DISSERTATION

COMBINING QUANTITATIVE GENETICS AND GENOMICS TO IDENTIFY POLYMORPHISMS ASSOCIATED WITH DROUGHT PHYSIOLOGY IN *ARABIDOPSIS* AND *BRASSICA NAPUS*

Submitted by

Richard Fletcher

Department of Bioagricultural Sciences and Pest Management

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2014

Doctoral Committee:

Advisor: John McKay

William Bauerle
Patrick Byrne
Jan Leach

ABSTRACT

COMBINING QUANTITATIVE GENETICS AND GENOMICS TO IDENTIFY POLYMORPHISMS ASSOCIATED WITH DROUGHT PHYSIOLOGY IN *ARABIDOPSIS* AND *BRASSICA NAPUS*

Isolating the functional genetic variants controlling important traits is a central goal in modern plant genetics. A great deal of progress has been made using mutant screens in many species to understand gene function. The current challenge is to translate this basic knowledge to traits important for agriculture. Yet, examples of success in this pursuit are rare due to complex patterns of inheritance for important traits such as yield. This is because the expression of the trait as a phenotype is under the influence of genetic and environmental factors as well as the interaction between gene and environment. With respect to environment, water availability remains the most influential in regards to crop productivity. The following research was performed using a combination of quantitative genetic and genomic approaches to gain a better understanding of each factor in relation to drought physiology in two members of the *Brassicaceae* family, *Arabidopsis thaliana* and *Brassica napus.*

Arabidopsis is the model dicotyledonous species and stands alone as the most well characterized of all angiosperms. An emerging question in agriculture is the degree to which breeding will be informed by results from the model plant *Arabidopsis*. *Brassica* crops are ideally suited for applying the extensive genetic and genomic information gained in *Arabidopsis* due to a recent common ancestor (~20 million years). DNA sequence comparisons between *A. thaliana* and several *Brassica* species has revealed a strong conservation of genic space (i.e. synteny) suggesting that the *Brassicas* are well-positioned for validating the potential of comparative genomics for crop improvement. *Brassica napus* (AACC) is an important crop species with an allopolyploid genome formed from the hybridization of the diploid species *B. rapa* (AA) and *B. oleracea* (CC). Consequently, genetics and genomics research are complicated by

the high degree of genomic redundancy of a polyploid genome. However, the documented synteny between *Arabidopsis* and *B. napus* provide the opportunity for genetic knowledge gained in *Arabidopsis* to be translated into hypotheses for testing in the *Brassicas*. Likewise, results from research in *B. napus* can be validated in the much simpler genetic system afforded by *Arabidopsis*. This system of reciprocal knowledge transfer between model and crop offers a powerful approach for understanding the genetics of complex traits such as drought physiology.

Near-isogenic lines (NILs) are inbred plant lines carrying one, or a few, small genomic introgressions from a "donor" line in a different and otherwise homogeneous genomic background (termed the recurrent parent). By standardizing genetic factors outside of the genomic region of interest, the true impact of the introgressed locus on the phenotype of interest can be estimated relative to the line lacking the chromosomal introgression. The creation of a near-isogenic line generally employs a backcross breeding strategy in combination with molecular markers to track the chromosomal introgression and recovery of the recurrent parent genome. Currently, three *Arabidopsis* NIL populations have been created where the number of molecular markers used to estimate recurrent parent recovery varies from 31 markers to 321. The true number of markers necessary to accurately estimate recurrent parent recovery is still unknown which has important consequences to research on NILs since unrealized chromosomal introgressions can impact the phenotype, creating incorrect estimates of the effect of the chromosomal introgression. To answer this question and create a powerful resource for researching the genetics of drought physiology, a new NIL population derived from two *Arabidopsis* lines which differ in many aspects of drought physiology was created. The study utilized a new genotyping-by-sequencing (GBS) method in which the recent advancements in DNA sequencing technology (i.e. "next-gen sequencing") are exploited to increase the number of genotyped loci. The result was a 3-fold increase in marker density compared to previously described NIL populations. Comparisons of chromosomal introgression estimates made using a "coarse" map of 81 markers relative to a "dense" map of 930 GBS markers found that the dense map identified 128 introgressions missed in the coarse map. A novel aspect of the NIL population is the maintenance of heterozygous NILs along with the traditional homozygous

NILs. The power of this new population is demonstrated as a locus underlying a drought physiology trait (night-time stomatal conductance) is validated using homozygous NILs. At the same time, a heterozygous NIL is used in combination with molecular markers to identify recombinants within the target interval to begin "fine-mapping" the locus with the end goal being the identification of the causal genetic variant.

Drought escape and dehydration avoidance represent the most prevalent drought coping strategies among annual crops. However, a tradeoff between them seems to exist as negative correlations in their underlying mechanisms has been reported in several studies. In order to understand if such a tradeoff exists in *B. napus*, and what the genetic basis of such a tradeoff might be, a quantitative trait locus (QTL) analysis was conducted for root mass and flowering time. Root mass represents a mechanism of dehydration avoidance (minimization of water loss and/or maximization of water uptake) and flowering time represents drought escape (completion of the life cycle prior to drought). Maximum segregating variation was achieved as the study population was derived from a cross between an annual (spring) and biennial (winter) variety, life-history traits which differentiate genetic and morphological classes. The experiment was conducted in the field and included irrigated and rainfed treatments so that trait plasticity could be studied. Evidence for a tradeoff was discovered in the form of strong genetic correlations among traits. In addition, several QTL co-localized at two regions, providing evidence that the tradeoff is genetically constrained. The mechanistic relationship between root mass, flowering time and QTL was specifically tested in conditional models which incorporated correlated traits as covariates. The results of these analyses are suggestive of a mechanistic model where QTL affect root mass directly, in addition to their impacts on the physiologically "upstream" trait, flowering time. Candidate genes underlying the identified QTL, including the well-characterized flowering time gene *FLC*, are discussed.

Next-gen sequencing technology is revolutionizing our ability to analyze genomes. These technologies were used to sequence the annual and biennial parent lines of the mapping population in order to characterize molecular variation in the parental genomes at the scale of the genome, QTL and gene. Sequence data was used to construct a reference based assembly where recently released *Brassica* draft genomes were used as references. On average, the overall single nucleotide polymorphism (SNP)

density was higher in the A genome whereas the rate of non-synonymous substitutions was elevated in the C genome. Non-synonymous substitutions were used to select gene ontology (GO) terms which distinguish the parent genomes and to select a list of candidate genes contained within each QTL. Marker assays developed for several of the discovered SNPs improved the density and overall quality of the genetic map. QTL analysis with the new map resulted in an improvement in model fit, suggesting the new markers were more tightly linked to the causal variants. Ultimately, this research provides a long list of molecular variants differentiating the parent lines that can be used in hypothesis development for research focused on the genetics of drought physiology in *B. napus*.

ACKNOWLEDGEMENTS

This research would not have been possible without contributions from a large group of people and institutions. Therefore, I would like to thank:

My advisor, Dr. John McKay, for setting high expectations, trusting me to be independent and expanding my view of plant biology. His guidance and friendship have changed my thought process as much as the technical knowledge he has transferred;

My committee members, Dr. Jan Leach, Dr. Pat Byrne and Dr. Bill Bauerle, for many insightful suggestions and thought-provoking questions. Each has a contagious enthusiasm for scientific research that I only hope becomes as strong a characteristic in me;

Dr. Lorin DeBonte and Dr. Matthew Ries, for encouraging me to pursue a Ph.D. This journey has changed my life and they were instrumental in my decision;

Dr. Zhizheng Chen, for showing me that when you love what you do, you never work a day in your life. His passion for scientific research is inspiring;

Dr. Judy Schnurr, for allowing flexibility in my schedule and continuous encouragement as this research evolved. I am extremely thankful for her trust and many levels of support during this project;

Dr. Benyuan Dang, for many helpful discussions and being a constant source of optimism;

Dr. Jack Mullen, for countless stimulating discussions, suggestions and assistance with data analysis.

Molly Lauterbach, Jason Rivest, Vanessa Dominguez and Kelly Sheff, for all of their assistance and hard work with many aspects of the genotyping conducted in this research. They are experts at their craft and I cannot express my appreciation enough;

Dave Herrmann, for all of his help with the analysis of the next-generation sequence data;

Daren Coonrod, for his encouragement and friendship during my graduate studies;

Dr. Honggang Zheng, for all of his assistance with my quantitative genetics analyses;

The entire research team at Cargill Specialty Seeds & Oils, for numerous helpful discussions, many direct contributions to this research and patience with my highly variable schedule;

Annie Heiliger, for her assistance collecting data and fighting birds in the field. All of those hard hours have paid off with this excellent data set;

Present and former members of the McKay Lab, for many thought-provoking discussions, suggestions, questions and, of course, their help with many aspects of this research including planting, harvesting, data entry and carbon isotope preparations;

All of my friends, to whom I have almost certainly canceled an engagement to work on this project;

The Apps, for letting me take over your basement during Christmas to study for my prelims, write papers, etc. They are an amazing family and their support during this research is truly valued;

Jeff, Lisa, Effie and Hank Fletcher, for taking my mind off all things work and school related during our visits. I always left with 'batteries re-charged';

My parents, for endless support and always letting me choose my own path;

And most of all, my wife Laura, for being there. We met when I started this journey and she has been a constant source of encouragement and positive energy. I couldn't have done it without her.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

Arabidopsis thaliana and Brassica napus: the relationship between model and crop

The *Brassica*ceae (Cruciferae), or mustard family, is a large and diverse angiosperm family comprised of 337 genera and 3,709 species which includes weeds (e.g. *Capsella*), ornamentals (e.g. *Hesperis*), crops (e.g. *Brassica* spp.) and the model organism *Arabidopsis thaliana* (Al-Shehbaz, 2006). No other plant species has been more studied or well characterized than *Arabidopsis* (Koornneef and Meinke, 2010) and only two oilseed crops are of higher global importance than *Brassica napus* (http://www.fas.usda.gov/oilseeds/Current/; 2013). *B.napus* is an allopolyploid formed from the hybridization of the diploids *B. rapa* (A genome) and *B. oleracea* (C genome) (U, 1935; Parkin *et al.*, 1995). Despite the complexity of the *B. napus* genome and nearly 20 million years since its divergence from *Arabidopsis*, extensive sequence collinearity between them was discovered over a decade ago (Cavell, 1998) suggesting that the transfer of genomics data across species would be a powerful reality. A landmark in our understanding of this relationship was the construction of a segmental comparative genetic map of *B. napus* which clearly demonstrated syntenic regions of the *Arabidopsis* genome replicated throughout the *B. napus* genome (Parkin *et al.*, 2005). This knowledge has been applied in, among other things, furthering our understanding of genome evolution in the *Brassicas* (Tang *et al.* 2012), annotation of draft *Brassica* genomes (Wang *et al.*, 2011; Cheng *et al.*, 2011; Zhao *et al.*, 2013; http://www.ocri-genomics.org; 2014) and selecting candidate genes within candidate genomic intervals (Long *et al.*, 2007; Shi *et al.*, 2009; Cai *et al.*, 2012; Shi *et al.*, 2013).

Each chapter of the following research took advantage of the *Arabidopsis*-*Brassica* relationship or contributes towards our ability to capitalize upon it. In Chapter 2, a new *Arabidopsis* population of near-isogenic lines (NILs) derived from parents differing in many aspects of drought physiology is introduced. The utility of this population in understanding the genetics of drought physiology and its impacts across members of the *Brassica*ceae is discussed. Chapter 3 capitalizes on the relationship in building hypotheses about candidate genes underlying a quantitative trait loci (QTL) associated with

drought coping strategies. Finally, Chapter 4 uses the well annotated *Arabidopsis* genome to select candidate genes for detailed sequence characterization within one of the aforementioned QTL.

Drought, the major limitation to modern crop productivity

Drought is the most common reason for crop yield reductions to levels which are frequently below half of their theoretical potential (Boyer, 1982). Several coping mechanisms have been associated with plant survival under drought leading to the proposal of three discrete strategies (Ludlow 1989): drought escape, dehydration avoidance, and dehydration tolerance. Dehydration tolerance is defined by mechanisms which enable survival at low internal water potentials and is not a prevalent strategy in vascular plants (Oliver, Cushman and Koster, 2010). Crop breeding has most commonly focused on drought escape, which refers to the completion of the crop life cycle prior to the onset of drought, thus avoiding moisture limitations altogether. Alternatively, dehydration avoidance is defined by the ability to maintain internal water status during moisture-limited conditions by minimizing water loss and/or maximizing water uptake. Correlations between drought escape and mechanisms of dehydration avoidance have been observed (Mitchell-Olds, 1996; McKay *et al*., 2003; Wu *et al*., 2010; Franks, 2011). However, the generality of this hypothesized constraint requires more research as suggested by contrary results (Sherrard and Maherali, 2006).

Changing climate conditions and expansion into new production geographies are increasing the exposure of *B. napus* oilseed production to drought stress. Yet, little research has focused on drought adaptation in *B. napus* or its inheritance. Chapter 2 focuses on the genetics of night-time stomatal conductance (a dehydration avoidance mechanism) in *Arabidopsis* which, as discussed previously, has potential applications in our understanding and improvement of *B. napus* drought adaptation. Chapter 3 directly studies the inheritance and relationship among traits related to drought escape and dehydration avoidance, flowering time and root mass. Root pulling force (RPF, the vertical force required to remove a plant from the soil; Hayes & Johnson, 1939) is used as the measure of root system variation which is eventually determined to be acting as a measure of taproot size.

In pursuit of the 'perfect marker': moving from quantitative trait locus to causal polymorphism

Identifying the causal genetic variants underlying important phenotypes is a critical goal in contemporary plant genetics. It offers answers to evolutionary questions about the genetics of adaptation in natural populations and provides crop breeders the so-called 'perfect marker' for developing crops in order to meet the world's growing demand for food, fuel and fiber. It can be a daunting task complicated by complex genetic bases which are clouded by the effects of environment and further by interactions between gene and environment (Glazier, Nadeau and Aitman, 2002). However, advancements in quantitative genetics and molecular biology are creating viable avenues for elucidating the impacts of each of these variables and drilling down to the genes and functional polymorphisms underlying them.

Quantitative trait locus (QTL) mapping is a common statistical approach in plant genetics where experimental populations generated from crosses between inbred parent lines are used, in combination with molecular markers, to identify loci associated with variation in continuously distributed traits (Sax, 1923; Soller and Brody, 1976; Lander and Botstein, 1989; Haley and Knott, 1992; Zeng, 1994). This approach has been immensely successful in improving our understanding of the inheritance of complex traits. Typically however, QTL intervals from these analyses span large confidence intervals and can contain hundreds of potential causal genes (Price, 2006) so that QTL mapping has become a reliable first step for coarsely identifying regions of the genome for further research. Chapters 2 and 3 both utilized QTL mapping to identify loci associated with drought coping strategies (night-time stomatal conductance, flowering time and root size) in *Arabidopsis* and *B. napus* during which the candidate gene search space was reduced to approximately 5% and 1% of the respective genome. In addition, Chapter 3 incorporates conditional analyses of root size using correlated traits as covariates in order to understand the mechanistic model underlying co-localizing QTL. Interestingly, these analyses suggest that a QTL controlling flowering time may also be impacting root size directly.

After QTL identification, movement towards the causal gene(s) is generally a laborious process which often involves the creation of NILs (Eshed and Zamir, 1995). NILs are lines carrying genomic introgression(s) from one line in the homogeneous background of another line. The power of these lines is

their ability to isolate the genetic factor(s) controlling a trait to only the focal genomic region, thus demonstrating the true phenotypic effect due to the locus relative to the line which is void of the chromosomal introgression (Frary *et al*., 2000; Takahashi *et al*., 2001; Gerald *et al*., 2006). Creation of a near-isogenic line commonly begins with recurrent crosses between a line carrying the targeted QTL region (the donor line) to one which does not (the recurrent parent line), thus creating a backcross population. Backcross progeny are genotyped across the genome to track recombination events so that progeny with a high proportion of the recurrent parent genome carrying the donor introgression can be identified. Depending upon the genomic resolution desired (i.e. the physical size of the introgression), large populations and densely spaced molecular markers may be required so that those rare recombination events occurring within small genomic physical distances can be identified (Dinka *et al*., 2007). Ultimately, a smaller introgression yields a smaller list of possible candidate genes. The NIL population described in Chapter 2 provides a valuable resource for fine-mapping QTL as it is the most densely genotyped *Arabidopsis* NIL population to date (Koumproglou *et al*., 2002; Keurentjes *et al*., 2007; Torjek *et al*., 2008). In addition, heterozygous NILs were selected and maintained during population development to enable rapid fine-mapping.

In Chapter 4, the challenge of moving from QTL to candidate gene was approached differently. Rather than creating a NIL population and narrowing the list of candidate genes by narrowing the size of the genomic interval, the entire QTL region was characterized at the molecular level. This approach is made possible by recent advances in DNA sequencing technology where billions of small DNA fragments are sequenced simultaneously (Edwards, Henry and Edwards, 2012). Whole-genome sequence data generated from the mapping population parent lines (Chapter 3) was used to create reference based assemblies. The recently released draft genomes of *B. rapa* (Wang *et al*., 2011; Cheng *et al*., 2011) and *B. oleracea* (http://www.OCRI-genomics.org; 2013), the progenitor species, were used as references. These assemblies were used to characterize molecular variation occurring across five QTL intervals and select a list of candidate genes. Additionally, a list of genomic variants differentiating these divergent lines was

created. Ultimately, the results of this research narrow the focal regions for future research aimed at identifying the causal genetic variants underlying drought physiology.

REFERENCES

Al-Shehbaz IA, Beilstein MA, Kellogg EA. 2006. Systematics and phylogeny of the *Brassica*ceae (Cruciferae): an overview. *Plant Systematics and Evolution* **259**, 89-120.

Boyer JS.1982. Plant productivity and environment. *Science* **218,** 443–448.

Cai G, Yang Q, Yang Q, Zhao Z, Chen H, Wu J, Fan C, Zhou Y. 2012. Identification of candidate genes of QTLs for seed weight in *Brassica napus* through comparative mapping among *Arabidopsis* and *Brassica* species. *BMC Genetics* **13**,105.

Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M. 1998. Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* **41**, 62-69.

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. 2011. BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biology* **11,** 136.

Dinka SJ, Campbell MA, Demers T, Raizada MN. 2007. Predicting the size of the progeny mapping population required to positionally clone a gene. *Genetics* **176**, 2035-2054.

Edwards D, Henry R J, Edwards K J. 2012. Preface: advances in DNA sequencing accelerating plant biotechnology. *Plant Biotechnology Journal* **10**, 621–622.

Eshed Y, Zamir D. 1994. A genomic library of *Lycopersicon pennellii* in *L. esculentum*: A tool for fine mapping of genes. *Euphytica* **79**,175-179.

Gerald JNF, Lehti-Shiu MD, Ingram PA, Deak KI, Biesiada T, Malamy JE. 2006. Identification of quantitative trait loci that regulate *Arabidopsis* root system size and plasticity. *Genetics* **172**, 485–498.

Franks SJ. 2011. Plasticity and evolution in drought avoidance and escape in the annual plant *Brassica rapa*. *New Phytologist* **190,** 249-257.

Frary A, Nesbitt TC, Grandillo S, van der Knaap E, Cong B, Liu JP, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88.

Glazier AM, Nadeau JH, Aitman TJ. 2002. Finding genes that underlie complex traits. *Science* **298**, 2345-2349.

Haley CS, Knott SA. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Hayes HK, Johnson IJ. 1939. The breeding of improved selfed lines of corn. *Journal of the American Society of Agronomy* **31,** 710-724.

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, EffgenS, Vreugdenhil D, Koornneef M. 2007. Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* **175**, 891-905.

Koornneef M, Meinke D. 2010. The development of *Arabidopsis* as a model plant. *The Plant Journal* **61**, 909-921.

Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, Pooni HS, Newbury HJ, Kearsey MJ. 2002. STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. *The Plant Journal* **31**, 355-364.

Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-199.

Long Y, Shi J, Qiu D, *et al*. 2007. Flowering time quantitative trait loci analysis of oilseed *Brassica* in multiple environments and genomewide alignment with *Arabidopsis. Genetics* **177,** 2433-2444.

McKay JK, Richards JH, Mitchell-Olds T. 2003. Genetics of drought adaptation in *Araabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Molecular Ecology* **12,** 1137-1151.

Mitchell-Olds T. 1996. Pleiotropy causes long-term genetic constraints on life-history evolution in *Brassica rapa*. *Evolution* **50,** 1849-1858.

Oliver MJ, Cushman JC, Koster KL. 2010. Dehydration tolerance in plants. In: Sunkar R, ed. *Plant Stress Tolerance: Methods in Molecular Biology*. New York: Humana Press, 3-24.

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ. 1995. Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* **38**, 1122-1131.

Parkin IAP, SM Gulden, AG Sharpe, L Lukens, M Trick, TC Osborn, DJ Lydiate. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* **171,** 765-781.

Price AH. 2006. Believe it or not, QTLs are accurate!. *Trends in Plant Science* **11**, 213-216.

Sax K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**,552-560.

Sherrard ME, Maherali H. 2006. The adaptive significance of drought escape in *Avena barbata*, an annual grass. *Evolution* **60,** 2478-2489.

Shi J, Li R, Qiu D, Jiang C, Long Y, Morgan C, Bancroft I, Zhao J, Meng J. 2009. Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. *Genetics* **182,** 851-861.

Shi T, Li R, Zhao Z, Ding G, Long Y, Meng J, Xu F, Shi L. 2013. QTL for Yield Traits and Their Association with Functional Genes in Response to Phosphorus Deficiency in *Brassica napus*. *PLoS One* **8**, e54559.

Soller M, Brody T. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* **47**, 35-39.

Takahashi Y, Shomura A, Sasaki T, Yano M. 2001. Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proceedings of the National Academy of Sciences, USA* **98**, 7922–7927.

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pederson BS, Conant G, Wang X, Freeling M, Pires JC. 2012. Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* **190**, 1563-1574.

Torjek O, Meyer RC, Zehnsdorf M, Teltow M, Strompen G, Witucka-Wall H, Blacha A, Altmann T. 2008. Construction and analysis of 2 reciprocal introgression line populations. *Journal of Heredity* **99**, 396-406.

U N. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **7,** 389-452.

Wang X, Wang H, Wang J, *et al*. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43,** 1035-1039.

Wu CA, Lowry DB, Nutter LI, Willis JH. 2010. Natural variation for drought-response traits in the *Mimulus guttatus* species complex. O*ceologica* **162,** 23-33.

Zeng *ZB*. 1994. Precision mapping of quantitative trait loci. *Genetics* **4**, 1457-1468.

Zhao M, Du J, Lin F, Tong C, Yu J, Huang S, Wang X, Liu S, Ma J. 2013. Shifts in the evolutionary rate and intensity of purifying selection between two *Brassica* genomes revealed by analyses of orthologous transposons and relics of whole genome triplication. *The Plant Journal* **76**, 211-222.

CHAPTER 2: DEVELOPMENT OF A NEXT-GENERATION NIL LIBRARY IN *ARABIDOPSIS THALIANA* FOR DISSECTING COMPLEX TRAITS

SUMMARY

For complex traits, moving from quantitative trait loci to causal alleles remains daunting. To facilitate this process, we developed a population of 75 *Arabidopsis thaliana* near-isogenic lines (NILs) composed of both homozygous and heterozygous introgressions spanning the genome. As a proof of concept, we utilize homozygous NILs to validate a QTL for stomatal conductance, a very low heritability trait. We then exploit a heterozygous NIL to fine map this QTL in a single generation. Having both homozygous and heterozygous introgressions combined with dense genotyping by sequencing makes this library a powerful tool for dissecting complex traits of agricultural and ecological significance.

INTRODUCTION

Linkage mapping of QTL is a common statistical approach in plant genetics where recombinant populations generated from crosses between inbred parent lines are used, in combination with molecular markers, to identify loci associated with variation in continuously distributed traits (Sax, 1923; Soller and Brody, 1976; Lander and Botstein, 1989; Haley and Knott, 1992; Jansen, 1993; Zeng 1993; Zeng 1994; Sen and Churchill, 2001). Mapping populations common to QTL analyses are many and include doubled haploids (DH), F2, backcross, advanced intercross, nested association mapping and RILs. Mapping QTL for complex traits is now routine, with the typical output being QTL spanning large confidence intervals encompassing many (hundreds or more) possible causal genes (Price, 2006).

The steps following QTL identification frequently involve functional validation of the QTL, and refinement of location (fine-mapping) towards the goal of identification of a causal gene – the major challenge in quantitative genetics today (Rockman, 2008). One of the most common approaches for accomplishing these objectives is through the development and phenotypic characterization of NILs

(Monforte and Tanksley, 2000a). The generation and phenotyping of NILs is considered a laborious and time consuming process, but the robust design leads to a minimal false positive rate.

NILs are lines containing a single or small number of genomic introgressions from a donor parent in a different and otherwise homogeneous genomic background. By homogenizing all genetic factors outside of the focal genomic region, the true effect of the QTL on the phenotype can be estimated relative to the line into which the introgression was introduced (i.e. void of the chromosomal introgression; Landi *et al.*, 2005). In addition to the simplification of genetic analyses, NILs are considered genetically 'immortal' (Iniguez-Luy *et al.*, 2009) which allows for replicated experiments across multiple environments resulting in more accurate estimates of effect size for complex traits. NILs have proven to be an effective resource for QTL validation and a logical starting point for the creation of fine-mapping populations (Eshed and Zamir, 1994; Frary *et al.,* 2000; Takahashi *et al.*, 2001; Lin *et al.*, 2003; Juenger *et al.*, 2005a; Gerald *et al.*, 2006; Xing *et al.*, 2008; Zhou *et al.*, 2010).

Creation of a single near-isogenic line generally starts by crossing a line carrying the targeted QTL region to one of the parental lines of the population, thus creating a backcross population. Genome-wide genotyping of the backcross progeny is performed to identify recombination events allowing for selection of progeny which carry the target chromosomal introgression derived from the donor and recurrent parent genome elsewhere. Subsequent generations of self-pollination (selfing) are normally required to achieve homozygosity of the introgressed region and the process can take several backcrossing cycles to produce a NIL carrying an introgression of acceptable size and genomic location. An alternative approach has been the use of heterogeneous inbred families (HIFs) where NILs are selected from incompletely inbred lines which still harbour a small amount of heterozygosity at random intervals across the genome (Tuinstra, Ejeta, and Goldsbrough, 1997; Loudet *et al.*, 2005). Analysis of a HIF population with molecular markers allows for the selection of lines heterozygous at a candidate genomic location, which in combination with further selfing and genotyping, enables selection of NILs derived from several heterogeneous genetic backgrounds. Producing NILs with smaller introgressions

requires greater effort. Large populations are needed to break up small chromosomal segments, and high-density genotyping is required to discover them.

A NIL library is a family of near-isogenic lines where each line carries a different donor parent fragment and the population carries introgressions spanning the entire genome (Eshed and Zamir, 1995). A NIL library is an ideal starting point for QTL validation, especially in cases where the library is derived from parent lines for which an immortal recombinant population (i.e. RILs, DH, etc.) already exists. In this case, QTL identified via traditional linkage mapping experiments performed on the mapping population can be immediately tested by selecting NIL(s) representing the QTL introgression and testing them for a phenotypic effect relative to the wild type recurrent parent. NIL libraries are also valuable starting material for fine-mapping QTL through the creation of sub-NILs (Monforte and Tanksley, 2000b), recombinant lines in which the original NIL introgression is broken into smaller genomic fragments. In this case, a candidate NIL is backcrossed to the recurrent parent and the progeny are genotyped using markers specific to the introgression region so that individuals carrying genomic fragments spanning the length of the original introgression can be identified. Subsequent phenotyping of the sub-NILs provides finer resolution of the region controlling the trait of interest, effectively narrowing the list of possible causal genes.

Several NIL populations are currently available to the *Arabidopsis* research community. Koumproglou *et al*. (2002), using 31 simple sequence repeat (SSR) markers, created a population of Chromosome Substitution Strains by replacing chromosomes from the accession Columbia (Col-0) with homologous chromosomes from the accessions Landsberg erecta (Ler) and Niederzenz (Nd). Additionally, a population of more traditional NILs were created in a systematic approach where increasing lengths of chromosomal introgressions were introduced from Ler into the Col-0 background. Keurentjes *et al*. (2007) generated a population of 92 NILs carrying genome-wide chromosomal introgression from the accession Cape Verde Islands (Cvi) into the Ler background. Selections were made from the genotyped RIL mapping population described by Alonso-Blanco *et al*. (1998) and used in backcrosses to create the NIL library. The RIL population has been mapped for QTL underlying

flowering time and carbon isotope ratio ($\delta^{13}C$; Juenger *et al.*, 2005b), recombination frequency (Esch *et al.*, 2007), seed germination (Laserna, Sanchez, and Botto, 2008), seed mineral concentration (Waters and Grusak, 2008) and fructose sensitivity (Li *et al.*, 2011). The same 321 AFLP (Amplified Fragment Length Polymorphism) markers used to build the RIL map were used in the NIL breeding scheme. Finally, Torjek *et al*. (2008) created a population of 140 reciprocal NILs from the accessions Col-0 and C24 (78 NILs in the Col-0 background and 62 lines in the C24 background) utilizing a total of 125 markers (Torjek *et al.*, 2003). This NIL library has been used in subsequent studies of epistasis (Reif *et al.*, 2009) and heterosis (Jan Lisec *et al.*, 2009).

Here we report the development of a new population of 75 NILs constituting genome-wide chromosomal introgressions. The NIL population exploited inbred lines selected from the RIL population described in McKay *et al*. (2008) as the starting material for backcrossing. Briefly, the RIL population is derived from a cross between the *A. thaliana* ecotypes Tsu-1 (CS1640), an accession originating from Tsushima, Japan and Kas-1 (CS903), an accession originating from Kashmir. These sites of collection are among the wettest and driest habitats, respectively, in the *A. thaliana* species range and differ in several aspects of drought physiology (McKay, Richards and Mitchell-Olds, 2003; Juenger *et al.*, 2010). Recombinant populations derived from these diverse accessions will therefore segregate alleles underlying variation in these physiological traits, providing a powerful resource for identifying functional genes.

We developed a population of 75 *Arabidopsis thaliana* NILs containing both homozygous and heterozygous introgressions, enabling simultaneous pursuit of QTL validation and fine-mapping. Genotyping the population with over 1,000 molecular markers has provided us with excellent resolution on the total number of introgressions existing in each NIL as well as their location and length. It is the most densely genotyped NIL population developed thus far by more than 3-fold. The utility of the NIL library is demonstrated in a simple case study where, in a single generation, we utilize a homozygous NIL to validate and localize a QTL for a low heritability physiological trait (g_0; night-time stomatal

conductance) while concurrently selfing heterozygous selections to create sub-NILs for further fine-mapping.

MATERIALS AND METHODS

Plant Material & Growth Conditions

The *A. thaliana* accessions Kas-1 and Tsu-1 were used as the original parent lines in developing the RIL population of 346 lines. Kas-1 and Tsu-1 were chosen as parents for developing this population as a result of their extreme differences in water use efficiency as measured by $\delta^{13}C$ (McKay, Richards and Mitchell-Olds, 2003; Juenger *et al.*, 2010). RILs from this population served as the starting point for the NIL breeding program described below.

For the QTL experiment, seed of the RILs along with the parents were sown on soil (Fafard 4P mix, Conrad Fafard Inc., Agawam, MA) in 3-inch pots. Seeds were planted in a randomized complete block design consisting of 2 blocks, and then the pots were refrigerated at 4°C in darkness for 5 d to cold-stratify the seeds prior to commencement of a 8:16 h (light: dark) photoperiod in Conviron ATC60 growth chambers (Controlled Environments, Winnipeg, MB) at 23°C and 40% humidity during the day and 20°C and 50% humidity during the dark period. Light intensity was approximately 330 μmol m^{-2} s^{-1}. Plants were grown for approximately 6 weeks prior to measurement. Stomatal conductance was measured in darkness on non-senescing leaves that were large enough to fully accommodate the leaf chamber (1 cm x 2 cm), using an infrared gas analyser (model Li-Cor 6400, LiCor Inc., Lincoln, NE). Prior to measurement the plants were dark adapted for 20 – 28 h. A humidifier was used to reduce variation in humidity over the course of the measurements. For each leaf 10 measurements were taken, with an interval of 10 s between measurements.

For the QTL validation experiment, plants were grown in a randomized complete block design consisting of 3 blocks where each genotype was replicated 6 times within each block. Plants were grown under exactly the same conditions as those described above except that the photoperiod was increased to 12;12 h (light:dark) to accommodate other experiments conducted in the same chamber. One major

difference between the two experiments was the use of leaf porometers (model SC-1, Decagon Devices, Inc., Pullman, WA) rather than an infrared gas analyser for stomatal conductance estimates. Two non-senescing leaves were measured on each plant following the manufacturer's recommended protocol.

Genetic analyses

Broad-sense heritability was estimated by calculating the ratio V_G:V_P, where V_G is the among-RIL component of variance and V_P is the total phenotypic variance. QTL mapping was performed in the R/qtl program of the R statistical package (Broman *et al.*, 2003; Broman and Sen, 2009) using Haley-Knott regression. Significance thresholds were determined using 1000 permutations. A penalized stepwise approach (Manichaikul *et al.*, 2009) was used for selection of a multiple-QTL model.

For the QTL validation experiment, data were analyzed with a linear mixed model using PROC MIXED in the SAS software package (SAS Institute Inc. 2003, Cary, NC) where block, row and column effects were treated as random.

Marker Assisted NIL Breeding Program

To start, 7 RILs were selected from the original population of 346 using the code supplied in an additional file (http://www.biomedcentral.com/1471-2164/14/655). These 7 represented lines homozygous for Kas-1 alleles across one of each of the 5 chromosomes and all were crossed to Tsu-1 at least 10 times. Some attempted crosses may result in self-pollination due to technical error, thus we genotyped progeny to confirm they were F1s. In general, the real F1s were several times larger than the midparent value, so genotyping was almost unnecessary. Confirmed F1s were crossed back to Tsu-1 and each fruit was collected separately and considered a BC1 family, ultimately creating 25 families. 24 plants from each family were genotyped at the chromosome of interest and selected for selfing to generate BC1S1 seed. In addition to culling the occasional plant generated due to self-pollination, it was also necessary to remove individuals sired by (haploid) pollen from the F1 carrying Tsu-1 alleles for the chromosome of interest. In the next generation, 690 BC1S1 plants were genotyped with the 48 genome-

wide SSRs described in Mckay *et al.* (2008). These were then ranked using an algorithm to find lines that were largely Tsu-1, but carrying Kas-1 introgressions spanning the genome. In the end, 75 lines were selected which we screened at an additional 149 loci using the Sequenom MassARRAY® platform, of which 41 were polymorphic. 930 polymorphic loci were added to this marker data set via 2b-RAD (Wang *et al.*, 2012) where class IIB restriction enzymes are used minimize genome complexity for a final total of 1011 genotyped.

DNA Extraction and Genotyping

Genomic DNA was isolated from lyophilized tissue collected from approximately 4-week-old, chamber grown plants using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The 48 polymorphic microsatellites used in this study were selected from the large number of those available in *A. thaliana* (Bell and Ecker, 1994; Symonds and Lloyd, 2003; arabidopsis.org.) due to easily distinguishable allele calls. Descriptions of the primers, PCR conditions and allele scoring are explained in McKay *et al*. (2008).

