles of FLC in pennycress that confer a spring annual growth habit identified through whole genome sequencing, segregation analyses, and comparative genomics. However, these specific alleles were only identified in a small percentage of spring annual lines tested, indicating there are likely additional FLC mutations not identified here, and/or mutations in other negative regulators such as FRIGIDA. These findings will provide important information for the ongoing development of pennycress as a winter annual cover crop and biodiesel feedstock, as an enhanced understanding of the regulation of flowering in this species should allow for the fine-tuning of flowering in commercial lines.

Introduction

Thlaspi arvense L. (field pennycress, pennycress herein) is currently a target for domestication as a new, cold-hardy winter oilseed crop that can fit within the corn/soybean rotation in the Midwestern United States (Dorn *et al.*, 2013, Dorn, 2015, Sedbrook *et al.*, 2014). As a winter annual crop, pennycress can be planted in the late summer to early autumn, either into standing corn or immediately after corn harvest. The pennycress crop establishes a robust vegetative cover prior to winter, providing important ecosystem services such as limiting soil erosion and nutrient runoff. In the spring, pennycress will flower and set seed, yielding upwards of 1600 pounds per acre of oil-rich seed in time for planting a crop of soybean (Phippen and Phippen, 2012, Sedbrook *et al.*, 2014). Pennycress seed is high in oils that can be readily converted to biodiesel or jet fuel, with the remaining high-protein seed meal having potential as animal feed or feedstock for industrial uses (Evangelista *et al.*, 2012, Fan *et al.*, 2013, Selling *et al.*, 2011, Moser *et al.*, 2009, Hojilla-Evangelista *et al.*, 2013).

Pennycress is a member of the Thlaspideae, a tribe of the Brassicaceae. The Brassicaceae family is divided into three lineages, and pennycress resides in lineage II, along with the Brassica genus, including the economically important oilseed crops Brassica rapa and Brassica napus, but not the model species *Arabidopsis thaliana* and Capsella rubella, which are members of lineage I (Franzke *et al.*, 2011). Pennycress is considered naturalized to most temperate to subarctic regions throughout the northern hemisphere, including the entire United States (excluding Hawaii and Alabama), all

provinces of Canada, as well as the southern hemisphere, including Australia, New Zealand, and Argentina (Warwick, 2002). Pennycress already exhibits numerous traits that make it an attractive winter rotation crop(Sedbrook *et al.*, 2014). Of particular importance in the upper Midwest is the existence of pennycress lines that complete their life cycle rapidly enough to fit within the corn/soybean rotation. Central to this rapid spring development of pennycress is the underlying growth habits of winter annuals. In wild populations, there are both winter and spring annual pennycress (McIntyre and Best, 1978 , Best and McIntyre, 1972), similar to many other Brassicaceae species such as *Arabidopsis thaliana* (Stinchcombe *et al.*, 2004, Shindo *et al.*, 2005), *Brassica rapa* (Wu *et al.*, 2012), *Camelina sativa* (Crowley, 1999), and *Brassica napus* (Tadege *et al.*, 2001).

Throughout the decades of research on basic developmental questions in Arabidopsis, and the expanding translational research in other Brassica crops, the underlying molecular mechanisms controlling these differences has been identified in many of these species (Jung and Muller, 2009, Simpson and Dean, 2002, Amasino, 2005, Kim *et al.*, 2009). In wild accessions of Arabidopsis, only a handful of mutations are responsible for variation of flowering time and vernalization requirement (Burn *et al.*, 1993, Clarke and Dean, 1994, Johanson *et al.*, 2000). Most notably, allelic variation in two key negative regulators, FLOWERING LOCUS C (FLC) and/or FRIGIDA (FRI), underlies the key difference between spring and winter annual plants. FLC encodes a MADS box transcription factor (Michaels and Amasino, 1999), and inhibits flowering prior to vernalization by repressing the expression of FLOWERING LOCUS T (FT)

(Searle *et al.*, 2006). In Arabidopsis, *flc* null mutations eliminate the vernalization requirement and impart the spring annual, rapid flowering phenotype (Michaels and Amasino, 1999). Allelic variation within FRI also impact flowering time and the vernalization requirement. Similar to FLC, Arabidopsis accessions harboring loss of function mutations in FRI flower rapidly without vernalization (Johanson *et al.*, 2000, Shindo *et al.*, 2005).

The expression of FLC is positively regulated by FRI (Michaels and Amasino, 1999), thus, mutations in either of the two main vernalization-responsive negative regulators, FLC and FRI, can lead to a loss of vernalization requirement and a spring annual growth habit (Michaels and Amasino, 2001, Michaels and Amasino, 1999). The vernalization signal provided by the cold of winter removes the repression on the transition to flowering through the epigenetic silencing of FLC. Specifically, vernalization increases histone 3 K27 trimethylation (H3K27me3) at FLC chromatin, reducing transcriptional activity (Sung *et al.*, 2006, Finnegan and Dennis, 2007, Greb *et al.*, 2007). The vernalization-induced silencing of FLC releases the repression on the transition to flowering, which promotes the transition to reproductive growth (Searle *et al.*, 2006).

While this extensive understanding of the molecular genetic pathways controlling flowering time and vernalization in Arabidopsis has informed similar studies in Brassica relatives, little is known about the underlying mechanisms controlling flowering time

variation in pennycress. Different accessions of pennycress have been reported to either act as early flowering or late flowering, with late flowering accessions growing as rosettes for as long as 150 days prior to flowering (Best and McIntyre, 1972). It was later found that vernalization increased the rate of flowering in the late flowering accessions (Best and Mc Intyre, 1976). Analyses of F2 progeny between the late and early flowering accessions determined the early flowering (spring annual) phenotype was determined by a single gene, with the late flowering allele being completely dominant (McIntyre and Best, 1978).

The development of pennycress is only in its infancy, and current available lines are only a few generations derived from wild collections. While comprehensive plant improvement programs are less than a decade old, there has been a rapid development of genetic and genomic resources for pennycress, including a transcriptome (Dorn *et al.*, 2013) and genome sequence (Dorn, 2015), a large scale EMS mutagenized population, and numerous groups now focused on the genetics of key agronomic traits. Foundational to this work has been the feasibility of using DNA sequencing technologies to generate these resources for a developing crop. While previous efforts to sequence the human (Lander *et al.*, 2001, Venter *et al.*, 2001) and Arabidopsis (S. Kaul, 2000) genomes involved massive teams and multi-billion dollar budgets, the advent and wide-scale adoption of next generations sequencing technologies (NGS) has allowed for the development of equivalent genomic resources for hundreds of species, including numerous crop species such as chickpea (Varshney *et al.*, 2013), cassava, cucumber,

sugar beet, flax, and camelina. In each of these cases, the assembly of even draft genomes for these species has allowed for the direct identification of important gene homologs potentially responsible for controlling traits of interest, as informed by both previous research in the species and by basic research in model species.

In this study, we utilized the power NGS and our previously developed draft genome for pennycress to rapidly identify a causal mutation for the spring annual growth habit. Through whole genome sequencing of a winter and spring annual pennycress accession and accompanying segregation analysis of the identified mutation, we have also identified one additional natural ‘spring annual’ allele of FLC, as well as an induced mutation identified from a large scale EMS-mutagenized population. Each of the two spring annual FLC alleles were identified in accessions throughout North America.

Materials and Methods

Thlaspi arvense accession MN108-SA, MN111, and MN111 x MN108-SA F2 population

The spring annual *Thlaspi arvense* line MN108-SA was derived from a wild Minnesota population containing both winter and spring annual plants. Three generations of single seed decent was performed on a spring annual plant from this collection and sequenced. A single MN111 plant was also carried through two generations single seed decent and sequenced. Plants were germinated on moist Berger BM2 germination mix (Berger, Inc., www.berger.ca), stratified at 4° C for 7 days, and grown in climate controlled growth chambers at the University of Minnesota (21° C, 16 h/8 h day/night cycles at 100 micromoles/m2/s PAR). The MN111 plant sequenced in this analysis was vernalized at six weeks post germination at 4°C for 30 days in the dark. After vernalization, this plant was returned to the growth chamber conditions described above. The MN108-SA plant was not vernalized as it flowered immediately.

DNA isolation and Illumina genomic DNA sequencing

DNA was isolated from a single MN108-SA and MN111 plant using the Omega Mag-Bind Plant DNA kit (Omega Bio-Tek, www.omegabioteck.com) according to the manufacturers recommended protocol. These DNA samples were submitted to the University of Minnesota Genomics Center for sequencing on the Illumina HiSeq 2000 platform (Illumina Inc, www.illumina.com). Sequencing libraries were prepared using

the Illumina TruSeq Nano DNA Sample Prep kit with an average library insert size of 460 base pairs. The Illumina Universal Adaptor and an Indexed Adaptor (MN111 – Illumina Indexed Adaptor #12 – barcode = CTTGTA, MN108 – Illumina Indexed Adaptor #11 – barcode = GGCTAC) were used to create the sequencing libraries. Each library was sequenced on a full lane of Illumina HiSeq 2000 (100 base pair, paired-end).

Sequencing data quality control, read mapping, and de novo assembly

FASTQ files from all sequencing runs were imported into CLC Genomics Workbench Version 7.5 (CLC Bio, www.clcbio.com) Illumina reads were subjected to quality control using the Sequencing QC Report tool. Reads were trimmed and filtered using the following parameters: Trim adapter list = Illumina TruSeq Universal Adaptor + Illumina Indexed Adaptor used for library preparation, Ambiguous Trim = Yes, Ambiguous Limit = 2, Quality Limit = 0.05, Use Colorspace = No, Search Reverse Sequence = Yes, Save discarded sequences = No, Remove 5' terminal nucleotides = Yes, Number of 5' Terminal Nucleotides to Remove = 8, Remove 3' terminal nucleotides = Yes, Number of 3' terminal nucleotides to remove = 2, Discard Long Reads = No, Save Broken Pairs = No.

Trimmed and filtered sequencing reads were mapped to the previously published draft pennycress genome (Dorn et al, 2015) with CLC Genomics Workbench Version 7.5 with the 'Map Reads to Reference' tool using the following parameters: Masking mode = No masking, Mismatch cost = 2, Insertion cost = 3, Deletion cost = 3, Length fraction =

0.9, Similarity fraction = 0.9, Global alignment = No, Auto-detect paired distances = Yes, Non-specific match handling = Ignore, Collect un-mapped reads = Yes.

Variants were called in CLC Genomics Workbench Version 7.5 using the Basic Variant Detection tool using the following parameters: Ploidy = 2, Ignore positions with coverage above = 100,000, Ignore broken pairs = yes, Ignore non-specific matches = Reads, Minimum coverage = 10, Minimum count = 2, Minimum frequency = 35%, Base quality filter = yes, Neighborhood radius = 5, minimum central quality = 20, minimum neighborhood quality = 15. A called variant table was exported and further filtered upon variant quality score, with variants below or equal to 30 being excluded from further analysis.

Independent de novo assemblies of the MN111 and MN108-SA reads were performed using the de novo assembly tool in CLC Genomics Workbench using the following parameters: Mapping mode = Map reads back to contigs, update contigs = Yes, Automatic bubble size = yes, Minimum contig length = 500, Word size = 63, Perform scaffolding = Yes, Auto-detect paired distances = Yes, Mismatch cost = 3, Insertion cost = 3, Deletion cost = 3, Length Fraction = 0.95, Similarity Fraction = 0.95. Variant detection was performed on the assembled scaffolds and corresponding read mappings using the above methods and parameters.

Sequence comparison of flowering time gene orthologs

Sequence comparison analyses were completed in CLC Genomics Workbench 7.5 using the 'Classical Sequence Analysis' toolkit. Consensus genomic sequences for pennycress FT, FRI, and FLC orthologs were derived from the corresponding read mappings for MN111 and MN108-SA. Extracted consensus sequences were manually examined to ensure appropriate base calling of variants based on read mapping information. Genomic DNA alignments and predicted peptide alignments for the FT, FRI, and FLC sequences from MN111, MN108-SA, and MN106 reference genome were generated with the CLC 'Create Alignment' tool using the following parameters: Gap open cost = 10, Gap extension cost = 1, End gap cost = as any other, Alignment = very accurate (slow).

Sanger sequencing analysis of c.6_7insG FLC mutation in MN111 x MN108-SA F2 population and global spring varieties

DNA was isolated from F2 progeny grown in the conditions described above using the Omega Bio-Tek Plant MagBind 96 kit according to the manufacturers recommended protocol. DNA oligos were designed to amplify the 5' end of the pennycress FLC gene, approximately 100 base pairs upstream of the transcriptional start site (TaFLC_1_Forw: 5' – CCGAGGAAGAAAAAGTAGATAGAGACA -3', TaFLC_1_Rev: 5' – GAAGCTTAAAGGGGGAAAAAGGAA – 3'). Polymerase Chain Reaction (PCR) was used to amplify this fragment, producing an approximately 450 base pair amplicon. New England Biolabs Q5 Hot-Start High-Fidelity PCR Kit with 2x Master

Mix was used, with the following thermal cycler conditions: 1.) 98°C for 30 seconds, 2.) 98°C for 10 seconds, 3.) 57°C for 20 seconds, 4.) 72°C for 20 seconds, 5.) Go to step #2 34 times , 6.) 72°C for 2 minutes, 7.) 4°C hold. Reactions were visualized using gel electrophoresis on a 1% agrose gel with ethidium bromide stain to confirm amplification of a single band using gel electrophoresis. PCR products were submitted to Beckman Coulter Genomics for PCR product purification and single pass Sanger sequencing. Amplicons were sequenced in both directions using the forward and reverse primers listed above. Sanger sequencing reads were analyzed in CLC Genomics Workbench, aligned against the pennycress MN106 reference genome at the FLC locus to identify sequence variants.

PCR and sequencing of TaFLC Amplicon 2, Amplicon 3, and c.456bp_del amplicon

Oligonucleotides were designed for the amplification of TaFLC Amplicon 2, encompassing exons 2 – 5 (TaFLC_2_Forw: 5' – TGTTTTCAAGGGTTAGCTCG – 3' and TaFLC_2_Rev: 5' – TGCCAGCTAATAAAGGATAC – 3').
Polymerase Chain Reaction (PCR) was used to amplify this fragment, producing an approximately 1000 base pair amplicon. New England Biolabs Q5 Hot-Start High-Fidelity PCR Kit with 2x Master Mix was used, with the following thermal cycler conditions: 1.) 98°C for 30 seconds, 2.) 98°C for 5 seconds, 3.) 60°C for 15 seconds, 4.) 72°C for 30 seconds, 5.) Go to step #2 34 times , 6.) 72°C for 2 minutes, 7.) 4°C hold. Reactions were visualized using gel electrophoresis on a 1% agrose gel with ethidium bromide stain to confirm amplification of a single band using gel electrophoresis. An

additional internal primer (TaFLC_2_internal_Forw 5'-GATCCACCTACGAGCTATTA – 3') was also used for sequencing.

Oligonucleotides were designed for the amplification of TaFLC Amplicon 3, encompassing exons 6 – 7 (TaFLC_3_Forw: 5' – TCTAGTCTTCTTGGGGATG – 3' and TaFLC_3_Rev: 5' – ACAACATGCACCCTTATCAA – 3').
Polymerase Chain Reaction (PCR) was used to amplify this fragment, producing an approximately 1000 base pair amplicon. New England Biolabs Q5 Hot-Start High-Fidelity PCR Kit with 2x Master Mix was used, with the following thermal cycler conditions: 1.) 98°C for 30 seconds, 2.) 98°C for 5 seconds, 3.) 60°C for 15 seconds, 4.) 72°C for 30 seconds, 5.) Go to step #2 34 times , 6.) 72°C for 2 minutes, 7.) 4°C hold. Reactions were visualized using gel electrophoresis on a 1% agrose gel with ethidium bromide stain to confirm amplification of a single band using gel electrophoresis.

Oligonucleotides were designed for the amplification of TaFLC c.456bp_del amplicon that encompasses the 3' end of the first intron, the 2nd exon, and the second intron (TaFLC_del_1_Forw: 5' – CGAACCATAGTTCAGAGCTT– 3' and TaFLC_del_1_Rev: 5' – ATAGTGTGCATCAACTGGTC– 3').
Polymerase Chain Reaction (PCR) was used to amplify this fragment, producing an approximately 1000 base pair amplicon. New England Biolabs Q5 Hot-Start High-Fidelity PCR Kit with 2x Master Mix was used, with the following thermal cycler conditions: 1.) 98°C for 30 seconds, 2.) 98°C for 5 seconds, 3.) 63°C for 15 seconds, 4.)

72°C for 60 seconds, 5.) Go to step #2 34 times , 6.) 72°C for 2 minutes, 7.) 4°C hold. Reactions were visualized using gel electrophoresis on a 1% agrose gel with ethidium bromide stain to confirm amplification of a single band using gel electrophoresis.

All PCR products were purified prior to Sanger sequencing using the Qiagen QiaQuick PCR Purification Kit (Qiagen N.V., www.qiagen.com) according to the manufacters recommended protocol. Following purification of PCR products, amplicons were sequenced by ACGT, Inc. (Wheeling, Illinois) for single pass Sanger sequencing. All raw Sanger reads were analyzed in CLC Genomics Workbench.

Results

Flowering time phenotypes of winter and spring annual pennycress

T. arvense line MN111 consistently exhibits a winter annual phenotype which requires a period of vernalization to induce flowering, whereas line MN108-SA completely lacks the vernalization requirement and flowers immediately after germination (Figure 1). After germination, MN108-SA plants develop two to three true leaves and internodes begin to elongate (Figure 1 – inset, white arrow). To examine the segregation of the winter and spring annual phenotypes, an F2 population was developed from a cross between MN111 and MN108-SA. A single F1 individual was self pollinated, and 50 F2 progeny were planted for further analysis. Of these F2 individuals, 38 exhibited a winter annual phenotype and 12 exhibited a spring annual phenotype (24%), suggesting the winter annual phenotype is caused by a single dominant locus (Table 2).