DNA samples were used to prepare 2b-RAD libraries as previously described in Wang *et al.* (2012). A detailed protocol is available at the Meyer laboratory website (http://people.oregonstate.edu/~meyere/). Briefly, library preparation for 2b-RAD genotyping began with digestion of gDNA samples with AlfI (Fermentas) for 37 °C for 3 h followed by ligation of adaptors at 4 °C for 16 h. Ligation products were amplified by PCR and barcodes introduced to gel-extracted products in a second PCR reaction. Finally, libraries were pooled for multiplex sequencing on the SOLiD sequencing platform (Applied Biosystems). Raw sequences were processed to exclude low-quality reads, and the HQ reads that remained aligned in color-space using the SHRiMP software package (Rumble *et al.*, 2009) to AlfI sites extracted from the *Arabidopsis* genome (TAIR9). A custom Perl script was applied to eliminate short, statistically weak and ambiguous alignments (reads matching multiple sites equally well). Finally, genotypes were determined from nucleotide frequencies using custom Perl scripts to classify each locus as homozygous (minor allele frequencies [MAF] <1%), heterozygous (MAF>25%), or

undetermined (1% > MAF >25%). 20x coverage was required in the parental genomes to identify these alleles with high confidence, and a relaxed threshold of 10x in all other samples to maximize marker densities. Each polymorphic locus identified in these genotypes was compared with the parental genotypes (Tsu-1 and Kas-1) to assign it to one of these backgrounds, a comparison that would obviously not be possible for any loci genotyped in one parent but not the other as a result of variation in sequencing coverage. To reduce the effects of such missing data, we imported genotypes for Tsu-1 and Kas-1 from resequencing data (McKay, unpublished results) for any loci genotyped in one parent but not the other.

KASP SNP genotyping assays (LGC Genomics, Teddington, Middlesex, UK) were used for sub-NIL development. Primer sequences (Supplementary Table 5.1) were designed using sequence data from TAIR10 (Lamesch *et al.,* 2012) for amplification of SNPs identified and validated on the SNPlex genotyping system (Applied Biosystems) as described in McKay *et al.* (2008). KASP is a novel allele-specific PCR assay that utilizes a FRET (Fluorescence Resonance Energy Transfer) system. In short, along with a common primer, allele-specific primers are designed to include a unique 18 bp sequence at the 5' end. The unique sequences are identical to a pair of oligonucleotides with 3' bound quenchers for a complement pair of 5' fluorescently labelled oligos inside the reaction mix. During PCR, allele specific amplification leads to the generated product(s) outcompeting the quencher containing oligos for binding to the fluorescently labelled oligos, allowing for an observable signal to be measured using a light reader. The intensity of the signal(s) allows for a quantitative measure of SNP copy number.

Estimating chromosomal introgression length and number

The physical length of introgressions in the final NIL library was estimated using graphical genotypes (Young and Tanksley, 1989). Physical length estimates of introgressions flanked by SSR markers were made using the location of the forward primers, SNP locations were determined by their location in the Col-0 reference genome. To avoid false-positives, an introgression was scored based on the presence of at least 3 consecutive markers with the Kas-1 genotype. Introgression boundaries were then defined by three consecutive markers with an alternative genotype. This helped avoid over-

estimating introgression numbers due to occasional incorrect allele calls or differences in the location of loci in this population relative to the Col-0 genome used as a reference for mapping sequence reads. For the analysis of introgression discovery at varying marker densities an Excel Macro was written to sum the number of heterozygous and homozygous introgressions discovered. The loci included in replicated sampling were selected randomly using Excel's RAND function.

Candidate Gene Identification

The full list of genes expected to lie within the QTL interval spanning physical positions 505,086 to 5,273,972 was assembled using TAIR10 (Lamesch *et al.*, 2012) GO annotations for the full gene list were downloaded using the Bulk Data Retrieval and Analysis tool on TAIR10 and searched using the terms abscisic acid (ABA), stomata and water. Gene enrichment analysis was performed using the GO enrichment analysis tool in AmiGO (Carbon *et al.*, 2009).

RESULTS

Marker-Assisted NIL Breeding Program

Figure 2.1 shows the breeding design for the NIL library. An algorithm was developed to select RILs homozygous for Kas-1 alleles across one of each of the 5 *Arabidopsis* chromosomes. The results found 7 such RILs from the population of 346. These RILs were crossed to Tsu-1 and progeny were genotyped to confirm they were truly F1s. These were then crossed back to Tsu-1, creating 25 BC1 families. Plants from each BC1 family were genotyped at the chromosome of interest to select individuals carrying Kas-1 alleles so they could be self-pollinated to generate BC1S1 seed. BC1S1 plants were genotyped using 48 genome-wide SSRs described in McKay *et al*. (2008). These data were analyzed using an algorithm designed to identify a subset of lines representing Kas-1 chromosomal introgressions spanning the genome in otherwise Tsu-1 backgrounds. The algorithm was used to select 103 BC1S1 plants which were screened at an additional 149 single nucleotide polymorphisms (SNPs) loci using the Sequenom MassARRAY® (Sequenom, San Diego, CA). Only 41 of 149 SNPs were

informative for the parental lines. Finally, an additional 930 polymorphic loci were revealed using 2b-RAD (Wang *et al.*, 2012) whereby genome complexity is reduced using class IIB restriction enzymes followed by sequencing on the SOLiD platform (Applied Biosystems, Foster City, CA).

Polymorphisms detected between Tsu-1 and Kas-1 by 2b-RAD genotyping

Restriction site–associated DNA (RAD) tag sequencing reduces genome complexity by focusing only on DNA flanking the recognition sites of the selected restriction endonuclease (Davey *et al.*, 2011). The RAD method used in this study, described in Wang *et al.* (2012) is a simple and effective means of discovering a large number of SNPs unique to the study population, avoiding the ascertainment bias associated with SNPs discovered via population surveys (Clark *et al.*, 2005). The 2b-RAD method utilizes the type IIB restriction enzyme, *AlfI*, which operates by cleaving DNA both upstream and downstream of the recognition site. The resulting tags are uniform in length, making them ideal for amplification and sequencing on next-generation platforms. Following digestion, tags were labelled with sample-specific oligonucleotide barcodes for multiplexed sequencing. Finally, reads were quality filtered and aligned to a collection of AlfI sites in the Col-0 *Arabidopsis* reference genome (TAIR9) in order to assign a physical location to each SNP.

Initially, 1319 polymorphisms were identified between the parent lines Tsu-1 and Kas-1 based on the 2b-RAD tags that were sequenced. Because these NILs are derived from a known pedigree of previously genotyped individuals, we were able to filter to include SNPs that would segregate in the progeny, resulting in a final set of 930 loci with high-confidence genotypes for use in subsequent population analyses. A non-trivial fraction of markers remained as missing data in each sample due to the stringent scoring criteria of our method. The majority of uncalled loci in typical 2b-RAD datasets are discarded because of low coverage (Meyer, unpublished observations) so this problem could be mitigated with deeper sequencing. However, the known pedigree of these samples and the low level of recombination made it possible to accurately reconstruct haplotypes despite these missing data. The filtered data were used to construct graphical genotypes (Young and Tanksely, 1989) of the NIL

population, a subset of which are represented in Figure 2.2. We also provide a database of the genotypes for the entire NIL population (http://www.biomedcentral.com/1471-2164/14/655). In addition, both parental accessions have been re-sequenced and the genome-wide reads have been deposited in the Short Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi) and posted on the 1001 Genomes Project website (www.1001genomes.org/) so the details of the 930 SNPs utilized in this study can be accessed at these resources.

Genomics of chromosomal introgressions in the NIL population and the added value of increased marker resolution

Across the 75 NILs, the average number of homozygous introgressions per NIL was 1.35 and ranged from 0 to 4 while the average number of heterozygous introgressions was 2.49 and ranged from 0 to 6 (Figure 2.3). The average number of introgressions per chromosome was 57.6, ranging from 34 on chromosome 2 to 79 on chromosome 1 (Supplementary Table 5.2). The total length of homozygous introgressions was 506 Mb compared to nearly 949 Mb of heterozygous chromosomal introgression which represent 4.3 and 8.0 times the total length of the *Arabidopsis* genome, respectively. Together these results suggest we have reached our goal- the entire genome is represented as a Kas-1 introgression for each genotypic state (i.e. zygosity) in at least one NIL, thus enabling QTL validation and fine-mapping for any locus of interest.

The additional loci accounted for by 2b-RAD genotyping resulted in a final marker density of 2.24 markers per cM, based on the estimated 450.8 cM map of the Kas-1 x Tsu-1 RIL population. This is a significant improvement in resolution from the 0.18 markers per cM when using only the original SSR and Sequenom marker set (hereafter referred to as the coarse map). In spite of the high frequency of uncalled alleles, 128 new introgressions were revealed which summed to nearly 539 Mb of DNA (164 Mb homozygous and 375 Mb heterozygous) that would have been missed without the additional markers from 2b-RAD genotyping. To illustrate this effect we re-sampled the dataset at varying marker densities (Figure 2.4). The exponential curve fit ($r^2 = 0.99$) used to estimate introgression detection begins to level

above 800 markers, suggesting diminishing introgression discovery with more extensive genotyping. In a comparison of the NILs using the coarse map relative to the dense map, the average size of a homozygous introgression in the coarse map was 18% larger (1.2 Mb) than in the dense map. Similarly, the average heterozygous introgression size in the coarse map was 19% larger (1.3 Mb), confirming that the additional markers were identifying smaller introgressions missed in the coarse map. This fact is highlighted by the total number of introgressions (Supplementary Table 5.2) discovered using the denser marker set. The result was a 1.8-fold (Figure 2.3) increase in the number of homozygous and heterozygous introgressions discovered.

Case Study: utilizing selections from the NIL library for QTL validation and sub-NIL development

To demonstrate the value of this new resource, we analyzed the RIL population (McKay *et al.*, 2008) for QTL for night-time leaf conductance (g_0). g_0 is a low-heritability, quantitative trait that is important for plant-water relations and mineral nutrition. While the adaptive value of g_0 has yet to be fully understood, incomplete stomatal closure during the night can lead to substantial transpirational water loss (Caird, Richards and Donovan, 2007). Variation in this trait has been found among and within species, and it correlates with some daytime gas-exchange traits such as water-use efficiency (the ratio of CO_2 assimilation to transpiration) (Christman *et al.*, 2008). Estimates of transpiration have been found to be particularly sensitive to g_0 (Bauerle and Bowden, 2011), making it an interesting candidate for studies on the physiology and genetics of plant drought adaptation. In view of that, intraspecific variation in observed g_0 has been found to have the largest effect on transpiration across a species' native habitat (Bauerle, unpublished observation).

Significant variation in night-time conductance was observed among the RILs. We identified a single QTL for g_0 on the top of chromosome 1 (Figure 2.5A), which explained 9% of the variance in g_0, and found the trait to have relatively low broad sense heritability ($H^2 = 0.21$) in this population. Lines having Kas-1 alleles of markers at the QTL had lower dark conductance, consistent with the dry habitat of the Kas-1 parent (Supplementary Figure 5.1). Additional loci were identified on chromosomes 2 and 4

below the threshold of significance, which may have had marginal effects on g_0 (Supplementary Figure 5.2).

To validate the QTL we selected two NILs homozygous for a Kas-1 introgression spanning the QTL and measured g_0 relative to Tsu-1 with the expectation that one or both would have a significantly lower g_0 value. NIL TK201_137_6 carries an introgression estimated to span physical positions 505,086 to 5,273,972 on chromosome 1 and KT116_63_15 is estimated to carry a much larger introgression between positions 2,040,091 and 19,225,223 (Figure 2.5B). KT116_63_15 also carried small heterozygous regions at either end of the homozygous introgression. Large and highly significant differences were found between both NILs and Tsu-1 (Table 2.1), providing strong evidence for the presence of the QTL and providing a surprisingly high estimate of the relative difference in g_0 conferred by the two alleles when compared to the results of the initial QTL experiment (Supplementary Figure 5.1). The region between 5,273,972 and 19,225,223 can be effectively eliminated from consideration for harbouring the causal locus since TK201_137_6 was significantly different from Tsu-1 and did not carry Kas-1 DNA in this interval (Eshed and Zamir, 1995). It is worth noting that both NILs carried introgressions on chromosomes other than the chromosome one focal area. However, none of them were common between the NILs and the difference in g_0 values between KT116_63_15 and TK201_137_6 was non-significant which suggests these introgressions were not impacting our results substantially.

Nearly 1,500 genes are predicted to lie within the region spanning physical positions 505,086 to 5,273,972 of chromosome one. We have assembled a list of candidate genes based upon hits to gene ontology (GO) terms relevant to stomatal conductance: abscisic acid (ABA), stomata and water (Supplementary Table 5.3).

To illustrate the power of deriving sub-NILs from heterozygote NILs, concurrent to the QTL validation experiment we planted seeds derived from a line heterozygous in the roughly 3Mb g_0 QTL interval (Figure 2.5). We selected 5 polymorphic loci from a panel of validated SNPs described in McKay *et al.* (2008) for genotyping a population of 286 BC1S3 individuals (BC1S1 graphical genotype is represented in Figure 2.6A). The marker representing the lower end of the interval at physical position

6,839,609 did not segregate and all individuals were homozygous for the Tsu-1 allele. The genotype for the 2b-RAD allele near this location was scored as "not genotyped" in the original BC1S1 genotyping so we were unsure exactly where this particular heterozygous introgression ended. In the end, we were left with 4 informative markers in the physical interval spanning positions 2,211,035 to 6,572,582. We selected 17 recombinants (Figure 2.6B) representing the majority of the recombination events possible. Unfortunately, no double recombinants were discovered so that a sub-NIL representing the Kas-1 alleles at the middle of the interval could be recovered. However, heterozygous individuals TK176_108_1_4_13 and TK176_108_1_4_38 were kept for selfing and will be available for re-planting to accomplish this since a crossover has already occurred at the lower end. Ultimately, individuals were recovered in this single selfing generation that could be used in the next generation for g_0 phenotyping experiments to effectively narrow the QTL interval down to, at most, the 1.9 Mb interval between markers C1_2211035 and C1_4142402, an interval predicted to carry about 590 genes.

DISCUSSION

Maintenance of homozygous and heterozygous NILs facilitates simultaneous QTL validation and fine-mapping efforts

Near-isogenic lines remain the ideal starting material for validation of QTL as well as breeding schemes designed for fine-mapping with the end goal being the identification of candidate genes (Yan *et al.*, 2011; Wang *et al.,* 2012; Patterson *et al.*, 1990; Ducrocq *et al.*, 2009). QTL validation is relatively straightforward and consists simply of phenotyping NILs with introgressions at the region of interest for the trait of interest. Creation of a suitable population for fine mapping is not as straightforward and is normally a three-generation process that starts with a cross between an inbred NIL and the recurrent parent. This is typically followed by a generation of self-pollination to allow for recombination in the introgression region. The seed harvested from these self-pollinated plants can then be genotyped with markers specific to the region so that homozygous sub-NILs can be identified. The process is fairly

straightforward and inexpensive in the context of physical resource, but there is a time cost of at least 3 generations (equivalent to a minimum of 18 weeks).

Our case study illustrates the advantages of maintaining both homozygotes and heterozygotes in the NIL population, combining the benefits of traditional homozygous NILs with the advantages of HIFs (Tuinstra, Ejeta , and Goldsbrough, 1997; Loudet *et al.*, 2005). For example, measuring g_0 on the homozygous NILs provided strong evidence for the presence of the QTL in a single generation, thus avoiding the process of generating homozygous lines that would be necessary in HIF populations. These results provided a better estimate of the QTL effect size relative to the results derived from our QTL mapping approach and have justified further investments in fine-mapping using heterozygous. This emphasizes the power NILs create by isolating the genetic factors controlling a phenotype to a single locus as there were other loci worthy of consideration as contributors to variation in g_0 in the RIL. Analysis of the genes predicted to lie within this interval revealed a majority of them had GO annotations related to ABA, the major signalling molecule in stomatal regulation (Acharya and Assmann, 2009; Nilson and Assman, 2007; Schroeder *et al.*, 2001), but examination of the entire region with the AmiGO enrichment analysis tool (Carbon *et al.,*2009) found it was not significantly enriched for ABA genes. Inspection of the physical location of these ABA-associated candidates reveals that they are clustered in a 1.2 Mb interval (At1 physical interval: 712,473-1,894,148) which represents a relatively small portion of the 4.8 Mb introgression tested, thus providing an interesting focal region during fine-mapping of the g_0 phenotype.

With regards to fine-mapping, selfing a heterozygous selection from the population yielded several sub-NILs suitable for phenotyping or additional genotyping in future generations, an attribute common with HIF populations and advantageous over traditional NILs. This was accomplished using a modest population size of BC1S3 plants (n=286) and the interval could be narrowed down further through genotyping at a higher number of loci and increasing the population size (Dinka *et al.*, 2007). Regardless, in a six-week period we have identified a population encompassing recombinants in the 4.8 Mb region

identified as causal during the QTL validation experiment, translating to a 3-fold change in total time versus a breeding scheme utilizing inbred NILs.

2b-RAD is an efficient method for dense genotyping of recombinant populations

Arabidopsis thaliana recently celebrated its 25th anniversary as a model organism and now stands alone as the most thoroughly studied plant species on record (http://www.arabidopsis.org/, Koorneef and Meinke, 2010). Recent efforts are producing comprehensive polymorphism databases (http://www.arabidopsis.org/, http://signal.salk.edu/cgi-bin/AtSFP1001genomes.org). To interpret the significance and functional consequences of this natural variation, we need to understand the multivariate phenotypic consequences of these variants. NIL libraries, mutants and complementation studies are the tools required for this mechanistic understanding.

The 2b-RAD method added an additional 930 high confidence genotypes to our map providing a level of resolution not yet achieved in any of the *Arabidopsis* NIL populations described to date. The value of these additional markers is obvious as we compare the coarse and dense maps. The discovery of an additional 129 introgressions is clearly important when making selections for QTL validation. For instance, three additional homozygous introgressions were discovered in KT154_2_3, changing the estimate from one to four. This is a clear illustration of the risks associated with utilizing NILs genotyped at low density in experiments aimed at QTL validation. These offsite introgressions may have effects on the phenotype of interest, potentially resulting in erroneous or uncertain conclusions regarding the QTL effect size and location.

The Kas-1 x Tsu-1 RIL and NIL populations are a valuable resource for research on the genetics of drought adaptation in the Brassicaceae

Substantial variation for several traits relevant to drought adaptation have been observed in the Kas-1 x Tsu-1 RIL population including $\delta^{13}C$, leaf water content, instantaneous transpiration rate, flowering time, abscisic acid content and root mass (McKay *et al.*, 2008; unpublished results).

Accordingly, the NIL population described herein is expected to vary for the same traits, providing a powerful resource for moving from QTL, encompassing thousands of genes, discovered in the RIL population towards a smaller list of putative functional candidates.

No other plant species has been more studied or characterized than *Arabidopsis thaliana* (Koornneef and Meinke, 2010). A high degree of sequence collinearity between it and members of the agriculturally significant *Brassica* genus was discovered over a decade ago (Cavell *et al.*, 1998). Similar levels of synteny have been found in comparisons with other taxa in the *Brassica*ceae (Boivin *et al.*, 2004; Yogeeswaran *et al*., 2005; Lysak *et al.*, 2006; Schranz *et al.*, 2007). These results suggest that translational genomics, that is utilizing basic research findings in model organisms to answer practical research questions in species of higher economic value or importance (Stacey and VandenBosch, 2005) could be a viable avenue in understanding complex traits. In this regard, we suggest the Kas-1 x Tsu-1 populations as the ideal starting point for basic research on the genetics and genomics of drought adaptation.

CONCLUSIONS

We have developed a population of 75 NILs that provides genetic resources for fine-mapping QTL as well as QTL corroboration. The high marker density used to construct the population provides a level of resolution not yet seen in a NIL population, thus minimizing ambiguity in fine-mapping and QTL validation studies caused by unidentified chromosomal introgressions elsewhere in the genome. The unique variation that exists between the parents used to construct this resource provides a valuable asset for research focused on identifying the genes responsible for drought adaptation in *Arabidopsis* and beyond.

TABLES AND FIGURES

Table 2.1 Results of QTL validation experiment comparing NIL g_0 values with Tsu-1.

Difference values with ** are highly significant ($P < 0.0001$)

Comparison	g_0 (mmol m-2 s-1)	Standard Error	Difference NIL – (Tsu-1)	t Value
Tsu-1	119.76	11.55	n/a	-0.94
TK201_137_6	52.42	10.99	-67.34 **	-4.02
KT116_63_15	67.45	11.40	-52.30 **	-3.15

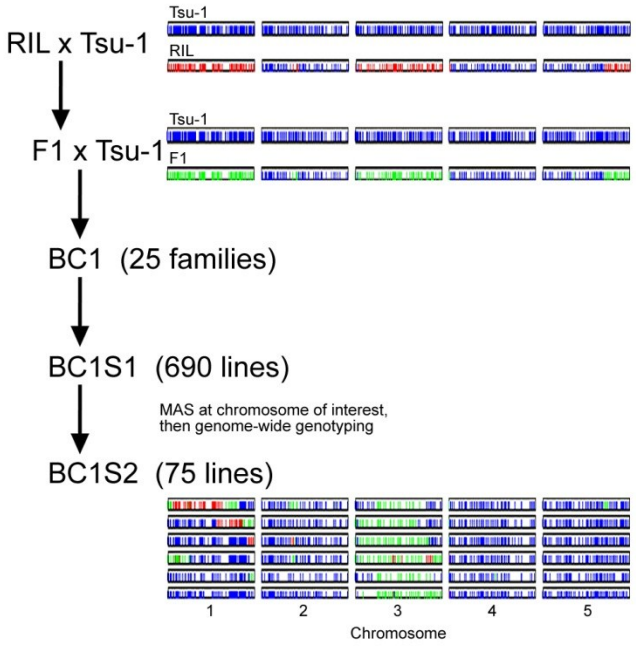


Figure 2.1 Breeding scheme of the NIL library.

Breeding scheme and graphical genotypes of a set of NILs containing both homozygous introgressions (Chromosome 1) and heterozygous introgressions (Chromosome 3) derived from a single RIL.

Each diploid breeding line is represented by a single row of 5 chromosomes where red coloring represents Kas-1 genotypes; Blue, Tsu-1; Green, heterozygous. Graphical genotypes of 6 of the 75 lines are shown.

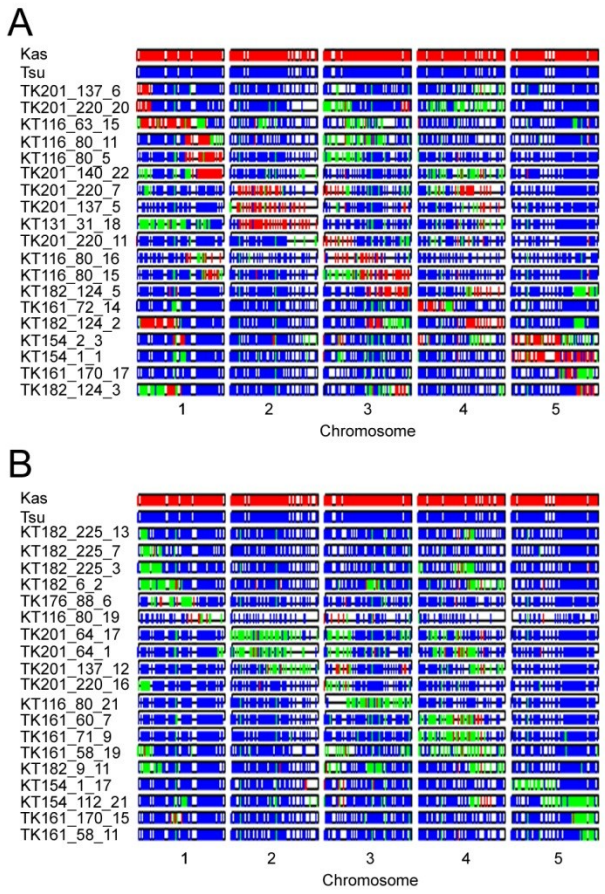


Figure 2.2 Graphical genotypes of NILs representing (A) homozygous and (B) heterozygous introgressions cumulatively spanning the length of the genome.

Red, Kas-1; Blue, Tsu-1; Green, heterozygous.

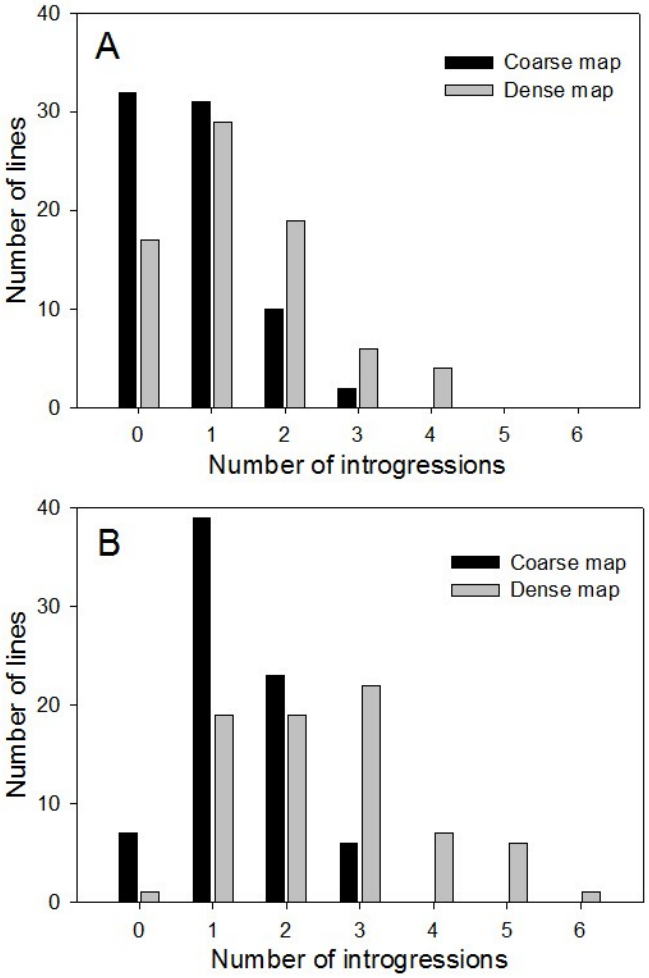


Figure 2.3 Distribution of (A) homozygous and (B) heterozygous introgression number, estimated using the coarse and dense maps.

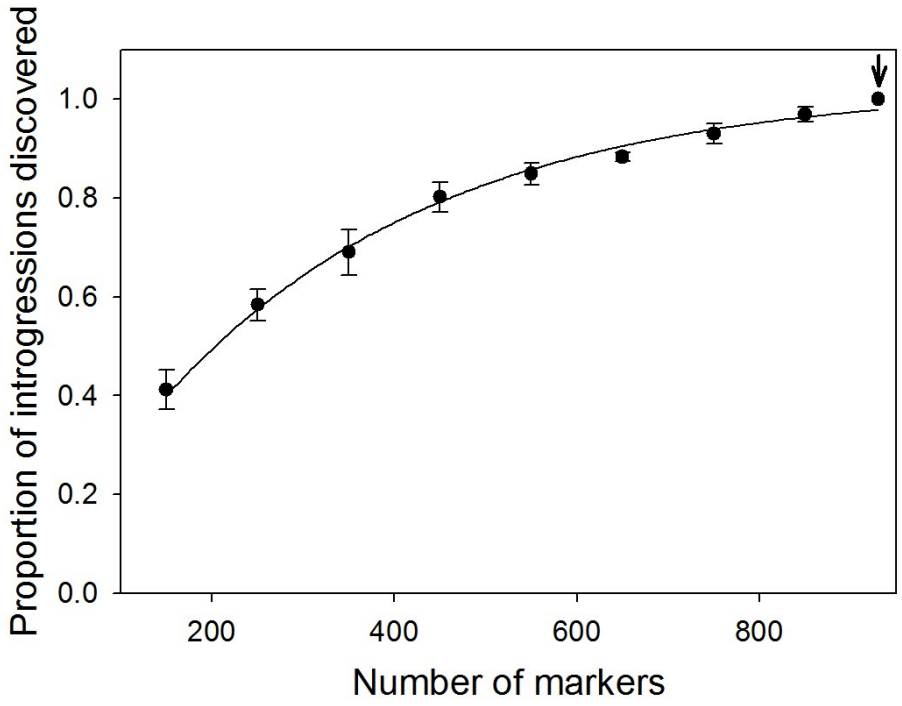


Figure 2.4 Heterozygous introgressions discovered at varying marker densities.

Estimates are based on 10 repetitions of markers selected at random from the dense map for each marker density (mean ± SD). The line is fitted based on an exponential rise function. The data point marked by an arrow represents the final estimates generated from the dense map. The plot of homozygous introgression discovery was nearly identical and excluded to simplify the figure.

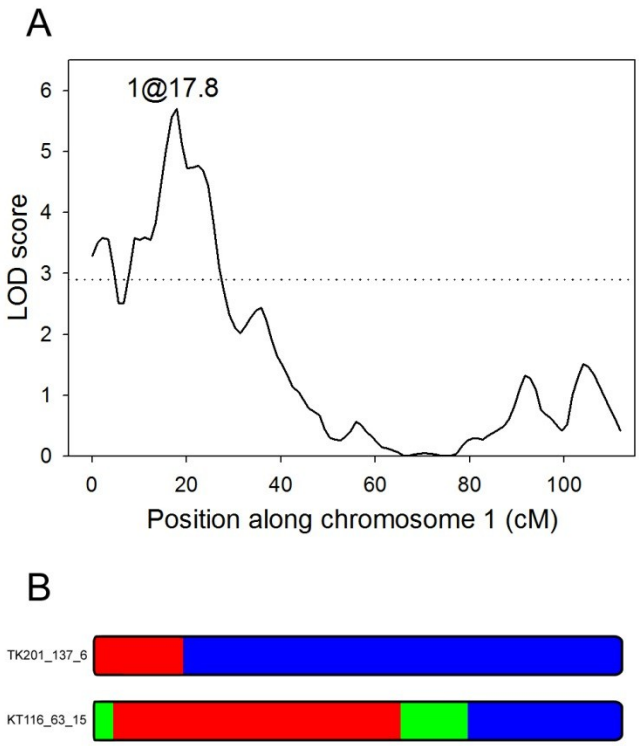


Figure 2.5 QTL location and graphical genotypes of the NILs used in the QTL validation.

(A) Localization of the dark-conductance QTL along chromosome 1, the dotted line indicates the threshold LOD score. (B) Graphical genotypes scaled to represent the genetic distance (cM) of the x-axis of panel A of chromosome 1 for the NILs used in the QTL validation.

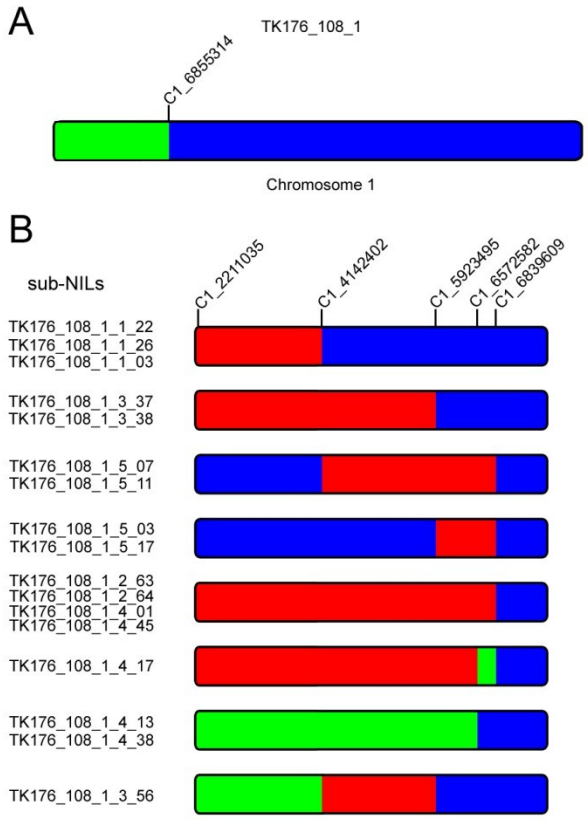


Figure 2.6 Diagram of genotype information for fine mapping lines.

Heterozygous NIL selected for selfing (top) and the detailed focal region of the selected sub-NILs (bottom). The numbers at the bottom of the sub-NILs indicate the physical position of markers and, therefore, represent estimated introgression boundaries. Red, Kas-1; Blue, Tsu-1; Green, heterozygous.

REFERENCES

Acharya BR, Assmann SM. 2009. Hormone interactions in stomatal function. Plant Molecular Biology **69**, 451-462.

Alonso-Blanco C, Peeters AJM, Koornneef M, Lister C, Dean C, van den Bosch N, Pot J, Kuiper MTR. 1998. Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. *The Plant Journal* **14**, 259–271.

Bauerle WL, Bowden JD. 2011. Separating foliar physiology from morphology reveals the relative roles of vertically structured transpiration factors within red maple crowns and limitations of larger scale models. *Journal of Experimental Botany* **62**, 4295-4307.

Bell C J, Ecker J. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**, 137–144.

Boivin K, Acarkan A, Mbulu R, Clarenz O, Schmidt R. 2004. The *Arabidopsis* genome sequence as a tool for genome sequence analysis in *Brassica*ceae. A comparison of the *Arabidopsis* and *Capsella rubella* genomes. *Plant Physiology* **135**, 735-744.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889-890.

Broman KW, Sen S. 2009. **A Guide to QTL Mapping with R/qtl**. New York: Springer 2009.

Caird MA, Richards JH, Donovan LA. 2007. Nighttime stomatal conductance and transpiration in C_3 and C_4 plants. *Plant Physiology* **143**, 4-10.

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, the AmiGO Hub, the Web Presence Working Group. S009. AmiGO: online access to ontology and annotation data. [http://bioinformatics.oxfordjournals.org/cgi/content/abstract/25/2/288]

Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M. 1998. Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* **41**, 62-69.

Christman MA, Richards JH, McKay JK, Stahl EA, Juenger TE, Donovan LA. 2008. Genetic variation in *Arabidopsis thaliana* for night-time leaf conductance. *Plant, Cell &Environment* **31**, 1170-1178.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research* 15, 1496–1502.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Dinka SJ, Campbell MA, Demers T, Raizada MN. 2007. Predicting the size of the progeny mapping population required to positionally clone a gene. *Genetics* **176**, 2035-2054.

Ducrocq S, Giauffret C, Madur D, Combes V, Dumas F, Jouanne S, Coubriche D, Jamin P, Moreau L, Charcosset A. 2009. Fine mapping and haplotype structure analysis of a major flowering time quantitative trait locus on maize chromosome 10. *Genetics* **183**, 1555-1563.

Esch E, Szymaniak JM, Yates H, Pawlowski WP, Buckler ES. 2007. Using crossover breakpoints in recombinant inbred lines to identify quantitative trait loci controlling the global recombination frequency. *Genetics* **177**, 1851-1858.

Eshed Y, Zamir D. 1994. A genomic library of *Lycopersicon pennellii* in *L. esculentum*: A tool for fine mapping of genes. *Euphytica* **79**, 175-179.

Eshed Y, Zamir D. 1995. An introgression line population of **Lycopersicon pennellii** in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, 1147-1162.

Frary A, Nesbitt TC, Grandillo S, van der Knaap E, Cong B, Liu JP, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88.

Gerald JNF, Lehti-Shiu MD, Ingram PA, Deak KI, Biesiada T, Malamy JE. 2006. Identification of quantitative trait loci that regulate *Arabidopsis* root system size and plasticity. *Genetics* **172**, 485–498.

Haley CS, Knott SA. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Iniguez-Luy FL, Lukens L, Farnham MW, Amasino RM, Osborn TC. 2009. Development of public immortal mapping populations, molecular markers and linkage maps for rapid cycling *Brassica rapa* and *B. Oleracea. Theoretical and Applied Genetics* **120**, 31-43.

Jansen RC. 1993. Interval mapping of multiple quantitative trait loci. *Genetics* **1**, 205-211.

Juenger TE, Wayne T, Boles S, Symonds VV, McKay J, Coughlan SJ. 2005a. Natural genetic variation in whole-genome expression in *Arabidopsis thaliana*: the impact of physiological QTL introgression. *Molecular Ecology* **15**, 1351–1365.

Juenger TE, Mckay JK, Hausmann N, Keurentjes JJB, Sen S, Stowe KA, Dawson TE, Simms EL, Richards JH. 2005b. Identification and characterization of QTL underlying whole-plant physiology in *Arabidopsis thaliana*: $\delta^{13}C$, stomatal conductance and transpiration efficiency. *Plant, Cell & Environment* **28**, 697–708.

Juenger TE, Sen S, Bray E, Stahl E, Wayne T, McKay J, Richards JH. 2010. Exploring genetic and expression differences between physiologically extreme ecotypes: comparative genomic hybridization and gene expression studies of Kas-1 and Tsu-1 accessions of *Arabidopsis thaliana*. *Plant, Cell & Environmen*t **33**, 1268-1284.

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, EffgenS, Vreugdenhil D, Koornneef M. 2007. Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* **175**, 891-905.

Koebener RMD, Varshney RK. 2006. Development and application of genomic models for large-crop plant genomes. In: Varshney RK, Koebner RMD eds. *Model Plants and Crop Improvement*. Boca Raton: CRC Press, 1-9.

Koornneef M, Meinke D. 2010. The development of *Arabidopsis* as a model plant. *The Plant Journal* **61**, 909-921.

Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, Pooni HS, Newbury HJ, Kearsey MJ. 2002. STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. *The Plant Journal* **31**, 355-364.

Lamesch P, Berardini TZ, Li D, *et al.* 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40,** D1201-D1210.

Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-199.

Landi P, Sanguineti MC, Salvi S, Giuliani S, Bellotti M, Maccaferri M, Conti S, Tuberosa R. 2005. Validation and characterization of a major QTL affecting leaf ABA concentration in maize. *Molecular Breeding* **15,** 291–303.

Laserna MP, Sanchez RA, Botto JF. 2008. Light-related loci controlling seed germination in Ler × Cvi and Bay-0 × Sha Recombinant Inbred-line Populations of *Arabidopsis thaliana*. *Annals of Botany* **102**, 631-642.

Li P, Wind JJ, Shi X, Zhang H, Hanson J, Smeekens SC, Teng S. 2011. Fructose sensitivity is suppressed in *Arabidopsis* by the transcription factor ANAC089 lacking the membrane-bound domain. *Proceedings of the National Academy of Sciences, USA* **108**, 3436-3441.

Lin H, Liang ZW, Sasaki T, YanoM. 2003. Fine mapping and characterization of quantitative trait loci Hd4 and Hd5 controlling heading date in rice. *Breeding Science* **53**, 51-59.

Loudet O, Gaudon V, Trubuil, A, Daniel-Vedele F. 2005. Quantitative trait loci controlling root growth and architecture in *Arabidopsis thaliana* confirmed by heterogeneous inbred family. *Theoretical and Applied Genetics* **110**, 742-753.

Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related *Brassica*ceae species. *Proceedings of the National Academy of Sciences, USA* **103**, 5224-5229.

Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW. 2009. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics* **181**, 1077-1086.

McKay JK, Richards JH, Mitchell-Olds T. 2003. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Molecular Ecology* **12**, 1137-1151.

McKay JK, Richards JH, Nemali KS, Sen S, Mitchell-Olds T, Boles S, Stahl EA, Wayne T, Juenger TE. 2008. Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population, Kas-1 x Tsu-1. *Evolution* **62-12,** 3014-3026.

Monforte AJ, Tanksley SD. 2000a. Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery. *Genome* **43**, 803–813.

Monforte AJ, Tanksley SD. 2000b. Fine mapping of a quantitative trait locus .QTL. from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. *Theoretical and Applied Genetics* **100**, 471-479.

Nilson SE, Assmann SM. 2007. The control of transpiration. Insights from *Arabidopsis*. *Plant Physiology* **143**, 19-27.

Paterson AH, DeVerna JW, Lanini B, Tanksley SD. 1990. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. *Genetics* **124**, 735-742.

Price AH. 2006. Believe it or not, QTLs are accurate!. *Trends in Plant Science* **11**, 213-216.

Reif JC, Kusterer B, Piepho HP, Meyer RC, Altmann T, Schon CC, Melchinger AE. 2009. Unraveling epistasis with triple testcross progenies of near-isogenic lines. *Genetics* **181**, 247-257.

Rockman MV. 2008. Reverse engineering the genotype–phenotype map with natural genetic variation. *Nature* **456**, 738-744.

Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, Brudno M. 2009. SHRiMP: Accurate mapping of short color-space reads. *PLoS Computational Biology* **5**, e1000386.

Sax K.1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**, 552-560.

Schranz ME, Windsor AJ, Song B, Lawton-Rauh A, Mitchell-Olds T. 2007. Comparative Genetic Mapping in *Boechera stricta*, a Close Relative of *Arabidopsis*. *Plant Physiology* **144**, 286-298.

Schroeder JI, Allen GJ, Hugouvieux V, Kwak JM, Waner D. 2001. Guard cell signal transduction. *Annual Review of Plant Physiology* **52**, 627-658.

Sen S, Churchill GA. 2001. A statistical framework for quantitative trait mapping. *Genetics* **159**, 371-387.

Soller M, Brody T. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* **47**, 35-39.

Stacey G, VandenBosch K. 2005. ‘‘Translational’’ legume biology. Models to crops. *Plant Physiology* **137**, 1173.

Symonds VV, Lloyd AM. 2003. An analysis of microsatellite loci in *Arabidopsis thaliana*: mutational dynamics and application. *Genetics* **165**, 1475–1488.

Takahashi Y, Shomura A, Sasaki T, Yano M. 2001. Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proceedings of the National Academy of Sciences, USA* **98**, 7922–7927.

Torjek O, Berger D, Meyer RC, Mussig C, Schmid KJ, Sorensen TR, Weisshaar B, Mitchell-Olds T, Altmann T. 2003. Establishment of a high-efficiency SNP-based framework marker set for *Arabidopsis*. *The Plant Journal* **36**, 122-140.

Torjek O, Meyer RC, Zehnsdorf M, Teltow M, Strompen G, Witucka-Wall H, Blacha A, Altmann T. 2008. Construction and analysis of 2 reciprocal introgression line populations. *Journal of Heredity* **99**, 396-406.

Tuinstra MR, Ejeta G, Goldsbrough P*B.* 1997. Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theoretical and Applied Genetics* **5**, 1005-1011.

Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Qian Q, Zhang G, Fu X. 2012. Control of grain size, shape and quality by OsSPL16 in rice. *Nature Genetics* **44**, 950-955.

Wang S, Meyer E, McKay JK, Matz MV. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods* **9**, 808-810.

Waters BM, Grusak MA. 2008. Quantitative trait locus mapping for seed mineral concentrations in two *Arabidopsis thaliana* recombinant inbred populations. *New Phytologist* **179**, 1033-1047.

Xing YZ, Tang WJ, Xue WY, Xu CG, Zhang Q. 2008. Fine mapping of a major quantitative trait loci, qSSP7, controlling the number of spikelets per panicle as a single Mendelian factor in rice. *Theoretical and Applied Genetics* **116**, 789-796.

Yan WH, Wang P, Chen HX, Zhou HJ, Li QP, Wang CR, Ding ZH, Zhang YS, Yu SB, Xing YZ, Zhang QF. 2011. A major QTL, Ghd8, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Molecular Plant* **4**, 319-330.

Yogeeswaran K, Frary A, York TL, Amenta A, Lesser AH, Nasrallah JB, Tanksley SD, Nasrallah ME. 2005. Comparative genome analyses of *Arabidopsis* spp.: Inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Research* **15**, 505-515.

Young ND, Tanksley SD. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theoretical and Applied Genetics* **77**, 95-101.

Zeng ZB. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences, USA* **23**, 10972-10976.

Zeng ZB. 1994. Precision mapping of quantitative trait loci. *Genetics* **4**, 1457-1468.

Zhou L, Zeng Y, Zheng W, Tang B, Yang S, Zhang H, Li J, Li Z. 2010. Fine mapping a QTL qCTB7 for cold tolerance at the booting stage on rice chromosome 7 using a near-isogenic line. *Theoretical and Applied Genetics* **121**, 895-905.

CHAPTER 3: QTL ANALYSIS OF ROOT SYSTEM MORPHOLOGY, FLOWERING TIME AND YIELD AND THEIR PLASTICITY TO DROUGHT IN *BRASSICA NAPUS*

SUMMARY

Drought escape and dehydration avoidance represent alternative strategies for drought adaptation in annual crops. The mechanisms underlying these two strategies are reported to have a negative correlation, suggesting a tradeoff. We conducted a quantitative trait locus (QTL) analysis of root mass and flowering time, traits representing each strategy, in *Brassica napus* to understand if a tradeoff exists and what the genetic basis might be. Our field experiment used a genotyped population of doubled haploid lines and included both irrigated and rainfed treatments, allowing analysis of plasticity in each trait. We found strong genetic correlations among all traits, suggesting a tradeoff among traits may exist. Summing across traits and treatments we found 20 QTL, but many of these co-localized to two major QTL, providing evidence that the tradeoff is genetically constrained. To understand the mechanistic relationship between root mass, flowering time and QTL, we analyzed the data by conditioning upon correlated traits. Our results suggest a causal model where such QTL affect root mass directly as well as through their impacts on flowering time. Additionally, we used draft *Brassica* genomes to identify orthologs of flowering time gene *FLC* as candidate genes. This research provides valuable clues to breeding for drought adaptation as it is the first to analyze the inheritance of the root system in *B. napus* in relation to drought. It also suggests that the role of *FLC* in plant development should be expanded to include roots.

INTRODUCTION

Nearly all aspects of terrestrial plant form and function depend upon adequate water availability. As a result, drought is the most common cause for reductions in crop yields, frequently causing reductions well below half of the crop's theoretical yield potential (Boyer, 1982). A variety of mechanisms have been associated with drought acclimation (plasticity) and adaptation (heritable

differences in traits) leading to the proposal of three distinct coping strategies (Ludlow, 1989): drought escape, dehydration avoidance, and dehydration tolerance. We focused on drought escape and dehydration avoidance, as dehydration tolerance is not prevalent in vascular plants, especially crops (Oliver, Cushman and Koster, 2010). The most common strategy exploited in crop breeding is drought escape, which refers to plants that complete their life cycle prior to the onset of drought, thus avoiding moisture limitations. The alternative strategy, dehydration avoidance, is the sustaining of internal water status during dry external conditions by minimizing water loss and/or maximizing water uptake.

Resource limitation creates a necessity for organisms to allocate energy to processes in a competitive manner such that relationships among processes are constrained (Levins, 1968; Obeso, 2002). In plants there is a major energetic tradeoff between investments in vegetative growth and reproduction, which can also be thought of as a life history tradeoff (Reznick, 1985). Many studies have reported a tradeoff between drought escape and mechanisms of dehydration avoidance, such as water use efficiency and root size (Mitchell-Olds, 1996; McKay *et al*., 2003; Wu *et al*., 2010; Franks, 2011). However, results reporting the absence of such a tradeoff (Sherrard and Maherali, 200) suggest that more research is needed to understand the generality of this hypothesized constraint.

Tradeoffs can be quantified as genetic correlation coefficients, which measure the degree to which genetic variation in one trait predicts variation in the other (Robertson, 1959). Genetic correlations among traits can impose significant constraints on the efficacy and response to selection (both natural and artificial). This is because the adaptive optimum of trait values may be orthogonal to the vector of trait covariation. Genetically correlated traits are mechanistically the result of either genetic linkage or pleiotropy (Wagner and Zhang, 2011). In the case of genetic linkage (Figure 3.1A), polymorphisms underlying variation at each trait are at different loci but are nearby physically, limiting recombination so that the trait value caused by the allele at one locus covaries with the trait value of the allele at the linked locus. Pleiotropy, on the other hand, refers to the effect an allele has on two or more phenotypes (Figure 3.1B). Finally, genetic correlations may be due to physiological interactions among traits where one trait

acts 'upstream' of another (Figure 3.1C and 3.1D). Ultimately, genetic correlations due to pleiotropy constrain the response to selection far more than those due to genetic linkage (Futuyama, 1998).

In crops, drought escape is often achieved through breeding by optimizing flowering time. Flowering time marks the transition from vegetative to reproductive growth and its influence on fitness and yield can be dramatic, making it perhaps the most important of all life history traits and the focus of extensive research in both crops and natural plant populations (reviewed in Michaels, 2009; Pose *et al*., 2012). The impact of flowering time on fitness may be due in large part to its many correlations with other diverse and potentially adaptive traits such as vegetative biomass (Shi *et al*., 2009; Edwards *et al*., 2012), vascular system development (Sibout *et al*., 2008), oxidative stress (Kurepa *et al*., 1998), water-use efficiency (McKay *et al*., 2003; Franks, 2011) and a variety of root characteristics (Bolaños and Edmeades, 1993; Mitchell-Olds, 1996; Lou *et al*., 2007). A study comparing isolines carrying mutant alleles in five loci annotated as "flowering time" genes showed significant differences in morphological traits such as leaf length, leaf number and auxillary shoot number, providing further evidence of pervasive pleiotropy at loci involved in flowering (van Tienderen *et al*., 1996). The recurring association between flowering time and other traits is perhaps not surprising, since variation at any genes related to environmental sensing, resource acquisition or resource allocation are also likely to lead to variation in flowering time (McKay *et al*., 2008).

Drought avoidance is less well characterized, but mechanisms include reduced stomatal conductance and increased water uptake by roots. The root system has been long recognized as a central component of crop productivity (Sharp and Davies, 1979). This is due to the role of roots in water and nutrient acquisition, anchorage, mechanical support and, perhaps most importantly, sensing and responding to the complex and often heterogeneous soil environment. A dehydration avoidance strategy through maximization of water uptake clearly involves the root system, making it a focal subject of breeding for low rainfall environments (Ludlow and Muchow, 1990). Associations between drought adaptation and increased root system size and/or rooting depth have been drawn across many species (Ekanayake, 1986; Cortes and Sinclair, 1986; White and Castillo, 1989; Johnson, 2000; Price *et al*., 2001;

Kirkegaard and Lilley, 2007; Lopes and Reynolds, 2010). However, the adaptive value of large or deep root systems varies by geography so that an applied breeding strategy must consider the climatic trends of the target production zone (Araus, 2002; Cativelli, 2008; Palta *et al*., 2011). Selection for root traits is hindered by generally low heritabilities and the difficulty of phenotyping large populations (Wasson *et al*., 2012; Topp *et al*., 2013). Root traits remain a relatively unexploited breeding target, but additional insight into the genetic architecture of root system variation will be necessary for engineering 'designer' root systems to meet the world's growing demand for food, fuel and fiber (Gregory *et al*., 2013).

In this study, we investigated variation and covariation in drought escape and avoidance traits in *Brassica napus*. Of the *Brassica* oilseed crops, *B. napus* is the most important and trails only soybean and oil palm in terms of global production (http://www.fas.usda.gov/oilseeds/Current/; 2013). Changing climate conditions and expansion into new production geographies are increasing the exposure of *B. napus* crop production to drought stress. Yet, little research has focused on its root system, especially regarding inheritance, genetics or relationship to drought.

We utilized a quantitative trait locus (QTL) mapping approach to better understand the genetic basis of root traits, flowering time and their impacts on grain yield in this global crop species. The QTL method is ideal for elucidating loci underlying trait correlations and, by including drought as an experimental treatment, loci associated with drought strategy tradeoffs and yield sensitivity. We focused on a doubled haploid (DH) population of 225 lines derived from a cross between IMC106RR, an annual cultivar, and Wichita, a biennial cultivar, to maximize genetic and phenotypic diversity. The vernalization requirement differentiating annual (spring) and biennial (winter) lines also defines genetic and morphologically distinct pools (Diers and Osborn, 1994; Lühs *et al*., 2003). We measured root pulling force (RPF, the vertical force required to remove a plant from the soil; Hayes and Johnson, 1939) as an index of root system size. Specifically, we asked whether a tradeoff exists for flowering time and RPF and, if so, which loci are involved. Further, we asked what specific root component RPF was measuring which, in a more focused study, we determined to be acting as a proxy for taproot mass.

MATERIALS AND METHODS

Plant Materials

This study utilized a DH population of 225 lines named SE1 that was produced from an F1 generation microspore donor plant derived from a cross made between the annual variety IMC106RR (Cargill Inc., National Registration No. 5118), and the biennial variety Wichita (Rife *et al*., 2001; Reg. no. CV-19, PI 612846) at Cargill (Fort Collins, CO) using the method of Palmer *et al*. (1996). The resulting population segregated for the requirement of vernalization to initiate flowering and consisted of approximately 1200 lines. From this, about 900 lines flowered in the greenhouse and, thus, were deemed to be annual growth habit. Of these 900 DH lines, we randomly selected 225 for use in this experiment.

Genotyping and Mapping

Genotyping was done using the Illumina (San Diego, CA) *Brassica* 60K Infinium array at DNA Landmarks (Quebec, Canada). The final list of 1,179 markers used in linkage map construction was selected based on GenTrain genotype scores above 0.75 as suggested by Illumina followed by selection for those which lack an inter-homoeologous polymorphism (Trick *et al*., 2009). The genetic linkage map was constructed in JoinMap3 (Van Ooijen *et al*., 2001) using a threshold recombination frequency of <0.25 and a minimum logarithm of the odds ratio (LOD) score of 6 for grouping loci into linkage groups. The Kosambi function (Kosambi, 1944) was used to calculate genetic distances. Each linkage group was named based on the nomenclature recommended by the Multinational *Brassica* Genome Project steering committee (http://www.*Brassica*.info/ resource/maps/lg-assignments.php). The map was analyzed further in the R/qtl program of the R statistical package (Broman *et al*., 2003; Broman and Sen, 2009) to confirm marker orders and assess general map quality.

Field Design

The DH lines and parents were planted at Colorado State University's Agricultural Research Education and Demonstration Center (40.66˚N/105˚W.) near Fort Collins, CO on 19 April, 2010. The

study was arranged in a Row-Column design (created with CycDesigN 3.0, www.cycdesign.co.nz) with three replicates per treatment. Plots were comprised of two rows separated by 0.23 m and 1 m in length. Plots were separated by a distance of 0.33 m and thinned to approximately 10 plants per plot. Irrigation was applied using a linear-move system at approximately 2.5 cm per week for the first month of development at which point it was discontinued in the rainfed (dry) treatment. Irrigation was maintained at the rate of 2.5 cm per week for the duration of the experiment in the irrigated (wet) treatment.

Phenotyping

DTF was recorded for each plot as the interval from sowing date to the date on which 50% of the plants in the plot had initiated flowering. To measure yield, plots were swathed by hand, allowed to dry, threshed using a Wintersteiger (Wintersteiger AG, Austria) combine harvester and weighed immediately thereafter. Yield sensitivity was calculated as the difference between the yield of a DH line in the wet environment and its yield in the dry environment: Yield(wet) – Yield(dry). Relative yield was calculated as a Z score to account for the large differences in mean yield and standard deviation between the wet and dry treatments. The RPF method designed and used extensively in maize (Hayes and Johnson, 1939; Spencer, 1940; Rogers *et al*., 1976; Lebreton *et al*., 1995) was modified in a manner suitable for *B. napus*. One end of a nylon rope was lassoed around the base of a plant and the other end was attached to a hand-held Imada DS2 dynamometer (Imada Inc., Northbrook, IL). The dynamometer was pulled vertically until the root system came out of the soil and the maximum force (Kgf) was recorded. RPF was measured within 1-2 days after grain was harvested. To optimize the phenotype, the field was watered 24 hours prior to measuring RPF.

Quantitative Genetic Analyses

A linear mixed model was used to analyze the data using the PROC MIXED procedure in the SAS software package (SAS Institute, 2004) with DH line treated as a fixed effect and row and/or column treated as random effects. The broad-sense heritability (H^2) for each trait was estimated using variance

components computed in the PROC VARCOMP procedure in SAS as the ratio V_G: (V_G + V_E), where V_G is the variance among DH lines and VE is the residual variance. Among-trait phenotypic correlations were computed as Pearson correlation coefficients using data points collected from individual plots and genetic correlations were computed from least square means estimated for each DH line within each environment (Falconer and Mackay, 1996).

QTL mapping was performed using Haley-Knott Regression (Haley and Knott, 1992) in R/qtl using 1 cM steps. QTL were selected using a step-wise model selection approach (Manichaikul *et al*., 2009) based on significance thresholds made from 1000 permutations (Churchill and Doerge, 1994). Genome-wide scans for QTL by environment interactions were conducted by comparing a model including the environment (moisture treatment) as a covariate along with a QTL-environment interaction to a model lacking the interaction. LOD 1.5 confidence intervals were determined in the R/qtl software package (Broman *et al*., 2003; Broman and Sen, 2009). QTL were named using the trait and treatment with which they were associated with ascending numbers based on linkage group location.

QTL confirmation in the field

In an effort to validate the effect of a QTL discovered in this study (*RPF.dry1*) and to gain a better understanding of the mechanisms underlying RPF, a second study was performed during the summer of 2012. It focused on a total of 40 lines of which each parental haplotype at *RPF.dry1* was represented by 20 lines. Haplotypes were defined by DH lines which carried marker alleles spanning the LOD1.5 confidence interval. The physical position of this interval was determined by comparing SNP sequence information with the *B. rapa* reference (Wang *et al*., 2011; Cheng *et al*., 2011) using BLAST (Altschul *et al*., 1990). The experiment was conducted at the same experimental farm in which each experimental unit consisted of a single plant per DH line watered by a precision drip irrigation emitter. Three randomized blocks were planted and thinned to a single plant 2 weeks after germination. All other factors of the experiment were conducted as they were in 2010. In addition to collecting data on RPF and DTF, aboveground biomass was weighed for each plant in the field as they were extracted and SFW was

also recorded. Root mass extracted during RPF measurement was oven dried at 80°C for 3 days prior to measurement of the taproot dry mass, lateral root dry mass, total root dry mass, tap root diameter (diameter of the basal portion of the dried taproot), tap root length (length of the extracted taproot), branching zone length (length of the taproot with primary laterals) and the number of coarse secondary roots (total number of secondary roots >1 mm in diameter).

RESULTS

The nineteen chromosomes of Brassica napus are recovered in the genetic map

The genetic map recovered nineteen linkage groups which represent the ten chromosomes of the A genome (*B. rapa*; 2n=20) and the nine chromosomes of the C genome (*B. oleracea*; 2n=18) which comprise the allopolyploid genome of *Brassica napus*. The map was constructed using 1179 markers and resulted in a total length of 2041 cM, had an average intermarker distance of 1.8 cM and carried one large gap of 46 cM located on A01 (Supplementary Figure 5.3). On average, segregation in the population met the expected 1:1 ratio (48.4% Wichita allele, 51.6% IMC106RR allele). Several regions showed segregation distortion in favor of the alleles from the Wichita parent, including most of linkage groups A01 and A08 with a maximum biases of 66.2% (chi-square = 24.4) and 66.1% (chi-square=24.1), respectively. This segregation in favor of alleles from the winter parent occurred despite selection for lines lacking vernalization requirement and is likely the product of gametic selection during the microspore culture process (Ferrie and Möllers, 2009). Other regions of lower segregation distortion were also observed on A06 and C03 in favor of the IMC106RR allele and on C01 in favor of the Wichita allele. This minor amount of segregation distortion did not have a noticeable impact on map construction or QTL analysis (Hackett and Broadfoot, 2003).

RPF, DTF and yield demonstrate strong genetic correlations

A significant treatment effect (Supplementary Table 5.4) and heritable variation for root pulling force (RPF), days to flower (DTF) and yield was observed. Much of this variation was attributable to

genetics with estimates of heritability ranging from 0.16 for RPF in the dry treatment to 0.83 for DTF in the wet (Table 3.1). In agreement with previous research (Udall *et al*., 2006; Shi *et al*., 2009), flowering time had the highest heritability estimates in both treatments. Despite selecting initially against vernalization, some lines didn't flower in the dry treatment resulting in a right-censored distribution (Leung *et al*., 1997). Those DH lines that flowered did so in a minimum of 59 days in both treatments and a maximum of 101 days in the dry treatment and 108 days in the wet treatment. RPF and yield demonstrated transgressive inheritance in the DH population in both treatments, a parameter that could only be measured relative to IMC106RR for DTF and yield since Wichita, a winter growth-habit line, neither flowered nor set seed during the field season.

Correlations were highly significant ($P < 0.0001$) among all traits (Figure 3.2). Yield had strong negative genetic correlations with both DTF and RPF under both treatments. Positive genetic and phenotypic correlations were observed between RPF and DTF so that late flowering lines required a larger force for removal (Supplementary Table 5.5).

QTL analysis identifies two major pleiotropic factors underlying the tradeoff between drought escape and avoidance

QTL analyses for DTF, RPF, yield and yield sensitivity in each environment were performed. Seven QTL for DTF were mapped; four in the wet and three in the dry treatments. For RPF, three QTL were identified in the wet environment and two in the dry. Analyses of yield found three QTL for each environment. Summing across all four traits, a total of 20 additive QTL were discovered (Figure 3.3). QTL co-localized to regions on linkage groups A03, A10 and C02 (Figure 3.3), thus implicating tight linkage or pleiotropy as the cause of the strong genetic correlations observed among traits. In particular, two regions on A10 and C02 (bracketed in red in Figure 3.3) explained a large proportion of the variation for each trait and their estimated effects were always larger than other QTL discovered for any particular trait. All of the QTL discovered for yield co-localized with QTL for DTF, further supporting the strong relationship between these traits where a flowering time of ~68 days increases the probability of higher

yield (Figure 3.2). These results also show that, unlike yield, the genetics underlying DTF and RPF did not overlap entirely. For example, *DTF.wet1* and *DTF.wet2* had no relationship to RPF where *RPF.wet3*, located on C07, had no relationship to DTF.

Analysis of variance showed the genotype by environment interaction to be significant for flowering time and yield (Supplementary Table 5.4) but significant QTL by environment interactions were only found for yield on chromosomes A10 and C02. The impact of the QTL on A10 and C02 in response to treatment was further supported by QTL which mapped to these two chromosomal locations for yield sensitivity (Figure 3.3; a detailed summary of all QTL results is provided in Supplementary Table 5.6). A closer examination of the allele effects at each locus shows that late flowering QTL alleles have a larger impact in the dry treatment than the wet (Figure 3.4).

Examination of the relationship between RPF and DTF using conditional QTL models shows evidence that RPF.dry1 may be acting directly on both traits

To better understand the genetic architecture between the traits, the data were analyzed with QTL models which conditioned upon flowering time, the hypothetically 'upstream' trait. The goal of this additional analysis was to infer the causal relationships among traits which share QTL (Li *et al*., 2006; Broman and Sen, 2009). More specifically, the objective was to elucidate whether a particular QTL is affecting a trait directly (Figure 3.1B), as a downstream effect of delayed flowering (Figure 3.1C), or a combination of both (Figure 3.1D).

All of the QTL for yield under both treatments disappeared when DTF was included as a covariate in the QTL scan (data not shown). This supports the intuitive notion that DTF is an upstream determinant of yield and the co-localizing QTL act indirectly on yield via their effects on flowering time.

Conditional genome-wide scans for RPF in the wet treatment identified a new QTL on A08 (*RPF.wet4*). Interestingly, the high RPF allele at *RPF.wet4* which did not affect flowering time originated from the spring parent, IMC106RR. Another QTL was mapped to the same location on C07 as *RPF.wet3*, identified previously in the unconditional scan (Figure 3.5A). These findings further support a genetic

basis to RPF that does not entirely overlap with that of DTF. The stepwise model selection used included the DTF term (1.09 Kg ± 0.08) which resulted in the disappearance of *RPF.wet1* on A10 and *RPF.wet2* on C02. This suggests that the QTL at these two loci may have been affecting RPF as a downstream result of their effect on DTF (Figure 3.1C) in this environment.

The conditional QTL scans of the dry treatment yielded a model with a single QTL on A10 (Figure 3.5B) co-localizing in the same location on A10 as *RPF.dry1* (0.50 Kg ± 0.09), along with a DTF effect. The significant impact of this QTL and the DTF covariate term in the model support a mode of causality similar to that of Figure 3.1D where the QTL affects RPF directly as well as indirectly through its impact on flowering time.

Single marker analysis of RPF using models conditional on DTF strata account for a right-censored distribution and further support the direct role of RPF.dry1 on both traits

In the dry treatment, 22 DH lines were censored (omitted) from the flowering time distribution because they did not flower and, therefore, had no observed flowering time to use as a covariate in the conditional QTL analysis. To account for these missing data, the population was stratified into five classes of approximately 45 lines based on their flowering times (Supplementary Table 5.7). Single marker analyses of RPF, conditional on DTF strata, was then performed at each of the QTL identified previously.

The stratification factor was highly significant ($P < 0.0001$) in all analyses, further supporting the strong effect of flowering time on RPF. In the dry treatment, only *RPF.dry1* (A10) remained significant as the estimated difference in the mean allele value changed only slightly between the conditional and unconditional analyses (Table 3.2). In contrast, *RPF.dry2* (C02) became insignificant in the conditional analysis despite the major difference in mean allele values estimated during the unconditional examination (Table 3.2). Analyses of the wet treatment provided further support for the presence of *RPF.wet3* (C07) and *RPF.wet4* (A08) and produced an insignificant result for *RPF.wet2*. *RPF.wet1* remained significant in the conditional model suggesting its effect may be constitutive across treatments.

To further illustrate that the alleles at *RPF.dry1* affect roots independently of flowering time, mean RPF values were plotted as a function of DTF strata where it is demonstrated that RPF values are higher for the Wichita allele across any of the five DTF strata than they are for the IMC106RR allele (Figure 3.6).

The effect of RPF.dry1 is validated in a second field experiment and determined to be acting on taproot size

To validate the effects of *RPF.dry1*, 20 lines representing each parental haplotype at the QTL were selected and the experiment was repeated. The haplotype was defined by the interval spanning the length of the LOD 1.5 confidence interval, a region encompassing a minimum of 1.0 Mb (physical positions 13,498,846 to 14,558,300) as estimated by the physical locations of the flanking markers (Cheng *et al.*, 2012) relative to the *B. rapa* reference genome V1.5 (Wang *et al.*, 2011; Cheng *et al.*, 2011).

The pleiotropic effect of *RPF.dry1* was confirmed, as lines carrying the Wichita haplotype flowered an average of 12 days later ($P < 0.0001$) and required nearly 17 Kgf more force to remove the roots ($P < 0.0001$) than lines carrying the IMC106RR haplotype. Conditional analyses accounting for DTF estimated a significant haplotype effect ($P < 0.05$), confirming that the effect of genotype on RPF at this locus is significant even after accounting for DTF.

In this experiment, the root system was harvested after RPF measurement and analyzed in an effort to gain a better understanding of the root qualities measured by RPF. RPF was most highly correlated with total root drymass but had significant correlations with DTF, shoot fresh weight, taproot drymass, lateral drymass, taproot diameter, taproot length and branching zone length (Table 3.3). No significant correlation was found between RPF and the number of coarse secondary roots. Analyses of the specific root components found the effect of haplotype was significant for all traits except lateral root dry mass, branching zone length and the number of coarse secondary roots.

Examination of the correlation matrix reveals significant and generally strong correlations between shoot fresh weight and all measured root traits except the number of coarse secondary roots.

Since aboveground biomass is expected to have a significant association with belowground biomass, the data were re-analyzed using models incorporating shoot fresh weight and DTF as covariates to further investigate the relationship between genotype and the measured root traits while accounting for these correlated and potentially confounding factors. We found that only taproot dry mass was significant in models conditioning on shoot fresh weight as well as those incorporating both shoot fresh weight and DTF as covariates (Table 3.4). It is remarkable that any trait remained significant after conditioning on two correlated traits; this suggests that the specific root trait which this locus is acting upon may be taproot size, as lines with the IMC106RR allele had an average taproot mass 74% as large as those with the Wichita allele. Thus, evidence for the direct effect of this QTL on taproot size provides a more detailed understanding of the genetics and specific root characteristics underlying the observed DTF:RPF correlation and the inferred tradeoffs between adaptive drought strategies.

DISCUSSION

Strong genetic correlations and conditional QTL models indicate that the tradeoff between drought escape and avoidance may be due to pleiotropy

The strong correlations observed among root traits, flowering time and yield in this study are concordant with previous research (Bolaños and Edmeades, 1993; Mitchell-Olds, 1996; Lou *et al*., 2007; Shi *et al*., 2009). Our QTL results provide first steps toward understanding the common and independent genomic regions contributing to variation in each of these traits, thus providing a better understanding of their inheritance and the genetic architecture of their covariance. Further, these results suggest a tradeoff between drought escape and avoidance strategies as there was a significant difference in yield between early flowering lines with small root systems and late flowering lines with larger root systems (Figure 3.2).

Negative correlations between drought escape and dehydration avoidance, where a trait conferring higher dehydration avoidance such as decreased transpiration and improved water use efficiency results in slower growth and development is not uncommon (McKay *et al*., 2003). Similarly,

reductions in growth rate and development have been found when investments in root are made in order to access more water (Gregory, 1978). Conversely, the "live fast, die young" approach (Wu *et al*., 2010) exercised in a drought escape strategy relies upon high stomatal conductance and photosynthetic capacity to complete the lifecycle prior to the onset of moisture limitation.

Co-localization of QTL discovered through traditional mapping approaches can be considered circumstantial evidence for pleiotropy (Lebreton *et al*., 1995; Tuberosa *et al*., 2003; Lanceras, 2004). Our results show that RPF and DTF are not invariably linked as four of 12 QTL show independent effects. Most QTL results in our study support a model of broad sense pleiotropy (i.e. an allele affecting more than one trait) underlying the correlations we observed between RPF, DTF and yield. The overall prevalence of genome-wide pleiotropy is expected to be rare but those genes demonstrating higher levels of pleiotropy (i.e. affecting a larger number of traits) are also expected to have larger effects on a per-trait basis (Wang *et al*., 2010). Therefore, we should expect that the effect sizes of pleiotropic genes should be larger than those due to genetic linkage. This is consistent with our QTL on A10 and C02 which showed larger effects, explained more variation on a per trait basis and had higher LOD support than the other QTL we discovered. However, further work to create and phenotype near-isogenic lines (NILs), mutants and transgenics will be necessary to conclusively rule out genetic linkage (Lovell *et al*., 2013; Uga *et al*., 2013; Huang *et al*., 2013). For breeding, this information would enable the design of breeding schemes to dissociate trait covariance should an increase or decrease in root investment be of value to the target production geography. For natural selection, it would facilitate our understanding of the genetics of adaptation in natural populations since pleiotropic genes have been shown to have both adaptive (Le Corre *et al*., 2002; Toomajian *et al*., 2006; Lovell *et al*., 2013) and maladaptive (Rose, 1982; Scarcelli *et al*., 2007) consequences.

In an effort to elucidate the functional pathway, we utilized the highly correlated, and putatively upstream, flowering time trait as a covariate in conditional analyses, essentially scanning for the significance of genetic effects using residual variation that is not explained by the correlated trait (Broman and Sen, 2009). These results provide support for a model where *RPF.dry1* impacts RPF

directly. The clear difference between the effect of the parent alleles at *RPF.dry1* (Figure 3.6) demonstrate that the Wichita allele increases RPF regardless of flowering date. In contrast, *RPF.dry2* appears to work indirectly through flowering time since RPF does not differ between alleles when flowering time is used as a covariate. The mean difference in RPF between alleles at the C02 locus may therefore be simply due to the fact that the majority of lines carrying the Wichita allele also flower later. The results of the 2012 QTL validation study suggest that the morphological characteristic underlying RPF may be taproot size as lines carrying the Wichita allele were consistently larger when analyzing the data using conditional models accounting for the correlated traits DTF and SFW.

The proposed mechanism of direct pleiotropy suggests that targeting root specific promoters might be an avenue for increasing root biomass without major effects on flowering time. However, root-specific reductions in cytokinin, a negative regulator of root system size, were shown to increase root biomass with minimal impacts on shoot growth except that bolting and flowering were delayed (Werner *et al*., 2010). These results may be indicative of inherent root to shoot feedback that would override the efficacy of such a strategy.

Discovery of root QTL independent of flowering time QTL suggest that root system size can be increased without impacts on flowering time

Despite the strong correlation between DTF and RPF across the population, trait values in some DH lines were contrary (i.e. high RPF and early flowering) to this expectation. Accordingly, we mapped two QTL in the wet environment that did not co-localize with flowering time QTL on linkage groups A08 and C07. The IMC106RR allele at the A08 QTL increases RPF, a result opposite to the rest of the QTL for RPF. This result partially explains the transgressive segregation we observed for RPF where some lines required more than 1.5 times more force than Wichita for root removal. Associations between loci on C07 and root traits such as root length and mass have been identified in other experiments conducted to understand the genetics of nutrient use efficiency (Hammond *et al*., 2009; Yang *et al*., 2010; Yang *et al*., 2011; Shi *et al*., 2012). Because the markers used in those studies do not overlap with ours, it is

difficult to draw strong conclusions about specific locational overlap but it suggests that this chromosome is a source of interesting variation in root biology across the *Brassica*s. These QTL and their associated markers could be valuable resources for breeding larger root systems without correlated responses in maturity.