Whole Genome Resequencing of MN111 and MN108-SA Individuals

Whole genome sequencing of the MN111 and MN108-SA parent plants was used to identify key genomic differences underlying the difference between winter annual and spring annual plants. DNA from each parent was sequenced on a full lane of the Illumina HiSeq 2000 platform (100 base pair, paired end). Over 334M sequencing reads were generated per parent (Table 1). A total of 29B high quality (post quality control) were generated per parent, representing >53X coverage of the predicted genome size of 539Mb (Johnston *et al.*, 2005).

Trimmed and filtered reads were mapped to the winter annual MN106 draft genome sequence (Dorn, 2015). Variant detection analysis identified 556,115 and 566,330 putative variants in the MN111 and MN108-SA parents, respectively. As shown in Table 1, the majority of these variants are single nucleotide polymorphisms (SNPs), with the 433,216 (MN111) and 443,713 (MN108) being homozygous SNPs. Of the total number of variants for each parent, less than 15% consisted of multiple nucleotide variants, insertions, deletions, and replacements (Table 1). The complete datasets listing all putative variants in the MN111 and MN108-SA parents are available at https://pennycressflowering.s3.msi.umn.edu/MN111_variants.xls.gz and https://pennycressflowering.s3.msi.umn.edu/MN108_variants.xls.gz, respectively.

Examination of flowering time orthologs in winter and spring annual pennycress

BLAST analyses were used to identify the genomic sequences of FLOWERING LOCUS T (FT), FRIGIDA (FRI), and FLOWERING LOCUS C (FLC) from MN106, MN111, and MN108. Read mappings were manually examined to identify potential mutations underlying the winter to spring annual transition. In the pennycress FT ortholog, there was one region with sequence differences among the three lines; however, this region is within an AT-rich region of the first intron. There were no sequence differences in the FT predicted peptides between the three lines. Comparison of the pennycress FRI homolog (Ta1.0_26225 on scaffold 1344 of the v1.0 pennycress genome) revealed a single SNP between these three lines, however, this SNP was shared by both the MN111 and MN108 individual, likely indicating a non-causative effect on the spring annual growth habit. This SNP was found to cause an amino acid change (Threonine to Serine) at position 553 of the TaFRI predicted peptide (position 2,085 of the gene model - c.2,085A>T).

BLASTp results of the Arabidopsis FLC peptide against pennycress predicted peptide set revealed several candidates. Seven of these top hits in the pennycress predicted proteome were extracted, and compared back to Arabidopsis via BLASTp. Each best hit from Arabidopsis was extracted and aligned with the pennycress predicted peptides (Figure 1). All of the corresponding BLAST hits contain MADS-box domains, and consist of SEPELLATA 1/AGAMOUS-LIKE 2 (SEP2-AT5G15800), SEPELLATA 3 (SEP3-AT1G24260), SEPALLATA 4/AGAMOUS-LIKE 3 (SEP4-AT3G03710),

APETALA 1/AGAMOUS-LIKE 7 (AP1-AT1G69120), MADS AFFECTING FLOWERING 2/AGAMOUS-LIKE 31 (MAF2-AT5G65050), MADS AFFECTING FLOWERING 3 (MAF3-AT5G65060), MADS AFFECTING FLOWERING 4 (MAF4-AT5G65070), MADS AFFECTING FLOWERING 5/AGAMOUS-LIKE 68 (MAF5-AT5G65080), AGAMOUS LIKE 17 (AGL17 – AT2G22630), and AGAMOUS LIKE 21 (AGL21-AT4G37940). Phylogenetic analyses revealed that the pennycress predicted peptide Ta00917 was most similar to TaFLC, with each of the other hits corresponding to distinct groupings of AGL-like, SEP-like, AP1-like, or MAF-like (Figure 1).

Investigation of WGS sequencing reads at the pennycress FLC locus revealed a single base pair (guanine) frame shift insertion after the second codon (c.6_7insG) in the MN108 spring annual line, which when theoretically translated, causes a precocious stop codon (TGA) in the 12th codon position. This mutation was not found in either winter annual line and thus investigated further.

Genetic analysis of co-segregation of c.6_7insG mutation with spring annual phenotype in an F2 population

Fifty F2 individuals from the previously described MN111 x MN108 cross were screened using polymerase chain reaction (PCR) and Sanger sequencing to examine co-segregation of the c.6_7insG mutation with the spring annual phenotype. Oligonucleotides were designed to amplify a 394 bp amplicon (TaFLC_1_Forw and TaFLC_2_Rev – Table 3) that encompasses the entire first exon of the FLC locus (Figure 3A, leftmost gray arrow). Sanger sequencing of these amplicons from all 12 spring annual F2s confirmed the co-segregation of the c.6_7insG mutation in this population (Table 4). Additionally, 14 randomly chosen winter annual F2 individuals were also screened, and showed a segregation ratio of 5 homozygous for the MN111 allele (no c.6_7insG), and 9 heterozygous for the mutation (as evidenced by examination of the Sanger sequencing trace data).

Additional oligonucleotides were generated to allow for PCR screening of two additional amplicons to cover the remaining six exons of TaFLC (Figure 3A, gray arrows on right). Sequencing of amplicon 2 (Figure 3A, middle arrow), which encompasses exons 2-5 of FLC, revealed an additional two SNPs. The first occurs at position 2,990, which was also evident in the WGS results of the MN111 and MN108 parents, where the MN111 parent posses a ‘C’ at this position, whereas the MN108 parent, along with the MN106 reference genome posses a ‘T’ at this position. Since this SNP resides in the predicted 2nd intron of the FLC gene, and that the spring annual

MN108 parent and the winter annual MN106 references accessions shared the same variant, it seemed unlikely this SNP was responsible for the spring/winter annual phenotype. A second variant at position 3,342 showed the identical relationship, with the reference MN106 and MN108 accessions sharing a "C" at this position, while the MN111 accession posses a "G". This particular SNP location, which is in the first position fourth exon of FLC, imparts a nonsynonymous substitution in the predicted peptide, changing the codon from "AGC" (in the MN106 and MN108 accession, encoding for serine) to "AGG" in the MN111 accession (encoding for arginine). Within the third amplicon, which covers exons 6 and 7 of FLC, one additional SNP was discovered immediately after the 6th exon in the final intron at position 3,897. As this SNP exists in an intron and does not impart a splice acceptor site, it was not analyzed further. The co-segregation of the SNPs at positions 7 (c.6_7insG), 2,990, and 3,342 within the F2 individuals was further analyzed, as shown in Table 3. All spring annual F2 progeny, which were already shown to possess the MN108 allele (the c.6_7insG mutation), also exhibited the MN108 variants at positions 2,990 ('T') and 3,342 ('C') (Table 3). The five winter annual F2 progeny homozygous for the MN111/MN106 allele at position 7 were also shown to be homozygous for MN111 variant at positions 2,990 and 3,342. These results suggest that the variants at positions 2,990 and 3,342 are not causal for the spring/winter annual phenotype.

Identification of a novel FLC allele from an EMS mutagenized population that confers the spring annual phenotype

In the fall of 2013, a large-scale population (30,000-40,000 seeds) of MN106 was treated with the mutagen ethylmethane sulfonate (EMS) and planted in St. Paul, Minnesota. EMS is a common mutagen that induces random point mutations through the alkylation of guanine residues in DNA, leading to abnormal base pairing with thymine, often causing a base pair change of G:C to A:T after replication. The use of mutagens like EMS has been a common practice to induce novel genetic variation within a population for both basic research and breeding. In the spring of 2014, 1,000 pools of 10 M1 plants were hand collected, and a portion of seed from each of these collections was planted into 1,000 individual rows the following fall. In the fall of 2014, a forward screen for phenotypes of interest was conducted. One individual spring annual plant was identified in the field (herein referred to as 'Squirt' – shown in Figure 4A). As this was the only spring annual plant within this row, a subpool of the remaining collected M1 seed that was used to seed this row was planted in the growth chamber in an attempt to identify an additional progeny of the field-identified Squirt mutant. One individual was found, shown in Figure 4B, which immediately began internode elongation and flowering, similar to MN108 (Figure 1) and the other spring lines examined in this study.

With the previous implication of mutations in TaFLC leading to the spring annual phenotype, we began PCR screening with the first exon (TaFLC_1_Forw and TaFLC_1_Rev – Fig 2A leftmost arrow). Sanger sequencing of the resulting PCR

amplicons confirmed a single nucleotide polymorphism in the Squirt mutant at the 55th base of TaFLC (first base of the 18th codon). This mutation (A to T) leads to a change of the 18th codon from CAA (in the MN106 and MN111 accessions, along with Arabidopsis) to the stop codon TAA (Figure 5). The identification of the c.55A>T allele of TaFLC, along with the previous finding the c.6_7insG mutation is shown to segregate in Mendelian fashion with the spring annual phenotype, we have concluded that the identify of the single dominant locus responsible for the spring annual phenotype in pennycress (McIntyre and Best, 1978) is TaFLC, and that each of these alleles alone is causal for the spring annual phenotype in these accessions.

Geographic distribution of FLC mutations in global spring annual pennycress accessions

With the finding that a single base pair insertion (c.6_7insG) or substitution (c.55A>T) can impart the spring annual flowering phenotype, we next examined 44 additional globally distributed spring annual lines collected from national germplasm repositories and the University of Minnesota collection. PCR screening and amplicon sequencing of these 44 accessions revealed that only three lines shared the c.6_7insG mutation originally identified in the MN108 accession, all originating in North America. These lines consisted of MN131 (Howard Springs, Montana), Ames 31489 (Saskatoon, Canada), and Ames 31491 (Lashburn, Canada).

With 41 remaining spring annual lines without an identified mutation, we expanded the search within these lines. Using previous findings in Arabidopsis, *Brassica rapa*, *Brassica napus*, *Capsella rubella*, and *Capsella bursa-pastoris* that have shown that intronic and promoter variation can impart flowering time variation, including vernalization requirement, we expanded our search to include the previous unsequenced, large first intron of TaFLC. Oligonucleotides (TaFLC_del_1_Forw and TaFLC_del_1_Rev, Table 3) were designed to amplify a 2,088 bp amplicon stretching from the middle of the first intron to the third exon of TaFLC (Figure 6A, gray arrow). The remaining 41 spring annual lines were PCR screened to examine if there were any large insertions or deletions in this region that would be evident by examining the size of the PCR amplicons. A total of 10 accessions were found to produce identically sized PCR

amplicons approximately 400 bp shorter than the MN106 reference amplicon (Figure 6B). These lines included Spring 32, a proprietary inbred spring line obtained from Dr. Win Phippen at Western Illinois State University under a Materials Transfer Agreement, along with four UMN accessions (MN121, 123, 124, and 125, all collected from near Roseau, Minnesota), and several accessions obtained from United States Department of Agriculture Germplasm Resources Information Network (USDA-GRIN), including Ames 29512, 31024, 31490, 31492, and 31500 (Figure 6B). The geographic origins of each of these additional 10 accessions containing this new allele, c.456bp_del, along with the accessions containing the c.6_7insG allele and MN111, MN106, and MN108 accessions are shown in Figure 7. Ames 29512 is not shown in Figure 4, as its precise origin is unknown, other than originating as a donated plant material from Canada, as listed in the GRIN database. The Spring 32 accession origin is arbitrarily shown in Macomb, Illinois (the location of Western Illinois University).

Global spring annual varieties of unknown genetic origin

Of the remaining 31 spring annual accessions that were not found to possess either the c.6_7insG or c.456bp_del alleles of TaFLC, 6 were from Europe, including Germany (PI633415, PI650284), PI650285, and PI650286), France (PI650287), and Poland (Ames 22461). These European spring lines were next examined for mutations in TaFRI (Ta1.0_26225 on scaffold 1344 of the draft genome), as it was plausible there could also exist *fri* null alleles that could also confer the spring annual phenotype, as is seen in Arabidopsis. Primers were designed to amplify TaFRI in two amplicons (Figure 7) and sequenced. In each of these lines, only a single SNP was found at position 2,085, an 'A' to 'T' substitution, which causes an amino acid change in the final exon (threonine to serine), as was described above. Following reconfirmation via Sanger sequencing of MN106, MN111, MN108, and several other spring annual lines, it was found that MN106 was the only line tested with an 'A' at position 2,085. As MN111, the winter annual parent of the F2 cross described above, possessed the c.2,085A>T allele, it is unlikely this allele of FRI is underlying the spring annual phenotype of the European spring lines.

Discussion

Use of NGS to quickly identify mutations of interest

Here we report the discovery of two natural alleles of FLC that confer the spring annual phenotype in pennycress, one of which was discovered via whole genome sequencing of the parents of an F2 cross. At the onset of this experiment, upon calculating the predicted costs associated with PCR-based cloning and sequencing of all candidate genes, we found it was less expensive to use this WGS approach to first identify a candidate locus in the parent accessions, and then proceed to test F2 progeny via PCR and Sanger sequencing. The added value with the WGS approach here is that we now also have 'reference' genome sequences for these two accessions (MN111 and MN108), along with an F2 population that is also being used to develop a linkage map via Genotype by Sequencing (Poland and Rife, 2012). Additionally, as the network controlling flowering time is known to be extremely complex, the gene variants in each of these parents that may have a minor affect on vernalization and flowering provide a wealth of untapped information for later investigation.

Towards an understanding of flowering time control in pennycress

Our previous efforts (Chapter 2 and 3 of this dissertation) to develop genomic resources for the domestication of pennycress identified likely homologs controlling flowering time and the vernalization response via DNA and RNA sequence homology (Dorn *et al.*, 2015, Dorn *et al.*, 2013). Here we present the first sequence-supported genetic information on the underlying mechanisms controlling flowering time in

pennycress. While the two new natural alleles of FLC reported here are sufficient to confer the spring annual phenotype, there is likely a host of interacting gene products also essential for the rapid flowering seen in the spring annual accessions investigated here. Of particular interest is FRIGIDA and members of the 'FRIGIDA Complex' (FRI-C), including FRIGIDA LIKE 1 (FRL1), FRIGIDA-ESSENTIAL 1 (FES1), SUPPRESSOR OF FRIGIDA4 (SUF4), and FLC EXPRESSOR (FLX) (Choi *et al.*, 2011). The FRIGIDA Complex acts as a transcriptional activation complex on FLC expression through a diverse range of functions of each complex member (Choi *et al.*, 2011). While no loss of function FRIGIDA alleles were identified in the accessions examined here, it remains possible an FLC-independent path to the spring annual phenotype exists. Future studies should focus on the collection of a more comprehensive set of globally distributed lines, including both spring and winter annuals, as important FRI alleles could otherwise be masked in heterozygotes. The interaction between the vernalization response pathways, photoperiodic, and autonomous pathways are also of great interest, as allelic variation and unique combinations of alleles from each pathway can contribute to quantitative variation in flowering time, which is the ultimate applied goal of the research described here.

Future directions for pennycress flowering time research

While the scope of this study was to define a key regulator behind the vernalization requirement in pennycress, the results presented here of a homologous spring/winter annual system controlled by loss of function alleles of FLC open up immense opportunity for applying knowledge from Arabidopsis to improve a new crop species. For example, gene expression patterns in accessions harboring each FLC allele identified here will help define the roles of FLC, members of the FRIGIDA Complex, and the downstream floral integration genes and provide important information understanding the flowering network in pennycress. Also currently unknown is the underlying genetic cause of the spring annual phenotype of the remaining 31 spring annual accessions tested here. While there is currently missing sequencing information covering the large first exon of FLC, as well as the promoter of FLC, ongoing experiments should be able to identify any new allelic variants of interest. Of particular interest is that neither of the two natural spring annual FLC alleles were found in European accessions, suggesting the potential these two alleles evolved after colonizing North America. However, with only a total of 6 spring annual accessions from Europe, this conclusion could easily be disproven with a large sampling of natural accession from a wider geographic range across northern Europe.

Work is currently ongoing to identify early flowering EMS-induced mutations from the same population from which the 'Squirt' FLC mutant (c.55A>T) was derived. While mutations in FLC will be among the first targets for sequencing in any new

mutants, the fact these mutants persisted as winter annuals throughout the course of the 2014-2015 winter season in Minnesota suggests any causative mutation is has a minor affect on earliness, opposed to the complete loss of vernalization requirement seen with the c.55A>T, c.6_7insG, and 2385_2841del mutants. While the guiding goal of this work is develop winter annual varieties of pennycress that rapidly flower in the spring, the identification of the importance of a functional FLC allele here and the identification of new genetic and phenotypic variance for flowering time should allow for the rapid development of an elite pennycress variety that fits within the short growing season of northern climates.

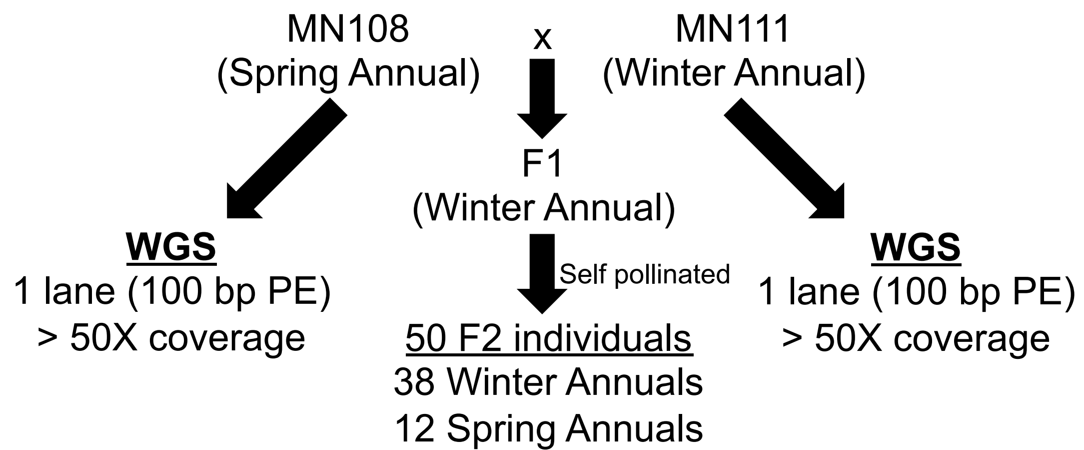


Figure 4-1: MN108 and MN111 accessions used in WGS and to develop F2 population found to segregate for the spring annual phenotype.

		MN111 (WA)	MN108 (SA)
Number of Unfiltered Reads		334,442,965	334,725,738
Number of Filtered and Trimmed Reads		328,837,468	329,656,802
Total Nucleotides in Filtered and Trimmed Reads		29,318,733,923	29,427,924,523
Variants Identified*			
Single Nucleotide Polymorphisms	Heterozygous	61,526	57,952
	Homozygous	433,216	443,713
Multiple Nucleotide Variant	Heterozygous	1,726	1,622
	Homozygous	7,400	7,659
Insertions	Heterozygous	5,334	5,581
	Homozygous	18,708	19,777
Deletions	Heterozygous	4,732	4,994
	Homozygous	22,122	23,590
Replacements	Heterozygous	178	207
	Homozygous	1,173	1,235
	Total Variants Identified	556,115	566,330

*Variants called with average quality > 30

Table 4-1: Summary of Whole Genome Sequencing Results and Variant Detection Analysis in MN108 and MN111

Type	#	Winter Annual (WA) /Spring Annual (SA) Phenotype	Type	#	Winter Annual (WA) /Spring Annual (SA) Phenotype
F2 progeny	1	SA	F2 progeny	26	WA
F2 progeny	2	SA	F2 progeny	27	SA
F2 progeny	3	WA	F2 progeny	28	WA
F2 progeny	4	WA	F2 progeny	29	WA
F2 progeny	5	WA	F2 progeny	30	SA
F2 progeny	6	SA	F2 progeny	31	WA
F2 progeny	7	WA	F2 progeny	32	SA
F2 progeny	8	WA	F2 progeny	33	WA
F2 progeny	9	SA	F2 progeny	34	SA
F2 progeny	10	SA	F2 progeny	35	WA
F2 progeny	11	WA	F2 progeny	36	WA
F2 progeny	12	WA	F2 progeny	37	WA
F2 progeny	13	WA	F2 progeny	38	WA
F2 progeny	14	WA	F2 progeny	39	WA
F2 progeny	15	WA	F2 progeny	40	WA
F2 progeny	16	WA	F2 progeny	41	SA
F2 progeny	17	WA	F2 progeny	42	WA
F2 progeny	18	WA	F2 progeny	43	WA
F2 progeny	19	WA	F2 progeny	44	WA
F2 progeny	20	SA	F2 progeny	45	WA
F2 progeny	21	WA	F2 progeny	46	WA
F2 progeny	22	SA	F2 progeny	47	WA
F2 progeny	23	WA	F2 progeny	48	WA
F2 progeny	24	WA	F2 progeny	49	WA
F2 progeny	25	WA	F2 progeny	50	WA

Table 4-2: Phenotypes of 50 F2 MN111 x MN108 progeny (WA: Winter Annual, SA: Spring Annual – flowered immediately after germination)

Name of oligo	Sequence (5' - 3')
TaFLC_1_Forw	CCGAGGAAGAAAAAGTAGATAGAGACA
TaFLC_1_Rev	GAAGCTTAAAGGGGGAAAAAGGAA
TaFLC_2_Forw	TGTTTTCAAGGGTTAGCTCG
TaFLC_2_internal_Forw	GATCCACCTACGAGCTATTA
TaFLC_2_Rev	TGCCAGCTAATAAAGGATAC
TaFLC_3_For	TCTAGTCTTCTTGGGGATG
TaFLC_3_Rev	ACAACATGCACCCTTATCAA
TaFLC_del_1_Forw	CGAACCATAGTTCAGAGCTT
TaFLC_del_1_Rev	ATAGTGTGCATCAACTGGTC

Table 4-3: Oligonucleotide sequences used for amplification and sequencing of various TaFLC amplicons

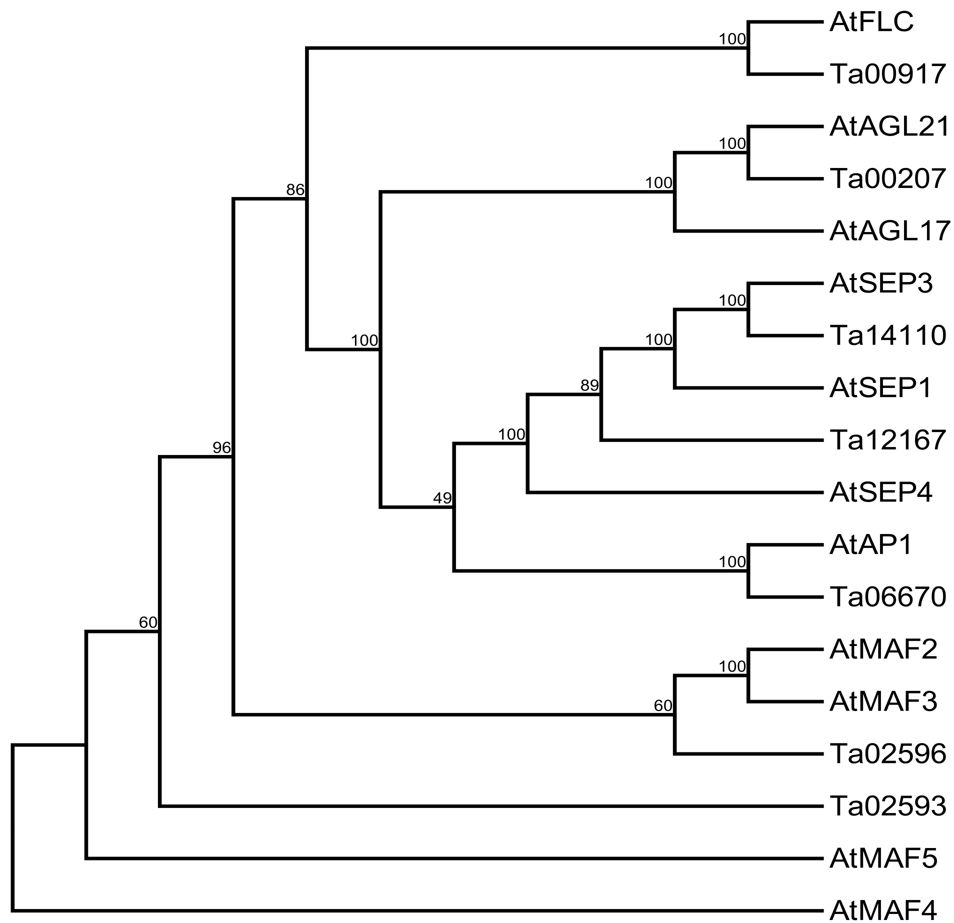


Figure 4-2: Phylogenetic analysis of FLOWERING LOCUS C-like predicted peptides in pennycress and Arabidopsis. Tree created using Neighbor Joining algorithm (Jukes-Cantor distance measure) with 100 bootstrap replicates. Bootstrap values shown at each node.

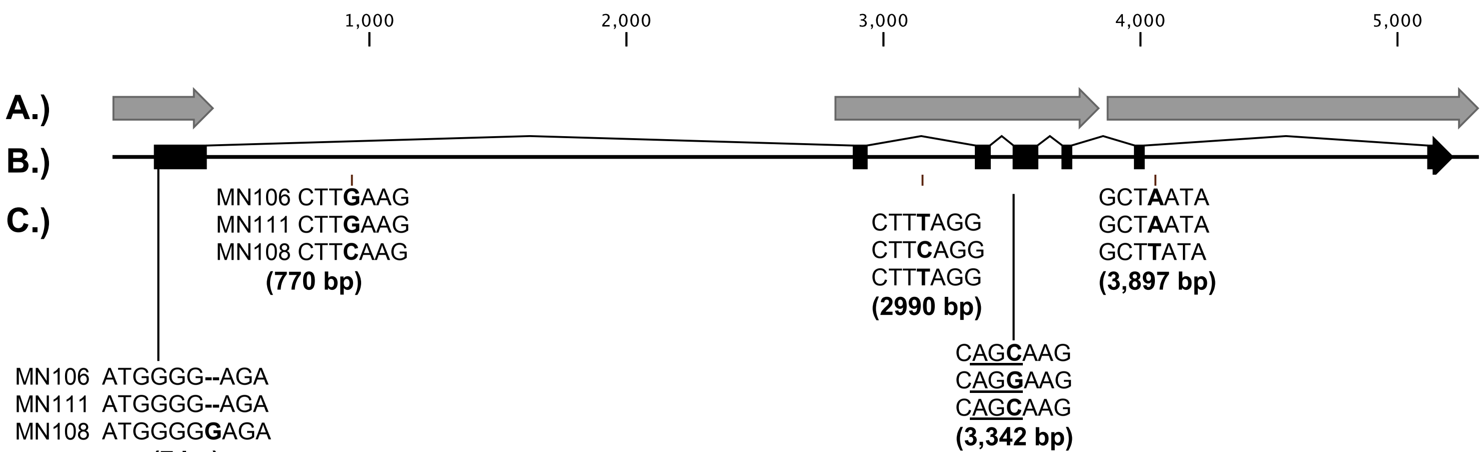

1,000
2,000
3,000
4,000
5,000
A.)
B.)
C.)
MN106 CTTGAAG
MN111 CTTGAAG
MN108 CTTCAAG
(770 bp)
CTTTAGG
CTTCAGG
CTTTAGG
(2990 bp)
GCTAATA
GCTAATA
GCTTATA
(3,897 bp)
CAGCAAG
CAGGAAG
CAGCAAG
(3,342 bp)
MN106 ATGGGG--AGA
MN111 ATGGGG--AGA
MN108 ATGGGGGAGA
(7 bp)
(c.6_7insG)
D.)
MN106 (W.A.) ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAG
M G R K K L E I K R I E
MN111 (W.A.) ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAG
M G R K K L E I K R I E
MN108 (S.A.) ATGGGGGAGAAAAAAACTAGAAATCAAGCGAATTGA
M G E K K T R N Q A N *

Figure 4-3: Overview of TaFLC amplicons and Variants.

A.) Location of amplicons along the TaFLC locus, starting with Amplicon 1 (leftmost arrow). B.) Gene structure of TaFLC, exons shown as black boxes, introns by thin black lines. C.) Variants identified in TaFLC in MN106, MN111, and MN108 at the corresponding positions along TaFLC. D.) Nucleotide and predicted peptide sequence alignment of 5’ end of TaFLC in MN106, MN111, and MN108.

Analysis of FLC mutations in MN108 x MN111 F2 population				
Individual	**SA/WA**	**Position 7**	**Position 2,990**	**Position 3,342**
MN106	WA	G--A	T	C
MN111	WA	G--A	C	G
MN108	SA	GGA	T	C
F2 #1	SA	GGA	T	C
F2 #2	SA	GGA	T	C
F2 #6	SA	GGA	T	C
F2 #9	SA	GGA	T	C
F2 #10	SA	GGA	T	C
F2 #20	SA	GGA	T	C
F2 #22	SA	GGA	T	C
F2 #27	SA	GGA	T	C
F2 #30	SA	GGA	T	C
F2 #32	SA	GGA	T	C
F2 #34	SA	GGA	T	C
F2 #41	SA	GGA	T	C
F2 #3	WA	G--A (Heterozygous)	T/C (Heterozygous)	C/G (Heterozygous)
F2 #4	WA	G--A (Homozygous)	C	G
F2 #5	WA	G--A (Homozygous)	C	G
F2 #7	WA	G--A (Homozygous)	C	G
F2 #8	WA	G--A (Heterozygous)	T/C (Heterozygous)	C/G (Heterozygous)
F2 #11	WA	G--A (Heterozygous)	NS	NS
F2 #18	WA	G--A (Heterozygous)	NS	NS
F2 #31	WA	G--A (Homozygous)	C	G
F2 #38	WA	G--A (Heterozygous)	NS	NS
F2 #39	WA	G--A (Heterozygous)	NS	NS
F2 #40	WA	G--A (Heterozygous)	NS	NS
F2 #42	WA	G--A (Heterozygous)	NS	NS
F2 #43	WA	G--A (Heterozygous)	NS	NS
F2 #44	WA	G--A (Homozygous)	C	G

Table 4: Analysis of FLC Mutations in MN108 x MN111 F2 population.

Summary of the identified variants at position 7 (c.6_7insG mutation), 2,990, and 3,342 in MN106, MN111, MN108, the 12 spring annual MN111 x MN108 progeny, and 14 randomly selected winter annual MN111 x MN108 progeny, with the confirmed sequence for each individual and location shown. All sequences homozygous unless otherwise noted. “NS” indicates individual not sequenced at that location.

B.)

A.)

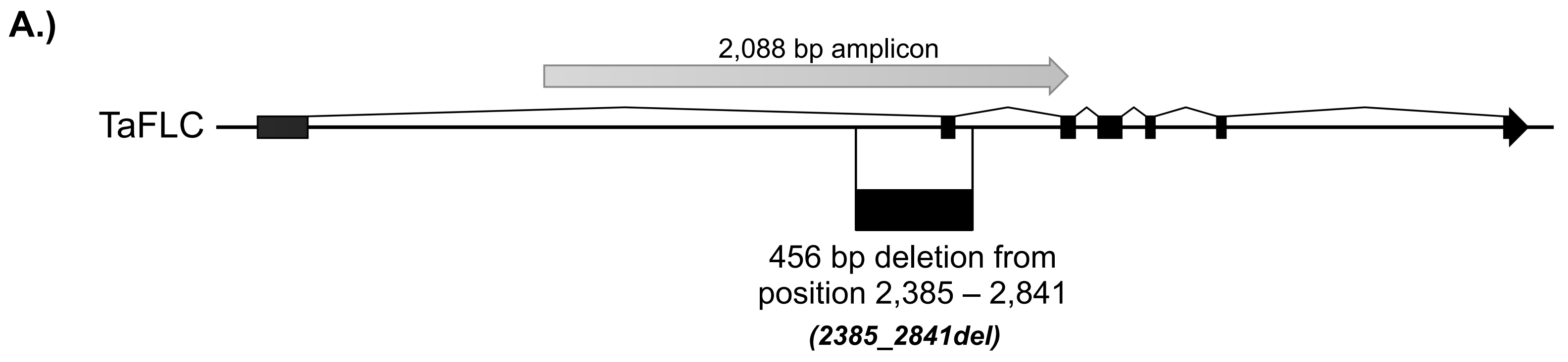
Figure 4-4: Identification of a novel, EMS-induced FLC allele conferring the spring annual phenotype.

A.) Photograph of field grown ‘Squirt’ mutant B.) Photograph of lab-grown ‘Squirt’ mutant identified from pooled M2 seeds. Both photographs courtesy of Evan Johnson and Erin Daniels.

```
AtFLC  ATGGGAAGAAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTC
        M  G  R  K  K  L  E  I  K  R  I  E  N  K  S  S  R  Q  V

MN106  ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTC
        M  G  R  K  K  L  E  I  K  R  I  E  N  K  S  S  R  Q  V

MN111  ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTC
        M  G  R  K  K  L  E  I  K  R  I  E  N  K  S  S  R  Q  V

MN108  ATGGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTC
        M  G  E  K  K  T  R  N  Q  A  N  *

Squirt ATGGGGAGAAAAAAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGATAA
        M  G  R  K  K  L  E  I  K  R  I  E  N  K  S  S  R  *
```

Figure 4-5: Alignments of 5’ end of Arabidopsis and various pennycress FLC alleles and predicted peptides

A.)



B.)

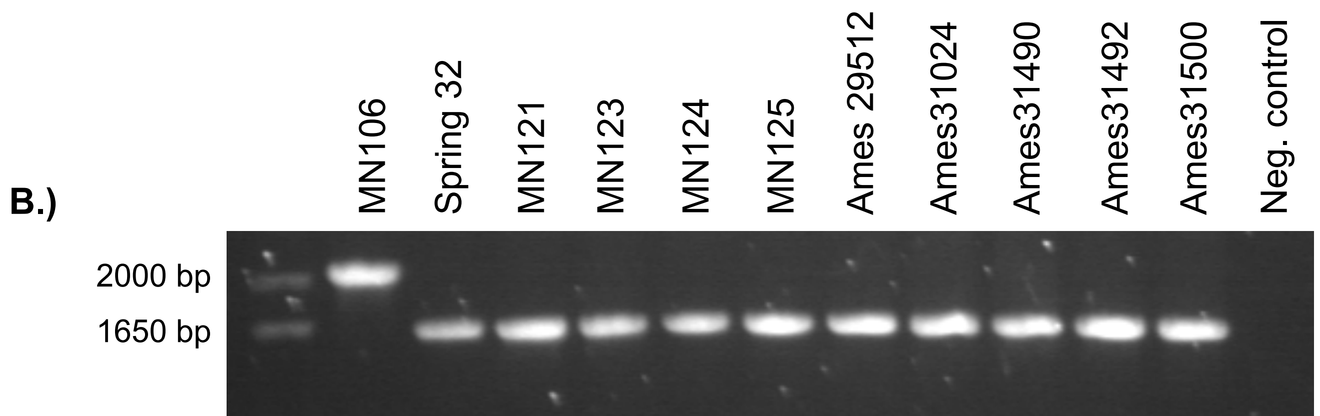
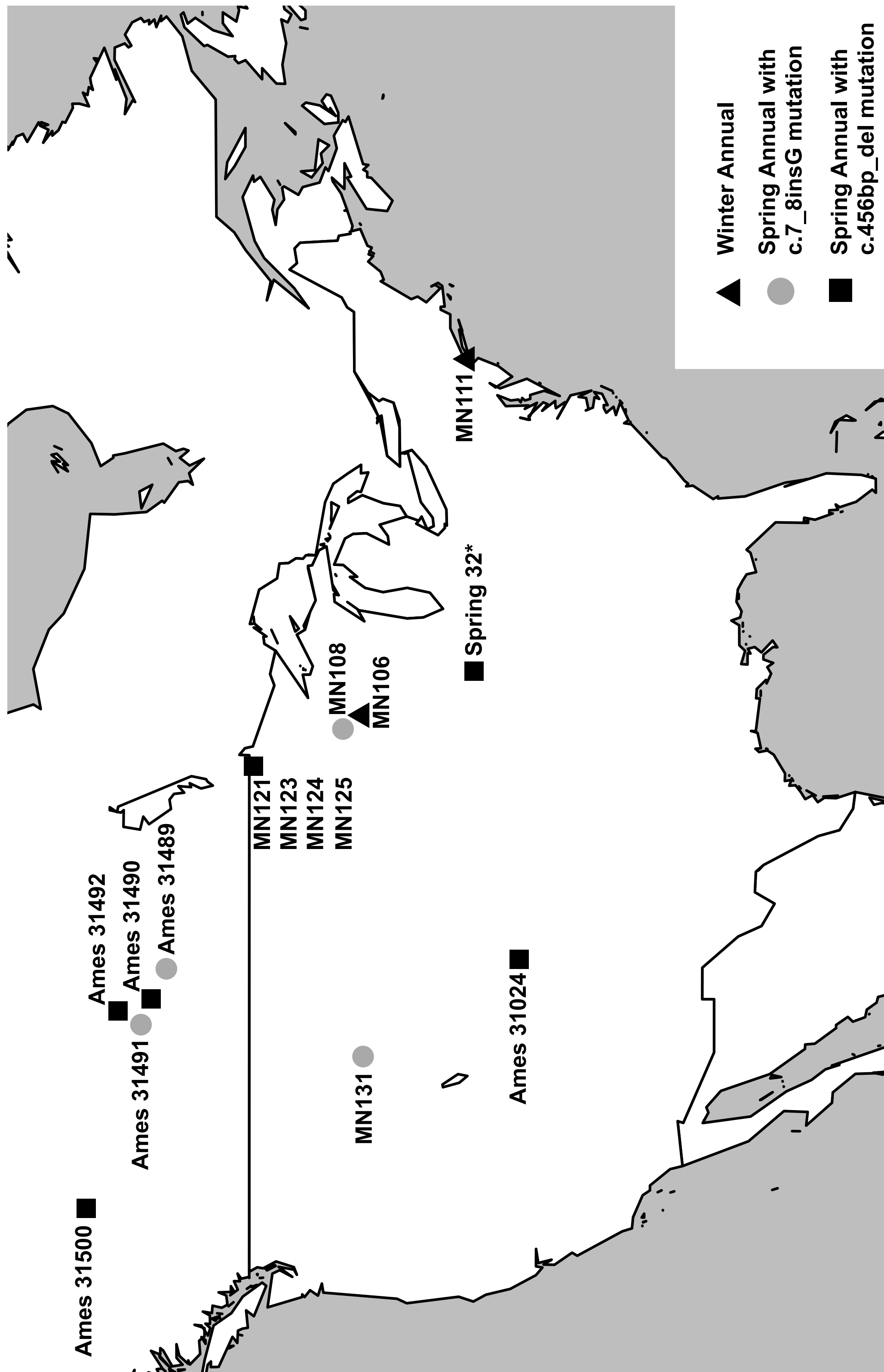