Of the many candidate genes predicted to line within the QTL discovered, FLC is the most promising

Flowering time QTL have been identified previously on A02, A03, A10, C02 and C03 in other *B. napus* and *B. rapa* populations (Osborn, 1997; Schranz *et al*., 2002; Osborn and Lukens, 2003; Udall *et al*., 2006; Long *et al*., 2007; Shi *et al*., 2009) and, with the exception of the locus on C03, entirely agree with our results. All of these chromosomal regions are syntenic to the top of Arabidopsis (*Arabidopsis thaliana*) chromosome 5 (Parkin *et al*., 2005), a region that contains the well characterized flowering time genes *CO*, *FY* and *FLC*, among others. Prior research (Schranz *et al*., 2002; Razi *et al*., 2008) as well as the draft genomes of *B. rapa* and *B. oleracea* (Cheng *et al*., 2011; Cheng *et al*., 2012; Zhao *et al*., 2013) indicates that *FLC* has been retained after two rounds of whole-genome duplication (Tang *et al*., 2012) and is present on all of the aforementioned chromosomes, making it a prime candidate as the functional gene underlying these QTL. Additionally, CO has been maintained on A02, A10 and C02 and FY on A02 and A03 so that these also cannot be ruled out as candidates for the flowering time QTL on these chromosomes.

Many mutants and QTL associated with root development have been identified in research using the model plant Arabidopsis (Benfey *et al*., 2010). In particular, several QTL related to root growth have been mapped to the top of Arabidopsis chromosome 1 (Kobayashi and Koyama, 2002; Reymond *et al*., 2006; Sergeeva *et al*., 2006; Kellermeier *et al*., 2013) and the bottom of chromosome 4 (Loudet *et al*., 2005; Gerald *et al*., 2006; Reymond *et al*., 2006; Kellermeier *et al*., 2013). The QTL we identified for RPF on A08 and C07 appear to be in regions of the *B. napus* genome that are homologous to these segments of chromosomes 1 and 4, respectively (Cheng *et al*., 2011; Zhao *et al*., 2013). This may be

suggestive of root-specific genetic mechanisms that have been conserved within the *Brassica*ceae, but more research is clearly necessary to confirm this.

Several studies in species of *Brassica* have mapped QTL for flowering time which overlap with those for primary root (taproot) traits such as fresh weight, length and width (Lou *et al*., 2007; Lu *et al*., 2008; Yang *et al*., 2010; Kubo *et al*., 2010). Concordant with the results of our study, at least one of the QTL identified in each of these experiments was in a location putatively harboring an *FLC* homolog. None of these experiments resolved *FLC* as the functional gene but there is plenty of support for the role of this transcription factor in flowering time (Michaels and Amasino, 1999; Sheldon *et al*., 2002; Yuan *et al*., 2009) through its regulation of the flowering time genes, *FLOWERING LOCUS T* (*FT*), *SUPPRESSOR OF OVEREXPRESSSON OF CONSTANS 1* (*SOC1*), and *FLOWERING LOCUS D* (*FD*) (Searle *et al*., 2006; Helliwell *et al*., 2006). It was recently revealed that the *FLC* protein has over 500 other potential binding sites in the Arabidopsis genome, sites which were enriched in several gene ontology (GO) categories including response to stress and abiotic stimulus (Deng *et al*., 2011). This is may be considered circumstantial support for the results of our conditional examinations of the QTL on A10, and its putatively direct role in root development, since it seems possible that *FLC* could be regulating genes involved in root biosynthesis in *trans*-. Similar to our results, a recent analysis of *FLC* in Arabidopsis found that it impacted leaf shape and trichome number independently of its impact on flowering time (Willmann and Poethig, 2011).

The impact of *FLC* on flowering time varies dramatically across Arabidopsis accessions which suggest that a functionally diverse suite of alleles exists at this locus in this species (Salomé *et al*., 2011). Such variation has been associated with adaptation across geographies (Méndez-Vigo *et al*., 2011). This is noteworthy in regard to the allotetraploid *Brassica napus* since the presence of up to nine potential *FLC* paralogs (Zou *et al*., 2012) opens the opportunity for neo-functionalization of some copies leading to the potential for an expanded role in adaptation (Ohno, 1970). Regardless of the mode of action, these results provide motivation for further research into the impacts of *FLC* on whole-plant physiology in the *Brassica*s.

The results of this research support a body of evidence in which traits relevant to differential drought coping strategies may be genetically constrained, thereby creating an inherent tradeoff (Mitchell-Olds, 1996; McKay *et al*., 2003; Heschel and Riginos, 2005; Wu *et al*., 2010; Franks, 2011). These results must be considered in the context of the *Brassica*s in which little work has been conducted on drought coping mechanisms and none has focused on the roots. We are currently developing NILs so that the many alleles residing within *RPF.dry1* and *RPF.dry2* can be separated and their impacts on root mass and flowering time may be unequivocally estimated. Additionally, these NILs should be grown under diverse growing conditions and different geographies to understand the role of the environment on these traits and its interaction with the underlying genetics. Results from these experiments will inform fine-mapping activities aimed at cloning the causal variant(s), a process that will require identification of many more molecular markers within the candidate QTL regions. This task will be greatly enabled by the recent release of the draft genomes of the *B. napus* progenitors, *B. rapa* and *B. oleracea* (Wang *et al*., 2011; Cheng *et al*., 2011; Zhao *et al*., 2013). These activities will answer whether the QTL co-localization observed in this study is the result of pleiotropy or genetic linkage, ultimately improving our understanding of the genetics of drought physiology and enabling breeding for drought adaptation.

TABLES AND FIGURES

Table 3.1 Descriptive statistics for Days to Flowering (DTF), Root Pulling Force (RPF) and Yield measured in the SE1 population in Fort Collins, CO in 2011

Trait	Treatment	N	Mean	SD	Min	Max	H^2
DTF	Wet	643	77.00	9.49	59	108	0.83
	Dry	544	75.05	7.79	59	101	0.73
RPF	Wet	651	36.71	20.65	1.9	127	0.25
	Dry	650	30.78	16.49	4.5	114.4	0.16
Yield	Wet	663	40.73	46.17	1	305	0.21
	Dry	667	6.90	9.35	0	55	0.53

Table 3.2 Mean RPF difference between parental alleles (Wichita – IMC106RR) estimated in unconditional and conditional (using DTF as covariate) single marker analyses.

QTL	Chr	Scan	*RPF.wet* (Kgf)	*RPF.dry* (Kgf)
4	A08	Unconditional	-4.74 [b]	-1.69
		Conditional	-4.57 [b]	-1.47
1	A10	Unconditional	13.81 [a]	10.07 [a]
		Conditional	6.80 [a]	5.29 [a]
2	C02	Unconditional	12.88 [a]	7.49 [a]
		Conditional	3.28	-0.28
3	C07	Unconditional	5.34 [b]	-0.58
		Conditional	6.68 [b]	-0.10

[a] Differences significant at $P < 0.01$

[b] Differences significant at $P < 0.05$

Table 3.3 Genetic correlation coefficients of traits measured in the 2012 field experiment (n=39).

[a] $P < 0.05$, [b] $P < 0.01$, [c] $P < 0.0001$

	DTF	SFW	DMT	DML	DM	TRD	TRL	BZL	NSR
RPF	0.40 [a]	0.71 [c]	0.76 [c]	0.82 [c]	0.88 [c]	0.76 [c]	0.67 [c]	0.64 [c]	0.22
DTF		0.48 [b]	0.31	0.26	0.30	0.40 [a]	0.30	0.28	0.14
SFW			0.60 [c]	0.70 [c]	0.73 [c]	0.63 [c]	0.38 [b]	0.40 [b]	-0.05
DMT				0.62 [c]	0.83 [c]	0.71 [c]	0.71 [c]	0.63 [c]	0.21
DML					0.95 [c]	0.82 [c]	0.52 [b]	0.52 [b]	0.05
DM						0.86 [c]	0.65 [c]	0.61 [c]	0.12
TRD							0.55 [b]	0.50 [b]	0.14
TRL								0.88 [c]	0.41 [b]
BZL									0.52 [b]

RPF root pulling force

DTF days to flower

SFW shoot fresh weight

DMT dry mass taproot

DML dry mass laterals

DM dry mass

TRD tap root diameter

TRL tap root length

BZL branching zone length

NSR number coarse secondary laterals

Table 3.4 F-values for haplotype (genotype at *RPF.dry1*) in models incorporating DTF and SFW as covariates. [a] $P < 0.05$, [b] $P < 0.01$

Trait	Haplotype	Haplotype + DTF	Haplotype + SFW	Haplotype + DTF + SFW
DMT	14.22 [b]	10.76 [b]	4.81 [a]	5.93 [a]
DM	6.62 [a]	3.88	0.16	0.54
TRD	8.11 [b]	3.04	1.22	0.60
TRL	4.69 [a]	1.58	1.32	0.52

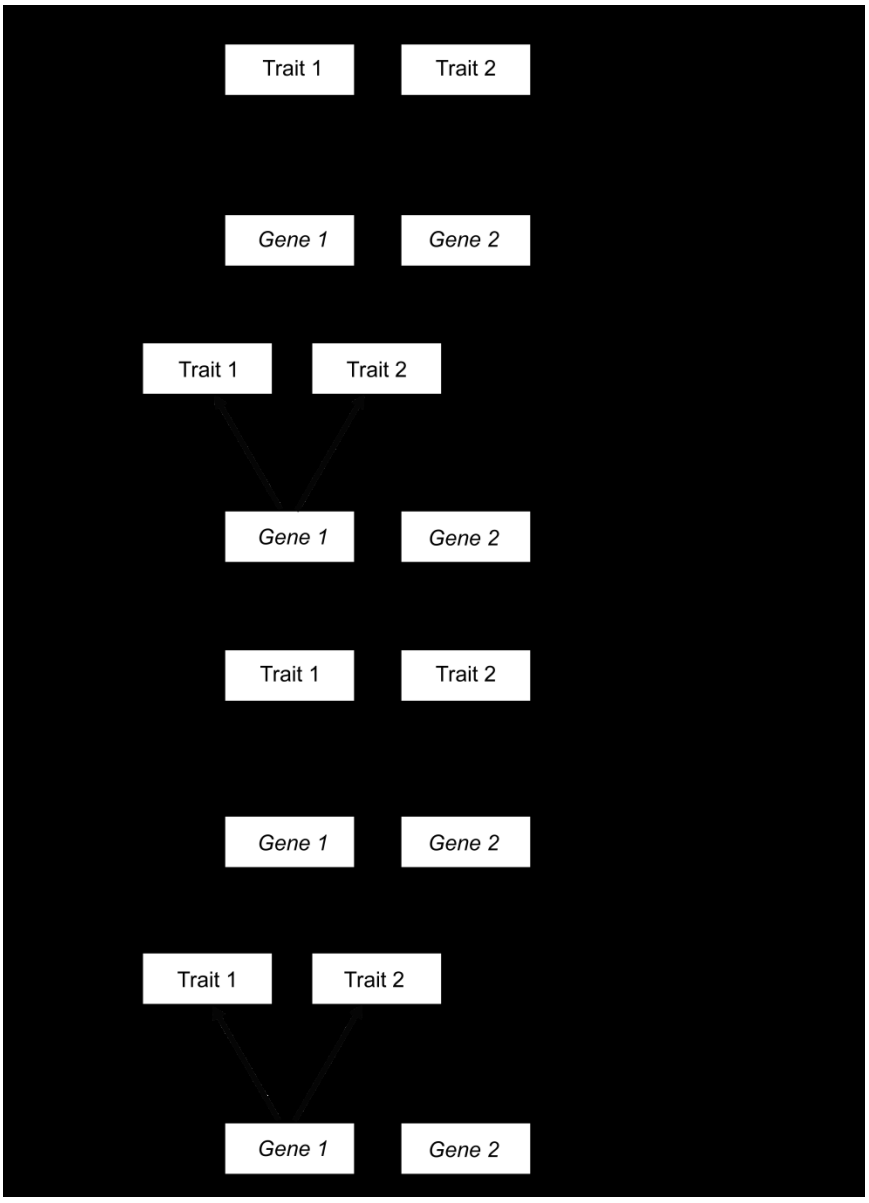


Figure 3.1 Diagram of putative mechanistic relationships for A, genetic linkage; B, pleiotropy; C, physiological interaction; D, combination of pleiotropy and physiological interaction.

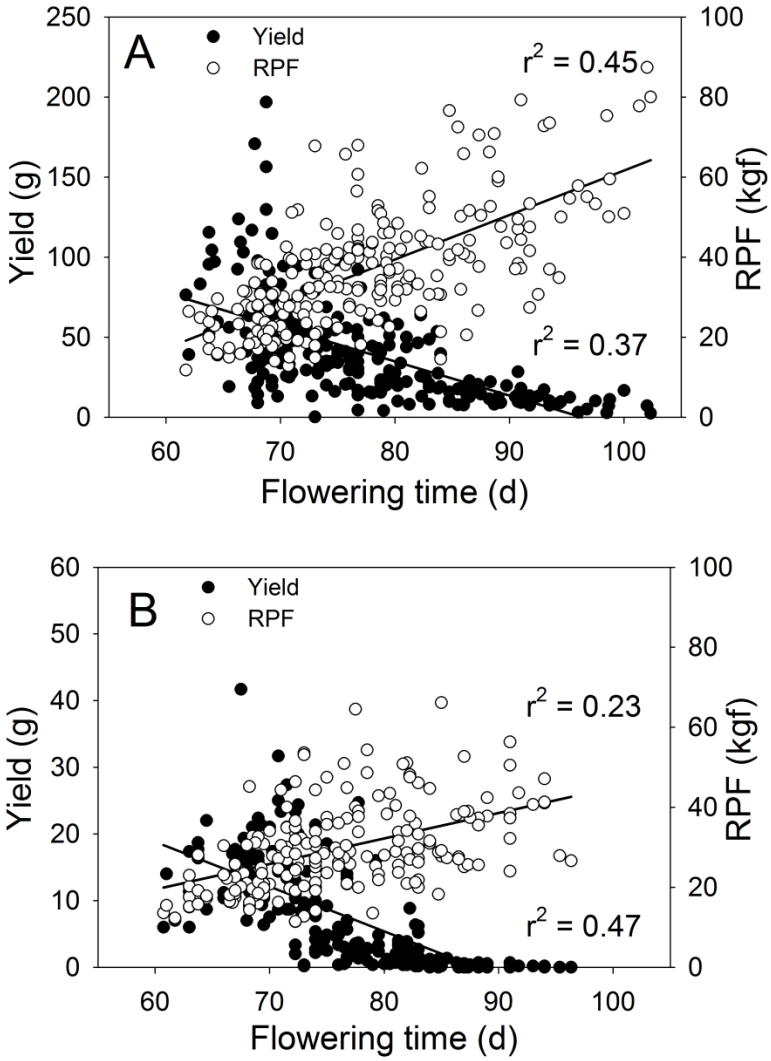


Figure 3.2 Genetic correlations among traits in the wet (A) and dry (B) treatments in the SE1 population (n=195 – 225, P < 0.0001).

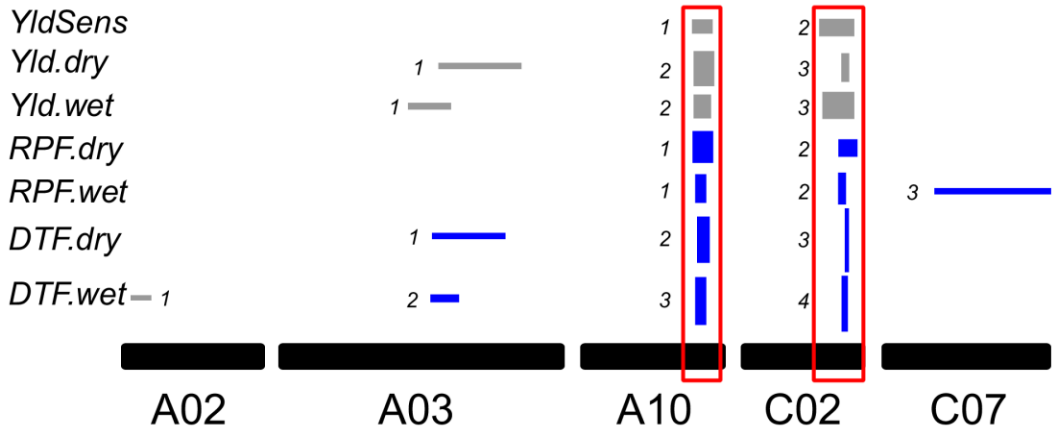


Figure 3.3 Localization and relative effect sizes of QTL for the six traits analyzed. Box widths indicate LOD 1.5 confidence intervals for the QTL. The box height represents the percent variance explained. Color indicates the directional effect of the Wichita allele (blue, positive; grey, negative). The pleiotropic QTL on chromosomes A10 and C02 are bracketed in red. Numbers next to boxes indicate QTL naming scheme.

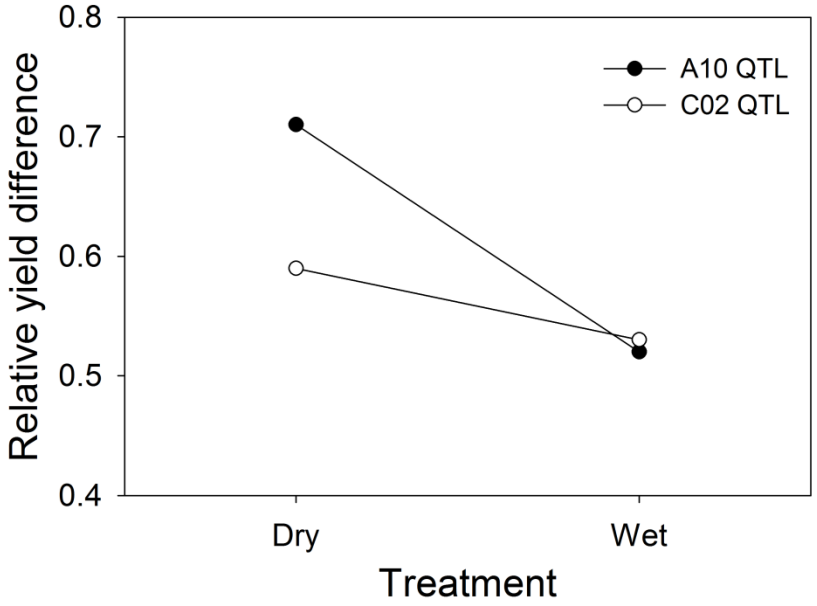


Figure 3.4 Difference between alleles (IMC106RR-Wichita) for relative yield (Z score) at QTL on A10 and C02 under wet and dry treatments.

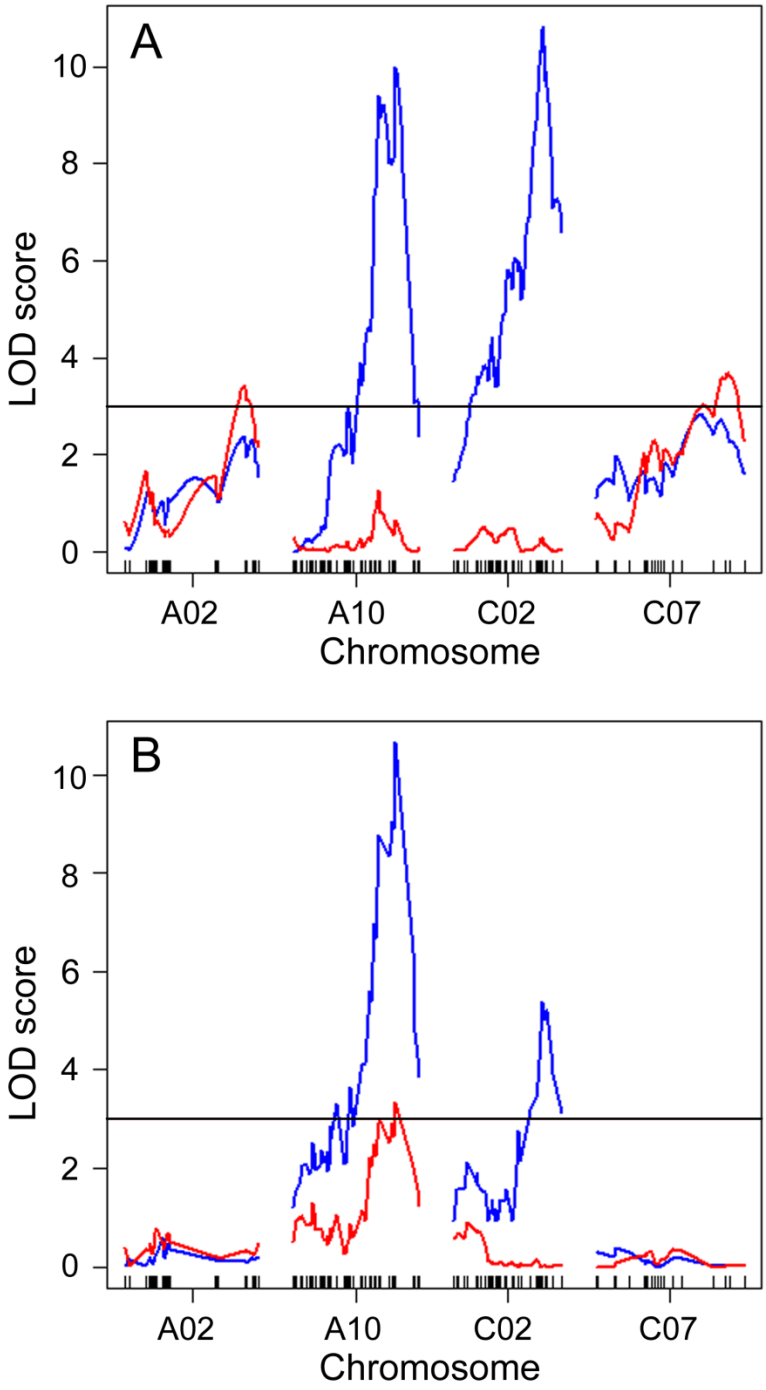


Figure 3.5 LOD profiles comparing conditional (incorporating DTF as a covariate; red) and unconditional (no covariate; blue) QTL scans in the wet (A) and dry (B) environments. The horizontal line indicates the LOD threshold based on 1000 permutations.

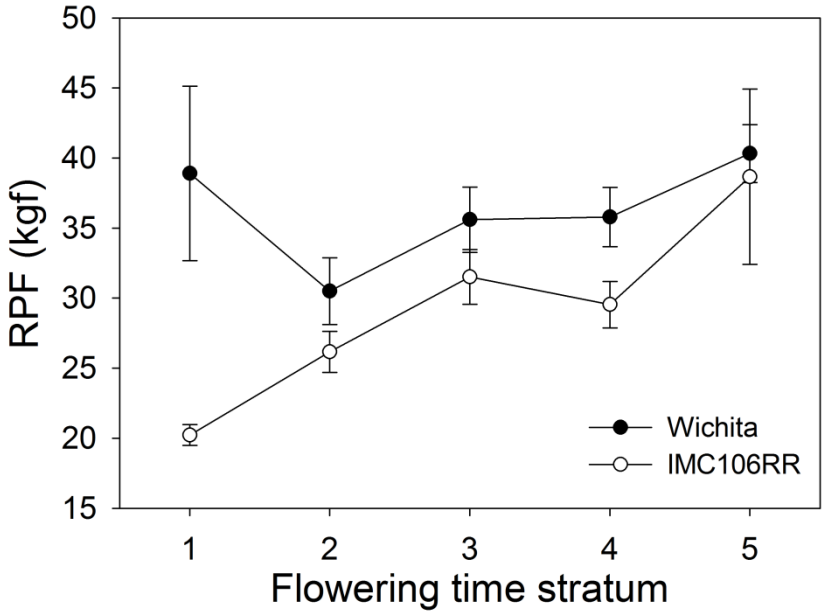


Figure 3.6 Dependence of RPF in the dry treatment on flowering time strata (1 = earliest, 5 = latest) for each allele at *RPF.dry* (mean ± SE). See Supplementary Table 5.7 for further description of strata.

REFERENCES

Araus JL, Slafer GA, Reynolds MP, Royo P. 2002. Plant breeding and drought in C_3 cereals: what should we breed for?. *Annals of Botany-London* **89,** 925-940.

Benfey PN, Bennett M, Schiefelbein J. 2010. Getting to the root of plant biology: impact of the *Arabidopsis* genome sequence on root research. *The Plant Journal* **61,** 992–1000.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19,** 889-890.

Broman KW, Sen S. 2009. *A guide to QTL mapping with R/qtl*. New York: Springer.

Bolaños J, Edmeades GO. 1993. Eight cycles of selection for drought tolerance in lowland tropical maize. II. Responses in reproductive behavior. *Field Crops Research* **31,** 258-268.

Boyer JS. 1982. Plant productivity and environment. *Science* **218,** 443–448.

Cativelli L, Rizza F, Badeck FW, Mazzucotelli E, Mastrangelo AM, Francia E, Maré C, Tondelli A, Stanca AM. 2008. Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Research* **105,** 1-14.

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. 2011. BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biology* **11,** 136.

Cheng F, Wu J, Fang L, Wang X. 2012. Syntenic gene analysis *between Brassica rapa* and other Brassicaceae species. *Frontiers in Plant Science* **3,** 198.

Cheverud JM. 1988. A comparison of genetic and phenotypic correlations. *Evolution* **42,** 958-968.

Churchill GA, Doerge R. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138,** 963-971.

Cortes PM, Sinclair TR.1986. Water relations of field-grown soybean under drought. *Science* **26,** 993–998.

Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES. 2011. FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **108,** 6680-6685.

Diers BW, Osborn TC. 1994. Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. *Theoretical and Applied Genetics* **88,** 662-668.

Edwards CE, Ewers BE, McClung CR, Lou P, Weinig C. 2012. Quantitative variation in water-use efficiency across water regimes and its relationship with circadian, vegetative, reproductive, and leaf gas-exchange traits. *Molecular Plant* **5,** 653-668.

Ekanayake IJ, Garrity DP, O'Toole JC. 1986. Influence of deep root density on root pulling resistance in rice. *Crop Science* **26,** 1181-1186.

Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics, Ed 4.* Harlow, Longman Group Ltd, 145-183.

Ferrie AMR, Möllers C. 2011. Haploids and doubled haploids in *Brassica* spp. for genetic and genomic research. *Plant Cell, Tissue and Organ Culture* **104,** 375-386.

Fitz Gerald JN, Lehti-Shiu MD, Ingram PA, Deak KI, Biesiada T, Malamy JE. 2006. Identification of quantitative trait loci that regulate Arabidopsis root system size and plasticity. *Genetics* **172,** 485-498.

Franks SJ. 2011. Plasticity and evolution in drought avoidance and escape in the annual plant *Brassica rapa*. *New Phytologist* **190,** 249-257.

Futuyma DJ. 1998. *Evolutionary biology*. Massachusetts: Sinauer Associates.

Gregory PJ, McGowan M, Biscoe PV, Hunter B. 1978. Water relations of winter wheat I. Growth of the root system. *The Journal of Agricultural Science* **91,** 91-102.

Gregory PJ, Atkinson CJ, Bengough AG, Else MA, Fernández-Fernández F, Harrison RJ, Schmidt S. 2013. Contributions of roots and rootstocks to sustainable, intensified crop production. *Journal of Experimental Botany* **64,** 1209-1222.

Hackett CA, **Broadfoot LB**. 2003. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* **90,** 33-38.

Haley CS, Knott SA. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69,** 315-324.

Hammond JP, Broadley MR, White PJ, *et al*. 2009. Shoot yield drives phosphorus use efficiency in *Brassica oleracea* and correlates with root architecture traits. *Journal of Experimental Botany* **60,** 1953-1968.

Hayes HK, Johnson IJ. 1939. The breeding of improved selfed lines of corn. *Journal of the American Society of Agronomy* **31,** 710-724.

Helliwell CA, Wood CC, Robertson M, Peacock WJ, Dennis ES. 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.

Heschel MS, Riginos C. 2005. Mechanisms of selection for drought stress tolerance and avoidance in *Impatiens capensis* (Balsaminaceae). *American Journal of Botany* **92,** 37-44.

Huang X, Ding J, Effgen S, Turck F, Koornneef M .2013. Multiple loci and genetic interactions involving flowering time genes regulate stem branching among natural variants of Arabidopsis. *New Phytologist* **199,** 843-857.

Johnson WC, Jackson LE, Ochoa O, van Wijk R, Peleman J, St. Clair DA, Michelmore RW. 2000. Lettuce, a shallow-rooted crop, and Lactuca serriola, its wild progenitor, differ at QTL determining root architecture and deep soil water exploitation. *Theoretical and Applied Genetics* **101,** 1066–1073.

Kellermeier F, Chardon F, Amtmann A. 2013. Natural variation of Arabidopsis root architecture reveals complementing adaptive strategies to potassium starvation. *Plant Physiology* **161,** 1421-1432.

Kirkegaard JA, Lilley JM. 2007. Root penetration rate – a benchmark to identify soil and plant limitations to rooting depth in wheat. *Australian Journal of Experimental Agriculture* **47,** 590–602.

Kobayashi Y, Koyama H. 2002. QTL analysis of Al tolerance in recombinant inbred lines of *Arabidopsis thaliana*. *Plant and Cell Physiology* **43,** 1526-1533.

Kosambi DD. 1944. The estimation of map distances from recombination values. *Annals of Eugenics* **12,** 172-175.

Kubo N, Saito M, Tsukazaki H, Kondo T, Matsumoto S. 2010. Detection of quantitative trait loci controlling morphological traits in *Brassica rapa* L. *Breeding Science* **60,** 164-171.

Kurepa J, Small J, Va M, Montagu N, Inzé D. 1998. Oxidative stress tolerance and longevity in Arabidopsis: the late-flowering mutant gigantea is tolerant to paraquat. *The Plant Journal* **14,** 759-764.

Lanceras JC, Pantuwan G, Jongdee B, Toojinda T. 2004. Quantitative trait loci associated with drought tolerance at reproductive stage in rice. *Plant Physiology* **135,** 384-399.

Le Corre V, Roux F, Reboud X. 2002. DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution* **19,** 1261–1271.

Lebreton C, Lazicjancic V, Steed A, Pekic S, Quarrie SA. 1995. Identification of QTL for drought responses in maize and their use in testing causal relationships between traits. *Journal of Experimental Botany* **46,** 853-865.

Leung K-M, Elashoff RM, Afifi AA. 1997. Censoring issues in survival analysis. *Annual Review of Public Health* **18,** 83-104.

Levins R. 1968. *Evolution in changing environments*. New Jersey: Princeton University Press.

Li R, Tsaih SW, Shockley K, Stylianou IM, Wergedal J, Paigen B, Churchill GA. 2006. Structural model analysis of several multiple traits. *Plos Genetics* **2,** e114.

Lopes MS, **Reynolds MP**. 2010. Partitioning of assimilates to deeper roots is associated with cooler canopies and increased yield under drought in wheat. *Functional Plant Biology* **37,** 147-156.

Lou P, Zhao J, Kim JS, Shen S, Del Carpio DP, Song X, Jin M, Vreugdenhill D, Wang X, Koornneef M, Bonnema G. 2007. Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. *Journal of Experimental Botany* **58,** 4005-4016.

Loudet O, Gaudon V, Trubuil A, Daniel-Vedele F. 2005. Quantitative trait loci controlling root growth and architecture in *Arabidopsis thaliana* confirmed by heterogeneous inbred family. *Theoretical and Applied Genetics* **110,** 742-753.

Long Y, Shi J, Qiu D, *et al.* 2007. Flowering time quantitative trait loci analysis of oilseed *Brassica* in multiple environments and genomewide alignment with *Arabidopsis*. *Genetics* **177,** 2433-2444.

Lovell JT, Juenger TE, Michaels SD, *et al.* 2013. Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. *Proceedings of the Royal Society B: Biological Sciences* **280,** 20131043.

Lu G, Cao J, Yu X, Xiang X, Chen H. 2008. Mapping QTLs for root morphological traits in *Brassica rapa* L. based on AFLP and RAPD markers. *Journal of Applied Genetics* **49,** 23-31.

Ludlow MM, **Muchow RC**. 1990. A critical evaluation of traits for improving crops in water-limited environments. *Advances in Agronomy* **43,** 107-153.

Ludlow MM. 1989. Strategies of response to water stress. In: Kreeb KH, Richter H, Hinckley TM, eds. *Structural and functional responses to environmental stresses*. The Hague: SPB Academic, 269-281.

Lühs W, Seyis F, Frauen M, Busch H, Frese L, Willner E, Friedt W, Gustafsson M Poulsen G. 2003. Development and evaluation of a *Brassica napus* core collection. In: Knüpffer H and Ochsmann J, eds, *Rudolf Mansfeld and Plant Genetic Resources.* Bonn: ZADI/IBV, 284–289.

Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW. 2009. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics* **181:** 1077-1086.

McKay JK, Richards JH, Mitchell-Olds T. 2003. Genetics of drought adaptation in *Araabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Molecular Ecology* **12,** 1137-1151.

McKay JK, Richards JH, Nemali KS, Sen S, Mitchell-Olds T, Boles S, Stahl EA, Wayne T, Juenger TE. 2008. Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population, Kas-1 x Tsu-1. *Evolution* **62,** 3014-3026.

Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C. 2011. Altitudinal and climatic adaptation is mediated by flowering traits and *FRI*, *FLC*, and *PHYC* genes in Arabidopsis. *Plant Physiology* **157,** 1942-1955.

Michaels SD, Amasino RM. 1999. *Flowering Locus C* encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant Cell* **11,** 949-956.

Michaels SD. 2009. Flowering time regulation produces much fruit. *Current Opinion in Plant Biology* **12,** 75-80.

Mitchell-Olds T. 1996. Pleiotropy causes long-term genetic constraints on life-history evolution in *Brassica rapa*. *Evolution* **50,** 1849-1858.

Obeso JR. 2002. The costs of reproduction in plants. *New Phytologist* **155,** 321-348.

Ohno S. 1970. *Evolution by Gene Duplication*. Berlin: Springer-Verlag.

Oliver MJ, Cushman JC, Koster KL. 2010. Dehydration tolerance in plants. In: Sunkar R, ed, *Plant Stress Tolerance: Methods in Molecular Biology*. New York: Humana Press, 3-24.

Osborn TC, Kole C, Parkin IAP, Sharpe AG, Kuiper M, Lydiate DJ, Trick M. 1997. Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics* **146,** 1123-1129.

Osborn T C, Lukens L. 2003. The molecular genetic basis of flowering time variation in *Brassica* species. In: Nagata T, Tabata S, eds, *Biotechnology in Agriculture and Forestry: Brassicas and Legumes, From Gene Structure to Breeding*. Berlin: Springer-Verlag, 69–86.

Palmer CE, Keller WA, Arnison PG. 1996. Experimental haploidy in *Brassica* species. In: Mohan Jain S, Sopory SK, Veilleux RE, eds, *In vitro Haploid Production in Higher Plants*. Dordrecht: Kluwer Academic Publishers, 143-164.

Palta JA, Chen X, Milroy SP, Rebetzke GJ, Dreccer MF, Watt M. 2011. Large root systems: are they useful in adapting wheat to dry environments?. *Functional Plant Biology* **38,** 347-354.

Parkin IAP, SM Gulden, AG Sharpe, L Lukens, M Trick, TC Osborn, DJ Lydiate. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* **171,** 765-781.

Price AH, Cairns JE, Horton P, Jones HG, Griffiths H. 2001. Linking drought-resistance mechanisms to drought avoidance in upland rice using a QTL approach: progress and new opportunities to integrate stomatal and mesophyll responses. *Journal of Experimental Botany* **53,** 989-1004.

Posé D, Yant L, Schmid M. 2012. The end of innocence: flowering networks explode in complexity. *Current Opinion in Plant Biology* **15,** 45-50.

Udall JA, Quijada PA, Lambert B, Osborn TC. 2006. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L): 2. Identification of alleles from unadapted germplasm**.** *Theoretical and Applied Genetics* **113,** 597-609.

Rahman M, McClean P. 2013. Genetic analysis on flowering time and root system in *Brassica napus* L. *Crop Science* **53,** 141-147.

Razi H, Howell EC, Newbury HJ, Kearsey MJ. 2008. Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in *Brassica oleracea*?. *Theoretical and Applied Genetics* **116,** 179-192.

Reymond M, Svistoonoff S, Loudet O, Nussaume L, Desnos T. 2006. Identification of QTL controlling root growth response to phosphate starvation in *Arabidopsis thaliana*. *Plant Cell & Environment* **29,** 115–125.

Reznick D. 1985. Costs of reproduction: an evaluation of the empirical evidence. *Oikos* **44,** 257-267.

Rife C, Auld DL, Sunderman HD, Heer WF, Baltensperger DD, Nelson LA, Johnson DL, Bordovsky D, Minor HC. 2001. Registration of 'Wichita' winter rapeseed. *Crop Science* **41,** 263-a-264.

Robertson A. 1959. The sampling variance of the genetic correlation coefficient. *Biometrics* **15,** 469-485.

Rogers RR, Russell WA, Owens JC. 1976. Evaluation of a vertical pull technique in population improvement of maize for corn rootworm tolerance. *Crop Science* **16,** 591-594.

Rose MR. 1982. Antagonistic pleiotropy, dominance, and genetic variation. Heredity **48,** 63–78.

Salomé PA, Bomblies K, Laitinen RAE, Yant L, Mott R, Weigel D. 2011. Genetic architecture of flowering-time variation in *Arabidopsis thaliana*. *Genetics* **188,** 421-433.

Scarcelli N, Cheverud JM, Schaal BA, Kover PX. 2007. Antagonistic pleiotropic effects reduce the potential adaptive value of the FRIGIDA locus. *Proceedings of the National Academy of Sciences, USA* **104,** 16986–16991.

Schranz ME, Quijada P, Sung SB, Lukens L, Amasino R, Osborn TC. 2002. Characterization and effects of the replicated flowering time gene *FLC* in *Brassica rapa*. *Genetics* **162,** 1457-1468.

Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, Amasino RA, Coupland G. 2006. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes & Development* **20,** 898-910.

Sergeeva LI, Keurentjes JJB, Bentsink L, Vonk J, Van der Plas LW, Koornneef M, Vreughenhil D. 2006. Vacuolar invertase regulates elongation of *Arabidopsis thaliana* roots as revealed by QTL and mutant analysis. *Proceedings of the National Academy of Sciences, USA* **103,** 2994–2999.

Sharp RE, Davies WJ. 1979. Solute regulation and growth by roots and shoots of water-stressed Maize plants. *Planta* **147,** 43-49.

Sheldon CC, Conn AB, Dennis ES, Peacock WJ. 2002. Different regulatory regions are required for the vernalization-induced repression of *Flowering Locus C* and for the epigenetic maintenance of repression. *The Plant Cell* **14,** 2527-2537.

Sherrard ME, Maherali H. 2006. The adaptive significance of drought escape in *Avena barbata*, an annual grass. *Evolution* **60,** 2478-2489.

Shi J, Li R, Qiu D, Jiang C, Long Y, Morgan C, Bancroft I, Zhao J, Meng J. 2009. Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. *Genetics* **182,** 851-861.

Shi L, Yang J, Liu J, Li R, Long Y, Xu F, Meng J. 2012. Identification of quantitative trait loci associated with low boron stress that regulate root and shoot growth in *Brassica napus* seedlings. *Molecular Breeding* **30,** 393-406.

Sibout R, Plantegenet S, Hardtke CS. 2008. Flowering as a condition for xylem expansion in *Arabidopsis* hypocotyl and root. *Current Biology* **18,** 458-463.

Spencer JT. 1940. Comparative study of the seasonal root development of some inbred lines and hybrids of maize. *Agricultural Research* **61,** 521-538.

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pederson BS, Conant G, Wang X, Freeling M, Pires JC. 2012. Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* **190,** 1563-1574.

Topp CN, lyer-Pascuzzi AS, Anderson JT, *et al.* 2013. 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. *Proceedings of the National Academy of Sciences, USA* **110,** E1695-E1704.

Toomajian C, Hu TT, Aranzana MJ, *et al.* 2006. A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *Plos Biology* **4,** 732–738.

Trick M, Long Y, Meng J, Bancroft I. 2009. Single nucleotide polymorphism (SNP) discovery in the poly-ploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnology Journal* **7,** 334-346.

Tuberosa R, Salvi S, Sanguineti MC, Maccaferri M, Giuliani S, Landi P. 2003. Searching for quantitative trait loci controlling root traits in maize: a critical appraisal. *Plant and Soil* **255,** 35-54.

Uga Y, Sugimoto K, Ogawa S, *et al.* 2013. Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nature Genetics* **45,** 1097–1102.

U N. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **7,** 389-452.

Van Ooijen JW, Voorrips RE. 2001. *Joinmap version 3.0: software for the calculation of genetic linkage maps*. Wageningen, the Netherlands: Plant Research International.

Van Tienderen PH, Hammad I, Zwaal FC. 1996. Effects of flowering time genes in the annual Crucifer *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany* **83,** 169-174.

Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics* **12,** 204-213.

Wang Z, Liao BY, Zhang J. 2010. Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Sciences, USA* **107,** 18034-18039.

Wang X, Wang H, Wang J, *et al.* 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43,** 1035-1039.

Wasson AP, Richards RA, Chatrath R, Misra SC, Sai Prasad SC, Rebetzke GJ, Kirkegaard JA, Christopher, Watt M. 2012. Traits and selection strategies to improve root systems and water uptake in water-limited wheat crops. *Journal of Experimental Botany* **63,** 3485-3498.

Werner T, Nehnevajova E, Köllmer I, Novák O, Strnad M, Krämer U, Schmülling T. 2010. Root-specific reduction of cytokinin causes enhanced root growth, drought tolerance, and leaf mineral enrichment in *Arabidopsis* and tobacco. *The Plant Cell* **22,** 3905-3920.

White JW, Castillo JA. 1989. Relative effect of root and shoot genotypes on yield of common bean under drought stress. *Crop Science* **29,** 360-362.

Willmann MR, Poethig RS. 2011. The effect of the floral repressor *FLC* on the timing and progression of vegetative phase change in *Arabidopsis*. *Development* **138,** 677-685.

Wu CA, Lowry DB, Nutter LI, Willis JH. 2010. Natural variation for drought-response traits in the *Mimulus guttatus* species complex. *Oecologia* **162,** 23-33.

Yang M, Ding G, Shi L, Feng j, Xu F, Meng J. 2010. Quantitative trait loci for root morphology in response to low phosphorus stress in *Brassica napus*. *Theoretical and Applied Genetics* **121,** 181-193.

Yang M, Ding G, Shi L, Xu F, Meng J. 2011. Detection of QTL for phosphorus efficiency at vegetative stage in *Brassica napus*. *Plant and Soil* **339,** 97-111.

Yuan YX, Wu J, Sun RF, Zhang XW, Xu DH, Bonnema G, Wang XW. 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.

Zou X, Suppanz I, Raman H, Hou J, Wang J, Long Y, Jung C, Meng J. 2012. Comparative analysis of *FLC* homologues in Brassicaceae provides insight into their role in the evolution of oilseed rape. *PLoS One* **7,** e45751

CHAPTER FOUR: DRAFTING A *BRASSICA NAPUS* GENOME TO CHARACTERIZE SEQUENCE VARIATION IN QUANTITATIVE TRAIT LOCI ASSOCIATED WITH DROUGHT ADAPTATION

SUMMARY

Brassica napus is a globally important oilseed which is lacking research on the genetics of drought adaptation. We previously mapped several quantitative trait loci (QTL) underlying drought adaptation traits (root mass and flowering time) in a bi-parental mapping population created from a cross between annual and spring *B. napus* cultivars. Here we create a draft genome and examine the genomic divergence of the parents of this mapping population. We sequenced each parental cultivar on the Illumina HiSeq platform to a minimum depth of 23X and performed a reference based assembly in order to describe the molecular variation differentiating them at the scale of the genome, QTL and gene. The draft genomes of *B. rapa* (AA) and *B. oleracea* (CC), the progenitors of the allopolyploid *B. napus* (AACC), were used as references. Genome-wide patterns of variation were characterized by an overall higher single nucleotide polymorphism (SNP) density in the A genome and a higher rate of nonsynonymous substitutions in the C genome. Nonsynonymous substitutions were used to categorize gene ontology (GO) terms differentiating the parent genomes along with a list of putative functional variants contained within each QTL. Marker assays were developed for several of the discovered SNPs and shown to improve the genetic map. QTL analysis with the new, denser map resulted in an improvement in model fit, suggesting the new markers were more tightly linked to the causal variants. Together, these results provide a draft of the genome of *B. napus* and a catalog of genomic variants differentiating annual and biennial growth types. In addition, they provide a list of candidate genes for future research into the functional genes underlying traits related to drought adaptation.

INTRODUCTION

Brassica napus is an amphidiploid formed from recursive and independent hybridizations between the diploids *B. rapa* (A genome) and *B. oleracea* (C genome) (U, 1935; Parkin *et al.*, 1995). It is now the third most important oilseed crop (http://www.fas.usda.gov/oilseeds/Current/; 2013), trailing only soybean and oil palm in global oil production. Much of this global acreage is considered “canola”,

defined by grain which produces low quantities of erucic acid (< 2% C22:1) and glucosinolates (< 30 μmol g $^{-1}$). *Brassica napus* is produced commercially as both annual and biennial flowering types which have been shown to represent genetically and morphologically distinct groups (Diers and Osborn, 1994; Lühs *et al*., 2003). The genetic diversity differentiating these morphotypes represents a valuable source of variation for improving adaptation and ultimately increasing grain yields (Quijada *et al*., 2004; 2006). It will likely be an important resource in creating stress-adapted cultivars in both annual and biennial flowering types.

Adequate water availability is crucial to almost all aspects of plant survival and reproduction. Drought frequently causes reductions in crop yields to levels well below half of their theoretical potential (Boyer, 1982). Plants cope with drought through physiological plasticity and/or adaptive mechanisms under heritable genetic controls (Juenger, 2013). Therefore, three drought coping strategies have been proposed to summarize these mechanisms: drought escape, dehydration avoidance, and dehydration tolerance (Ludlow, 1989). Dehydration tolerance is achieved through mechanisms which enable survival of low internal water potentials and is a strategy which exists in only 0.1% of angiosperm species (Oliver, Cushman and Koster, 2010). Drought escape has been the focal strategy in breeding for low rainfall environments and focuses on rapid life cycling so that reproduction is completed prior to the onset of drought. Dehydration avoidance, in contrast, is the ability to maintain internal water status during drought conditions by minimizing water loss (i.e. reducing transpiration) and/or maximizing water uptake (i.e. larger root systems). The adaptive value of these strategies varies by geography and a deeper understanding of their genetic bases is of clear value to plant breeders and evolutionary biologists alike.

Continuous expansion of *B. napus* production into new geographies along with changing climatic trends is increasing its exposure to drought stress. Still, little research has focused on the genetics of drought physiology in *B. napus*. In a previous study, we mapped QTL for drought escape (flowering time) and dehydration avoidance (root mass) in a doubled haploid (DH) population of 225 lines (Chapter 3). The microspore donor plant used to create the DH mapping population originated from a cross made between an annual and a biennial variety so that the DH population exhibited large variation for all of the

phenotypes analyzed. We discovered a total of 20 QTL, some of which co-localized for each of the measured traits (Supplementary Table 5.6). For those that co-localized, we were unable to resolve whether the underlying mechanism was the result of genetic linkage or pleiotropy, a common outcome of QTL mapping studies (Tuberosa *et al*., 2003; Lanceras, 2004).

The past decade has seen revolutionary advances in DNA sequencing technology which have transformed the way genetics and genomics research is conducted (Edwards, Henry and Edwards, 2012). These "next-generation" methods allow for the sequencing of billions of small DNA fragments simultaneously, a departure from the single fragment sequencing approach of the original Sanger technology. This fundamental change in strategy has reduced the cost of data acquisition to levels where whole-genome sequencing is a feasible approach for most organisms. The major challenge has become the development of efficient and accurate algorithms for assembling the massive number of short reads into an order that accurately represents the target genome. This step is simplified in species for which a draft genome exists where data analysis becomes the much simpler process of aligning reads to a reference (alignment-consensus approach) rather than *de novo* assembly (Flicek and Birney, 2009).

The draft genomes of the *B. napus* progenitor A and C genomes have become available recently and now serve as suitable "pseudo-references" (Edwards, Batley and Snowdon, 2013), prompting us to re-sequence the parent lines of our mapping population on the Illumina HiSeq (San Diego, CA) sequencing platform. We aligned these sequence data to the reference genomes in order to characterize genome-wide patterns of genetic variation differentiating the parent lines of the DH population. Gene ontology (GO) terms enriched for nonsynonymous substitutions were used to speculate on the major categorical differences among these divergent cultivars. In addition, the SNPs discovered were used to characterize the extent of variation existing within QTL regions and to identify putative candidate genes contained therein. In a more focused study, we developed molecular markers for a subset of the SNPs discovered in the pleiotropic QTL named *RPF.dry1* (Chapter 3) to compute a new genetic map. This was used in new QTL scans which resulted in an improvement of model fit and a more refined estimate of QTL location.

MATERIALS AND METHODS

Sequencing

This study analyzed the canola cultivars IMC106RR (Cargill Inc., National Registration No. 5118) and Wichita (Rife *et al*., 2001; Reg. no. CV-19, PI 612846) as well as a DH population of 225 lines derived from a cross between them as described in Chapter 3. High quality DNA was extracted from the fifth leaf of each parent using the standard methods described in the Qiagen (Valencia, CA) column extraction kit. The extracted DNA was run on a 1% agarose gel to confirm DNA quality and concentrated to 50 ng/µl. Parental DNA libraries were sent to the University of Missouri DNA Core Facility (http://biotech.rnet.missouri.edu/dnacore/) and prepped for an average insert size of 200 base pairs. DNA libraries of each parent were sequenced on one lane of an Illumina HiSeq 2000 (San Diego, CA) sequencer to generate 2 x 100 paired-end reads.

Alignment and polymorphism analysis

Mapping of the genomic sequencing data (fastq files) to the *B. rapa* (Wang *et al*., 2011; Cheng *et al*., 2011) and *B. olearacea* (Zhao *et al*., 2013; http://www.ocri-genomics.org; 2014) reference genomes was performed using SeqMan NGen v4 (DNAStar, Madison, WI). The alignment was performed using default settings for read mapping and SNP calling. The SNP report created by SeqMan NGen was exported to ArrayStar v4 (DNAStar, Madison, WI) for further filtering. The final list of SNPs was generated using the following filter criteria: quality call score ≥ 30 (Phred scale), SNP frequency ≥ 5%, depth ≥ 5 and "p not ref" ≥ 90 (probability that the base is different than the reference base). Because SNPs were called relative to the reference genome, three lists were generated: SNPs present in both parents and SNPs that existed in one parent or the other (Trick *et al*., 2009). Thus, SNPs called in one parent and not the other were deemed to be polymorphic among parents. SNPs called in candidate genes within the *RPF.dry1* QTL interval were manually reviewed to confirm or correct the results of the automated SNP caller. The results of these manual scores were used to estimate the SNP caller's error rate which was then used in correcting genome-wide SNP calls. To be conservative, we did not include 'hemi-

SNPs' (i.e. polymorphisms originating from homologous reads; Trick *et al.*, 2009) in our analyses as they may not represent genetic variants originating from the region of interest. Hemi-SNPs are the result of reads originating from a homologous locus which map to the incorrect locus due to a high degree of sequence homology. In the alignment, they appear as heterozygous loci since two alleles are present for one parent but not the other (see Bancroft *et al.*, 2011).

The Burrows-Wheeler Alignment (BWA) tool (Li and Durbin, 2009) was used as part of a pipeline developed to estimate the number of loci to which each NGS read would map in the A and C genomes. First, fastq and reference genome data were input to BWA to generate read mapping information. The SAM files output from this analysis were analyzed using a program written in Python to generate a data array containing the number of hits per read.

Polymorphisms were recorded as non-coding if they did not appear in a coding region of any gene model. SNPs in coding regions were then classified as either synonymous or nonsynonymous. The length of aligned sequence was calculated using the alignment summary report generated from SeqMan NGen. SNP density was calculated as the number of SNPs existing within a particular length of aligned sequence. In order to test Gene Ontology (GO) terms for an excess of nonsynonymous polymorphism, the number of nonsynonymous SNPs and the total number of codons in each gene model was obtained. For each GO term, the total density of nonsynonymous SNPs per codon among all gene models annotated with that term was calculated. Enrichment P-values were then calculated via permutation by randomly shuffling the numbers of nonsynonymous SNPs and codons across gene models while preserving GO annotations, and then asking whether the number of nonsynonymous SNPs per codon associated with the GO term in the permuted set was greater than or equal to the corresponding value from the true data set. 1000 such permutations were performed.

For the detailed analysis of *RPF.dry1*, SNPs were scored based upon their location in an exon or intron as well as their location within 1 kilobase of the start or stop codon. SNPs in these flanking regions are presumed to capture putative regulatory regions. Divergence for candidate gene regions were

calculated as the total number of SNPs discovered across the entire gene sequence plus the putative 1 kb regulatory regions.

Candidate Gene Identification and Analysis

The LOD 1.5 confidence intervals of the genetic map positions of the QTL described in Chapter 3 (Supplementary Table 5.6) were determined in the R/qtl software package (Broman *et al*., 2003; Broman and Sen, 2009). Markers spanning these intervals were compared with the *B. rapa* and *B. oleracea* references using BLAST (Altschul *et al*., 1990) to identify their physical map positions along with a list of genes expected to lie within them. Candidate genes were defined as those annotated with any term related to root and/or flowering. Sequence and amino acid identities were analyzed using the ClustalW method in the Lasergene software package (DNAstar, Madison, WI).

Linkage analysis

KASP SNP genotyping assays (LGC Genomics, Teddington, Middlesex, UK) were developed for candidate SNPs discovered in comparison of the reference based assembly of the parents and used to genotype the original DH population. The parent genome alignments were used to design primer sequences for the KASP assays. A genetic linkage map was constructed which incorporated the molecular markers using JoinMap3 (Van Ooijen *et al*., 2001) with a threshold recombination frequency of <0.25 and a minimum logarithm of the odds ratio (LOD) score of 6 for creating linkage groups. Genetic distances were calculated using the Kosambi function (Kosambi, 1944).

Analyses of the association between discovered polymorphisms and drought adaptive phenotypes were performed by regressing quantitative phenotype on to molecular markers (Whittaker, Thompson and Visscher, 1996) using PROC REG in the SAS software package (SAS Institute 2004). Genome-wide QTL scans were performed using Haley-Knott Regression (Haley and Knott, 1992) in R/qtl with 1 cM steps. QTL were selected using a step-wise model selection approach (Manichaikul *et al*., 2009) based on significance thresholds made from 1000 permutations (Churchill and Doerge, 1994).

RESULTS

Re-sequencing

Whole-genome sequencing of IMC106RR and Wichita produced paired-end read data sets summing to 300 million reads (27.8 Gb passed Illumina filter) and 349 million reads (32.2 Gb passed Illumina filter), respectively. The mean Q-score was over 35 with 91% of reads in each data set considered to be of high-quality (Q-score ≥ 30). This equates to an average high-quality read coverage of 26.6X for Wichita and 23X for IMC106RR for the estimated 1100 Mbp *B. napus* genome (Johnston *et al.*, 2005).

Across the *B. rapa* (A genome) and *B. oleracea* (C genome) reference genomes, 69% of reads (> 447 million) were mapped which summed to nearly 565 Mb of aligned sequence. The effective genome size (excluding N's) of the combined *Brassica* draft genomes is 670 Mb so that 84% of the estimated genome was represented in our alignment. However, nearly 3% more of the C genome alignment was missing than the A genome, likely due to a higher frequency of N's in the C genome (data not shown).

B. napus is an amphidiploid shown to carry an average of three homologous regions within each homeologous sub-genome (Osborn *et al.*, 1997; Cavell *et al.*, 1998; Parkin *et al.*, 2005). In an effort to understand the effects of such genomic redundancy on alignments that rely on single best match of short-reads, we estimated the number of possible genomic regions that a single 100-bp read would map to at a minimum threshold of 95% sequence identity. The distribution of this analysis is characterized by a heavy right-skew where a diminishing number of reads map to many genomic regions (Figure 4.1). However, the median number of genomic regions to which a read maps is one and suggests that most reads originate from a unique region of the genome. Expectedly, reads on the far right side of the distribution that mapped to hundreds of loci tended to be simple repetitive elements.

As an empirical measure of the sequence alignment, we manually assessed all of the SNPs called across the 42 candidate genes selected within the nearly 146 kb interval of *RPF.dry1*. Out of the total 455 SNPs called, 89 were incorrect (20%). The majority of these false positives were due to insufficient coverage in one parent so that a well-supported SNP from the other parent appeared to be unique. A more

genome-wide assessment was performed by searching for 100 SNPs randomly selected from the nearly 1,200 used in the construction of the original linkage map. Of these, 97 were successfully identified where one of the missing SNPs was present as a hemi-SNP but supported by only one read and the other two were located in regions of zero coverage.

Genome-wide SNP variation

A total of 1,080,412 SNPs differentiating the parental genomes were identified after corrections made using the empirically-derived estimates. This represents an average difference between the parents of 0.20%. However, the extent of this difference varied substantially between the homeologous genomes of *B. rapa* and *B. oleracea* and across chromosomes within them (Table 4.1; Supplementary Figures 5.4-22). Divergence of the parent lines was 47% higher ($P \leq 0.01$) in the A genome than it was in the C genome. These results are similar to those reported in previous studies comparing SNP variation among *B. napus* cultivars (Trick *et al.*, 2009; Durstewitz *et al.*, 2010; Bus *et al.*, 2012; Tollenaere *et al.*, 2012). Similarly, the synonymous substitution rate of 0.13% in the A genome was significantly higher than that observed in the C genome (0.10%; $P < 0.01$). However, the nonsynonymous substitution rate in each genome was identical so that the ratio of nonsynonymous to synonymous substitutions was significantly higher in the C genome than it was in the A genome. This ratio can be a metric for the selective pressure on a locus (Nei and Gojobori, 1986). Our results suggest that purifying selection has been stronger in the A genome than it has in the C genome, a result that is concordant with a recent study comparing divergence of *B. rapa* and *B. oleracea* (Zhao *et al.*, 2013). Comparisons of SNP variation in coding regions relative to the non-coding intronic and intergenic regions revealed expectedly lower substitution rates in coding regions (Table 4.1).

Finally, we tested all genes carrying nonsynonymous variation for enrichment of particular GO terms in hopes of identifying particular classes that may broadly differentiate annual and biennial flowering types. A total of 25 GO terms were determined to be significant at $P < 0.01$ (Supplementary Table 5.8). Of these, 4 were related to transcriptional regulation and the second lowest P-value was

annotated as "response to gibberellin stimulus". Both of these gene classes are well documented as being involved in the vernalization and flowering pathways so the enrichment in nonsynonymous variation in genes classified by these GO terms may well describe some of the fundamental genetics underlying differences in the growth habit of these cultivars (Zanewich and Rood, 1995; Pose *et al*., 2012).

SNP variation within QTL intervals

We characterized SNP densities inside of QTL intervals in order to understand if molecular variation within these regions was elevated relative to the genome-wide average. The SNP density averaged across the five QTL regions was higher than the genome-wide average but there was considerable variation in density among QTL (Table 4.2). Additionally, the average ratio of nonsynonymous to synonymous substitutions in QTL regions was not significantly higher than the genome-wide average.

Across all QTL, 1,582 genes were determined to carry at least one nonsynonymous substitution out of a total of 6,008 genes predicted to lie within the QTL intervals. Candidate genes were selected from this list based on GO annotations related to roots and/or flowering (Table 4.3; Supplementary Table 5.9). The largest number of candidates remaining in any of the QTL for a single trait was the 28 genes related to flowering located on A03. This represents 53% of the total genes annotated to flowering within this QTL. The QTL on C02 carried the largest number of total candidate genes with a final number of 36 and this number was almost equal between flowering and root candidates (Table 4.3). In addition, the proportion of genes qualifying as candidates within this interval was identical (67%) for both flowering and roots. Conversely, only three candidates for root development in the QTL on C07 carried nonsynonymous substitutions which represent 18% of the root-related genes contained within this interval. Ultimately, the final list of candidates was narrowed down to 62 flowering genes and 25 root genes.

We chose to conduct a more detailed analysis of the pleiotropic QTL on A10 (named *RPF.dry1*) to better understand the genetic factors potentially underlying the multiple traits this QTL is controlling.

Of the 42 candidate genes in *RPF.dry1*, 17 were associated with roots, 22 were associated with flowering and 3 were associated with both terms (Table 4.4). The average SNP density across these 42 candidate gene regions was 0.24% and ranged from 0 to 0.83%. Almost half of the initial list (19) showed no molecular variation and only 12 had one or more nonsynonymous mutations. Across these 12 genes, a total of 37 nonsynonymous polymorphisms were discovered. Bra009156 had a SNP density of 0.83%, the highest of any of the candidates analyzed. With respect to the ratio of nonsynonymous to synonymous substitutions, two genes had ratios above one where Bra008802 had the highest ratio of 3:1 and Bra008971 had a ratio of 2:1, indicating that these loci may have experienced positive selection. Analyses of the parent sequence and protein identities, relative to their *Arabidopsis* and *B. rapa* orthologs, found no significant difference in any of the three genes. This suggests both parents have undergone equal rates of molecular evolution at these loci and that the selection pressure on them has been equivalent (Tajima, 1993). Finally, 12 consecutive genes spanning from Bra008955 to Bra009327 showed no genetic variation.

Mapping in RPF.dry1

To confirm genomic locations, we developed a KASP marker assay (LGC Genomics, Teddington, Middlesex, UK) for a nonsynonymous SNP in every gene within the *RPF.dry1* QTL that carried one. In addition, we developed assays for several loci within a region which lacked markers in our original map between genetic map positions 72.7 and 83.5 cM, equating to a physical distance of over 940 kb. Twelve loci passed the criteria for suitable primer and SNP marker development and were used to genotype the DH population. Eight SNP markers ended up passing the primer design criteria and segregated at the expected 1:1 ratio for each parental allele. We re-constructed the genetic map and validated the location of these eight SNPs as all of them mapped to the predicted chromosome A10 and all segregated in the expected order based on their physical locations (Table 4.5). In addition, the original gap of nearly 11 cM between 72.7 and 83.5 cM was reduced to less than 6 cM with the addition of four markers that mapped in the interval.

We re-analyzed the association between genotype and drought adaptation phenotypes in order to understand if the additional markers increased the amount of phenotypic variation explained, relative to the original QTL results. In the original analyses, the most highly correlated SNP was located in Bra008998 at position 13,586,650. The new analyses revealed two markers located at 13,663,563 and 13,897,294 that improved the total amount of explained variation in root mass but did not, however, change the results for flowering time (Table 4.5). These results were confirmed in genome-wide QTL scans conducted using the new map (data not shown). In addition to its contribution to filling in the gap in our genetic map, the SNP at 13,897,294 was selected because of its close proximity to *FLC* (41 kb). The addition of this marker increased the R^2 value for root mass and shifted the peak LOD score away from the flowering time QTL peak. Despite separation in the location of the QTL peaks, the LOD 1.5 QTL confidence intervals for both root mass and flowering time are identical (73-80 cM).

As mentioned, *RPF.dry1* was selected as the focal QTL for this study due to the possibility that it is directly involved in the control of root mass. This conclusion is based on QTL scans which determined this locus to be significant even after conditioning upon flowering time, a genetically correlated trait (r_g = 0.48). We re-performed these analyses and, again, found an improvement in the fit of the model using the map incorporating NGS-derived markers as the LOD score for this locus increased by 1.2, resulting in a reduction in the LOD 1.5 confidence interval of over 20 cM to an interval that largely overlaps with the confidence intervals of QTL results gathered from analyses of the traits independently (Table 4.6). This reduces the number of candidate genes within the new confidence interval down to nine, all of which reside within the region of 12 consecutive invariant genes.

DISCUSSION

We previously mapped several quantitative trait loci (QTL) underlying drought-related traits in a bi-parental mapping population created from a cross between annual and spring *B. napus* cultivars. Here we create a draft genome and examine the genomic divergence of the parents of this mapping population. These data reveal genome-wide molecular variation differentiating the parent lines and identified major

gene classes which may broadly underlie the differential growth habit of the cultivars studied. In addition, candidate genes were identified at five QTL explaining variation in traits of importance to drought adaptation. In our focused analysis of *RPF.dry1*, we developed molecular markers from a subset of the SNPs we discovered and mapped them back onto the genetic map. QTL analyses were re-conducted using the new map, resulting in a 40% reduction in the QTL confidence interval.

Genome-wide variation differentiating the parent lines

Our study provides a first draft of the *B. napus* genome and the results provide a first glimpse of the genetic variants distinguishing the distinct genetic pools of annual and biennial *B. napus* (Hasan *et al.*, 2006; Bus *et al.*, 2011). The A genome showed higher SNP divergence than the C genome, a result that is consistent with previous research (Bancroft *et al.*, 2011; Bus *et al.*, 2011; Delourme *et al.*, 2013; Zhao *et al.*, 2013). These previous studies have all included Asian cultivars which are known to have been bred with diverse *B. rapa* types. It is hypothesized that this breeding history may be the basis of the higher genetic diversity observed in the A genome of *B. napus* (Qian *et al.*, 2006). However, no Asian or *B.rapa* parent lines exist in the pedigrees of IMC106RR or Wichita so this repeated discovery may be due to generally higher levels of diversity contributed from the A genomes involved in the multiple independent allpolyploidy events that created modern *B. napus* (Song and Osborn, 1992; Allender and King, 2010). SNP diversity also varied greatly among chromosomes, particularly in the C genome. This may be an artifact of the severe bottleneck invoked by selection against erucic acid and glucosinolate accumulation during the creation of modern canola quality varieties (Downey and Röbbelen, 1998; Sharpe and Lydiate, 2003). QTL have been repeatedly mapped onto A09 and C09 for glucosinolates content in *Brassica* species (Howell *et al.*, 2003; Feng *et al.*, 2012) which are among the least divergent chromosomes in IMC106RR and Wichita (Table 4.1).

Contrary to the lower overall divergence discovered in the C genome, the ratio of nonsynonymous substitutions to synonymous substitutions was significantly higher in the C genome. This result is in agreement with the recent study by Zhao *et al.* (2013) comparing the diploid *B. rapa* and *B.*

oleracea genomes. This is evidence of a higher rate of purifying selection in the A genome than the C genome. Zhao *et al*. (2013) also found evidence that a higher rate of recombination in the A genome may be one of the mechanisms underlying this phenomenon and provides an interesting hypothesis worthy of future research (Hill and Robertson, 1968).

Gene annotations captured in GO terms provide a useful framework for identifying candidate genes and gene classes associated with particular phenotypes. We have identified a list of 25 GO terms associated with genes enriched for nonsynonymous substitutions. The GO terms in this list are likely to capture many of the genes broadly differentiating annual and biennial growth habit types. For instance, "response to gibberellin stimulus" was identified as an enriched term and the role of gibberellins in flower induction in biennial species is well established where induction of flowering is possible through the application of gibberellic acid (Zeevaart, 1983; Mutasa-Göttgens and Hedden, 2009). Further, research in *Brassica* species has shown that vernalization increases GA biosynthesis and metabolism (Zanewich and Rood, 1995) and that flowering is delayed in GA deficient mutants (Rood *et al*., 1989).

Candidate genes

Genome sequencing was recently shown to be an effective means of rapidly identifying candidate genes contained within QTL of *B. napus* (Tollenaere *et al*., 2012). In our study, we narrowed an initial list of 6,008 genes contained across five QTL intervals down to 1,582 genes carrying one or more nonsynonymous substitutions which was narrowed further to a final list of 87 genes annotated to flowering or root (Table 4.3).

Flowering time has major impacts on fitness and yield under almost all conditions and an understanding of its genetic controls has implications to all facets of plant biology (Stinchcombe *et al*., 2004; Jung and Müller, 2009). In our study, we have identified a total of 62 candidate flowering time genes across four QTL and several of these represent strong candidates. Bra020249 and Bol008947 are located in the flowering time QTL on A02 and C02 and are orthologs of At5g60410 (a.k.a *ATSIZ1*, *SIZ1*), a negative regulator of flowering (Jin *et al*., 2008). Of the 28 candidates located in the QTL interval on

A03, the *Arabidopsis* orthologs of Bra001357 (At3g10390: *FLD*; He *et al*., 2003), Bra001729 (At3g18990: *VRN1*, *REM39*; Bastow *et al*., 2004), Bra013162 (At2g06210: *ELF8*, *VIP6*; Oh *et al*., 2004), Bra000392 and Bra000393 (At2g45650: *AGL6*; Yoo *et al*., 2011) all regulate the clade of genes known as *FLC*/*MAF* which, in turn, impact flowering time significantly (Posé *et al*., 2012). Of the eight candidate flowering genes on A10, Bra008802 is the most interesting based on what is known of its *Arabidopsis* ortholog (At5g13790: *AGL15*). *AGL15* is a MADS domain transcription factor involved in the initiation of flowering through suppression of the floral integrator FT in the photoperiod flowering pathway (Adamczyk *et al*., 2007). In addition to flowering time, it has been shown that *AGL15* is involved in the regulation of cold shock proteins (CSPs) expression (Nakaminami *et al*., 2009). Cold shock proteins are involved in the acclimation to cold temperatures and freezing tolerance (Sasaki and Imai, 2009) and are likely to have been important in the evolution of annual and biennial growth types, as these types are exposed to cold temperatures with different probabilities and at different developmental stages. Finally, the *Arabidopsis* orthologs of Bol018187 (At1g79280: *NUA*) and Bol036052 (At5g22290: *anac089*) contained within the QTL on C02 have been shown to have a direct role in the regulation of flowering time (Jacob *et al*., 2007; Li *et al*., 2010). Ultimately, this study has provided a well-supported list of candidate flowering time genes for further research.

The root system plays a vital role in water and nutrient acquisition and, therefore, has an essential role in crop productivity (Sharp and Davies, 1979). We have identified a total of 25 candidate genes involved in root development across three QTL of *B. napus*. Of the three candidates on A10, Bra009156 is noteworthy because the mutant phenotype of its *Arabidopsis* ortholog (At5g05980: *ATDFB*, *DFB*) is characterized by significant decreases in seedling root growth rate (Srivastava *et al*., 2011). We measured seedling root growth rate of 40 DH lines selected based on parental haplotype at locus (20 lines representing each parent) using the method described in Mullen *et al*. (1998). A previous field study has shown that the parental haplotypes, as represented by these lines, differ in several aspects of root morphology including taproot dry mass, diameter and length (Chapter 3). There was no difference in seedling root growth rates of these 40 lines due to the effect of haplotype ($P = 0.53$), indicating that this

locus may not be impacting root growth during this early phase of development. Of the list of 19 root related candidates identified in the QTL region on C02, Bol014250 (At5g57090: *EIR1*, *WAV6*, *PIN2*, *AGR*, *AGR1*) represents the most interesting due to its demonstrated role in growth and development of the primary root through its control of auxin distribution (Blilou *et al*., 2005). None of the three genes identified in the QTL on C07 have a strong body of literature to support them as exceptional candidates. However, QTL have been repeatedly mapped to C07 for traits such as root length and mass (Hammond *et al*., 2009; Yang *et al*., 2010; Yang *et al*., 2011; Shi *et al*., 2012) suggesting there are genes of clear importance to root biology harbored on this chromosome.