Figure 4-6: Screening and identification of new allele (2385_2841del) conferring the spring annual phenotype

A.) Schematic of TaFLC gene structure and 2 kilobase amplicon used to identify 2385_2841del allele (gray arrow). Bottom – black box indicating position of 456 bp deletion from position 2385 to 2841.

B.) Gel electrophoresis of 2385_2841del amplicons from MN106 and 10 spring annual lines. Predicted fragment size of MN106 amplicon is 2,088 bp. Sanger sequencing of all spring annual lines shown identified a 456 bp deletion, resulting in a 1,632 bp amplicon in the spring lines. Negative control shown in rightmost lane.



Winter Annual
Spring Annual with c.7_8insG mutation
Spring Annual with c.456bp_del mutation
MN111
Spring 32*
MN108
MN106
MN121
MN123
MN124
MN125
Ames 31492
Ames 31490
Ames 31489
Ames 31491
Ames 31024
MN131
Ames 31500

Figure 7: Geographic distribution of winter and spring annual lines analyzed in this study and confirmed FLC alleles.

Table 4-5: Sequence of oligonucleotides used for amplification and sequencing of TaFRI

Name of oligo	Sequence (5' - 3')
TaFRI_1_Fow	TGGTTGGTGCTAATGCTACG
TaFRI_1_Rev	GCCATGTGACCTCCTAGTTC
TaFRI_1_int_Forw	TAAACGCCAGTACGATGACC
TaFRI_2_Forw	AAGCCTTGAGCAATGTTGGT
TaFRI_2_Rev	CAGTTAGCCGCGTTATCGTC
TaFRI_2_int_Forw	AGATGATGAAGAGCCGGAGG

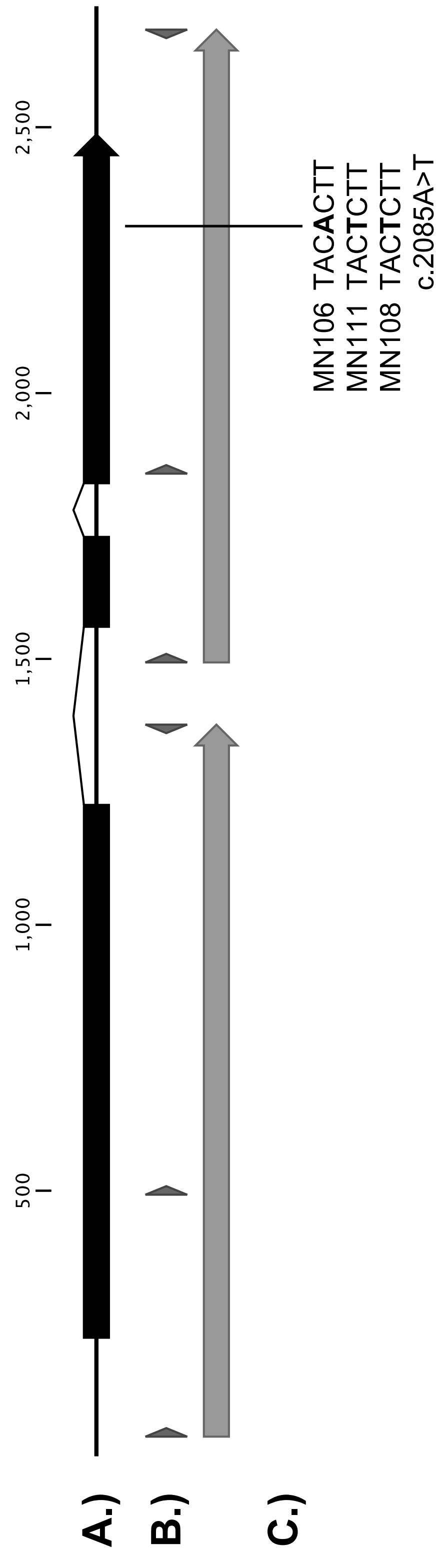

A.)
B.)
C.)
500
1,000
1,500
2,000
2,500
MN106 TACACTT
MN111 TACTCTT
MN108 TACTCTT
c.2085A>T

Figure 8: Pennycress FRIGIDA gene structure and cloning strategy to identify c.2085A>T variant

A.) Gene structure of TaFRI, exons shown as black boxes, introns by thin black lines. B.) Location of primers used for amplification and sequencing C.) Variants identified in TaFRI in MN106, MN111, and MN108

References

R. M. Amasino 2005. Vernalization and flowering time. *Curr Opin Biotechnol,* 16**,** 154-8.

Kf Best & Gi Mc Intyre 1976. Studies on the flowering of Thlaspi arvense L. III. The influence of vernalization under natural and controlled conditions. *Botanical Gazette***,** 121-127.

Kf Best & Gi Mcintyre 1972. Studies on the Flowering Time of Thlaspi arvense L. I. The Influence of some Environmental and Genetic Factors. *Botanical Gazette,* 133**,** 454 - 459.

Je Burn, Dr Smyth, Wj Peacock & Es Dennis 1993. Genes conferring late flowering inArabidopsis thaliana. *Genetica,* 90**,** 147-155.

K. Choi, J. Kim, H. J. Hwang, S. Kim, C. Park, S. Y. Kim & I. Lee 2011. The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *Plant Cell,* 23**,** 289-303.

J. H. Clarke & C. Dean 1994. Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. *Mol Gen Genet,* 242**,** 81-9.

James Gerard Crowley 1999. *Evaluation of Camelina sativa as an alternative oilseed crop*, Teagasc.

K. M. Dorn, J. D. Fankhauser, D. L. Wyse & M. D. Marks 2013. De novo assembly of the pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant J,* 75**,** 1028-38.

K. M. Dorn, J. D. Fankhauser, D. L. Wyse & M. D. Marks 2015. A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *DNA Res,* 22**,** 121-131.

Roque L. Evangelista, Terry A. Isbell & Steven C. Cermak 2012. Extraction of pennycress (Thlaspi arvense L.) seed oil by full pressing. *Industrial Crops and Products,* 37**,** 76-81.

Jiqing Fan, David R. Shonnard, Tom N. Kalnes, Peter B. Johnsen & Serin Rao 2013. A life cycle assessment of pennycress (Thlaspi arvense L.) -derived jet fuel and diesel. *Biomass and Bioenergy,* 55**,** 87-100.

E Jean Finnegan & Elizabeth S Dennis 2007. Vernalization-induced trimethylation of histone H3 lysine 27 at FLC is not maintained in mitotically quiescent cells. *Current Biology,* 17**,** 1978-1983.

A. Franzke, M. A. Lysak, I. A. Al-Shehbaz, M. A. Koch & K. Mummenhoff 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci,* 16**,** 108-16.

Thomas Greb, Joshua S Mylne, Pedro Crevillen, Nuno Geraldo, Hailong An, Anthony R Gendall & Caroline Dean 2007. The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. *Current Biology,* 17**,** 73-78.

Mila P. Hojilla-Evangelista, Roque L. Evangelista, Terry A. Isbell & Gordon W. Selling 2013. Effects of cold-pressing and seed cooking on functional properties of protein in pennycress (Thlaspi arvense L.) seed and press cakes. *Industrial Crops and Products,* 45**,** 223-229.

U. Johanson, J. West, C. Lister, S. Michaels, R. Amasino & C. Dean 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science,* 290**,** 344-7.

J. S. Johnston, A. E. Pepper, A. E. Hall, Z. J. Chen, G. Hodnett, J. Drabek, R. Lopez & H. J. Price 2005. Evolution of genome size in Brassicaceae. *Ann Bot,* 95**,** 229-35.

C. Jung & A. E. Muller 2009. Flowering time control and applications in plant breeding. *Trends Plant Sci,* 14**,** 563-73.

D. H. Kim, M. R. Doyle, S. Sung & R. M. Amasino 2009. Vernalization: winter and the timing of flowering in plants. *Annu Rev Cell Dev Biol,* 25**,** 277-99.

E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. Fitzhugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. Levine, P. Mcewan, K. Mckernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. Mcmurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. Mcpherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C.

Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. Mccombie, M. De La Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. Mclysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. De Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen & Consortium International Human Genome Sequencing 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Gi Mcintyre & Kf Best 1978 Studies on the Flowering of Thlaspi arvense IV. Genetic and Ecological Differences between Early and Late Flowering Strains *Botanical Gazette,* 139**,** 190-195.

S. D. Michaels & R. M. Amasino 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell,* 11**,** 949-56.

S. D. Michaels & R. M. Amasino 2001. Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell,* 13**,** 935-41.

B. R. Moser, G. Knothe, S. F. Vaughn & T. A. Isbell 2009. Production and Evaluation of Biodiesel from Field Pennycress (Thlaspi arvense L.) Oil. *Energy & Fuels,* 23**,** 4149-4155.

W. B. Phippen & M. E. Phippen 2012. Soybean Seed Yield and Quality as a Response to Field Pennycress Residue. *Crop Science,* 52**,** 2767-2773.

J. A. Poland & T. W. Rife 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome,* 5**,** 92-102.

H.L. Koo S. Kaul, J. Jenkins, M. Rizzo, T. Rooney, L.J. Tallon, T. Feldblyum, W. Nierman, M.-I. Benito, M.-I.X. Lin, C.D. Town, J.C. Venter, C.M. Fraser, S.Tabata, Y. Nakamura, T. Kaneko, S. Sato, E. Asamizu, T. Kato, H. Kotani, S. Sasamoto, J.R. Ecker, A. Theologis, N.A. Federspiel, C.J. Palm, B.I. Osborne, P. Shinn, A.B. Conway, V.S. Vysotskaia, K. Dewar, L. Conn, C.A. Lenz, C.J. Kim, N.F. Hansen, S.X. Liu, E. Buehler, H. Altafi, H. Sakano, P. Dunn, B. Lam, P.K. Pham, Q. Chao, M. Nguyen, G. Yu, H. Chen, A. Southwick, J.M. Lee, M. Miranda, M.J. Toriumi, R.W. Davis, R. Wambutt, G. Murphy, A. Düsterhöft, W. Stiekema, T. Pohl, K.-D. Entian, N. Terryn, G. Volckaert, M. Salanoubat, N. Choisne, M. Rieger, W. Ansorge, M. Unseld, B. Fartmann, G. Valle, F. Artiguenave, J. Weissenbach, F. Quetier, R.K. Wilson, M. De La Bastide, M. Sekhon, E. Huang, L. Spiegel, L. Gnoj, K. Pepin, J. Murray, D. Johnson, K.

Habermann, N. Dedhia, L. Parnell, R. Preston, L. Hillier, E. Chen, M. Marra, R. Martienssen, W.R. Mccombie, K. Mayer, O. White, M. Bevan, K. Lemcke, T.H. Creasy, C. Bielke, B. Haas, D. Haase, R. Maiti, S. Rudd, J. Peterson, H. Schoof, D. Frishman, B. Morgenstern, P. Zaccaria, M. Ermolaeva, M. Pertea, J. Quackenbush, N. Volfovsky, D. Wu, T.M. Lowe, S.L. Salzberg, H.-W. Mewes, S. Rounsley, D. Bush, S. Subramaniam, I. Levin, S. Norris, R. Schmidt, A. Acarkan, I. Bancroft, F. Quetier, A. Brennicke, J.A. Eisen, T. Bureau, B.-A. Legault, Q.-H. Le, N. Agrawal, Z. Yu, R. Martienssen, G.P. Copenhaver, S. Luo, C.S. Pikaard, D. Preuss, I.T. Paulsen, M. Sussman, A.B. Britt, J.A. Eisen. D.A. Selinger, R. Pandey, D.W. Mount, V.L. Chandler, R.A. Jorgensen, C. Pikaard, G. Juergens, E.M. Meyerowitz, J.R. Ecker, A. Theologis, J. Dangl, J.D.G. Jones, M. Chen, J. Chory, C. Somerville. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408**,** 796-815.

Iain Searle, Yuehui He, Franziska Turck, Coral Vincent, Fabio Fornara, Sandra Kröber, Richard A Amasino & George Coupland 2006. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes & Development,* 20**,** 898-912.

J. C. Sedbrook, W. B. Phippen & M. D. Marks 2014. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (Thlaspi arvense L.). *Plant Sci,* 227**,** 122-32.

G. Selling, R. Evangelista, T. Isbell & M. H. Evangelista 2011. Extraction of proteins from pennycress. *Abstracts of Papers of the American Chemical Society,* 241.

C. Shindo, M. J. Aranzana, C. Lister, C. Baxter, C. Nicholls, M. Nordborg & C. Dean 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant Physiol,* 138**,** 1163-73.

G. G. Simpson & C. Dean 2002. Arabidopsis, the Rosetta stone of flowering time? *Science,* 296**,** 285-9.

J. R. Stinchcombe, C. Weinig, M. Ungerer, K. M. Olsen, C. Mays, S. S. Halldorsdottir, M. D. Purugganan & J. Schmitt 2004. A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *Proc Natl Acad Sci U S A,* 101**,** 4712-7.

Sibum Sung, Yuehui He, Tifani W Eshoo, Yosuke Tamada, Lianna Johnson, Kenji Nakahigashi, Koji Goto, Steve E Jacobsen & Richard M Amasino 2006. Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1. *Nat Genet,* 38**,** 706-710.

Million Tadege, Candice C Sheldon, Chris A Helliwell, Peter Stoutjesdijk, Elizabeth S Dennis & W James Peacock 2001. Control of flowering time by FLC orthologues in Brassica napus. *The Plant Journal,* 28**,** 545-553.

R. K. Varshney, C. Song, R. K. Saxena, S. Azam, S. Yu, A. G. Sharpe, S. Cannon, J. Baek, B. D. Rosen, B. Tar'an, T. Millan, X. Zhang, L. D. Ramsay, A. Iwata, Y. Wang, W. Nelson, A. D. Farmer, P. M. Gaur, C. Soderlund, R. V. Penmetsa, C. Xu, A. K. Bharti, W. He, P. Winter, S. Zhao, J. K. Hane, N. Carrasquilla-Garcia, J. A. Condie, H. D. Upadhyaya, M. C. Luo, M. Thudi, C. L. Gowda, N. P. Singh, J. Lichtenzveig, K. K. Gali, J. Rubio, N. Nadarajan, J. Dolezel, K. C. Bansal, X. Xu, D. Edwards, G. Zhang, G. Kahl, J. Gil, K. B. Singh, S. K. Datta, S. A. Jackson, J. Wang & D. R. Cook 2013. Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nat Biotechnol,* 31**,** 240-6.

J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos,

C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. Mckusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. Mccawley, T. Mcintosh, I. Mcmullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski,

K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. Mcdaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh & X. Zhu 2001. The sequence of the human genome. *Science,* 291**,** 1304-51.

S.I. Warwick, Francis, A., Susko, D.J. 2002 The biology of Canadian weeds. 9. Thlaspi arvense L. (updated) *Canadian Journal of Plant Science,* 82**,** 803-823.

Jian Wu, Keyun Wei, Feng Cheng, Shikai Li, Qian Wang, Jianjun Zhao, Guusje Bonnema & Xiaowu Wang 2012. A naturally occurring InDel variation in BraA. FLC. b (BrFLC2) associated with flowering time variation in Brassica rapa. *BMC Plant Biol,* 12**,** 151.

Comprehensive Bibliography

A. E. Agee, M. Surpin, E. J. Sohn, T. Girke, A. Rosado, B. W. Kram, C. Carter, A. M. Wentzell, D. J. Kliebenstein, H. C. Jin, O. K. Park, H. Jin, G. R. Hicks & N. V. Raikhel 2010. MODIFIED VACUOLE PHENOTYPE1 is an Arabidopsis myrosinase-associated protein involved in endomembrane protein trafficking. *Plant Physiol,* 152**,** 120-32.