It must be acknowledged that genes specific to the *B. napus* lineage may also be carried within these QTL which are not captured by the *B. rapa* draft genome as it has been shown that nearly 6% of gene families within the *B. rapa* genome are lineage-specific when compared to *Arabidopsis* (Wang *et al*., 2011). This number will likely be lower when comparing the much more closely related *B. napus* and *B. rapa* genomes. It is also probable that gene translocations have occurred during hybridization of the *B. napus* progenitor species (Town *et al*., 2006; Gaeta *et al*., 2007), resulting in the presence of other root and/or flowering related candidates within the *B. napus* genome that are unaccounted for by the *B. rapa* reference.

Re-mapping at RPF.dry1

The most highly correlated SNP identified in the re-mapping of *RPF.dry1* is only 41 kb away from Bra009055 (At5g10140: *FLC*). *FLC* is a MADS-box transcription factor that has been identified as an important regulator of the vernalization and autonomous flowering pathways in Brassica (Kole *et al*., 2001) due to its location within a network of genes (Michaels, 2009). In addition, it was recently shown that the *FLC* protein binds to over 500 other sites throughout the *Arabidopsis* genome and that those sites were enriched in gene ontology (GO) categories annotated with stress response and abiotic stimulus (Deng *et al*., 2011). Orthologous QTL regions encompassing *FLC* have been recurrently associated with

flowering time QTL (Osborn *et al*., 1997; Schranz *et al*., 2002; Osborn and Lukens, 2003; Long *et al*., 2007; Shi *et al*., 2009) as well as root QTL (Lou *et al*., 2007; Lu *et al*., 2008; Yang *et al*., 2010; Kubo *et al*., 2010) in the genus *Brassica*. Cumulatively, this leads us to hypothesize that the gene underlying these pleiotropic QTL may be *FLC*. We found no molecular variation in the coding region of *FLC* which was not surprising since most genetic variation in *FLC* exists in regulatory regions and introns (Gazzani *et al*., 2003; Yuan *et al*., 2009; Zou *et al*., 2012). We extended our search 200 bp beyond the putative 1 kb coding region upstream of *FLC* to look for the presence of a transposable element associated with the vernalization response in different *B. napus* genotypes as reported by Hou *et al*. (2012) but found the region to be invariate. The most likely explanation for the lack of variation in *FLC* is that it is not the candidate gene as it has been reported that even when genetic variation exists in *FLC*, it may not associate with flowering time (Razi *et al*., 2008). An alternative explanation is that the original Wichita source used to make the DH population wasn't fully inbred and was segregating at this locus. To test this, we re-planted seed from the original source and genotyped ten resulting plants with markers at the *RPF.dry1* locus. The results of this analysis confirmed the *RPF.dry1* locus to be segregating among individuals in the original seed source, a result that is possible given the breeding history for this cultivar (Rife *et al*., 2001). We are currently re-sequencing *FLC* in DH lines which we expect to represent the haplotypes of the original individual parents selected for crossing. These results should allow us to conclusively rule out *FLC* as a candidate or provide reason for more research on its function.

Conclusions and next steps

The prevalence and severity of global drought is expected to increase over the next 50 years (Dai, 2010). However, very little research has concentrated on drought adaptation in *Brassica napus* despite the increasing global importance of its oil and meal. An understanding of the genetic architecture underlying the mechanisms of drought adaptation in *B. napus*, including roots, has practical applications to breeding programs focused on sustaining yield gains into the future. Ongoing advancements in the genetics and genomics tools of the *Brassica* species are enabling such efforts (Snowdon and Inguez Luy, 2012; Edwards, Batley and Snowdon, 2013).

Sequencing and assembly of the *B. napus* genome, including both parent lines of our mapping population, has provided us a genome-scale view of the molecular variation differentiating them. The higher SNP density observed in the A genome suggests that additional C genome diversity accessed from other *Brassica* species may improve genetic gains of *B. napus* breeding and selection programs (Rahman *et al*., 2011). Similarly, the large variation discovered among chromosomes suggests that particular chromosomes could be specifically targeted for increasing genetic diversity. The conclusion of a significantly higher ratio of nonsynonymous to synonymous substitutions in the C genome is in agreement with previous research (Zhao *et al*., 2013) and certainly deserves further research into the extent of this phenomenon and the factors driving it.

In addition to the genome-wide analyses, we have also gained a high resolution view of the specific variation occurring within five drought adaptation QTL. This variation allowed us to create a list of 87 candidate genes annotated to flowering and/or roots. These genes and the nonsynonymous substitutions they harbor will provide clear hypotheses for future studies on mutant and transgenic lines in understanding the functional genes underlying these QTL (Lovell *et al*., 2013; Uga *et al*., 2013; Huang *et al*., 2013). As we have demonstrated in our detailed analysis of *RPF.dry1*, they can also be used in developing high-throughput molecular marker assays which can be used in fine-mapping of each of these loci. With regard to *RPF.dry1*, the additional molecular markers added to the genetic map resulted in an improvement in the fit of the QTL model which appears to be largely the result of additional SNPs in close physical proximity to *FLC*. This result certainly merits further characterization of this gene region using DH lines which we expect to directly represent the molecular signature of the individuals used to create the population.

TABLES AND FIGURES

Table 4.1 Average non-coding, synonymous, nonsynonymous SNPs, total SNP density (± s.d.) and ratio of nonsynonymous to synonymous substitutions (dN/dS) occurring between IMC106RR and Wichita across chromosomes of the A and C genomes.

Chromsome	SNP Density			% Total	dN/dS
	Non-coding	Synonymous	Nonsynonymous		
A01	0.31%	0.18%	0.10%	0.30%	0.53
A02	0.18%	0.09%	0.05%	0.17%	0.62
A03	0.28%	0.16%	0.09%	0.28%	0.56
A04	0.26%	0.14%	0.09%	0.26%	0.59
A05	0.28%	0.15%	0.08%	0.27%	0.56
A06	0.22%	0.13%	0.07%	0.22%	0.56
A07	0.27%	0.14%	0.08%	0.26%	0.59
A08	0.27%	0.12%	0.08%	0.26%	0.66
A09	0.19%	0.08%	0.05%	0.18%	0.64
A10	0.30%	0.19%	0.10%	0.30%	0.53
A Genome Avg	**0.25 ± 0.04%**	**0.13 ± 0.04%**	**0.08 ± 0.02%**	**0.25 ± 0.04%**	**0.57 ± 0.04**
C01	0.31%	0.12%	0.09%	0.30%	0.77
C02	0.27%	0.15%	0.11%	0.27%	0.75
C03	0.15%	0.08%	0.06%	0.15%	0.76
C04	0.19%	0.07%	0.06%	0.18%	0.86
C05	0.10%	0.06%	0.04%	0.10%	0.77
C06	0.15%	0.07%	0.06%	0.15%	0.86
C07	0.18%	0.06%	0.05%	0.17%	0.85
C08	0.13%	0.06%	0.05%	0.13%	0.89
C09	0.10%	0.05%	0.04%	0.10%	0.78
C Genome Avg	**0.22 ± 0.07%**	**0.10 ± 0.03%**	**0.08 ± 0.02%**	**0.17 ± 0.07%**	**0.80 ± 0.06**
Genome-wide Avg	**0.21 ± 0.07%**	**0.11 ± 0.04%**	**0.07 ± 0.02%**	**0.20 ± 0.07%**	**0.65 ± 0.13**

Table 4.2 Average non-coding, synonymous, nonsynonymous SNPs, total SNP density and ratio of nonsynonymous to synonymous substitutions (dN/dS) occurring between IMC106RR and Wichita across QTL regions associated with flowering time and/or root mass.

QTL Chromosome	Trait	SNP Density			% Total	dN/dS
		Noncoding	Synonymous	Nonsynonymous		
A02	Flowering	0.30%	0.07%	0.06%	0.22%	0.84
A03	Flowering	0.32%	0.10%	0.07%	0.23%	0.69
A10	Both	0.25%	0.15%	0.07%	0.20%	0.48
C02	Both	0.44%	0.21%	0.16%	0.39%	0.74
C07	Root Mass	0.20%	0.04%	0.05%	0.18%	1.29
	Mean	0.30%	0.11%	0.08%	0.24%	0.81

Table 4.3 Summary of the total number of genes carrying nonsynonymous substitutions and the number of flower and/or root candidates carried within each QTL interval.

QTL Chromosome	Trait	No. genes w/ nonsynonymous variation		
		Across QTL	Flower candidates	Root candidates
A02	Flowering	243	9	n/a
A03	Flowering	602	28	n/a
A10	Both	212	8	3
C02	Both	381	17	19
C07	Root Mass	144	n/a	3
	Total	316	62	25

Table 4.4 Summary of sequence coverage, polymorphism results and gene information for the 42 *Brassica rapa* candidate genes selected within the *RPF.dry1* QTL.

Genes in bold carried polymorphisms for which KASP assays were designed and used in mapping.

Bra Gene#	At ortholog	GO Annotation	CDS size (bp)	No. SNPs					
				5'	Synonymous	Nonsynonymous	Intronic	3'	SNP Density (%)
Bra008668	AT5G15850	Flower	720	0	0	0	0	0	0.00
Bra008669	AT5G15840	Flower	1101	0	0	0	0	0	0.00
Bra008674	AT5G15800	Flower	759	0	0	0	0	10	0.24
Bra008689	AT5G15580	Root	2691	0	0	0	0	0	0.00
Bra008740	AT5G14750	Root	594	0	0	0	2	11	0.39
Bra008781	**AT5G14080**	**Flower**	849	**9**	**4**	**4**	**0**	**8**	**0.71**
Bra008787	AT5G14010	Flower	462	2	1	0	0	1	0.80
Bra008802	AT5G13790	Flower	795	13	1	3	10	0	0.44
Bra008838	AT5G13300	Both	2481	0	6	0	8	4	0.28
Bra008839	AT5G13290	Both	1188	0	0	0	0	14	0.43
Bra008853	AT5G13140	Root	702	0	0	0	0	0	0.00
Bra008858	AT5G13080	Root	444	13	2	0	2	0	0.20
Bra008888	AT5G12430	Root	3657	18	8	8	1	1	0.48
Bra008889	AT5G12410	Flower	1098	3	5	1	8	0	0.46
Bra008890	AT5G12400	Root	5184	0	0	0	0	0	0.00
Bra008898	AT5G12330	Root	918	1	0	0	0	0	0.03
Bra008903	AT5G12250	Root	1347	2	0	0	0	0	0.06

Bra008937	AT5G11780	Flower	1479	0	3	2	0	1	0.15
Bra008955	AT5G11530	Flower	3321	0	0	0	3	0	0.03
Bra008966	**AT5G11390**	**Root**	1971	**5**	**8**	**2**	**0**	**0**	**0.26**
Bra008971	AT5G11320	Flower	1236	0	1	2	2	2	0.17
Bra008976	AT5G11260	Flower	495	4	0	0	1	0	0.16
Bra008979	AT5G11240	Flower	1851	19	0	1	1	1	0.32
Bra008995	AT5G11030	Root	1755	0	0	0	0	0	0.00
Bra008998	AT5G11000	Root	1167	0	0	0	0	0	0.00
Bra009011	AT5G10720	Root	2697	0	0	0	0	0	0.00
Bra009026	AT5G10510	Both	1725	0	0	0	0	0	0.00
Bra009029	AT5G10480	Root	531	0	0	0	0	0	0.00
Bra009047	AT5G10250	Root	1893	0	0	0	0	0	0.00
Bra009053	AT4G27010	Flower	1071	0	0	0	0	0	0.00
Bra009055	AT5G10140	Flower	621	0	0	0	0	0	0.00
Bra009075	AT2G19930	Flower	1038	0	0	0	0	0	0.00
Bra009081	AT5G09810	Root	1062	0	0	0	0	0	0.00
Bra009353	AT5G08580	Flower	1140	0	0	0	0	0	0.00
Bra009327	AT5G08180	Flower	471	0	0	0	0	0	0.00
Bra009287	**AT5G07180**	**Flower**	2811	**0**	**2**	**1**	**15**	**1**	**0.27**
Bra009221	AT5G06720	Flower	1077	0	0	0	0	0	0.00
Bra009209	AT5G06590	Flower	924	0	0	0	0	0	0.00

Bra009205	AT5G06550	Flower	1539	1	11	0	2	5	0.53
Bra009167	**AT5G06170**	**Flower**	1479	**0**	**7**	**2**	**0**	**0**	**0.24**
Bra009166	AT5G06140	Root	1173	1	8	0	3	7	0.32
Bra009156	**AT5G05980**	**Root**	2463	**7**	**17**	**10**	**32**	**10**	**0.83**
		Average	1475.7	2.3	2.0	0.9	2.1	1.8	0.24

Table 4.5 Gene name, genetic location, physical location and percent variance explained (R^2) between markers and drought adaptation phenotypes (Root Pulling Force, RPF; Days to Flowering, DTF).

Markers in bold indicate genes carrying SNPs for which molecular markers were designed.

SNP Location			Trait R^2	
Bra Gene No.	Genetic Location (cM)	Physical Location	RPF	DTF
Bra008781	***54.4***	***12,625,712***	***11.9%***	***8.5%***
Bra008742	56.6	12,464,249	12.0%	8.7%
n/a	56.6	12,641,628	12.1%	8.7%
Bra008852	61.3	12,947,450	15.7%	11.8%
n/a	62.2	12,966,544	15.6%	11.9%
Bra008945	69.3	13,329,042	17.7%	13.7%
Bra008963	70.7	13,443,382	18.7%	14.4%
Bra008966	***70.7***	***13,462,920***	***17.8%***	***12.1%***
Bra008968	70.7	13,469,370	18.7%	14.4%
Bra008970	71.1	13,475,553	19.1%	14.5%
Bra008998	72.2	13,586,650	19.0%	15.5%
n/a	72.7	13,613,487	18.5%	15.3%
Bra009011	***73.3***	***13,663,563***	***19.9%***	***14.7%***
n/a	***75.6***	***13,897,294***	***21.3%***	***14.1%***
Bra009323	***82.1***	***14,311,147***	***15.2%***	***11.6%***
Bra009287	***82.6***	***14,479,147***	***10.3%***	***12.3%***
Bra009274	83.5	14,557,194	10.3%	12.3%
Bra009253	83.5	14,643,017	7.8%	11.7%
Bra009203	84.8	14,858,181	7.8%	9.3%
Bra009164	86.2	14,985,132	7.3%	8.3%
Bra009167	***86.6***	***14,976,123***	***6.7%***	***7.7%***
Bra009156	***87.5***	***15,021,599***	***6.3%***	***7.4%***

Markers are ordered based on genetic map location

Table 4.6 Comparison of model fit parameters for *RPF.dry1* generated from genome-wide QTL scans using the original map (Chapter 3) versus the NGS map (incorporating additional markers).

	Type III SS	LOD Value	% Variation Explained	F-value	LOD 1.5 C.I.
Original Map	1229	3.27	5.7	15.3	55.8-86.3
NGS Map	1412	3.74	6.5	17.7	73.3-82.1

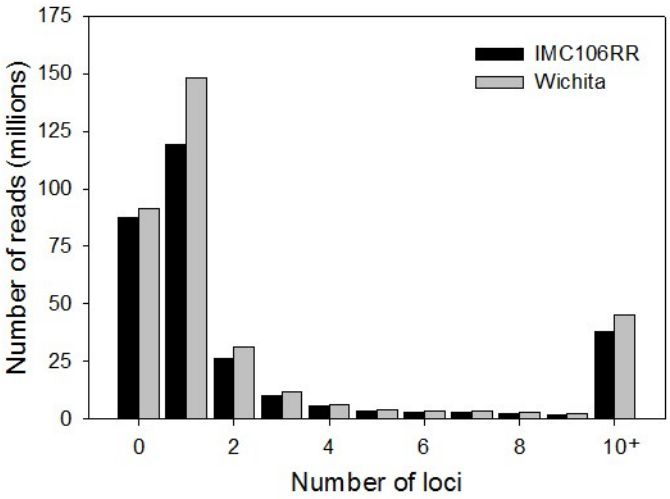


Figure 4.1 Histogram of the number of loci to which 100-bp Illumina HiSeq reads originating from IMC106RR and Wichita map to.

REFERENCES

Adamczyk BJ, Lehti-Shiu MD, Fernandez DE. 2007. The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in Arabidopsis. *The Plant Journal* **50,** 1007–1019.

Allender CJ, King GJ. 2010. Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. *BMC Plant Biology* **10,** 54.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Bancroft I, Morgan C, Fraser F, *et al.* 2011. Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nature Biotechnology* **29**, 762-766.

Bastow R, MyIne JS, Lister C, Lippman Z, Martienssen RA, Dean C. 2004. Vernalization requires epigenetic silencing of *FLC* by histone methylation. *Nature* **427,** 164-167.

Blilou I, Xu J, Wildwater M, Willemsen V, Paponov I, Friml J, Heidstra R, Aida M, Palme K, Scheres B. 2005. The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. *Nature* **433,** 39–44.

Boyer JS. 1982. Plant productivity and environment. *Science* **218**, 443–448.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889-890.

Broman KW, Sen S. 2009. A guide to QTL mapping with R/qtl. New York: Springer.

Bus A, Hecht J, Huettel B, Reinhardt R, Stich B. 2012. High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *BMC Genomics*, **13**, 281.

Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M. 1998. Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* **41**, 62-69.

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. 2011. BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biology* **11,** 136.

Churchill GA, Doerge R. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-971.

Dai A. 2011. Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change* **2**, 45–65.

Delourme R, Falentin C, Fomeju BF, *et al.* 2013. High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC Genomics* **14,** 120.

Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES. 2011. FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **108,** 6680-6685.

Diers BW, Osborn TC. 1994. Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. *Theoretical and Applied Genetics* **88**, 662-668.

Downey RK, Röbbelen G. 1989. Brassica Species. In: Downey RK, Röbbelen G and Ashri A eds. *Oil Crops of the World*. New York: McGraw-Hill, 339-362.

Durstewitz G, Polley A, Plieske J, Luerssen H, Graner EM, Wieseke R, Ganal MW. 2010. SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species *Brassica napus*. *Genome* **53**: 948–956.

Edwards D, Henry R J, Edwards K J. 2012. Preface: advances in DNA sequencing accelerating plant biotechnology. *Plant Biotechnology Journal* **10**, 621–622.

Edwards D, Batley J, Snowdon RJ. 2013. Accessing complex crop genomes with next-generation sequencing. *Theoretical and Applied Genetics* **126**, 1-11.

Feng J, Long Y, Shi L, Shi J, Barker G, Meng J. 2012. Characterization of metabolite quantitative trait loci and metabolic networks that control glucosinolate concentration in the seeds and leaves of *Brassica napus*. **193**, 96–108.

Flicek P, Birney E. 2009. Sense from sequence reads: methods for alignment and assembly. *Nature Methods* **6**, S6–12.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *The Plant Cell* **19,** 3403–3417.

Gazzani S, Gendall AR, Lister C, Dean C. 2003. Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. *Plant Physiology* **132,** 1107-1114.

Haley CS, Knott SA. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315-324.

Hammond JP, Broadley MR, White PJ, *et al*. 2009. Shoot yield drives phosphorus use efficiency in *Brassica oleracea* and correlates with root architecture traits. *Journal of Experimental Botany* **60**, 1953-1968.

He Y, Michaels SD, Amasino RA. 2006. Regulation of flowering time by histone acetylation in *Arabidopsis*. *Science* **302,** 1751-1754.

Hill WG, Robertson A. 1968. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38,** 226:-231.

Hirota A, Kato T, Fukaki H, Aida M, Tasaka M. 2007. The Auxin-Regulated AP2/EREBP Gene PUCHI Is Required for Morphogenesis in the Early Lateral Root Primordium of Arabidopsis. *The Plant Cell* **19**, 2156–2168.

Hou J, Long Y, Raman H, *et al.* 2012. A *Tourist*-like MITE insertion in the upstream region of the *BnFLC.A10* gene is associated with vernalization requirement in rapeseed (*Brassica napus* L.). *BMC Plant Biology* **12,** 238.

Howell PM, Sharpe AG, Lydiate DJ. 2003. Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*Brassica napus*). *Genome* **46,** 454-460.

Huang X, Ding J, Effgen S, Turck F, Koornneef M. 2013. Multiple loci and genetic interactions involving flowering time genes regulate stem branching among natural variants of Arabidopsis. *New Phytologist* **199,** 843-857.

Jacob Y, Mongkolsiriwatana C, Veley KM, Kim SY, Michaels SD. 2007. The nuclear pore protein AtTPR is required for RNA homeostasis, flowering time, and auxin signaling. *Plant Physiology* **144,** 1383-1390.

Jin JB, Jin YH, Lee J, *et al.* 2008. The SUMO E3 ligase, *AtSIZ1*, regulates flowering by

controlling a salicylic acid-mediated floral promotion pathway and through affects on *FLC* chromatin structure. *The Plant Journal* **53**, 530–540.

Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ. 2005. Evolution of Genome Size in Brassicaceae. *Annals of Botany* **95**, 229-235.

Juenger JE. 2013. Natural variation and genetic constraints on drought tolerance. *Current Opinion in Plant Biology* **16**, 274–281.

Jung C, Müller AE. 2009. Flowering time control and applications in plant breeding. *Trends in Plant Science*. **14,** 563-573.

Kole C, Quijada P, Michaels SD, Amasino RM, Osborn TC. 2001. Evidence for homology of flowering-time genes *VFR2* from *Brassica rapa* and *FLC* from *Arabidopsis thaliana*. *Theoretical and Applied Genetics* **102,** 425–430.

Kosambi DD. 1944. The estimation of map distances from recombination values. *Annals of Eugenics* **12**, 172-175.

Kubo N, Saito M, Tsukazaki H, Kondo T, Matsumoto S. 2010. Detection of quantitative trait loci controlling morphological traits in *Brassica rapa* L. Breeding Science **60,** 164-171.

Lanceras JC, Pantuwan G, Jongdee B, Toojinda T. 2004. Quantitative trait loci associated with drought tolerance at reproductive stage in rice. *Plant Physiology* **135**, 384-399.

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

Li JQ, Zhang J, Wang XC, Chen J. 2010. A membrane-tethered transcription factor ANAC089 negatively regulates floral initiation in *Arabidopsis thaliana. Science China Life Sciences*. **53,** 1299-1306.

Long Y, Shi J, Qiu D, *et al.* 2007. Flowering time quantitative trait loci analysis of oilseed *Brassica* in multiple environments and genomewide alignment with *Arabidopsis*. *Genetics* **177,** 2433-2444.

Lou P, Zhao J, Kim JS, Shen S, Del Carpio DP, Song X, Jin M, Vreugdenhill D, Wang X, Koornneef M, Bonnema G. 2007. Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. Journal of Experimental Botany **58,** 4005-4016.

Lovell JT, Juenger TE, Michaels SD, *et al.* 2013. Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. *Proceedings of the Royal Society B: Biological Sciences* **280,** 20131043.

Lu G, Cao J, Yu X, Xiang X, Chen H. 2008. Mapping QTLs for root morphological traits in Brassica rapa L. based on AFLP and RAPD markers. *Journal of Applied Genetics* **49,** 23-31.

Ludlow MM. 1989. Strategies of response to water stress. In: Kreeb KH, Richter H, Hinckley TM, eds. *Structural and functional responses to environmental stresses*. The Hague: SPB Academic, 269-281.

Lühs W, Seyis F, Frauen M, Busch H, Frese L, Willner E, Friedt W, Gustafsson M Poulsen G. 2003. Development and evaluation of a *Brassica napus* core collection. In: Knüpffer H and Ochsmann J eds. *Rudolf Mansfeld and Plant Genetic Resources*. Bonn: ZADI/IBV, 284–289**.**

Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW. 2009. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics* **181**, 1077-1086.

Michaels SD. 2009. Flowering time regulation produces much fruit. *Current Opinion in Plant Biology* **12,** 75-80.

Mullen JL, Turk E, Johnson K, Wolverton C, Ishikawa H, Simmons C, Soll D, Evans ML. 1998. Root-growth behavior of the Arabidopsis mutant *rgr1*: Roles of gravitropism and circumnutation in the waving/coiling phenomenon. *Plant Physiology* **118**, 1139–1145.

Mutasa-Göttgens E, Hedden P. 2009. Gibberellin as a factor in floral regulatory networks. *Journal of Experimental Botany* **60,** 1979–1989.

Nakaminami K, Hill K, Perry SE, Sentoku N, Long JA, Karlson DT. 20009. Arabidopsis cold shock domain proteins: relationships to floral and silique development. *Journal of Experimental Botany* **60**, 1047–1062.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3,** 418-426.

Oh S, Zhang H, Ludwig P, van Nocker S. 2004. A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the Arabidopsis *FLC/MAF* MADS box gene family. *The Plant Cell* **16,** 2940–2953.

Oliver MJ, Cushman JC, Koster KL. 2010. Dehydration tolerance in plants. In: Sunkar R, ed. *Plant Stress Tolerance: Methods in Molecular Biology*. New York: Humana Press, 3-24.

Osborn TC, Kale C, Parkin IAP, Sharpe AG, Kuiper M. Lydiate DJ, Trick M. 1997. Comparison of flowering time genes in *Brassica rapa, B. napus and Arabidopsis thaliana*. *Genetics* **146**, 1123–1129.

Osborn T C, Lukens L. 2003. The molecular genetic basis of flowering time variation in *Brassica* species. In: Nagata T, Tabata S, eds, *Biotechnology in Agriculture and Forestry: Brassicas and Legumes, From Gene Structure to Breeding*. Berlin: Springer-Verlag, 69–86.

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ. 1995. Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* **38**, 1122-1131.

Parkin IAP, SM Gulden, AG Sharpe, L Lukens, M Trick, TC Osborn, DJ Lydiate. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. G*enetics* **171,** 765-781.

Posé D, Yant L, Schmid M. 2012. The end of innocence: flowering networks explode in complexity. *Current Opinion in Plant Biology* **15**, 45-50.

Qian W, Meng J, Li M, Frauen M, Sass O, Noack J, Jung C. Introgression of genomic components from Chinese *Brassica rapa* contributes to widening the genetic diversity in rapeseed (*B. napus* L.), with emphasis on the evolution of Chinese rapeseed. *Theoretical and Applied Genetics* **113,** 49-54.

Quijada PA, Udall JA, Polewicz H, Vogelzang RD, Osborn TC. 2004. Phenotypic effects of introgressing French winter germplasm into hybrid spring canola. *Crop Science* **44**,1982–1989.

Quijada PA, Udall JA, Lambert B, Osborn TC. 2006. Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm. *Theoretical and Applied Genetics* **113**, 549-561.

Rahman MH, Bennett RA, Yang RC, Kebede B, Thiagarajah MR. 2011. Exploitation of the late flowering species *Brassica oleracea* L. for the improvement of earliness in *B. napus* L.: an untraditional approach. *Euphytica* **177,** 365–374.

Razi H, Howell EC, Newbury HJ, Kearsey MJ. 2008. Does sequence polymorphism of *FLC* paralogues underlie fowering time QTL in *Brassica oleracea*? *Theoretical and Applied Genetics* **116,** 179-192.

Rife C, Auld DL, Sunderman HD, Heer WF, Baltensperger DD, Nelson LA, Johnson DL, Bordovsky D, Minor HC. 2001. Registration of 'Wichita' winter rapeseed. *Crop Science* **41**, 263-a-264.

Rood SB, Pearce D, Williams PH, Pharis RP. 1989. A gibberellin deficient *Brassica* mutant-rosette. *Plant Physiology* **89,** 482-487.

Sasaki K, Imai R. 2012. Pleiotropic roles of cold shock domain proteins in plants. *Frontiers in Plant Science.* **2,** 1-6.

Schranz ME, Quijada P, Sung SB, Lukens L, Amasino R, Osborn TC. 2002. Characterization and effects of the replicated flowering time gene *FLC* in *Brassica rapa*. *Genetics* **162,** 1457-1468.

Sharp RE, Davies WJ. 1979. Solute regulation and growth by roots and shoots of water-stressed Maize plants. *Planta* **147,** 43-49.

Sharpe AG, Lydiate DJ. 2003. Mapping the mosaic of ancestral genotypes in a cultivar of oilseed rape (*Brassica napus*) selected via pedigree breeding. *Genome* **46,** 461-468.

Shi J, Li R, Qiu D, Jiang C, Long Y, Morgan C, Bancroft I, Zhao J, Meng J. 2009. Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. *Genetics* **182,** 851-861.

Shi L, Yang J, Liu J, Li R, Long Y, Xu F, Meng J. 2012. Identification of quantitative trait loci associated with low boron stress that regulate root and shoot growth in *Brassica napus* seedlings. Molecular Breeding **30,** 393-406.

Snowdon RJ, Inguez Luy FL. 2012. Potential to improve oilseed rape and canola breeding in the genomics era. *Plant Breeding* **131**, 351—360.

Song K, Osborn TC. 1992. Polyphyletic origins of *Brassica napus*: new evidence based on organelle and nuclear RFLP analyses. *Genome* **35,** 992-1001.

Srivastava AC, Ramos-Parra PA, Bedair M, Robledo-Hernández AL, Tang Y, Sumner LW, de la Garza RI, Blancaflor EB. 2011. The folylpolyglutamate synthetase plastidial isoform is required for postembryonic root development in Arabidopsis. *Plant Physiology* **155,** 1237–1251.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences, USA* **101**, 4712-4717.

Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599-607.

Tollenaere R, Hayward A, Dalton-Morgan J, *et al.* 2012. Identification and characterization of candidate Rlm4 blackleg resistance genes in *Brassica napus* using next-generation sequencing. *Plant Biotechnology Journal* **10**, 709-715.

Town CD, Cheung F, Maiti R, *et al.* 2006. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *The Plant Cell* **18,** 1348–1359.

Trick M, Long Y, Meng J, Bancroft I. 2009. Single nucleotide polymorphism (SNP) discovery in the poly-ploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnology Journal* **7**, 334-346.

Tuberosa R, Salvi S, Sanguineti MC, Maccaferri M, Giuliani S, Landi P. 2003. Searching for quantitative trait loci controlling root traits in maize: a critical appraisal. *Plant and Soil* **255**, 35-54.

Uga Y, Sugimoto K, Ogawa S, *et al.* 2013. Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nature Genetics* **45,** 1097–1102.

U N. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **7,** 389-452.

Van Ooijen JW, Voorrips RE. 2001. *Joinmap version 3.0: software for the calculation of genetic linkage maps*. Wageningen: Plant Research International.

Wang X, Wang H, Wang J, *et al*. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43,** 1035-1039.

Whittaker JC, Thompson R, Visscher PM. 1996. On the mapping of QTL by regression of phenotype on marker-type. *Heredity* **77,** 23-32.

Yang M, Ding G, Shi L, Feng j, Xu F, Meng J .2010. Quantitative trait loci for root morphology in response to low phosphorus stress in *Brassica napus*. *Theoretical and Applied Genetics* **121,** 181-193.

Yang M, Ding G, Shi L, Xu F, Meng J .2011. Detection of QTL for phosphorus efficiency at vegetative stage in *Brassica napus*. *Plant and Soil* **339,** 97-111.

Yuan YX, Wu J, Sun RF, Zhang XW, Xu DH, Bonnema G, Wang XW. 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.

Yoo SK, Wu X, Lee JS, Ahn JH. 2011. *AGAMOUS-LIKE 6* is a floral promoter that negatively regulates the *FLC/MAF* clade genes and positively regulates *FT* in Arabidopsis. *The Plant Journal* **65,** 62-76.

Zanewich KP, Rood SB .1995. Vernalization and gibberellin physiology of winter canola. *Plant Physiology* **108,** 615-621.

Zeevaart JAD. 1983. Gibberellins and flowering. In: Crozier A, ed. *The biochemistry and physiology of gibberellins*. New York: Praeger, 333–374.

Zhao M, Du J, Lin F, Tong C, Yu J, Huang S, Wang X, Liu S, Ma J. 2013. Shifts in the evolutionary rate and intensity of purifying selection between two *Brassica* genomes revealed by analyses of orthologous transposons and relics of whole genome triplication. *The Plant Journal* **76**, 211-222.

Zou X, Suppanz I, Raman H, Hou J, Wang J, Long Y, Jung C, Meng J. 2012. Comparative analysis of *FLC* homologues in Brassicaceae provides insight into their role in the evolution of oilseed rape. *PLoS One* **7,** e45751.

CHAPTER 5: Supplementary Materials

Supplementary Table 5.1 Primer sequences used for KASP assays

Marker	Chr	Loc (cM)	Col-0 position (bp)	Sequence
1_112908	1	0	112908	gcctggatctctagttagcccaaaagaagcccaaga[a/g]tttttgcacataactcacaaattgttttggctgttgggt
1_2211035	1	5.145	2211035	tcaaaatatttaagttttatgttatcgacgttgacgcaa[a/g]caataaactattcccacaagtttactgattg
1-4142402	1	10.1	4142402	ggatngctaatgaataannattttgaaaagata[c/t]ttggcagtcccgaatgcgttgaacttntaactngtttccttcag
1_5923549	1	15.09	5923495	gattagctcatactttacaattgaatagtttattgtc[a/g]aggaacagttctgcatcttaaatctagctttagtgactgatgttttgtttggt
1_6572582	1	19.48	6572582	caaagagacgtggctgcagattcgcaccgcaaagc[t/c]attgatgggagaactacaacggaatctggttcaagcactcgactccgaa
1_6839609	1	20.16	6839609	agtactccggttattgtttatgttaccaccactcggga[g/c]catccaaaagagtagagatttgttggagcagccgcatgattttgttgc

tsu allele/kas allele

Supplementary Table 5.2 Summary of the number of introgressions discovered on each chromosome in the coarse and dense maps.

Chromosome	homozygous		heterozygous		Total		Δ
	Dense	Coarse	Dense	Coarse	Dense	Coarse	
1	33	21	44	23	77	44	33
2	10	8	24	19	34	27	7
3	27	9	39	11	66	20	46
4	22	14	57	32	79	46	33
5	9	5	23	18	32	23	9
Total	101	57	187	103	288	160	128

Supplementary Table 5.3 Summary of candidate genes in QTL interval selected based on relevant GO terms.

Locus Identifier	GO term
AT1G03060	response to abscisic acid stimulus
	(GO:0009737)
AT1G03880	cellular response to abscisic acid stimulus
	(GO:0071215)
AT1G05470	response to abscisic acid stimulus
	(GO:0009737)
AT1G05510	response to abscisic acid stimulus
	(GO:0009737)
AT1G05575	abscisic acid mediated signaling pathway
	(GO:0009738)
AT1G05630	response to abscisic acid stimulus
	(GO:0009737)
AT1G05850	response to abscisic acid stimulus
	(GO:0009737)
	response to water deprivation
	(GO:0009414)
AT1G06040	response to abscisic acid stimulus
	(GO:0009737)
AT1G06190	response to abscisic acid stimulus
	(GO:0009737)
AT1G04110	stomatal complex morphogenesis
	(GO:0010103)
AT1G04400	stomatal movement
	(GO:0010118)
AT1G05230	stomatal complex morphogenesis
	(GO:0010103)
AT1G05180	response to water deprivation
	(GO:0009414)
AT1G05680	cellular response to water deprivation
	(GO:0042631)
AT1G06620	response to water deprivation
	(GO:0009414)

Supplementary Table 5.4 ANOVA results for days to flowering (DTF), root pulling force (RPF) and Yield

	DTF		RPF		Yield	
	F-Value	P	F-Value	P	F-Value	P
Genotype	31.3	<0.0001	2.7	<0.0001	2.4	<0.0001
Treatment	23.2	<0.0001	40.8	<0.0001	469.0	<0.0001
Treatment (Rep)	17.9	<0.0001	2.1	0.08	18.8	<0.0001
DH x Treatment	1.6	<0.0001	1.0	0.33	1.5	<0.0001

Supplementary Table 5.5 Genetic and phenotypic correlations among traits. Genetic correlations are shown above the diagonal, while phenotypic correlations are below.