R. M. Amasino 2005a. Vernalization and flowering time. *Curr Opin Biotechnol,* 16**,** 154-8.

Richard M. Amasino 2005b. Vernalization and flowering time. *Current Opinion in Biotechnology,* 16**,** 154-158.

T. G. Andersen, H. H. Nour-Eldin, V. L. Fuller, C. E. Olsen, M. Burow & B. A. Halkier 2013. Integration of biosynthesis and long-distance transport establish organ-specific glucosinolate profiles in vegetative Arabidopsis. *Plant Cell,* 25**,** 3133-45.

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

Sureshkumar Balasubramanian, Sridevi Sureshkumar, Janne Lempe & Detlef Weigel 2006. Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. *PLoS Genet,* 2**,** e106.

Jm Baskin & Cc Baskin 1989. Role of temperature in regulating timing of germination in soil seed reserves of Thlaspi arvense L. *Weed Research,* 29**,** 317-326.

Ruth Bastow, Joshua S Mylne, Clare Lister, Zachary Lippman, Robert A Martienssen & Caroline Dean 2004. Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature,* 427**,** 164-167.

S. Baud & L. Lepiniec 2009. Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. *Plant Physiol Biochem,* 47**,** 448-55.

K. Belhaj, A. Chaparro-Garcia, S. Kamoun & V. Nekrasov 2013. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods,* 9**,** 39.

L. Bentsink, J. Jowett, C. J. Hanhart & M. Koornneef 2006. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proc Natl Acad Sci U S A,* 103**,** 17042-7.

K. F. Best & G. I. Mcintyre 1975a. Biology of Canadian Weeds, 9. Thlaspi arvense L. . *Canadian Journal of Plant Science,* 55**,** 279-292.

K. F. Best & G. I. Mcintyre 1976. Studies on Flowering of Thlaspi-Arvense L .3. Influence of Vernalization under Natural and Controlled Conditions. *Botanical Gazette,* 137**,** 121-127.

Kf Best & Gi Mcintyre 1975b. The Biology of Canadian Weeds 9. Thlaspi arvense L. . *Canadian Journal of Plant Science,* 55**,** 279-292.

Kf Best & Gi Mc Intyre 1976. Studies on the flowering of Thlaspi arvense L. III. The influence of vernalization under natural and controlled conditions. *Botanical Gazette***,** 121-127.

Kf Best & Gi Mcintyre 1972. Studies on the Flowering Time of Thlaspi arvense L. I. The Influence of some Environmental and Genetic Factors. *Botanical Gazette,* 133**,** 454 - 459.

David Biello 2011. The false promise of biofuels. *Scientific American,* 305**,** 58-65.

Miguel A Blázquez, Ji Hoon Ahn & Detlef Weigel 2003. A thermosensory pathway controlling flowering time in Arabidopsis thaliana. *Nat Genet,* 33**,** 168-171.

Miguel A Blázquez, Roland Green, Ove Nilsson, Michael R Sussman & Detlef Weigel 1998. Gibberellins promote flowering of Arabidopsis by activating the LEAFY promoter. *The Plant Cell Online,* 10**,** 791-800.

Miguel A Blázquez, Lara N Soowal, Ilha Lee & Detlef Weigel 1997. LEAFY expression and flower initiation in Arabidopsis. *Development,* 124**,** 3835-3844.

A. A. Boateng, C. A. Mullen & N. M. Goldberg 2010a. Producing Stable Pyrolysis Liquids from the Oil-Seed Presscakes of Mustard Family Plants: Pennycress (Thlaspi arvense L.) and Camelina (Camelina sativa). *Energy & Fuels,* 24**,** 6624-6632.

A. A. Boateng, C. A. Mullen & N. M. Goldberg 2010b. Producing Stable Pyrolysis Liquids from the Oil-Seed Presscakes of Mustard Family Plants: Pennycress (Thlaspi arvense L.) and Camelina (Camelina sativa)†. *Energy & Fuels,* 24**,** 6624-6632.

M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler & W. Pirovano 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics,* 27**,** 578-9.

M. Boetzer & W. Pirovano 2012. Toward almost closed genomes with GapFiller. *Genome Biol,* 13**,** R56.

A. M. Bones & J. T. Rossiter 1996. The myrosinase-glucosinolate system, its organisation and biochemistry. *Physiologia Plantarum,* 97**,** 194-208.

John L Bowman, John Alvarez, Detlef Weigel, Elliot M Meyerowitz & David R Smyth 1993. Control of flower development in Arabidopsis thaliana by APETALA 1 and interacting genes. *DEVELOPMENT-CAMBRIDGE-,* 119**,** 721-721.

Je Burn, Dr Smyth, Wj Peacock & Es Dennis 1993. Genes conferring late flowering inArabidopsis thaliana. *Genetica,* 90**,** 147-155.

Mario Caccamo & Erich Grotewold 2013. Turning over a new leaf in plant genomics. *Genome Biol,* 14**,** 403.

B. L. Cantarel, I. Korf, S. M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sanchez Alvarado & M. Yandell 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res,* 18**,** 188-96.

T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove & D. F. Voytas 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res,* 39**,** e82.

M Chailakhyan 1936. New facts in support of the hormonal theory of plant development *Proceedings of the USSR Academy of Sciences,* 4**,** 79-83.

C. H. Chen, T. C. Kuo, M. H. Yang, T. Y. Chien, M. J. Chu, L. C. Huang, C. Y. Chen, H. F. Lo, S. T. Jeng & L. F. Chen 2014. Identification of cucurbitacins and assembly of a draft genome for Aquilaria agallocha. *BMC Genomics,* 15**,** 578.

Hyun Jung Cho, Jae Joon Kim, Jeong Hwan Lee, Wanhui Kim, Jae-Hoon Jung, Chung-Mo Park & Ji Hoon Ahn 2012. SHORT VEGETATIVE PHASE (SVP) protein negatively regulates miR172 transcription via direct binding to the pri-miR172a promoter in Arabidopsis. *FEBS Lett,* 586**,** 2332-2337.

K. Choi, J. Kim, H. J. Hwang, S. Kim, C. Park, S. Y. Kim & I. Lee 2011. The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *Plant Cell,* 23**,** 289-303.

J. H. Clarke & C. Dean 1994. Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. *Mol Gen Genet,* 242**,** 81-9.

A. Conesa, S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon & M. Robles 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics,* 21**,** 3674-3676.

Laurent Corbesier, Coral Vincent, Seonghoe Jang, Fabio Fornara, Qingzhi Fan, Iain Searle, Antonis Giakountis, Sara Farrona, Lionel Gissot & Colin Turnbull 2007. FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science,* 316**,** 1030-1033.

James Gerard Crowley 1999. *Evaluation of Camelina sativa as an alternative oilseed crop*, Teagasc.

S. M. Dabney, J. A. Delgado & D. W. Reeves 2001a. Using Winter Cover Crops to Improve Soil and Water Quality. *Communications in Soil Science and Plant Analysis,* 32**,** 1221-1250.

Sm Dabney, Ja Delgado & Dw Reeves 2001b. Using winter cover crops to improve soil and water quality. *Communications in Soil Science and Plant Analysis,* 32**,** 1221-1250.

Michel Delseny, Bin Han & Yue Ie Hsing 2010. High throughput DNA sequencing: The new sequencing revolution. *Plant Science,* 179**,** 407-422.

John F Doebley, Brandon S Gaut & Bruce D Smith 2006. The molecular genetics of crop domestication. *Cell,* 127**,** 1309-1321.

K. M. Dorn, J. D. Fankhauser, D. L. Wyse & M. D. Marks 2013. De novo assembly of the pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant J,* 75**,** 1028-38.

K. M. Dorn, J. D. Fankhauser, D. L. Wyse & M. D. Marks 2015. A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *DNA Res,* 22**,** 121-131.

K. M.; Fankhauser Dorn, J.D.; Wyse, D.L., Marks, M.D. 2015. A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *DNA Research*.

Roque L. Evangelista, Terry A. Isbell & Steven C. Cermak 2012. Extraction of pennycress (Thlaspi arvense L.) seed oil by full pressing. *Industrial Crops and Products,* 37**,** 76-81.

J. Q. Fan, D. R. Shonnard, T. N. Kalnes, P. B. Johnsen & S. Rao 2013a. A life cycle assessment of pennycress (Thlaspi aruense L.) -derived jet fuel and diesel. *Biomass & Bioenergy,* 55**,** 87-100.

J. Fan, Shonnard, D., Kalnes, T., Johnsen, P., Rao, S. 2013. A life cycle assessment of pennycress (Thlaspi arvense L.) -derived jet fuel and diesel. *Biomass and Bioenergy,* In Press.

Jiqing Fan, David R. Shonnard, Tom N. Kalnes, Peter B. Johnsen & Serin Rao 2013b. A life cycle assessment of pennycress (Thlaspi arvense L.) -derived jet fuel and diesel. *Biomass and Bioenergy,* 55**,** 87-100.

J. Fargione, J. Hill, D. Tilman, S. Polasky & P. Hawthorne 2008. Land clearing and the biofuel carbon debt. *Science,* 319**,** 1235-1238.

Cristina Ferrándiz, Qing Gu, Robert Martienssen & Martin F Yanofsky 2000. Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development,* 127**,** 725-734.

Catherine Feuillet, Jan E Leach, Jane Rogers, Patrick S Schnable & Kellye Eversole 2011. Crop genome sequencing: lessons and rationales. *Trends Plant Sci,* 16**,** 77-88.

E Jean Finnegan & Elizabeth S Dennis 2007. Vernalization-induced trimethylation of histone H3 lysine 27 at FLC is not maintained in mitotically quiescent cells. *Current Biology,* 17**,** 1978-1983.

Fabio Fornara, Kishore Cs Panigrahi, Lionel Gissot, Nicolas Sauerbrunn, Mark Rühl, José A Jarillo & George Coupland 2009. Arabidopsis DOF transcription factors

act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev Cell,* 17**,** 75-86.

A. Franzke, M. A. Lysak, I. A. Al-Shehbaz, M. A. Koch & K. Mummenhoff 2011a. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci,* 16**,** 108-16.

A. Franzke, M. A. Lysak, I. A. Al-Shehbaz, M. A. Koch & K. Mummenhoff 2011b. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci,* 16**,** 108-116.

Vinicius C Galvão, Daniel Horrer, Frank Küttner & Markus Schmid 2012. Spatial control of flowering by DELLA proteins in Arabidopsis thaliana. *Development,* 139**,** 4072-4082.

Ww Garner & Ha Allard 1922. Photoperiodism, the response of the plant to relative length of day and night. *Science***,** 582-583.

Tara Garnett, Mc Appleby, A Balmford, Ij Bateman, Tg Benton, P Bloomer, B Burlingame, M Dawkins, L Dolan & D Fraser 2013. Sustainable intensification in agriculture: premises and policies. *Science,* 341**,** 33-34.

S. Gazzani 2003. Analysis of the Molecular Basis of Flowering Time Variation in Arabidopsis Accessions. *Plant Physiology,* 132**,** 1107-1114.

H Charles J Godfray, John R Beddington, Ian R Crute, Lawrence Haddad, David Lawrence, James F Muir, Jules Pretty, Sherman Robinson, Sandy M Thomas & Camilla Toulmin 2010. Food security: the challenge of feeding 9 billion people. *Science,* 327**,** 812-818.

E. Gongora-Castillo & C. R. Buell 2013. Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Nat Prod Rep*.

Thomas Greb, Joshua S Mylne, Pedro Crevillen, Nuno Geraldo, Hailong An, Anthony R Gendall & Caroline Dean 2007. The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. *Current Biology,* 17**,** 73-78.

R and Virtanen Grmelin, A 1959 A new type of enzymatic cleavage of mustard oil glucosides. Formation of allylthiocyanate in Thlasip arvense L and Benzylthiocynate in Lepidium ruderale L and Lepidium sativum L. . *Acta Chem. Scand,* 13.

C. D. Grubb & S. Abel 2006. Glucosinolate metabolism and its control. *Trends Plant Sci,* 11**,** 89-100.

N. Gruenheit, O. Deusch, C. Esser, M. Becker, C. Voelckel & P. Lockhart 2012. Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants. *BMC Genomics,* 13.

S. Guo, J. Zhang, H. Sun, J. Salse, W. J. Lucas, H. Zhang, Y. Zheng, L. Mao, Y. Ren, Z. Wang, J. Min, X. Guo, F. Murat, B. K. Ham, Z. Zhang, S. Gao, M. Huang, Y. Xu, S. Zhong, A. Bombarely, L. A. Mueller, H. Zhao, H. He, Y. Zhang, Z. Zhang, S. Huang, T. Tan, E. Pang, K. Lin, Q. Hu, H. Kuang, P. Ni, B. Wang, J. Liu, Q. Kou, W. Hou, X. Zou, J. Jiang, G. Gong, K. Klee, H. Schoof, Y. Huang, X. Hu, S. Dong, D. Liang, J. Wang, K. Wu, Y. Xia, X. Zhao, Z. Zheng, M. Xing, X. Liang, B. Huang, T. Lv, J. Wang, Y. Yin, H. Yi, R. Li, M. Wu, A. Levi, X. Zhang, J. J. Giovannoni, J. Wang, Y. Li, Z. Fei & Y. Xu 2013. The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. *Nat Genet,* 45**,** 51-8.

B. A. Halkier & J. Gershenzon 2006. Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol,* 57**,** 303-33.

J. P. Hamilton & C. R. Buell 2012. Advances in plant genome sequencing. *Plant J,* 70**,** 177-90.

J. P. Hammond, H. C. Bowen, P. J. White, V. Mills, K. A. Pyke, A. J. M. Baker, S. N. Whiting, S. T. May & M. R. Broadley 2006. A comparison of the Thlaspi caerulescens and Thlaspi arvense shoot transcriptomes. *New Phytologist,* 170**,** 239-260.

Erika Check Hayden 2014. Technology: the $1,000 genome. *Nature,* 507**,** 294-5.

Jan P Hazebroek & James D Metzger 1990a. Environmental control of seed germination in Thlaspi arvense (Cruciferae). *Am J Bot***,** 945-953.

Jan P Hazebroek & James D Metzger 1990b. Seasonal pattern of seedling emergence, survival, and reproductive behavior in Thlaspi arvense (Cruciferae). *Am J Bot***,** 954-962.

Yuehui He, Scott D Michaels & Richard M Amasino 2003. Regulation of flowering time by histone acetylation in Arabidopsis. *Science,* 302**,** 1751-1754.

Emily A Heaton, Lisa A Schulte, Marisol Berti, Hans Langeveld, Walter Zegada-Lizarazu, David Parrish & Andrea Monti 2013. Managing a second-generation crop portfolio through sustainable intensification: Examples from the USA and the EU. *Biofuels, Bioproducts and Biorefining,* 7**,** 702-714.

Chris A Helliwell, Craig C Wood, Masumi Robertson, W James Peacock & Elizabeth S Dennis 2006. The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. *The Plant Journal,* 46**,** 183-192.

J. Hill, E. Nelson, D. Tilman, S. Polasky & D. Tiffany 2006a. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 11206-11210.

J. Hill, E. Nelson, D. Tilman, S. Polasky & D. Tiffany 2006b. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc Natl Acad Sci U S A,* 103**,** 11206-10.

J. A. Hofberger, E. Lyons, P. P. Edger, J. Chris Pires & M. Eric Schranz 2013. Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family. *Genome Biol Evol,* 5**,** 2155-73.

M. P. Hojilla-Evangelista, R. L. Evangelista, T. A. Isbell & G. W. Selling 2013. Effects of cold-pressing and seed cooking on functional properties of protein in pennycress (Thlaspi arvense L.) seed and press cakes. *Industrial Crops and Products,* 45**,** 223-229.

Mila P. Hojilla-Evangelista, Roque L. Evangelista, Terry A. Isbell, and Gordon W. Selling 2013. Effects of cold-pressing and seed cooking on functional properties of protein in pennycress (< i> Thlaspi arvense</i> L.) seed and press cakes. *Industrial Crops and Products,* 45**,** 223-229.

M. J. Holdsworth, L. Bentsink & W. J. Soppe 2008. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytol,* 179**,** 33-54.

T. T. Hu, P. Pattyn, E. G. Bakker, J. Cao, J. F. Cheng, R. M. Clark, N. Fahlgren, J. A. Fawcett, J. Grimwood, H. Gundlach, G. Haberer, J. D. Hollister, S. Ossowski, R. P. Ottilar, A. A. Salamov, K. Schneeberger, M. Spannagl, X. Wang, L. Yang, M. E. Nasrallah, J. Bergelson, J. C. Carrington, B. S. Gaut, J. Schmutz, K. F. Mayer, Y. Van De Peer, I. V. Grigoriev, M. Nordborg, D. Weigel & Y. L. Guo 2011a. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat Genet,* 43**,** 476-81.

T. T. Hu, P. Pattyn, E. G. Bakker, J. Cao, J. F. Cheng, R. M. Clark, N. Fahlgren, J. A. Fawcett, J. Grimwood, H. Gundlach, G. Haberer, J. D. Hollister, S. Ossowski, R.

P. Ottilar, A. A. Salamov, K. Schneeberger, M. Spannagl, X. Wang, L. Yang, M. E. Nasrallah, J. Bergelson, J. C. Carrington, B. S. Gaut, J. Schmutz, K. F. X. Mayer, Y. V. De Peer, I. V. Grigoriev, M. Nordborg, D. Weigel & Y. L. Guo 2011b. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat Genet,* 43**,** 476-+.

Eva Huala & Ian M Sussex 1992. LEAFY interacts with floral homeotic genes to regulate Arabidopsis floral development. *The Plant Cell Online,* 4**,** 901-913.

P. Huijser & M. Schmid 2011. The control of developmental phase transitions in plants. *Development,* 138**,** 4117-29.

L Hume 1990. Influence of emergence date and strain on phenology, seed production, and germination of Thlaspi arvense L. *Botanical Gazette***,** 510-515.

L. Hume, M. D. Devine & S. Shirriff 1995. The Influence of Temperature Upon Physiological Processes in Early-Flowering and Late-Flowering Strains of Thlaspi-Arvense L. *International Journal of Plant Sciences,* 156**,** 445-449.

Larry Hume 1994. Maternal environment effects on plant growth and germination of two strains of Thlaspi arvense L. *International journal of plant sciences***,** 180-186.

Takato Imaizumi, Thomas F Schultz, Frank G Harmon, Lindsey A Ho & Steve A Kay 2005. FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. *Science,* 309**,** 293-297.

Consortium International Barley Genome Sequencing, K. F. Mayer, R. Waugh, J. W. Brown, A. Schulman, P. Langridge, M. Platzer, G. B. Fincher, G. J. Muehlbauer, K. Sato, T. J. Close, R. P. Wise & N. Stein 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature,* 491**,** 711-6.

J. A. Irwin, C. Lister, E. Soumpourou, Y. W. Zhang, E. C. Howell, G. Teakle & C. Dean 2012. Functional alleles of the flowering time regulator FRIGIDA in the Brassica oleracea genome. *BMC Plant Biology,* 12.

Terry A. Isbell & Steven C. Cermak 2012. Extraction of pennycress (Thlaspi arvense L.) seed oil by full pressing. *Industrial Crops & Products,* 37**,** 6p.

Steven E Jacobsen & Neil E Olszewski 1993. Mutations at the SPINDLY locus of Arabidopsis alter gibberellin signal transduction. *The Plant Cell Online,* 5**,** 887-896.

Jose A Jarillo & Manuel Piñeiro 2011. Timing is everything in plant development. The central role of floral repressors. *Plant Science,* 181**,** 364-378.

U. Johanson 2000. Molecular Analysis of FRIGIDA, a Major Determinant of Natural Variation in Arabidopsis Flowering Time. *Science,* 290**,** 344-347.

U. Johanson, J. West, C. Lister, S. Michaels, R. Amasino & C. Dean 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science,* 290**,** 344-7.

Gregg A Johnson, Michael B Kantar, Kevin J Betts & Donald L Wyse 2015. Field Pennycress Production and Weed Control in a Double Crop System with Soybean in Minnesota. *Agronomy Journal*.

J. S. Johnston 2005. Evolution of Genome Size in Brassicaceae. *Annals of Botany,* 95**,** 229-235.

J. S. Johnston, A. E. Pepper, A. E. Hall, Z. J. Chen, G. Hodnett, J. Drabek, R. Lopez & H. J. Price 2005. Evolution of genome size in Brassicaceae. *Ann Bot,* 95**,** 229-35.

N Jordan, G Boody, W Broussard, Jd Glover, D Keeney, Bh Mccown, G Mcisaac, M Muller, H Murray & J Neal 2007. Sustainable development of the agricultural bio-economy. *SCIENCE-NEW YORK THEN WASHINGTON-,* 316**,** 1570.

C. Jung & A. E. Muller 2009. Flowering time control and applications in plant breeding. *Trends Plant Sci,* 14**,** 563-73.

Christian Jung & Andreas E. Müller 2009. Flowering time control and applications in plant breeding. *Trends in Plant Science,* 14**,** 563-573.

J. H. Jung, P. J. Seo, S. K. Kang & C. M. Park 2011. miR172 signals are incorporated into the miR156 signaling pathway at the SPL3/4/5 genes in Arabidopsis developmental transitions. *Plant Mol Biol,* 76**,** 35-45.

Jae-Hoon Jung, Pil Joon Seo, Ji Hoon Ahn & Chung-Mo Park 2012. Arabidopsis RNA-binding protein FCA regulates microRNA172 processing in thermosensory flowering. *Journal of Biological Chemistry,* 287**,** 16007-16016.

S. Kagale, C. S. Koh, J. Nixon, V. Bollina, W. E. Clarke, R. Tuteja, C. Spillane, S. J. Robinson, M. G. Links, C. Clarke, E. E. Higgins, T. Huebert, A. G. Sharpe & I. A. P. Parkin 2014. The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nat Commun,* 5.

H. ; Taab Karimmojeni, A.; Rashidi, B.; Bazrafshan, Ah. 2014. Dormancy breaking and seed germination of the annual weeds Thlaspi arvense, Descurainia sophia and Malcolmia africana (Brassicaceae). *JOURNAL OF PLANT PROTECTION RESEARCH,* 54**,** 179-187.

S. Kaul, H. L. Koo, J. Jenkins, M. Rizzo, T. Rooney, L. J. Tallon, T. Feldblyum, W. Nierman, M. I. Benito, X. Y. Lin, C. D. Town, J. C. Venter, C. M. Fraser, S. Tabata, Y. Nakamura, T. Kaneko, S. Sato, E. Asamizu, T. Kato, H. Kotani, S. Sasamoto, J. R. Ecker, A. Theologis, N. A. Federspiel, C. J. Palm, B. I. Osborne, P. Shinn, A. B. Conway, V. S. Vysotskaia, K. Dewar, L. Conn, C. A. Lenz, C. J. Kim, N. F. Hansen, S. X. Liu, E. Buehler, H. Altafi, H. Sakano, P. Dunn, B. Lam, P. K. Pham, Q. Chao, M. Nguyen, G. X. Yu, H. M. Chen, A. Southwick, J. M. Lee, M. Miranda, M. J. Toriumi, R. W. Davis, R. Wambutt, G. Murphy, A.

Dusterhoft, W. Stiekema, T. Pohl, K. D. Entian, N. Terryn, G. Volckaert, M. Salanoubat, N. Choisne, M. Rieger, W. Ansorge, M. Unseld, B. Fartmann, G. Valle, F. Artiguenave, J. Weissenbach, F. Quetier, R. K. Wilson, M. De La Bastide, M. Sekhon, E. Huang, L. Spiegel, L. Gnoj, K. Pepin, J. Murray, D. Johnson, K. Habermann, N. Dedhia, L. Parnell, R. Preston, L. Hillier, E. Chen, M. Marra, R. Martienssen, W. R. Mccombie, K. Mayer, O. White, M. Bevan, K. Lemcke, T. H. Creasy, C. Bielke, B. Haas, D. Haase, R. Maiti, S. Rudd, J. Peterson, H. Schoof, D. Frishman, B. Morgenstern, P. Zaccaria, M. Ermolaeva, M. Pertea, J. Quackenbush, N. Volfovsky, D. Y. Wu, T. M. Lowe, S. L. Salzberg, H. W. Mewes, S. Rounsley, D. Bush, S. Subramaniam, I. Levin, S. Norris, R. Schmidt, A. Acarkan, I. Bancroft, F. Quetier, A. Brennicke, J. A. Eisen, T. Bureau, B. A. Legault, Q. H. Le, N. Agrawal, Z. Yu, R. Martienssen, G. P. Copenhaver, S. Luo, C. S. Pikaard, D. Preuss, I. T. Paulsen, M. Sussman, A. B. Britt, D. A. Selinger, R. Pandey, D. W. Mount, V. L. Chandler, R. A. Jorgensen, C. Pikaard, G. Juergens, E. M. Meyerowitz, A. Theologis, J. Dangl, J. D. G. Jones, M. Chen, J. Chory, M. C. Somerville & Ar Gen In 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408**,** 796-815.

D. H. Kim, M. R. Doyle, S. Sung & R. M. Amasino 2009a. Vernalization: winter and the timing of flowering in plants. *Annu Rev Cell Dev Biol,* 25**,** 277-99.

Dong-Hwan Kim, Mark R. Doyle, Sibum Sung & Richard M. Amasino 2009b. Vernalization: Winter and the Timing of Flowering in Plants. *Annual Review of Cell and Developmental Biology,* 25**,** 277-299.

J. J. Kim, J. H. Lee, W. Kim, H. S. Jung, P. Huijser & J. H. Ahn 2012. The microRNA156-SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3 module regulates ambient temperature-responsive flowering via FLOWERING LOCUS T in Arabidopsis. *Plant Physiol,* 159**,** 461-78.

Lj Klebesadel 1969. Life cycles of field pennycress in the subarctic as influenced by time of seed germination. *Weed Science*, 563-566.

Daniel C Koboldt, Karyn Meltz Steinberg, David E Larson, Richard K Wilson & Elaine R Mardis 2013. The next-generation sequencing revolution and its impact on genomics. *Cell,* 155, 27-38.

M and Al-Shehbaz Koch, I 2004. Taxonomic and Phylogenetic Evaluation of the American "Thlaspi" species: Identity and relationship to the Eurasian genus Noccaea (Brassicaceae). *Systematic Botany,* 29, 375-384.

M. A. Koch & D. A. German 2013. Taxonomy and systematics are key to biological information: Arabidopsis, Eutrema (Thellungiella), Noccaea and Schrenkiella (Brassicaceae) as examples. *Front Plant Sci,* 4.

A. Konieczny & F. M. Ausubel 1993. A Procedure for Mapping Arabidopsis Mutations Using Codominant Ecotype-Specific Pcr-Based Markers. *Plant Journal,* 4, 403-410.

I. Korf 2004. Gene finding in novel genomes. *BMC Bioinformatics,* 5, 59.

J. C. Kuchernig, A. Backenkohler, M. Lubbecke, M. Burow & U. Wittstock 2011a. A thiocyanate-forming protein generates multiple products upon allylglucosinolate breakdown in Thlaspi arvense. *Phytochemistry,* 72, 1699-709.

J. C. Kuchernig, A. Backenkohler, M. Lubbecke, M. Burow & U. Wittstock 2011b. A thiocyanate-forming protein generates multiple products upon allylglucosinolate breakdown in Thlaspi arvense. *Phytochemistry,* 72, 1699-1709.

J. C. Kuchernig, M. Burow & U. Wittstock 2012. Evolution of specifier proteins in glucosinolate-containing plants. *BMC Evol Biol,* 12, 127.

S Vinod Kumar, Doris Lucyshyn, Katja E Jaeger, Enriqueta Alós, Elizabeth Alvey, Nicholas P Harberd & Philip A Wigge 2012. Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature,* 484**,** 242-245.

M. Kurowska, A. Daszkowska-Golec, D. Gruszka, M. Marzec, M. Szurman, I. Szarejko & M. Maluszynski 2011. TILLING - a shortcut in functional genomics. *Journal of Applied Genetics,* 52**,** 371-390.

P. Lamesch, T. Z. Berardini, D. H. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander, M. Garcia-Hernandez, A. S. Karthikeyan, C. H. Lee, W. D. Nelson, L. Ploetz, S. Singh, A. Wensel & E. Huala 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research,* 40**,** D1202-D1210.

E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. Fitzhugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. Levine, P. Mcewan, K. Mckernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. Mcmurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. Mcpherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C.

Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. Mccombie, M. De La Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. Mclysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. De Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen & Consortium International Human Genome Sequencing 2001a. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Es Lander, Linton Lm, Birren B, Nusbaum C, Zody Mc, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, Mcewan P, Mckernan K, Meldrim J, Mesirov Jp,

Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, Mcmurray A, Matthews L, Mercer S, Milne S, Mullikin Jc, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston Rh, Wilson Rk, Hillier Lw, Mcpherson Jd, Marra Ma, Mardis Er, Fulton La, Chinwalla At, Pepin Kh, Gish Wr, Chissoe Sl, Wendl Mc, Delehaunty Kd, Miner Tl, Delehaunty A, Kramer Jb, Cook Ll, Fulton Rs, Johnson Dl, Minx Pj, Clifton Sw, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T Doggett N, Cheng Jf, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs Ra, Muzny Dm, Scherer Se, Bouck Jb, Sodergren Ej, Worley Kc, Rives Cm, Gorrell Jh, Metzker Ml, Naylor Sl, Kucherlapati Rs, Nelson Dl, Weinstock Gm, Sakaki Y Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith Dr, Doucette-Stamm L, Rubenfield M Weinstock K, Lee Hm, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis Rw, Federspiel Na, Abola Ap, Proctor Mj, Myers Rm, Schmutz J, Dickson M, Grimwood J, Cox Dr, Olson Mv, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans Ga, Athanasiou M, Schultz R, Roe Ba, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, Mccombie Wr, De La Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G Agarwala R, Aravind L, Bailey Ja, Bateman A, Batzoglou S, Birney E, Bork P, Brown Dg, Burge Cb, Cerutti L, Chen Hc, Church D, Clamp M, Copley Rr, Doerks T, Eddy Sr, Eichler Ee, Furey Ts, Galagan J, Gilbert Jg, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson Ls, Jones Ta, Kasif S, Kaspryzk A, Kennedy S, Kent Wj, Kitts P, Koonin Ev, Korf I, Kulp D, Lancet D, Lowe Tm, Mclysaght A, Mikkelsen T, Moran Jv, Mulder N, Pollara Vj, Ponting Cp, Schuler G, Schultz J,

Slater G, Smit Af, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf Yi, Wolfe Kh, Yang Sp, Yeh Rf, Collins F, Guyer Ms, Peterson J, Felsenfeld A, Wetterstrand Ka, Patrinos A, Morgan Mj, De Jong P Catanese Jj, Osoegawa K, Shizuya H, Choi S & Chen Yj 2001b. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

Anton Lang 1957. The effect of gibberellin upon flower formation. *Proceedings of the National Academy of Sciences,* 43**,** 709-717.

Hanna Lee, Seong Jeon Yoo, Jeong Hwan Lee, Wanhui Kim, Seung Kwan Yoo, Heather Fitzgerald, James C Carrington & Ji Hoon Ahn 2010. Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis. *Nucleic Acids Res***,** gkp1240.

Jeong Hwan Lee, Seong Jeon Yoo, Soo Hyun Park, Ildoo Hwang, Jong Seob Lee & Ji Hoon Ahn 2007. Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. *Genes & Development,* 21**,** 397-402.

Sarah J Liljegren, Cindy Gustafson-Brown, Anusak Pinyopich, Gary S Ditta & Martin F Yanofsky 1999. Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *The Plant Cell Online,* 11**,** 1007-1018.

Chang Liu, Hongyan Chen, Hong Ling Er, Hui Meng Soo, Prakash P Kumar, Jin-Hua Han, Yih Cherng Liou & Hao Yu 2008. Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis. *Development,* 135**,** 1481-1491.

Fuquan Liu, Victor Quesada, Pedro Crevillén, Isabel Bäurle, Szymon Swiezewski & Caroline Dean 2007. The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. *Mol Cell,* 28**,** 398-407.

S. Liu, Y. Liu, X. Yang, C. Tong, D. Edwards, I. A. Parkin, M. Zhao, J. Ma, J. Yu, S. Huang, X. Wang, J. Wang, K. Lu, Z. Fang, I. Bancroft, T. J. Yang, Q. Hu, X. Wang, Z. Yue, H. Li, L. Yang, J. Wu, Q. Zhou, W. Wang, G. J. King, J. C. Pires, C. Lu, Z. Wu, P. Sampath, Z. Wang, H. Guo, S. Pan, L. Yang, J. Min, D. Zhang, D. Jin, W. Li, H. Belcram, J. Tu, M. Guan, C. Qi, D. Du, J. Li, L. Jiang, J. Batley, A. G. Sharpe, B. S. Park, P. Ruperao, F. Cheng, N. E. Waminal, Y. Huang, C. Dong, L. Wang, J. Li, Z. Hu, M. Zhuang, Y. Huang, J. Huang, J. Shi, D. Mei, J. Liu, T. H. Lee, J. Wang, H. Jin, Z. Li, X. Li, J. Zhang, L. Xiao, Y. Zhou, Z. Liu, X. Liu, R. Qin, X. Tang, W. Liu, Y. Wang, Y. Zhang, J. Lee, H. H. Kim, F. Denoeud, X. Xu, X. Liang, W. Hua, X. Wang, J. Wang, B. Chalhoub & A. H. Paterson 2014. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun,* 5**,** 3930.

M Alejandra Mandel, Cindy Gustafson-Brown, Beth Savidge & Martin F Yanofsky 1992. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1.

S Marquardt, Pk Boss, J Hadfield & C Dean 2006. Additional targets of the Arabidopsis autonomous pathway members, FCA and FY. *J Exp Bot,* 57**,** 3379-3386.

Klaus Fx Mayer, Jane Rogers, Jaroslav Doležel, Curtis Pozniak, Kellye Eversole, Catherine Feuillet, Bikram Gill, Bernd Friebe, Adam J Lukaszewski & Pierre Sourdille 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science,* 345**,** 1251788.

Gi Mc Intyre & Kf Best 1975. Studies on the flowering of Thlaspi arvense L. II. A comparative study of early-and late-flowering strains. *Botanical Gazette***,** 151-158.

C. M. Mccallum, L. Comai, E. A. Greene & S. Henikoff 2000a. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol,* 123**,** 439-42.

C. M. Mccallum, L. Comai, E. A. Greene & S. Henikoff 2000b. Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiology,* 123**,** 439-442.

G. I. Mcintyre & K. F. Best 1978. Studies on Flowering of Thlaspi-Arvense L .4. Genetic and Ecological Differences between Early-Flowering and Late-Flowering Strains. *Botanical Gazette,* 139**,** 190-195.

Gi Mcintyre & Kf Best 1978 Studies on the Flowering of Thlaspi arvense IV. Genetic and Ecological Differences between Early and Late Flowering Strains *Botanical Gazette,* 139**,** 190-195.

J. D. Metzger & A. T. Hassebrock 1990. Selection and Characterization of a Gibberellin-Deficient Mutant of Thlaspi arvense L. *Plant Physiol,* 94**,** 1655-62.

Jin Miao, Dongshu Guo, Jinzhe Zhang, Qingpei Huang, Genji Qin, Xin Zhang, Jianmin Wan, Hongya Gu & Li-Jia Qu 2013. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res,* 23**,** 1233-1236.

S. D. Michaels & R. M. Amasino 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell,* 11**,** 949-56.

S. D. Michaels & R. M. Amasino 2001a. Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell,* 13**,** 935-941.

S. D. Michaels & R. M. Amasino 2001b. Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell,* 13**,** 935-41.

M. J. Milner & L. V. Kochian 2008. Investigating heavy-metal hyperaccumulation using Thlaspi caerulescens as a model system. *Ann Bot,* 102**,** 3-13.