All correlations are significant at $P < 0.0001$

	RPF_{dry}	DTF_{dry}	$Yield_{dry}$	RPF_{wet}	DTF_{wet}	$Yield_{wet}$
RPF_{dry}		0.48 (n=195)	-0.41 (n=225)	0.46 (n=225)	0.52 (n=225)	-0.36 (n=225)
DTF_{dry}	0.31 (n=544)		-0.69 (n=195)	0.52 (n=195)	0.89 (n=195)	-0.49 (n=195)
$Yield_{dry}$	-0.23 (n=650)	-0.55 (n=544)		-0.45 (n=225)	-0.68 (n=225)	0.52 (n=225)
RPF_{wet}					0.66 (n=225)	-0.49 (n=225)
DTF_{wet}				0.51 (n=643)		-0.61 (n=225)
$Yield_{wet}$				-0.27 (n=651)	-0.48 (n=643)	

Supplementary Table 5.6 Summary of QTL identified with a genome-wide significance threshold of P ≤ 0.05 based on 1000 permutations. (Days to Flowering, DTF; Root Pulling Force, RPF)

Phenotype	Treatment	QTL	Linkage Group	Position (cM)	LOD	R^2	Additive Effect (Wichita allele)
DTF	Wet	*DTF.wet1*	A02	15.4	4.01	2.21	-1.41 ± 0.33
		DTF.wet2	A03	102.0	8.11	4.67	2.04 ± 0.32
		DTF.wet3	A10	75.0	37.69	30.05	5.10 ± 0.32
		DTF.wet4	C02	65.0	41.43	34.50	5.48 ± 0.32
	Dry	*DTF.dry1*	A03	104.2	4.19	3.92	1.62 ± 0.37
		DTF.dry2	A10	75.0	24.19	28.97	4.62 ± 0.38
		DTF.dry3	C02	66.0	33.58	45.51	5.56± 0.37
RPF	Wet	*RPF.wet1*	A10	73.6	13.73	18.46	6.55 ± 0.80
		RPF.wet2	C02	63.3	14.78	20.09	6.75 ± 0.80
		RPF.wet3	C07	95.0	3.34	4.02	3.21 ± 0.85
	Dry	*RPF.dry1*	A10	76.0	12.15	19.71	5.32 ± 0.66
		RPF.dry2	C02	63.3	7.22	11.12	3.92 ± 0.65
Yield	Wet	*Yld.wet1*	A03	85.4	3.14	4.42	-6.78 ± 1.77
		Yld.wet2	A10	74.0	10.17	15.40	-12.84 ± 1.80
		Yld.wet3	C02	64.0	10.95	16.72	-13.23 ± 1.78
	Dry	*Yld.dry1*	A03	141.7	3.12	3.82	-1.52 ± 0.40
		Yld.dry2	A10	78.0	15.57	21.73	-3.84 ± 0.42
		Yld.dry3	C02	66.0	12.98	17.62	-3.24± 0.40
	Sensitivity	*YldSens.1*	A10	73.6	5.19	9.00	-8.73 ± 1.75
		YldSens.2	C02	56.0	6.36	11.17	-9.87 ± 1.78

Supplementary Table 5.7 Flowering time strata ranges and the number of lines included in each stratification class.

Strata	DTF Range	# Lines
1	60.75-69.3	43
2	69.5-74	51
3	74.75-80.5	41
4	81-88.25	45
5	89-DNF	45

Supplementary Table 5.8 Nonsynonymous SNP density and term enrichment P-Value of GO terms enriched for genes carrying nonsynonymous substitutions.

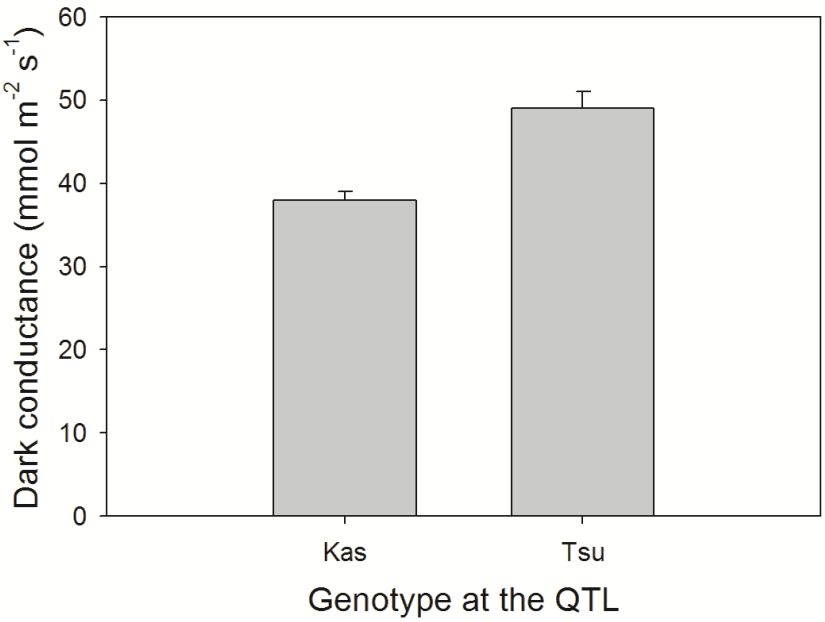
GO_Id	nonSyn_density	P-Value
GO:0008150|biological_process	0.002786885	0
GO:0009739|response to gibberellin stimulus	0.00886918	0
GO:0016481|negative regulation of transcription	0.017563869	0
GO:0003700|transcription factor activity	0.003641079	0.001
GO:0009944|polarity specification of adaxial/abaxial axis	0.024216524	0.001
GO:0000784|nuclear chromosome, telomeric region	0.01810585	0.002
GO:0004712|protein serine/threonine/tyrosine kinase activity	0.005380314	0.002
GO:0006649|phospholipid transfer to membrane	0.020942408	0.002
GO:0010196|nonphotochemical quenching	0.034653465	0.002
GO:0017169|CDP-alcohol phosphatidyltransferase activity	0.013503909	0.002
GO:0004519|endonuclease activity	0.022759602	0.003
GO:0000741|karyogamy	0.01653944	0.004
GO:0006446|regulation of translational initiation	0.018394649	0.004
GO:0008219|cell death	0.034482759	0.004
GO:0004129|cytochrome-c oxidase activity	0.031007752	0.005
GO:0016125|sterol metabolic process	0.020661157	0.005
GO:0009626|plant-type hypersensitive response	0.013759214	0.006
GO:0030528|transcription regulator activity	0.004840042	0.006
GO:0050801|ion homeostasis	0.01810585	0.006
GO:0051555|flavonol biosynthetic process	0.018292683	0.006
GO:0009535|chloroplast thylakoid membrane	0.013116802	0.007
GO:0004148|dihydrolipoyl dehydrogenase activity	0.025575448	0.008
GO:0010475|galactose-1-phosphate guanylyltransferase (GDP) activity	0.015548282	0.008
GO:0009073|aromatic amino acid family biosynthetic process	0.009605123	0.009
GO:0051743|red chlorophyll catabolite reductase activity	0.025396825	0.009

Supplementary Table 5.9 Full list of candidate genes contained within the five QTL intervals examined in this study.

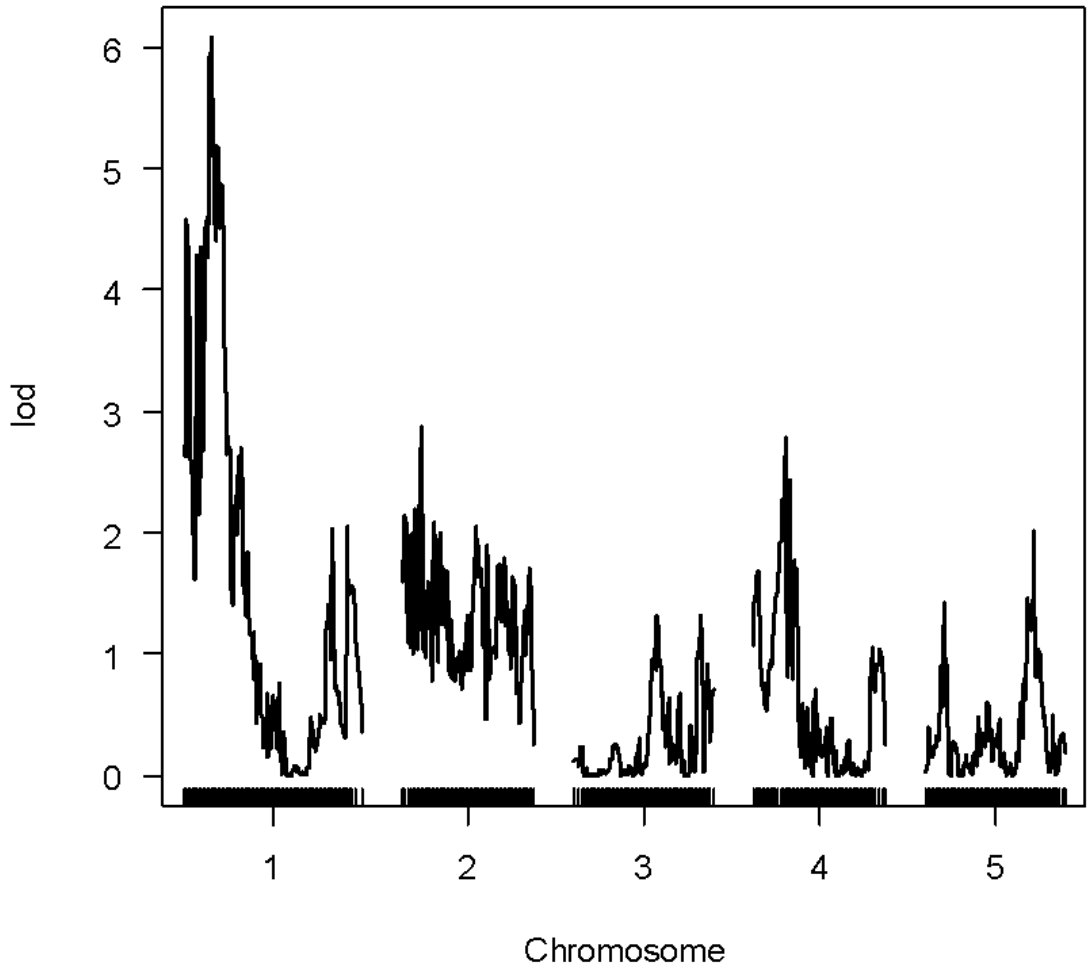
Gene Name	Chr	Start	Stop	At ortholog
Bra020185	A02	4419877	4422543	AT5G22330
Bra020236	A02	4707043	4711118	AT5G60690
Bra020249	A02	4835102	4839772	AT5G60410
Bra022500	A02	8393132	8398623	AT1G65010
Bra033931	A02	9846695	9847093	AT1G68480
Bra020463	A02	5886504	5892850	AT5G57130
Bra033987	A02	9415061	9418532	AT1G67630
Bra035571	A02	6453413	6454408	AT5G55390
Bra022711	A02	6845133	6848530	AT5G54310
Bra000413	A03	11014181	11015479	AT1G14620
Bra013198	A03	19941578	19944133	AT2G04740
Bra013162	A03	20142067	20148351	AT2G06210
Bra013169	A03	20104360	20108047	AT2G07690
Bra013145	A03	20229377	20235054	AT2G13540
Bra000304	A03	10454215	10455881	AT2G43370
Bra000305	A03	10460100	10461894	AT2G43430
Bra000392	A03	10914526	10916635	AT2G45650
Bra000393	A03	10918286	10920672	AT2G45650
Bra000420	A03	11040524	11044132	AT2G46340
Bra000446	A03	11189256	11190743	AT2G47310
Bra001058	A03	14576888	14578579	AT3G02650
Bra001357	A03	16013218	16016354	AT3G10390
Bra001825	A03	18694015	18695906	AT3G21320
Bra000795	A03	13181182	13182375	AT4G04480
Bra001256	A03	15565258	15566720	AT3G07540
Bra001532	A03	16907639	16909024	AT3G13960
Bra001538	A03	16935771	16937508	AT3G14190
Bra001708	A03	17991398	17994761	AT3G18520
Bra001709	A03	17995348	17999260	AT3G18524
Bra001713	A03	18028431	18031485	AT3G18600
Bra001729	A03	18117731	18126862	AT3G18990
Bra001782	A03	18431989	18433396	AT3G20260
Bra001792	A03	18507750	18510529	AT3G20550
Bra001927	A03	19407620	19409410	AT3G24120
Bra001944	A03	19524788	19528164	AT3G24340
Bra000891	A03	13763273	13765906	AT4G02150
Bra013186	A03	19998618	20003928	AT5G35750

Bra008781	A10	12625327	12626276	AT5G14080
Bra008802	A10	12699925	12701540	AT5G13790
Bra008888	A10	13100861	13106392	AT5G12430
Bra008889	A10	13110497	13112199	AT5G12410
Bra008937	A10	13293421	13295301	AT5G11780
Bra008966	A10	13461950	13464166	AT5G11390
Bra008971	A10	13484080	13486119	AT5G11320
Bra008979	A10	13508303	13511612	AT5G11240
Bra009156	A10	15021517	15027009	AT5G05980
Bra009167	A10	14975130	14976901	AT5G06170
Bra009287	A10	14474474	14479590	AT5G07180
Bol015506	C02	8260872	8261117	AT1G01030
Bol021328	C02	4129673	4131560	AT1G44110
Bol018187	C02	10203544	10204605	AT1G79280
Bol036006	C02	6676293	6676469	AT3G55770
Bol021303	C02	4001911	4006559	AT5G16750
Bol021333	C02	4146680	4148533	AT5G17300
Bol021358	C02	4295699	4297343	AT5G17690
Bol036052	C02	6264054	6265682	AT5G22290
Bol036025	C02	6498005	6500778	AT5G22650
Bol036024	C02	6515011	6516039	AT5G22650
Bol036023	C02	6516817	6518550	AT5G22650
Bol036022	C02	6519200	6521019	AT5G22650
Bol036021	C02	6522211	6524282	AT5G22650
Bol036020	C02	6532841	6534383	AT5G22650
Bol018202	C02	10343465	10350998	AT5G55390
Bol008947	C02	7243745	7248379	AT5G60410
Bol035990	C02	6843719	6847836	AT5G60690
Bol021336	C02	4168308	4171696	AT5G17330
Bol021338	C02	4189202	4190569	AT5G17400
Bol021369	C02	4406240	4407844	AT5G17820
Bol007160	C02	4737567	4738889	AT5G18520
Bol007157	C02	4752244	4752948	AT5G18560
Bol007155	C02	4768328	4771078	AT5G18580
Bol007148	C02	4845326	4849430	AT5G18830
Bol036139	C02	5494486	5503941	AT5G20490
Bol036044	C02	6318637	6319912	AT5G22410
Bol018183	C02	10185853	10187566	AT5G55730
Bol018177	C02	10107519	10110976	AT5G55920
Bol014266	C02	10048508	10049628	AT5G56860
Bol014250	C02	9722497	9725803	AT5G57090
Bol014244	C02	9598338	9602535	AT5G57130
Bol014226	C02	9352424	9355325	AT5G57390

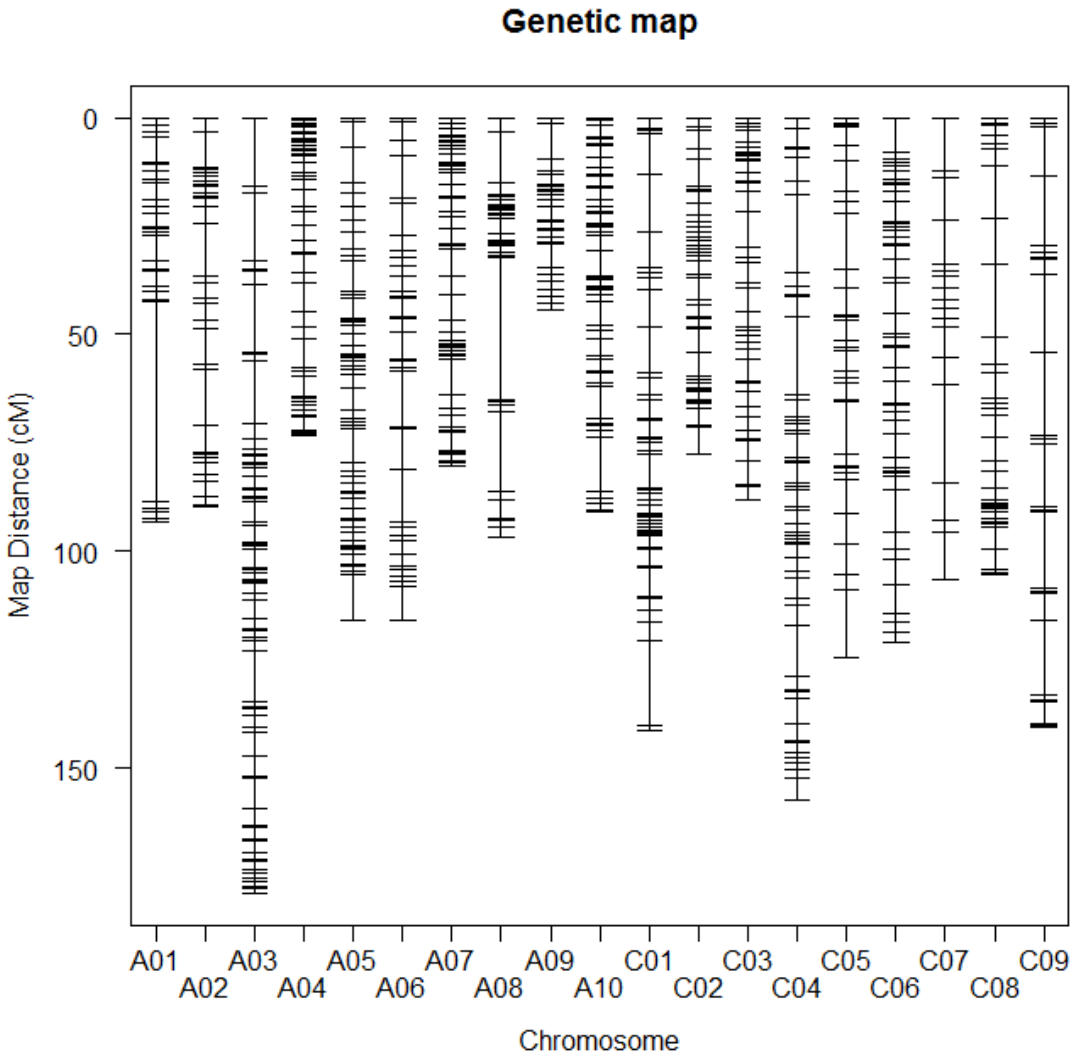
Bol015588	C02	8946516	8948250	AT5G57740
Bol015571	C02	8809296	8810392	AT5G58010
Bol008934	C02	7314733	7315446	AT5G60200
Bol008961	C02	7041051	7042479	AT5G60660
Bol042756	C07	31723010	31724505	AT1G45688
Bol004697	C07	30139585	30144027	AT2G01830
Bol042745	C07	31572735	31574033	AT3G25710



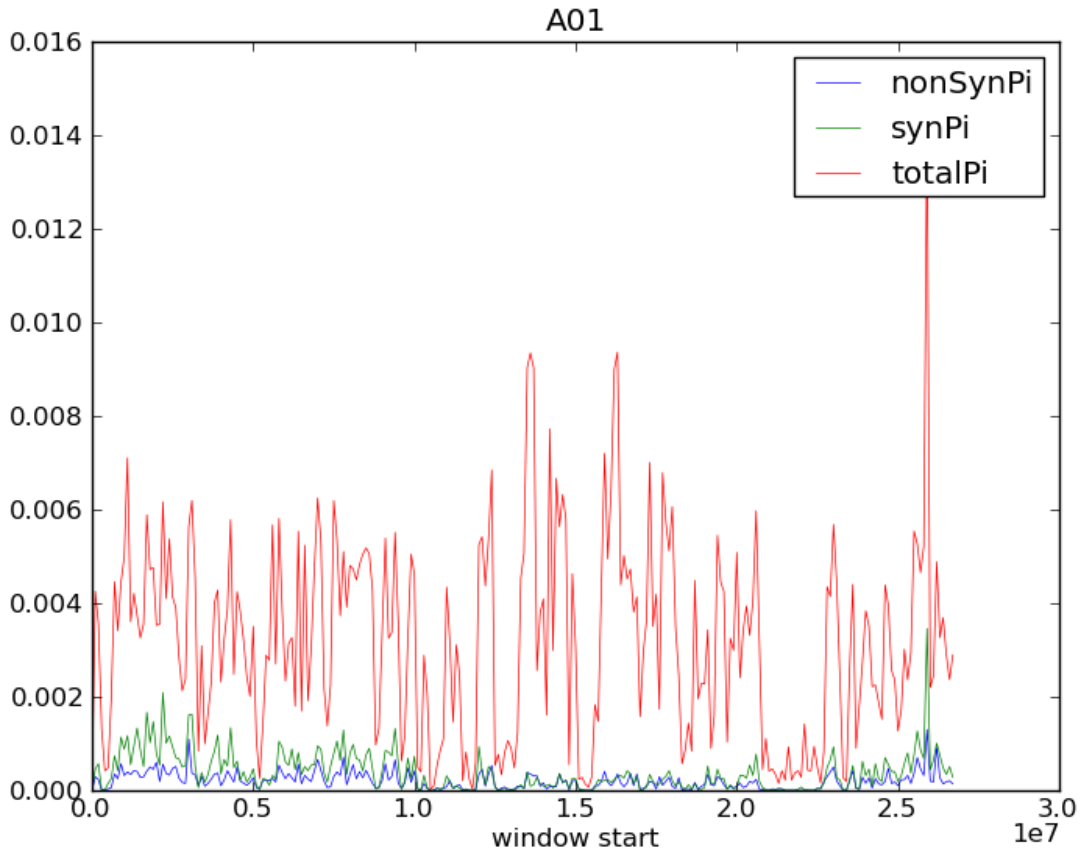
Supplementary Figure 5.1 Bar graph of mean dark conductance of RILs carrying the Kas-1 allele at the QTL on chromosome 1 relative to RILs carrying the Tsu-1 allele.



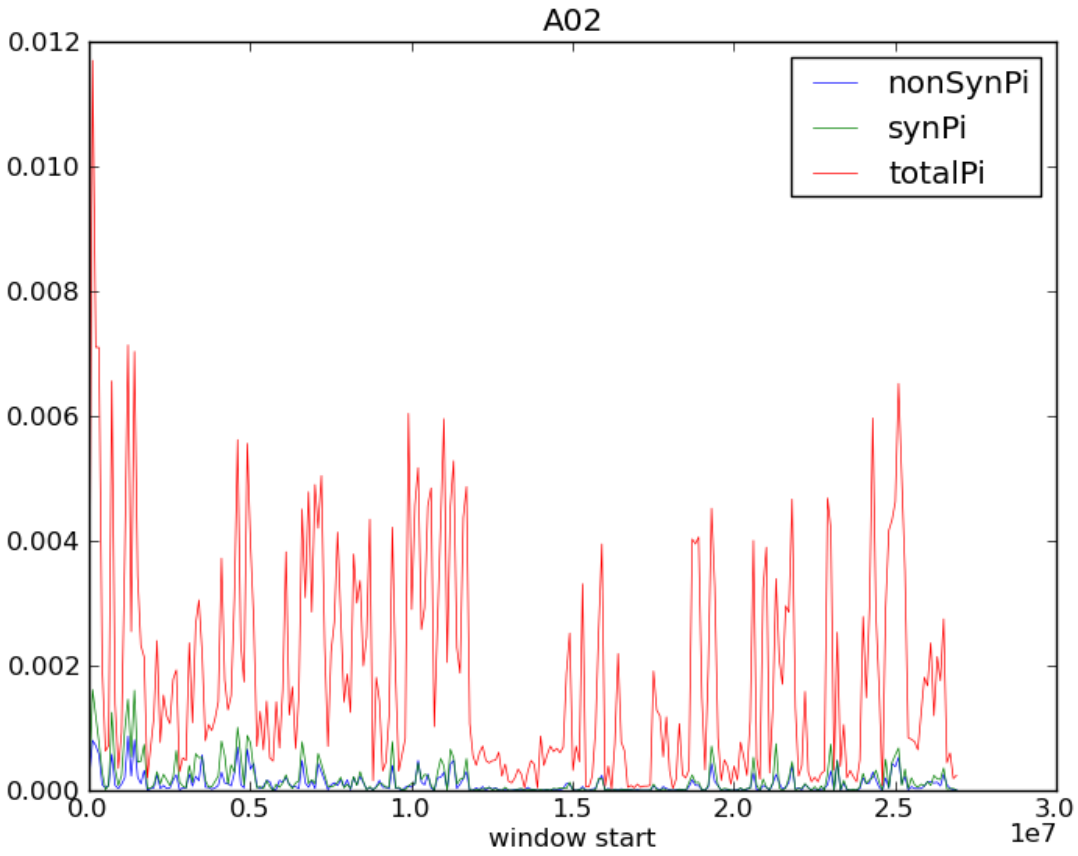
Supplemenatry Figure 5.2 Genome-wide LOD graph of g_0 QTL scans.



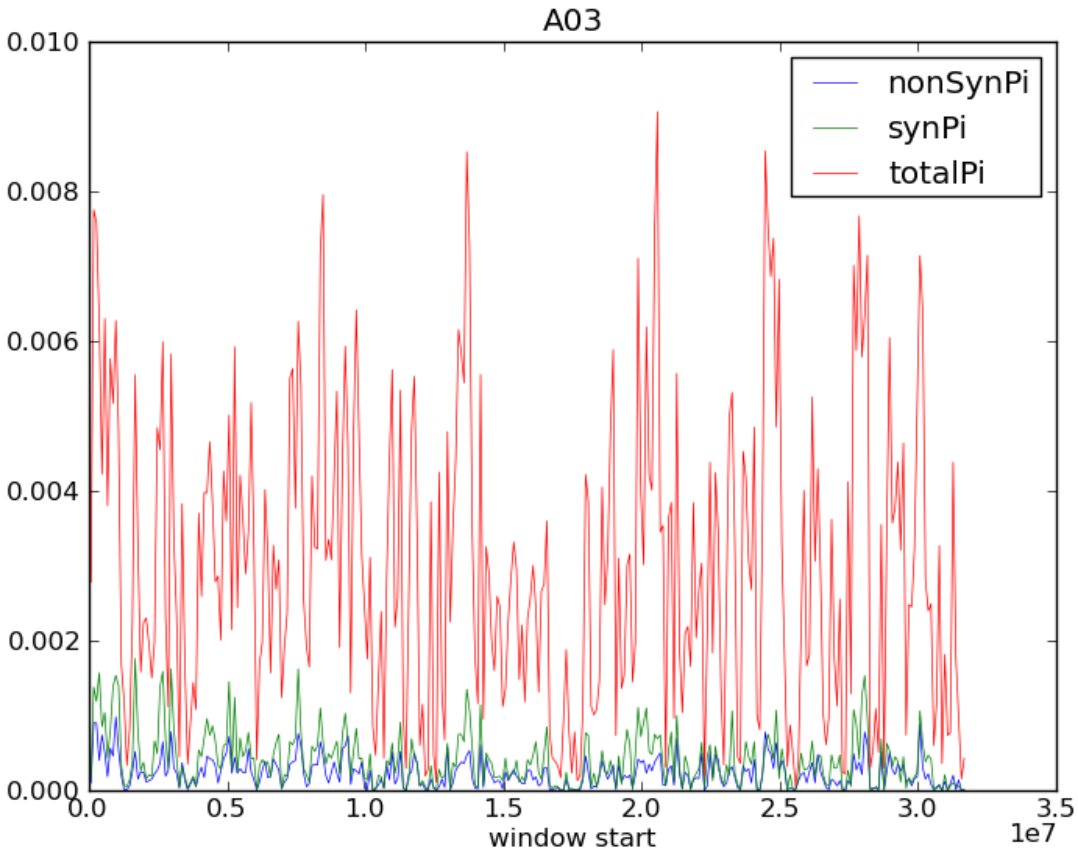
Supplementary Figure 5.3 Genetic map of SE-1 (IMC106RR x Wichita) doubled haploid population.



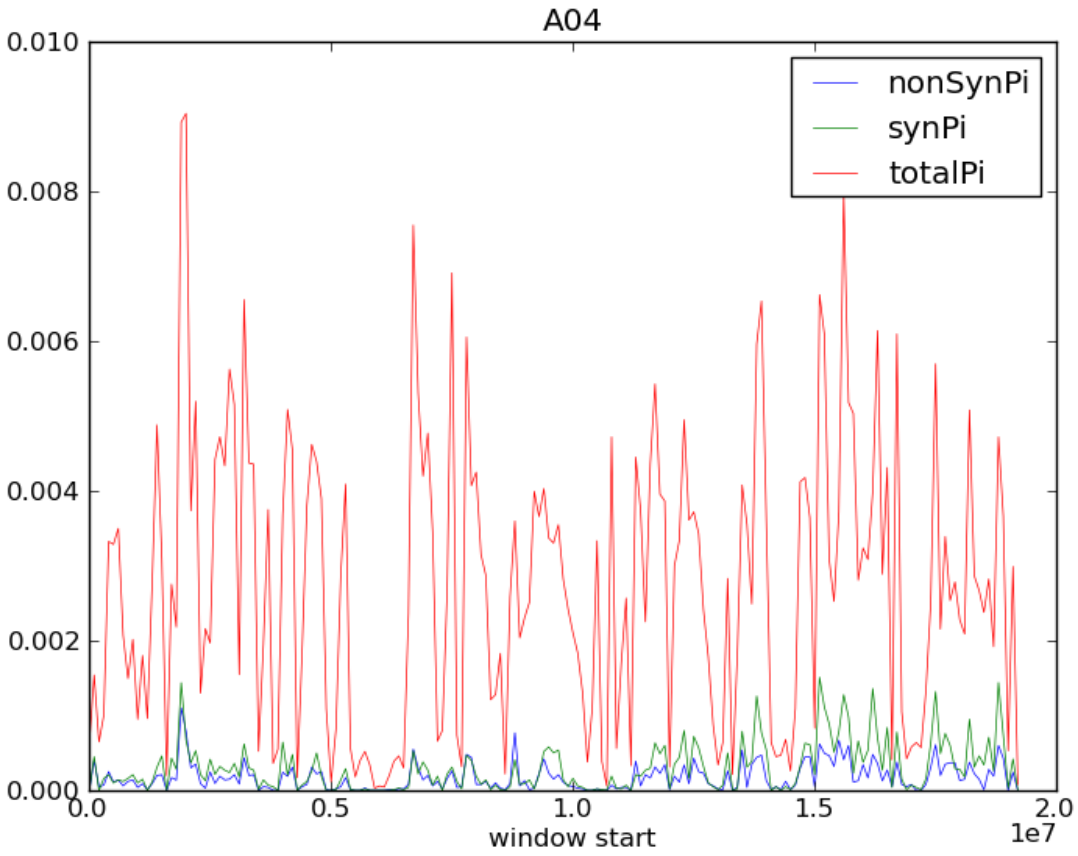
Supplementary Figure 5.4 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A01.



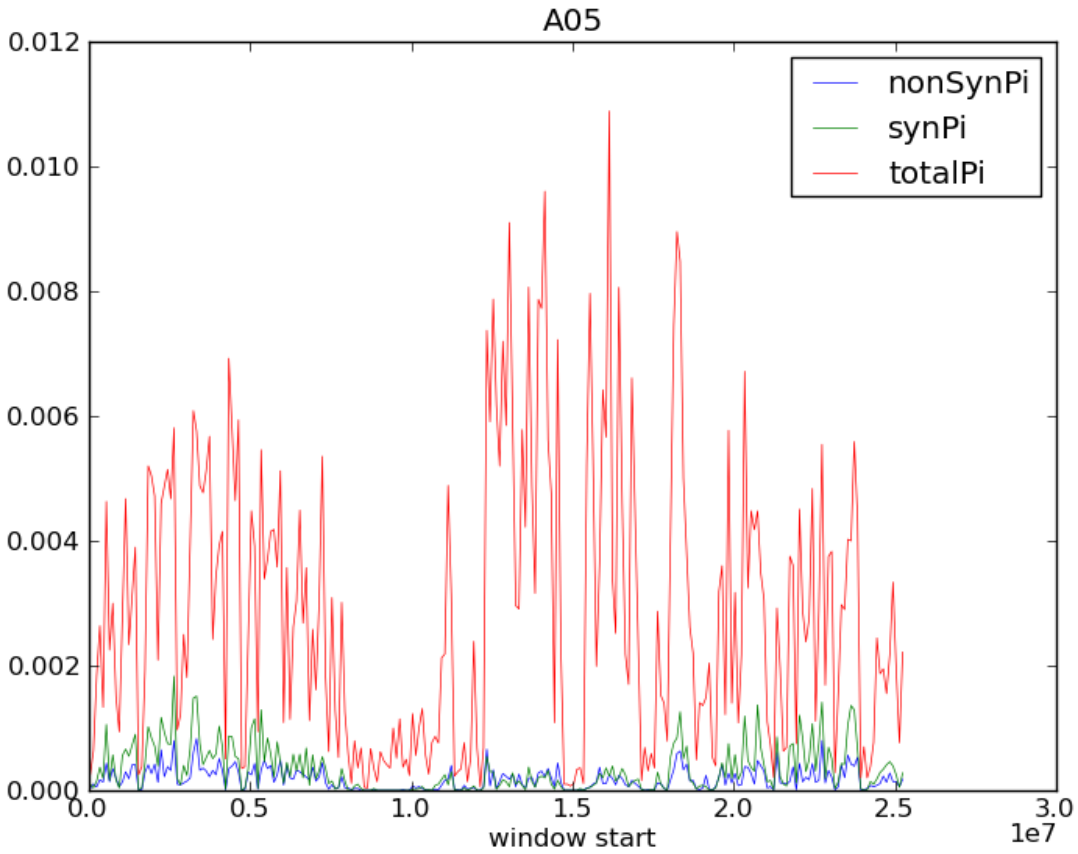
Supplementary Figure 5.5 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A02.



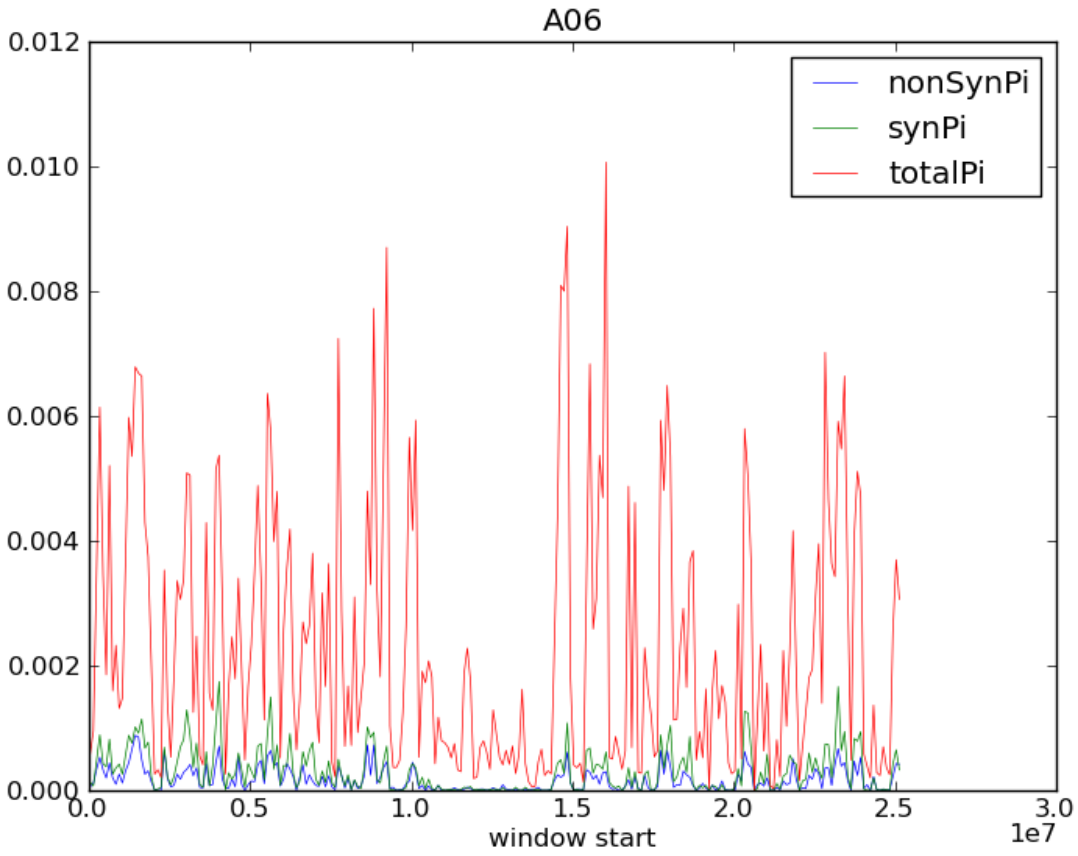
Supplementary Figure 5.6 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A03.