M. J. Milner, N. Mitani-Ueno, N. Yamaji, K. Yokosho, E. Craft, Z. J. Fei, S. Ebbs, M. C. Zambrano, J. F. Ma & L. V. Kochian 2014. Root and shoot transcriptome analysis of two ecotypes of Noccaea caerulescens uncovers the role of NcNramp1 in Cd hyperaccumulation. *Plant Journal,* 78**,** 398-410.

L. W. Mitich 1996. Field pennycress (Thlaspi arvense L) - The stinkweed. *Weed Technology,* 10**,** 675-678.

Jihyun Moon, Sung-Suk Suh, Horim Lee, Kyu-Ri Choi, Choo Bong Hong, Nam-Chon Paek, Sang-Gu Kim & Ilha Lee 2003. The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. *The Plant Journal,* 35**,** 613-623.

B. R. Moser, G. Knothe, S. F. Vaughn & T. A. Isbell 2009a. Production and Evaluation of Biodiesel from Field Pennycress (Thlaspi arvense L.) Oil. *Energy & Fuels,* 23**,** 4149-4155.

B. R. Moser, S. N. Shah, J. K. Winkler-Moser, S. F. Vaughn & R. L. Evangelista 2009b. Composition and physical properties of cress (Lepidium sativum L.) and field pennycress (Thlaspi arvense L.) oils. *Industrial Crops and Products,* 30**,** 199-205.

B. R. Moser, Shailesh N. Shah, Jill K. Winkler-Moser, Steven F. Vaughn & Roque L. Evangelista 2009c. Composition and physical properties of cress (Lepidium sativum L.) and field pennycress (Thlaspi arvense L.) oils. *Industrial Crops & Products,* 30**,** 7p.

K. Mummenhoff, A. Franzke & M. Koch 1997. Molecular data reveal convergence in fruit characters used in the classification of Thlaspi s. l. (Brassicaceae). *Botanical Journal of the Linnean Society,* 125**,** 183-199.

Joshua S Mylne, Lynne Barrett, Federico Tessadori, Stéphane Mesnage, Lianna Johnson, Yana V Bernatavichute, Steven E Jacobsen, Paul Fransz & Caroline Dean 2006. LHP1, the Arabidopsis homologue of HETEROCHROMATIN PROTEIN1, is

required for epigenetic silencing of FLC. *Proc Natl Acad Sci U S A,* 103**,** 5012-5017.

K. Nakabayashi, M. Bartsch, Y. Xiang, E. Miatton, S. Pellengahr, R. Yano, M. Seo & W. J. J. Soppe 2012. The Time Required for Dormancy Release in Arabidopsis Is Determined by DELAY OF GERMINATION1 Protein Levels in Freshly Harvested Seeds. *Plant Cell,* 24**,** 2826-2838.

H. H. Nour-Eldin, T. G. Andersen, M. Burow, S. R. Madsen, M. E. Jorgensen, C. E. Olsen, I. Dreyer, R. Hedrich, D. Geiger & B. A. Halkier 2012. NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature,* 488**,** 531-4.

N. E. Olszewski, F. B. Martin & F. M. Ausubel 1988. Specialized binary vector for plant transformation: expression of the Arabidopsis thaliana AHAS gene in Nicotiana tabacum. *Nucleic Acids Res,* 16**,** 10765-82.

N. S. Pence, P. B. Larsen, S. D. Ebbs, D. L. D. Letham, M. M. Lasat, D. F. Garvin, D. Eide & L. V. Kochian 2000. The molecular physiology of heavy metal transport in the Zn/Cd hyperaccumulator Thlaspi caerulescens. *Proc Natl Acad Sci U S A,* 97**,** 4956-4960.

B. K. Peterson, J. N. Weber, E. H. Kay, H. S. Fisher & H. E. Hoekstra 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS One,* 7.

W. B. Phippen & M. E. Phippen 2012a. Soybean Seed Yield and Quality as a Response to Field Pennycress Residue. *Crop Science,* 52**,** 2767-2773.

Winthrop B. Phippen & Mary E. Phippen 2012b. Soybean Seed Yield and Quality as a Response to Field Pennycress Residue. *Crop Science,* 52**,** 7p.

Winthrop B. Phippen & Mary E. Phippen 2012c. Soybean Seed Yield and Quality as a Response to Field Pennycress Residue. *Crop Science,* 52**,** 2767.

J. A. Poland & T. W. Rife 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome,* 5**,** 92-102.

Marcel Cg Proveniers & Martijn Van Zanten 2013. High temperature acclimation through PIF4 signaling. *Trends Plant Sci,* 18**,** 59-64.

A. Y. Rahman, A. O. Usharraj, B. B. Misra, G. P. Thottathil, K. Jayasekaran, Y. Feng, S. Hou, S. Y. Ong, F. L. Ng, L. S. Lee, H. S. Tan, M. K. Sakaff, B. S. Teh, B. F. Khoo, S. S. Badai, N. A. Aziz, A. Yuryev, B. Knudsen, A. Dionne-Laporte, N. P. Mchunu, Q. Yu, B. J. Langston, T. A. Freitas, A. G. Young, R. Chen, L. Wang, N. Najimudin, J. A. Saito & M. Alam 2013. Draft genome sequence of the rubber tree Hevea brasiliensis. *BMC Genomics,* 14**,** 75.

Deepak K Ray & Jonathan A Foley 2013. Increasing global crop harvest frequency: recent trends and future directions. *Environmental Research Letters,* 8**,** 044041.

S. Ray & P. Satya 2014. Next generation sequencing technologies for next generation plant breeding. *Front Plant Sci,* 5.

George Redington 1929. THE EFFECT OF THE DURATION OF LIGHT UPON THE GORWTH AND DEVELOPMENT OF THE PLANT. *Biological Reviews,* 4**,** 180-208.

Jose Luis Riechmann, Beth Allyn Krizek & Elliot M Meyerowitz 1996. Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proceedings of the National Academy of Sciences,* 93**,** 4793-4798.

C. Rogers, J. Wen, R. Chen & G. Oldroyd 2009. Deletion-based reverse genetics in Medicago truncatula. *Plant Physiol,* 151**,** 1077-86.

H.L. Koo S. Kaul, J. Jenkins, M. Rizzo, T. Rooney, L.J. Tallon, T. Feldblyum, W. Nierman, M.-I. Benito, M.-I.X. Lin, C.D. Town, J.C. Venter, C.M. Fraser, S.Tabata, Y. Nakamura, T. Kaneko, S. Sato, E. Asamizu, T. Kato, H. Kotani, S. Sasamoto, J.R. Ecker, A. Theologis, N.A. Federspiel, C.J. Palm, B.I. Osborne, P. Shinn, A.B. Conway, V.S. Vysotskaia, K. Dewar, L. Conn, C.A. Lenz, C.J. Kim, N.F. Hansen, S.X. Liu, E. Buehler, H. Altafi, H. Sakano, P. Dunn, B. Lam, P.K. Pham, Q. Chao, M. Nguyen, G. Yu, H. Chen, A. Southwick, J.M. Lee, M. Miranda, M.J. Toriumi, R.W. Davis, R. Wambutt, G. Murphy, A. Düsterhöft, W. Stiekema, T. Pohl, K.-D. Entian, N. Terryn, G. Volckaert, M. Salanoubat, N. Choisne, M. Rieger, W. Ansorge, M. Unseld, B. Fartmann, G. Valle, F. Artiguenave, J. Weissenbach, F. Quetier, R.K. Wilson, M. De La Bastide, M. Sekhon, E. Huang, L. Spiegel, L. Gnoj, K. Pepin, J. Murray, D. Johnson, K. Habermann, N. Dedhia, L. Parnell, R. Preston, L. Hillier, E. Chen, M. Marra, R. Martienssen, W.R. Mccombie, K. Mayer, O. White, M. Bevan, K. Lemcke, T.H. Creasy, C. Bielke, B. Haas, D. Haase, R. Maiti, S. Rudd, J. Peterson, H. Schoof, D. Frishman, B. Morgenstern, P. Zaccaria, M. Ermolaeva, M. Pertea, J. Quackenbush, N. Volfovsky, D. Wu, T.M. Lowe, S.L. Salzberg, H.-W. Mewes, S. Rounsley, D. Bush, S. Subramaniam, I. Levin, S. Norris, R. Schmidt, A. Acarkan, I. Bancroft, F. Quetier, A. Brennicke, J.A. Eisen, T. Bureau, B.-A. Legault, Q.-H. Le, N. Agrawal, Z. Yu, R. Martienssen, G.P. Copenhaver, S. Luo, C.S. Pikaard, D. Preuss, I.T. Paulsen, M. Sussman, A.B. Britt, J.A. Eisen. D.A. Selinger, R. Pandey, D.W. Mount, V.L. Chandler, R.A. Jorgensen, C. Pikaard, G. Juergens, E.M. Meyerowitz, J.R. Ecker, A. Theologis, J. Dangl, J.D.G. Jones, M. Chen, J. Chory, C. Somerville. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408**,** 796-815.

H. S. Saini, P. K. Bassi, J. S. Goudey & M. S. Spencer 1987. Breakage of Seed Dormancy of Field Pennycress (Thlaspi-Arvense) by Growth-Regulators, Nitrate, and Environmental-Factors. *Weed Science,* 35**,** 802-806.

Alon Samach, Hitoshi Onouchi, Scott E Gold, Gary S Ditta, Zsuzsanna Schwarz-Sommer, Martin F Yanofsky & George Coupland 2000. Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science,* 288**,** 1613-1616.

Mariko Sawa, Dmitri A Nusinow, Steve A Kay & Takato Imaizumi 2007. FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. *Science,* 318**,** 261-265.

Jeremy Schmutz, Steven B Cannon, Jessica Schlueter, Jianxin Ma, Therese Mitros, William Nelson, David L Hyten, Qijian Song, Jay J Thelen, Jianlin Cheng, Dong Xu, Uffe Hellsten, Gregory D May, Yeisoo Yu, Tetsuya Sakurai, Taishi Umezawa, Madan K Bhattacharyya, Devinder Sandhu, Babu Valliyodan, Erika Lindquist, Myron Peto, David Grant, Shengqiang Shu, David Goodstein, Kerrie Barry, Montona Futrell-Griggs, Brian Abernathy, Jianchang Du, Zhixi Tian, Liucun Zhu, Navdeep Gill, Trupti Joshi, Marc Libault, Anand Sethuraman, Xue-Cheng Zhang, Kazuo Shinozaki, Henry T Nguyen, Rod A Wing, Perry Cregan, James Specht, Jane Grimwood, Dan Rokhsar, Gary Stacey, Randy C Shoemaker & Scott A. Jackson 2010. Genome sequence of the palaeopolyploid soybean. *Nature,* 463**,** 178-183.

Ps Schnable, D Ware, Rs Fulton, Jc Stein, F Wei, S Pasternak, C Liang, J Zhang, L Fulton, Ta Graves, P Minx, Ad Reily, L Courtney, Ss Kruchowski, C Tomlinson, C Strong, K Delehaunty, C Fronick, B Courtney, Sm Rock, E Belter, F Du, K Kim, Rm Abbott, M Cotton, A Levy, P Marchetto, K Ochoa, Sm Jackson, B Gillam, W Chen, L Yan, J Higginbotham, M Cardenas, J Waligorski, E Applebaum, L Phelps, J Falcone, K Kanchi, T Thane, A Scimone, N Thane, J Henke, T Wang, J Ruppert, N Shah, K Rotter, J Hodges, E Ingenthron, M Cordes, S Kohlberg, J Sgro, B Delgado, K Mead, A Chinwalla, S Leonard, K Crouse, K Collura, D Kudrna, J Currie, R He, A Angelova, S Rajasekar, T Mueller, R Lomeli, G Scara, A Ko, K Delaney, M Wissotski, G Lopez, D Campos, M

Braidotti, E Ashley, W Golser, H Kim, S Lee, J Lin, Z Dujmic, W Kim, J Talag, A Zuccolo, C Fan, A Sebastian, M Kramer, L Spiegel, L Nascimento, T Zutavern, B Miller, C Ambroise, S Muller, W Spooner, A Narechania, L Ren, S Wei, S Kumari, B Faga, Mj Levy, L Mcmahan, P Van Buren, Mw Vaughn, K Ying, Ct Yeh, Sj Emrich, Y Jia, A Kalyanaraman, Ap Hsia, Wb Barbazuk, Rs Baucom, Tp Brutnell, Nc Carpita, C Chaparro, Jm Chia, Jm Deragon, Jc Estill, Y Fu, Ja Jeddeloh, Y Han, H Lee, P Li, Dr Lisch, S Liu, Z Liu, Dh Nagel, Mc Mccann, P Sanmiguel, Am Myers, D Nettleton, J Nguyen, Bw Penning, L Ponnala, Kl Schneider, Dc Schwartz, A Sharma, C Soderlund, Nm Springer, H Sun Q Wang, M Waterman, R Westerman, Tk Wolfgruber, L Yang, Y Yu, L Zhang, S Zhou, Q Zhu, Jl Bennetzen, Rk Dawe, J Jiang, N Jiang, Gg Presting, Sr Wessler, S Aluru, Ra Martienssen, Sw Clifton, Wr Mccombie, Ra Wing & Rk Wilson 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science,* 326**,** 1112-1115.

M. E. Schranz, P. Quijada, S. B. Sung, L. Lukens, R. Amasino & T. C. Osborn 2002. Characterization and effects of the replicated flowering time gene FLC in Brassica rapa. *Genetics,* 162**,** 1457-1468.

Elizabeth A Schultz & George W Haughn 1991. LEAFY, a homeotic gene that regulates inflorescence development in Arabidopsis. *The Plant Cell Online,* 3**,** 771-781.

Iain Searle, Yuehui He, Franziska Turck, Coral Vincent, Fabio Fornara, Sandra Kröber, Richard A Amasino & George Coupland 2006. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes & Development,* 20**,** 898-912.

J. C. Sedbrook, W. B. Phippen & M. D. Marks 2014a. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (Thlaspi arvense L.). *Plant Sci,* 227**,** 122-32.

J. C. Sedbrook, W. B. Phippen & M. D. Marks 2014b. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: Example pennycress (Thlaspi arvense L.). *Plant Sci,* 227C**,** 122-132.

Jc Sedbrook, Phippen, Wb, Marks, Md 2014. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (Thlaspi arvense L.). *Plant Sci,* 227**,** 122-132.

G. Selling, R. Evangelista, T. Isbell & M. H. Evangelista 2011. Extraction of proteins from pennycress. *Abstracts of Papers of the American Chemical Society,* 241.

G. W. Selling, M. P. Hojilla-Evangelista, R. L. Evangelista, T. Isbell, N. Price & K. M. Doll 2013. Extraction of proteins from pennycress seeds and press cake. *Industrial Crops and Products,* 41**,** 113-119.

N. Sharma, D. Cram, T. Huebert, N. Zhou & I. A. Parkin 2007a. Exploiting the wild crucifer Thlaspi arvense to identify conserved and novel genes expressed during a plant's response to cold stress. *Plant Mol Biol,* 63**,** 171-84.

N. Sharma, D. Cram, T. Huebert, N. Zhou & I. A. P. Parkin 2007b. Exploiting the wild crucifer Thlaspi arvense to identify conserved and novel genes expressed during a plant's response to cold stress. *Plant Mol Biol,* 63**,** 171-184.

C. C. Sheldon, D. T. Rouse, E. J. Finnegan, W. J. Peacock & E. S. Dennis 2000. The molecular basis of vernalization: The central role of FLOWERING LOCUS C (FLC). *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 3753-3758.

C. Shindo, M. J. Aranzana, C. Lister, C. Baxter, C. Nicholls, M. Nordborg & C. Dean 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant Physiol,* 138**,** 1163-73.

G. G. Simpson & C. Dean 2002. Arabidopsis, the Rosetta stone of flowering time? *Science,* 296**,** 285-9.

Gordon G Simpson, Paul P Dijkwel, Victor Quesada, Ian Henderson & Caroline Dean 2003. FY is an RNA 3′ end-processing factor that interacts with FCA to control the Arabidopsis floral transition. *Cell,* 113**,** 777-787.

Jw Singer, Sm Nusser & Cj Alf 2007. Are cover crops being used in the US corn belt? *Journal of Soil and Water Conservation,* 62**,** 353-358.

T. Slotte, K. M. Hazzouri, J. A. Agren, D. Koenig, F. Maumus, Y. L. Guo, K. Steige, A. E. Platts, J. S. Escobar, L. K. Newman, W. Wang, T. Mandakova, E. Vello, L. M. Smith, S. R. Henz, J. Steffen, S. Takuno, Y. Brandvain, G. Coop, P. Andolfatto, T. T. Hu, M. Blanchette, R. M. Clark, H. Quesneville, M. Nordborg, B. S. Gaut, M. A. Lysak, J. Jenkins, J. Grimwood, J. Chapman, S. Prochnik, S. Shu, D. Rokhsar, J. Schmutz, D. Weigel & S. I. Wright 2013. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nat Genet,* 45**,** 831-5.

T. Slotte, Hazzouri, K. M., Ågren, J. A., Koenig, D., Maumus, F., Guo, Y.-L., Steige, K., Platts, A. E., Escobar, J. S., Newman, L., Wang, W., Mandáková, T., Vello, E., Smith, L. M., Henz, S. R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., Andolfatto, P., Hu, T. T., Blanchette, M., Clark, R. M., Quesneville, H., Nordborg, M., Gaut, B. S., Lysak, M. A., Jenkins, J., Grimwood, J., Chapman, J., Prochnick, S., Shu, S., Rokhsar, D., Schmutz, J., Weigel, D., and Wright, S. I. 2013. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nature Genetics,* accepted for publication.

Afa Smit, Hubley, R & Green, P. 1996-2010. RepeatMasker Open-3.0

S. S. Snapp, S. M. Swinton, R. Labarta, D. Mutch, J. R. Black, R. Leep, J. Nyiraneza & K. O'neil 2005a. Evaluating cover crops for benefits, costs and performance within cropping system niches. *Agronomy Journal,* 97**,** 322-332.

Ss Snapp, Sm Swinton, R Labarta, D Mutch, Jr Black, R Leep, J Nyiraneza & K O'neil 2005b. Evaluating cover crops for benefits, costs and performance within cropping system niches. *Agronomy Journal,* 97**,** 322-332.

Y. H. Song, I. Lee, S. Y. Lee, T. Imaizumi & J. C. Hong 2012a. CONSTANS and ASYMMETRIC LEAVES 1 complex is involved in the induction of FLOWERING LOCUS T in photoperiodic flowering in Arabidopsis. *Plant J,* 69**,** 332-42.

Young Hun Song, Shogo Ito & Takato Imaizumi 2013. Flowering time regulation: photoperiod-and temperature-sensing in leaves. *Trends Plant Sci,* 18**,** 575-583.

Young Hun Song, Robert W Smith, Benjamin J To, Andrew J Millar & Takato Imaizumi 2012b. FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. *Science,* 336**,** 1045-1049.

J. R. Stinchcombe, C. Weinig, M. Ungerer, K. M. Olsen, C. Mays, S. S. Halldorsdottir, M. D. Purugganan & J. Schmitt 2004. A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *Proc Natl Acad Sci U S A,* 101**,** 4712-7.

G. Sun 2011. MicroRNAs and their diverse functions in plants. *Plant Mol Biol*.

Sibum Sung, Yuehui He, Tifani W Eshoo, Yosuke Tamada, Lianna Johnson, Kenji Nakahigashi, Koji Goto, Steve E Jacobsen & Richard M Amasino 2006. Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1. *Nat Genet,* 38**,** 706-710.

M. Tadege, C. C. Sheldon, C. A. Helliwell, P. Stoutjesdijk, E. S. Dennis & W. J. Peacock 2001a. Control of flowering time by FLC orthologues in Brassica napus. *Plant Journal,* 28**,** 545-553.

Million Tadege, Candice C Sheldon, Chris A Helliwell, Peter Stoutjesdijk, Elizabeth S Dennis & W James Peacock 2001b. Control of flowering time by FLC orthologues in Brassica napus. *The Plant Journal,* 28**,** 545-553.

Shojiro Tamaki, Shoichi Matsuo, Hann Ling Wong, Shuji Yokoi & Ko Shimamoto 2007. Hd3a protein is a mobile flowering signal in rice. *Science,* 316**,** 1033-1036.

R Development Core Team 2008. *R: A language and environment for statistical computing,* Vienna, Austria R Foundation for Statistical Computing.

D. Tilman, K. G. Cassman, P. A. Matson, R. Naylor & S. Polasky 2002. Agricultural sustainability and intensive production practices. *Nature,* 418**,** 671-7.

D. Tilman, R. Socolow, J. A. Foley, J. Hill, E. Larson, L. Lynd, S. Pacala, J. Reilly, T. Searchinger, C. Somerville & R. Williams 2009. Beneficial Biofuels-The Food, Energy, and Environment Trilemma. *Science,* 325**,** 270-271.

David Tilman, Christian Balzer, Jason Hill & Belinda L Befort 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences,* 108**,** 20260-20264.

Shiv B Tiwari, Yu Shen, Han-Chang Chang, Yanli Hou, Amanda Harris, Siu Fong Ma, Megan Mcpartland, Graham J Hymus, Luc Adam & Colleen Marion 2010. The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. *New Phytologist,* 187**,** 57-66.

Helen Tsai, Tyson Howell, Rebecca Nitcher, Victor Missirian, Brian Watson, Kathie J Ngo, Meric Lieberman, Joseph Fass, Cristobal Uauy & Robert K Tran 2011.

Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiol,* 156**,** 1257-1268.

Franziska Turck, François Roudier, Sara Farrona, Marie-Laure Martin-Magniette, Elodie Guillaume, Nicolas Buisine, Séverine Gagnot, Robert A Martienssen, George Coupland & Vincent Colot 2007. Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet,* 3**,** e86.

Usda. *Global Crop Production Analysis* [Online]. United States Department of Agriculture Foreign Agricultural Service. Available: http://www.pecad.fas.usda.gov/ [Accessed 4/30/15 2015].

R. K. Varshney, S. N. Nayak, G. D. May & S. A. Jackson 2009a. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol,* 27**,** 522-30.

R. K. Varshney, C. Song, R. K. Saxena, S. Azam, S. Yu, A. G. Sharpe, S. Cannon, J. Baek, B. D. Rosen, B. Tar'an, T. Millan, X. Zhang, L. D. Ramsay, A. Iwata, Y. Wang, W. Nelson, A. D. Farmer, P. M. Gaur, C. Soderlund, R. V. Penmetsa, C. Xu, A. K. Bharti, W. He, P. Winter, S. Zhao, J. K. Hane, N. Carrasquilla-Garcia, J. A. Condie, H. D. Upadhyaya, M. C. Luo, M. Thudi, C. L. Gowda, N. P. Singh, J. Lichtenzveig, K. K. Gali, J. Rubio, N. Nadarajan, J. Dolezel, K. C. Bansal, X. Xu, D. Edwards, G. Zhang, G. Kahl, J. Gil, K. B. Singh, S. K. Datta, S. A. Jackson, J. Wang & D. R. Cook 2013. Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nat Biotechnol,* 31**,** 240-6.

R. K. Varshney, R. Terauchi & S. R. Mccouch 2014. Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLoS Biol,* 12.

Rajeev K Varshney, Wenbin Chen, Yupeng Li, Arvind K Bharti, Rachit K Saxena, Jessica A Schlueter, Mark Ta Donoghue, Sarwar Azam, Guangyi Fan & Adam M

Whaley 2012. Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. *Nat Biotechnol,* 30**,** 83-89.

Rajeev K. Varshney, Spurthi N. Nayak, Gregory D. May & Scott A. Jackson 2009b. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology,* 27**,** 522-530.

S. F. Vaughn, T. A. Isbell, D. Weisleder & M. A. Berhow 2005a. Biofumigant compounds released by field pennycress (Thlaspi arvense) seedmeal. *J Chem Ecol,* 31**,** 167-77.

S. F. Vaughn, T. A. Isbell, D. Weisleder & M. A. Berhow 2005b. Biofumigant compounds released by field pennycress (Thlaspi arvense) seedmeal. *Journal of Chemical Ecology,* 31**,** 167-177.

S. Vaughn, T. Isbell, D. Weisleder & M. Berhow 2004. Biofumigation Potential of Field Pennycress (Thlaspi arvense) Seedmeal. *Hortscience,* 39**,** 745-745.

Steven F Vaughn, Debra E Palmquist, Sandra M Duval & Mark A Berhow 2009. Herbicidal activity of glucosinolate-containing seedmeals.

J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. Mckusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J.

Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. Mccawley, T. Mcintosh, I. Mcmullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. Mcdaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A.

Zandieh & X. Zhu 2001. The sequence of the human genome. *Science,* 291**,** 1304-51.

Johann Vollmann & Istvan Rajcan 2009. *Oil Crops*, Springer Science & Business Media.

A. Voskoboynik, N. F. Neff, D. Sahoo, A. M. Newman, D. Pushkarev, W. Koh, B. Passarelli, H. C. Fan, G. L. Mantalas, K. J. Palmeri, K. J. Ishizuka, C. Gissi, F. Griggio, R. Ben-Shlomo, D. M. Corey, L. Penland, R. A. White, I. L. Weissman & S. R. Quake 2013. The genome sequence of the colonial chordate, Botryllus schlosseri. *Elife,* 2.

Doris Wagner, Robert Wm Sablowski & Elliot M Meyerowitz 1999. Transcriptional activation of APETALA1 by LEAFY. *Science,* 285**,** 582-584.

K. Wang, Z. Wang, F. Li, W. Ye, J. Wang, G. Song, Z. Yue, L. Cong, H. Shang, S. Zhu, C. Zou, Q. Li, Y. Yuan, C. Lu, H. Wei, C. Gou, Z. Zheng, Y. Yin, X. Zhang, K. Liu, B. Wang, C. Song, N. Shi, R. J. Kohel, R. G. Percy, J. Z. Yu, Y. X. Zhu, J. Wang & S. Yu 2012. The draft genome of a diploid cotton Gossypium raimondii. *Nat Genet,* 44**,** 1098-103.

L. Wang, S. Yu, C. Tong, Y. Zhao, Y. Liu, C. Song, Y. Zhang, X. Zhang, Y. Wang, W. Hua, D. Li, D. Li, F. Li, J. Yu, C. Xu, X. Han, S. Huang, S. Tai, J. Wang, X. Xu, Y. Li, S. Liu, R. K. Varshney, J. Wang & X. Zhang 2014. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol,* 15**,** R39.

N. A. Wang, W. Qian, I. Suppanz, L. J. Wei, B. Z. Mao, Y. Long, J. L. Meng, A. E. Muller & C. Jung 2011a. Flowering time variation in oilseed rape (Brassica napus L.) is associated with allelic variation in the FRIGIDA homologue BnaA.FRI.a. *Journal of Experimental Botany,* 62**,** 5641-5658.

X. W. Wang, H. Z. Wang, J. Wang, R. F. Sun, J. Wu, S. Y. Liu, Y. Q. Bai, J. H. Mun, I. Bancroft, F. Cheng, S. W. Huang, X. X. Li, W. Hua, J. Y. Wang, X. Y. Wang, M.

Freeling, J. C. Pires, A. H. Paterson, B. Chalhoub, B. Wang, A. Hayward, A. G. Sharpe, B. S. Park, B. Weisshaar, B. H. Liu, B. Li, B. Liu, C. B. Tong, C. Song, C. Duran, C. F. Peng, C. Y. Geng, C. S. Koh, C. Y. Lin, D. Edwards, D. S. Mu, D. Shen, E. Soumpourou, F. Li, F. Fraser, G. Conant, G. Lassalle, G. J. King, G. Bonnema, H. B. Tang, H. P. Wang, H. Belcram, H. L. Zhou, H. Hirakawa, H. Abe, H. Guo, H. Wang, H. Z. Jin, I. A. P. Parkin, J. Batley, J. S. Kim, J. Just, J. W. Li, J. H. Xu, J. Deng, J. A. Kim, J. P. Li, J. Y. Yu, J. L. Meng, J. P. Wang, J. M. Min, J. Poulain, J. Wang, K. Hatakeyama, K. Wu, L. Wang, L. Fang, M. Trick, M. G. Links, M. X. Zhao, M. N. Jin, N. Ramchiary, N. Drou, P. J. Berkman, Q. L. Cai, Q. F. Huang, R. Q. Li, S. Tabata, S. F. Cheng, S. Zhang, S. J. Zhang, S. M. Huang, S. Sato, S. L. Sun, S. J. Kwon, S. R. Choi, T. H. Lee, W. Fan, X. Zhao, X. Tan, X. Xu, Y. Wang, Y. Qiu, Y. Yin, Y. R. Li, Y. C. Du, Y. C. Liao, Y. Lim, Y. Narusaka, Y. P. Wang, Z. Y. Wang, Z. Y. Li, Z. W. Wang, Z. Y. Xiong & Z. H. Zhang 2011b. The genome of the mesopolyploid crop species Brassica rapa. *Nature Genetics,* 43**,** 1035-U157.

X. Wang, H. Wang, J. Wang, R. Sun, J. Wu, S. Liu, Y. Bai, J. H. Mun, I. Bancroft, F. Cheng, S. Huang, X. Li, W. Hua, J. Wang, X. Wang, M. Freeling, J. C. Pires, A. H. Paterson, B. Chalhoub, B. Wang, A. Hayward, A. G. Sharpe, B. S. Park, B. Weisshaar, B. Liu, B. Li, B. Liu, C. Tong, C. Song, C. Duran, C. Peng, C. Geng, C. Koh, C. Lin, D. Edwards, D. Mu, D. Shen, E. Soumpourou, F. Li, F. Fraser, G. Conant, G. Lassalle, G. J. King, G. Bonnema, H. Tang, H. Wang, H. Belcram, H. Zhou, H. Hirakawa, H. Abe, H. Guo, H. Wang, H. Jin, I. A. Parkin, J. Batley, J. S. Kim, J. Just, J. Li, J. Xu, J. Deng, J. A. Kim, J. Li, J. Yu, J. Meng, J. Wang, J. Min, J. Poulain, J. Wang, K. Hatakeyama, K. Wu, L. Wang, L. Fang, M. Trick, M. G. Links, M. Zhao, M. Jin, N. Ramchiary, N. Drou, P. J. Berkman, Q. Cai, Q. Huang, R. Li, S. Tabata, S. Cheng, S. Zhang, S. Zhang, S. Huang, S. Sato, S. Sun, S. J. Kwon, S. R. Choi, T. H. Lee, W. Fan, X. Zhao, X. Tan, X. Xu, Y. Wang, Y. Qiu, Y. Yin, Y. Li, Y. Du, Y. Liao, Y. Lim, Y. Narusaka, Y. Wang, Z. Wang, Z. Li, Z. Wang, Z. Xiong, Z. Zhang & Consortium Brassica Rapa Genome

Sequencing Project 2011c. The genome of the mesopolyploid crop species Brassica rapa. *Nat Genet,* 43**,** 1035-9.

S. I. Warwick, A. Francis & D. J. Susko 2002. The biology of Canadian weeds. 9. Thlaspi arvense L. (updated). *Canadian Journal of Plant Science,* 82**,** 803-823.

S.I. Warwick, Francis, A., Susko, D.J. 2002 The biology of Canadian weeds. 9. Thlaspi arvense L. (updated) *Canadian Journal of Plant Science,* 82**,** 803-823.

Detlef Weigel, John Alvarez, David R Smyth, Martin F Yanofsky & Elliot M Meyerowitz 1992. LEAFY controls floral meristem identity in Arabidopsis. *Cell,* 69**,** 843-859.

O. Westesson, M. Skinner & I. Holmes 2013. Visualizing next-generation sequencing data with JBrowse. *Brief Bioinform,* 14**,** 172-7.

Ruth N Wilson, John W Heckman & Chris R Somerville 1992. Gibberellin is required for flowering in Arabidopsis thaliana under short days. *Plant Physiol,* 100**,** 403-408.

N. Winchester, Mcconnachie, D., Wollersheim, C, Waitz, I. 2013a. Market cost of renewable jet fuel adopiton in the United States Massachusetts Institute of Technology

N. Winchester, Mcconnachie, D., Wollersheim, C. And Waitz, I. 2013b. Market cost of renewable jet fuel adoption in the United States. *MIT Joint Program on the Science and Policy of Global Change*.

U. Wittstock & M. Burow 2010. Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. *Arabidopsis Book,* 8**,** e0134.

G. Wu, M. Y. Park, S. R. Conway, J. W. Wang, D. Weigel & R. S. Poethig 2009. The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell,* 138**,** 750-9.

H. J. Wu, Z. Zhang, J. Y. Wang, D. H. Oh, M. Dassanayake, B. Liu, Q. Huang, H. X. Sun, R. Xia, Y. Wu, Y. N. Wang, Z. Yang, Y. Liu, W. Zhang, H. Zhang, J. Chu, C. Yan, S. Fang, J. Zhang, Y. Wang, F. Zhang, G. Wang, S. Y. Lee, J. M. Cheeseman, B. Yang, B. Li, J. Min, L. Yang, J. Wang, C. Chu, S. Y. Chen, H. J. Bohnert, J. K. Zhu, X. J. Wang & Q. Xie 2012a. Insights into salt tolerance from the genome of Thellungiella salsuginea. *Proc Natl Acad Sci U S A,* 109**,** 12219-24.

Jian Wu, Keyun Wei, Feng Cheng, Shikai Li, Qian Wang, Jianjun Zhao, Guusje Bonnema & Xiaowu Wang 2012b. A naturally occurring InDel variation in BraA. FLC. b (BrFLC2) associated with flowering time variation in Brassica rapa. *BMC Plant Biol,* 12**,** 151.

L. Yang, S. R. Conway & R. S. Poethig 2011. Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. *Development,* 138**,** 245-9.

R. Yang, D. E. Jarvis, H. Chen, M. A. Beilstein, J. Grimwood, J. Jenkins, S. Shu, S. Prochnik, M. Xin, C. Ma, J. Schmutz, R. A. Wing, T. Mitchell-Olds, K. S. Schumaker & X. Wang 2013. The Reference Genome of the Halophytic Plant Eutrema salsugineum. *Front Plant Sci,* 4**,** 46.

Sha Yu, Vinicius C Galvão, Yan-Chun Zhang, Daniel Horrer, Tian-Qi Zhang, Yan-Hong Hao, Yu-Qi Feng, Shui Wang, Markus Schmid & Jia-Wei Wang 2012. Gibberellin regulates the Arabidopsis floral transition through miR156-targeted SQUAMOSA PROMOTER BINDING–LIKE transcription factors. *The Plant Cell Online,* 24**,** 3320-3332.

Y. X. Yuan, J. Wu, R. F. Sun, X. W. Zhang, D. H. Xu, G. Bonnema & X. W. Wang 2009. A naturally occurring splicing site mutation in the Brassica rapa FLC1 gene is associated with variation in flowering time. *Journal of Experimental Botany,* 60**,** 1299-1308.

Jan Ad Zeevaart 1976. Physiology of flower formation. *Annual Review of Plant Physiology,* 27**,** 321-348.

Xiaoyu Zhang, Sophie Germann, Bartlomiej J Blus, Sepideh Khorasanizadeh, Valerie Gaudin & Steven E Jacobsen 2007. The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat Struct Mol Biol,* 14**,** 869-871.

Z. Zhang, J. A. Ober & D. J. Kliebenstein 2006. The gene controlling the quantitative trait locus EPITHIOSPECIFIER MODIFIER1 alters glucosinolate hydrolysis and insect resistance in Arabidopsis. *Plant Cell,* 18**,** 1524-36.

Q. Y. Zhao, Y. Wang, Y. M. Kong, D. Luo, X. Li & P. Hao 2011. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *Bmc Bioinformatics,* 12.

N. Zhou, S. J. Robinson, T. Huebert, N. J. Bate & I. A. Parkin 2007. Comparative genome organization reveals a single copy of CBF in the freezing tolerant crucifer Thlaspi arvense. *Plant Mol Biol,* 65**,** 693-705.