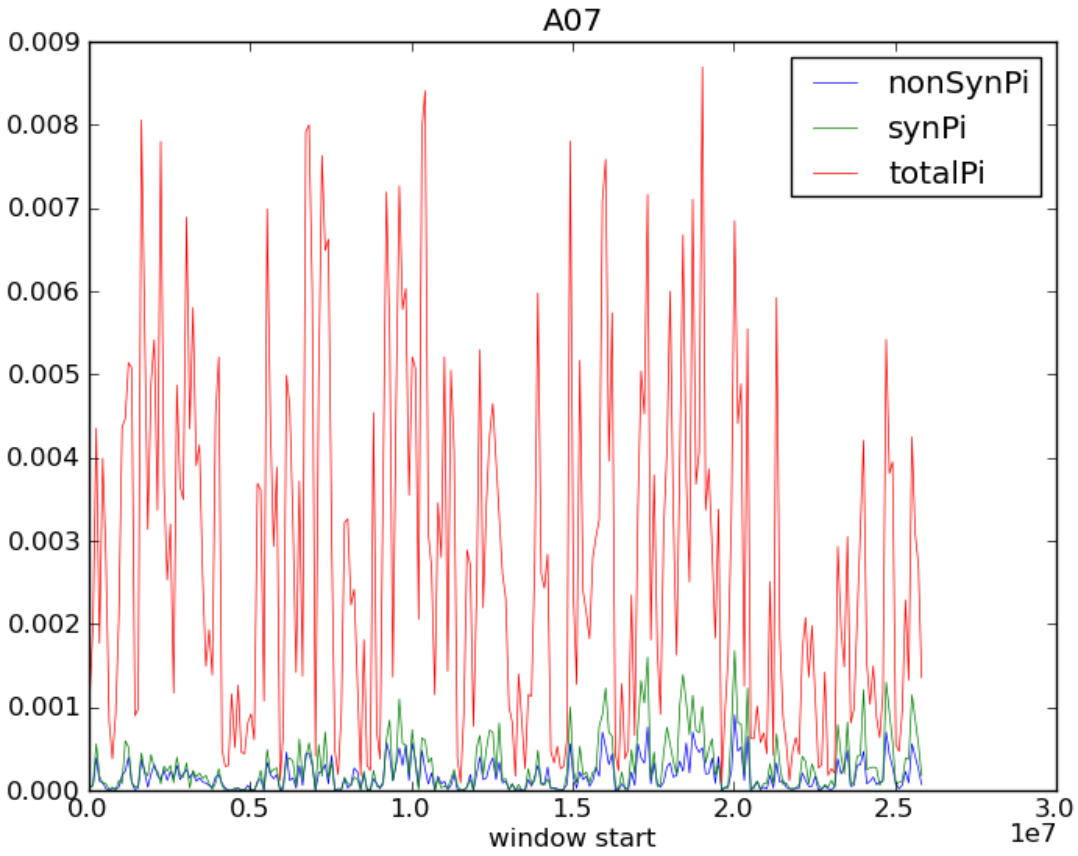
Supplementary Figure 5.7 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A04.



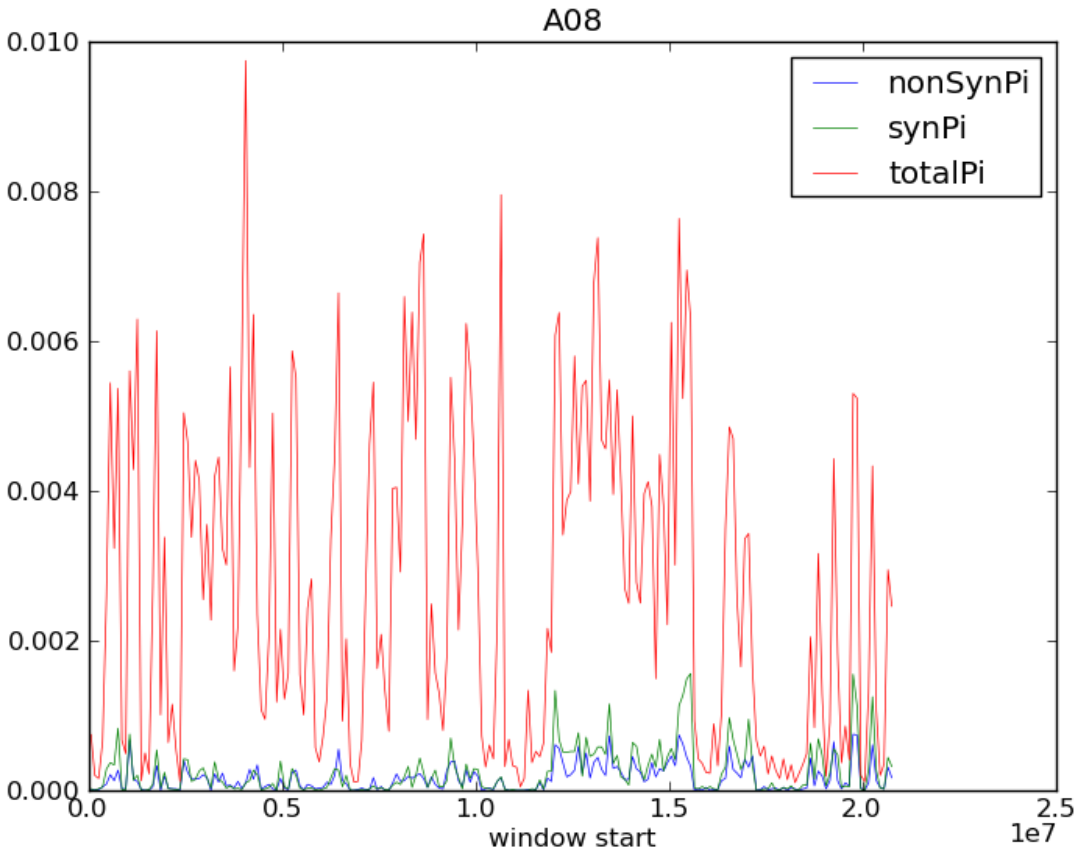
Supplementary Figure 5.8 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A05.



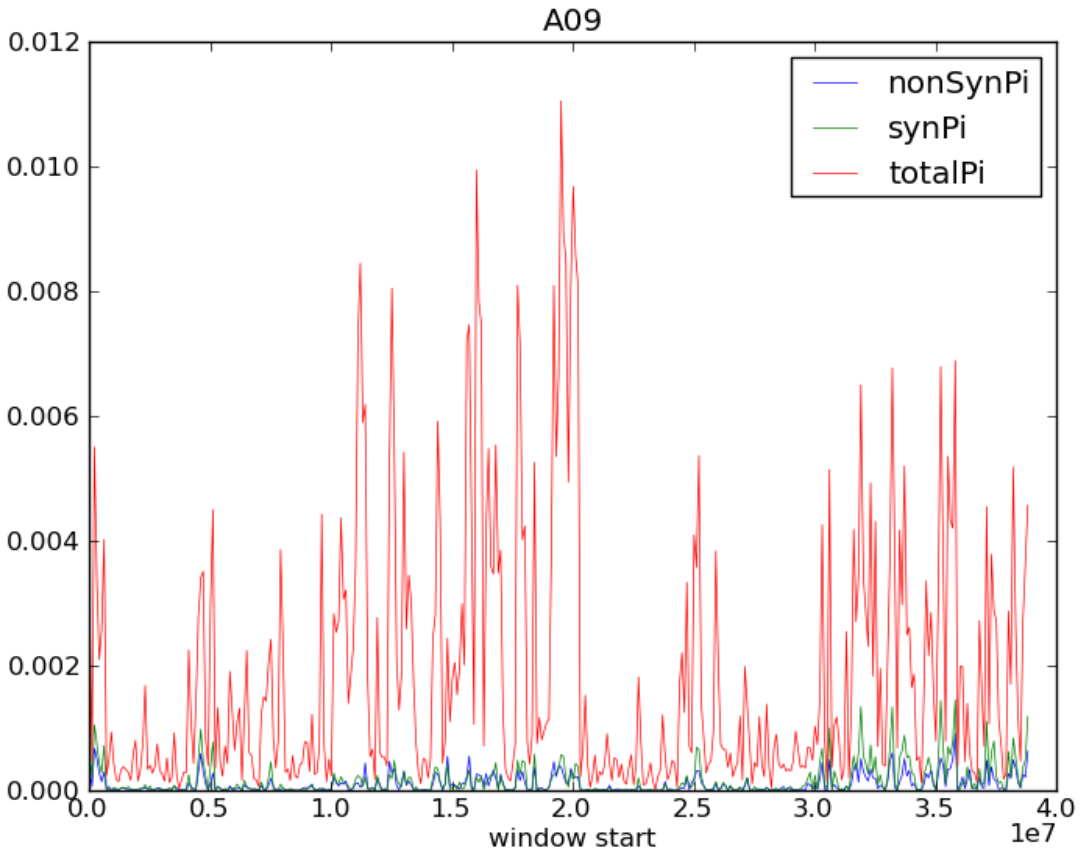
Supplementary Figure 5.9 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A06.



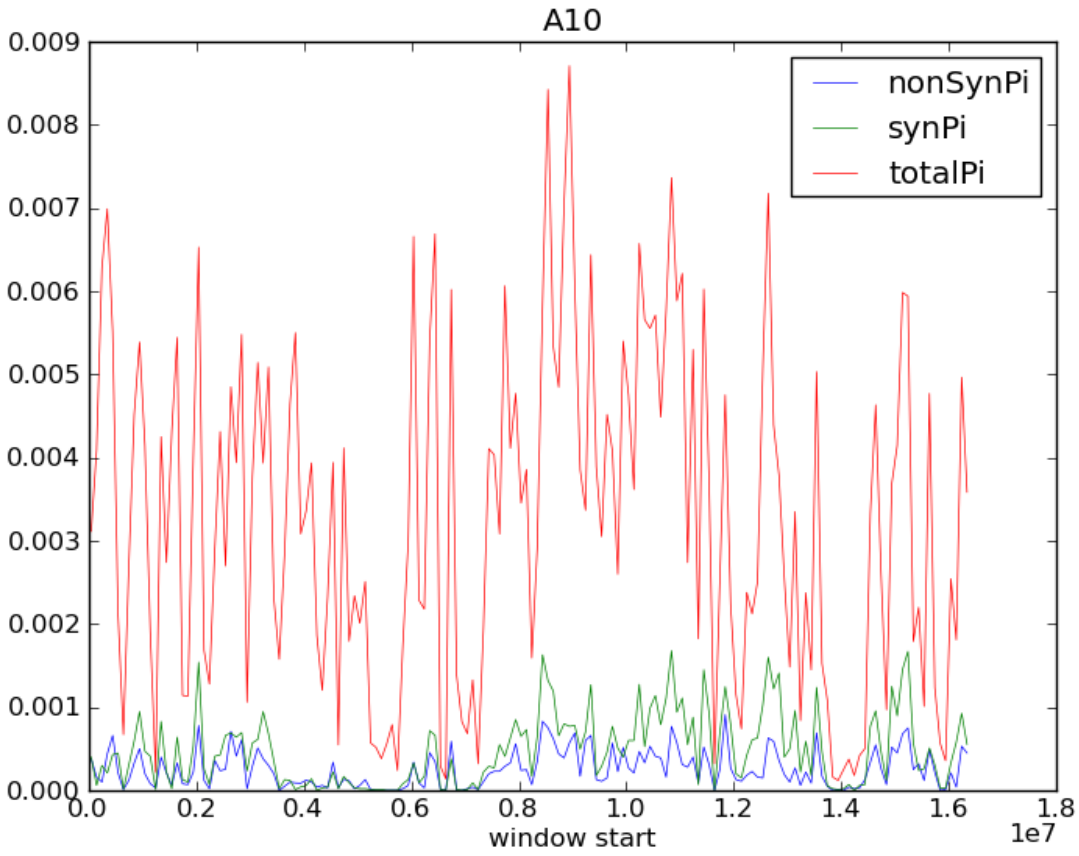
Supplementary Figure 5.10 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A07.



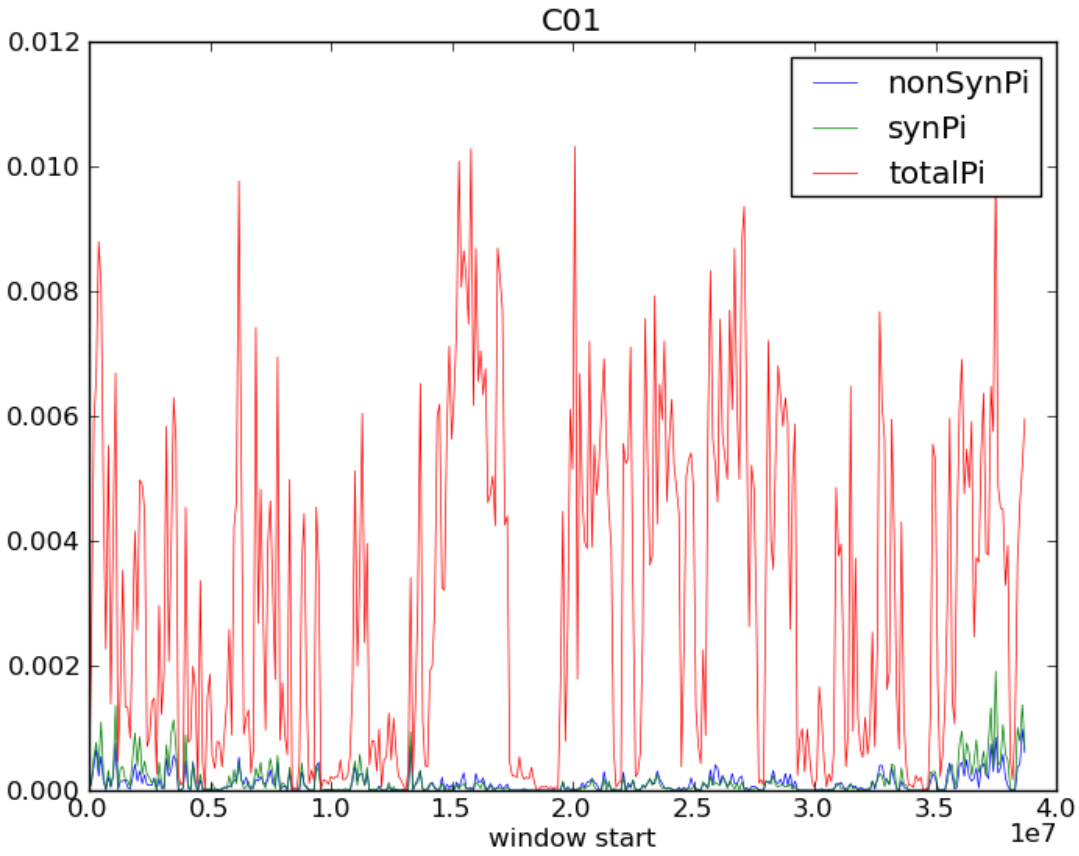
Supplementary Figure 5.11 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A08.



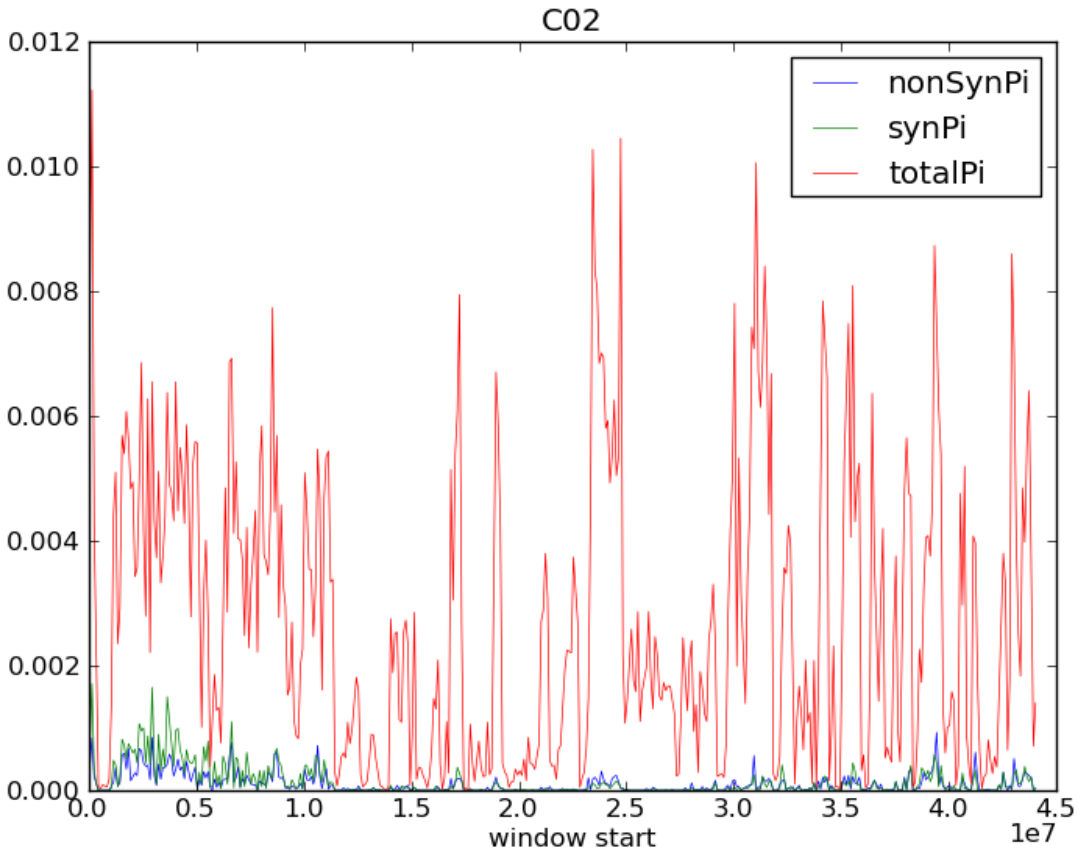
Supplementary Figure 5.12 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A09.



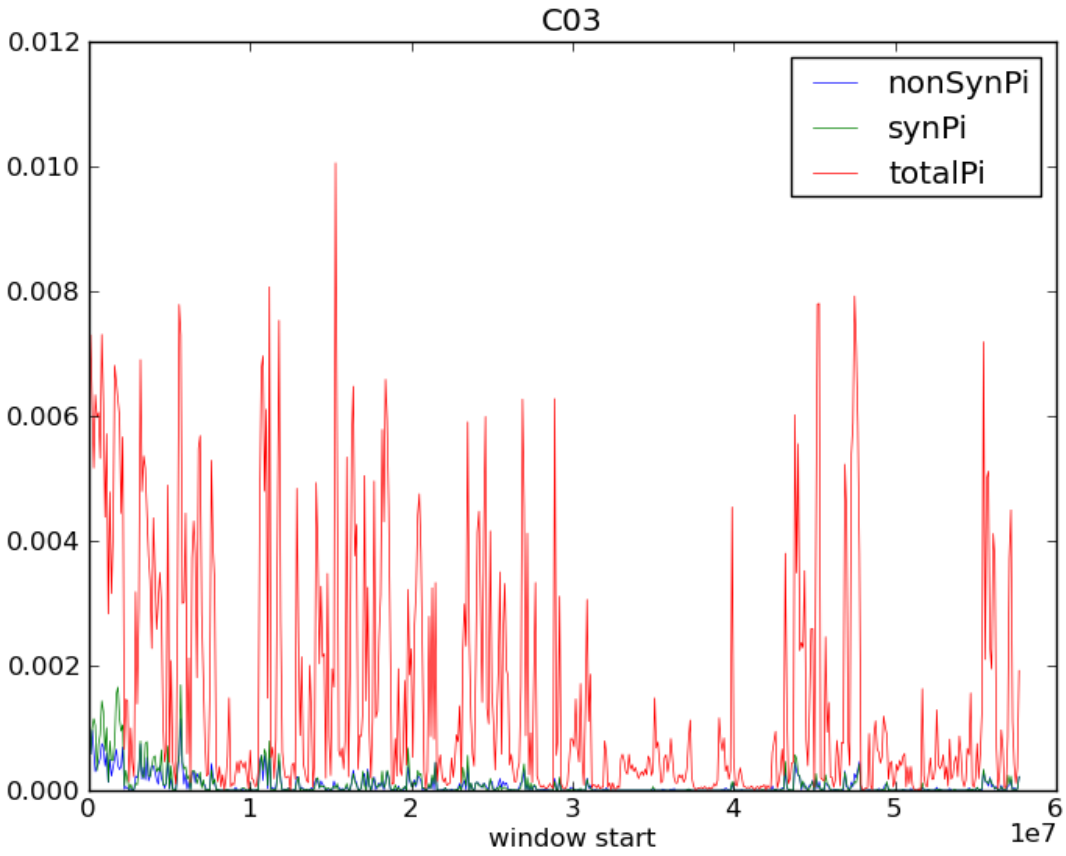
Supplementary Figure 5.13 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome A10.



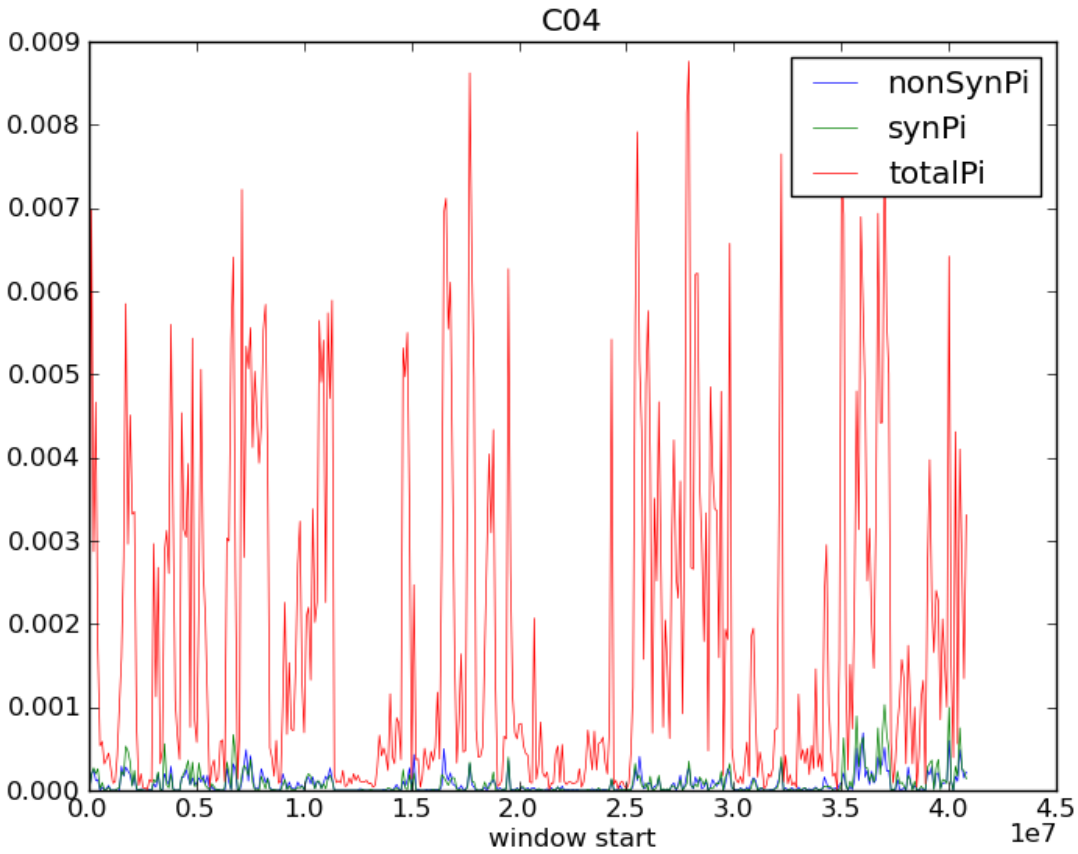
Supplementary Figure 5.14 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C01.



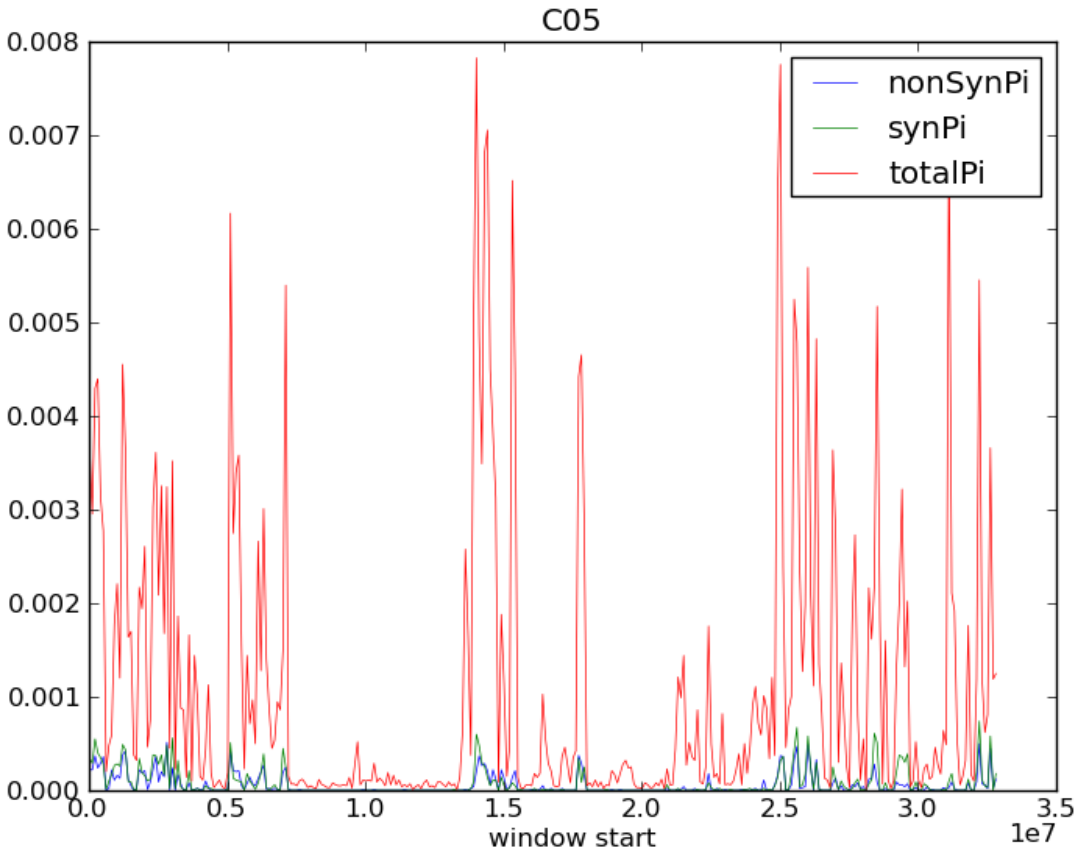
Supplementary Figure 5.15 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C02.



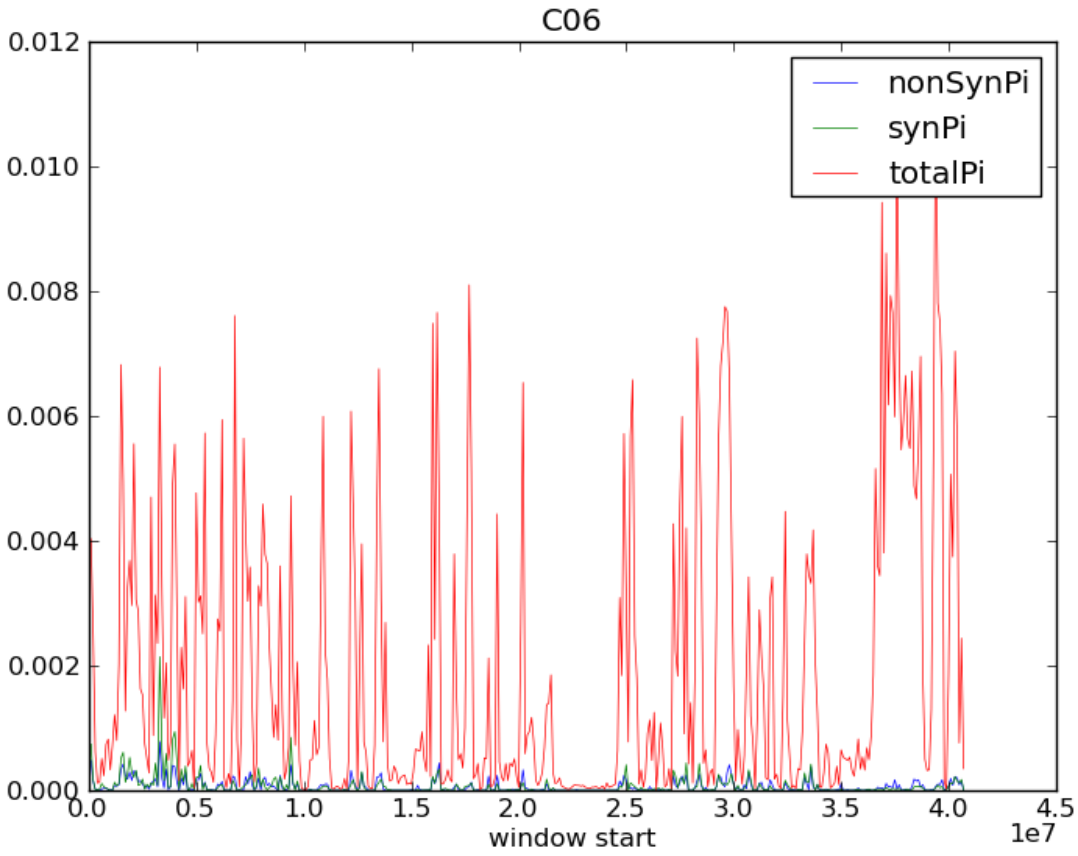
Supplementary Figure 5.16 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C03.



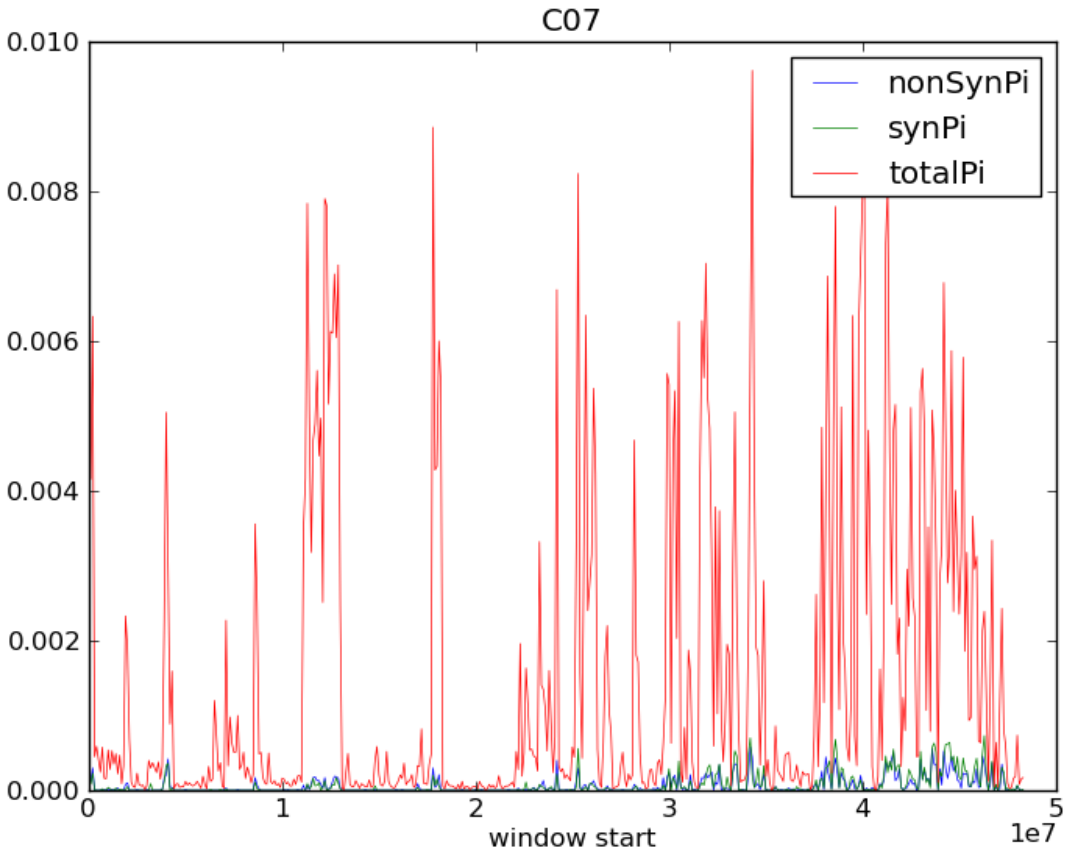
Supplementary Figure 5.17 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C04.



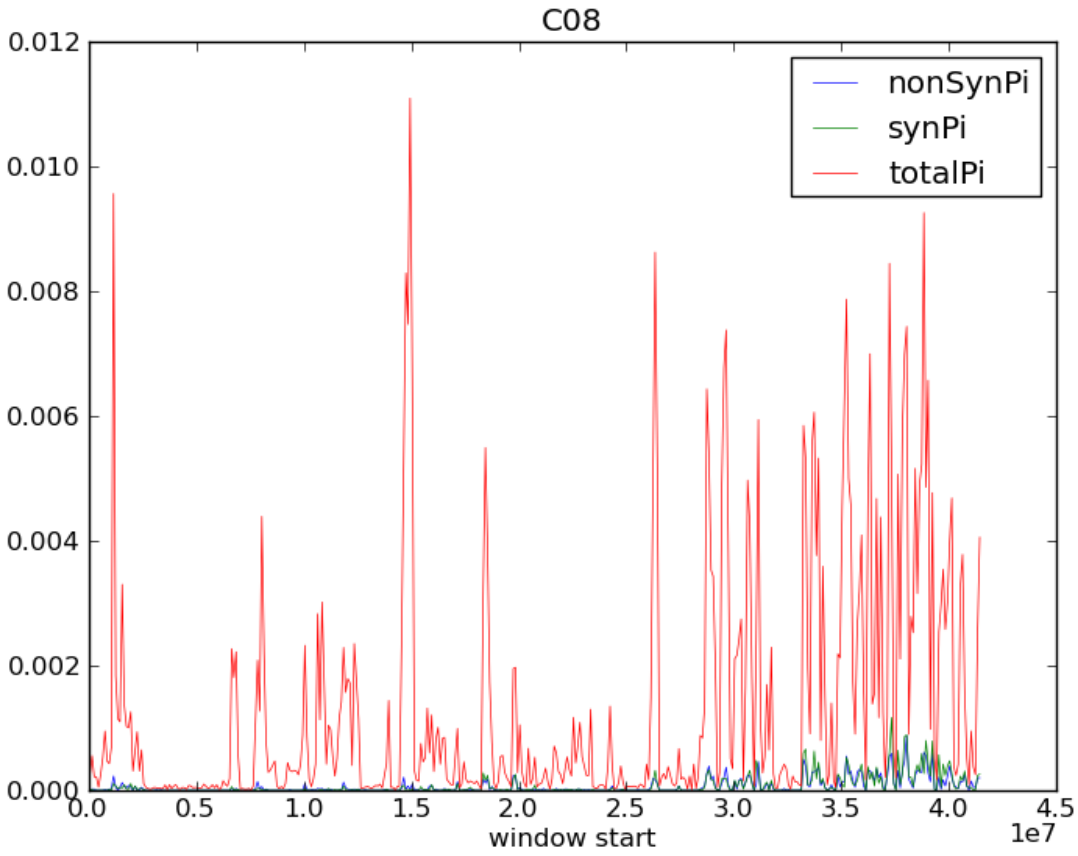
Supplementary Figure 5.18 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C05.



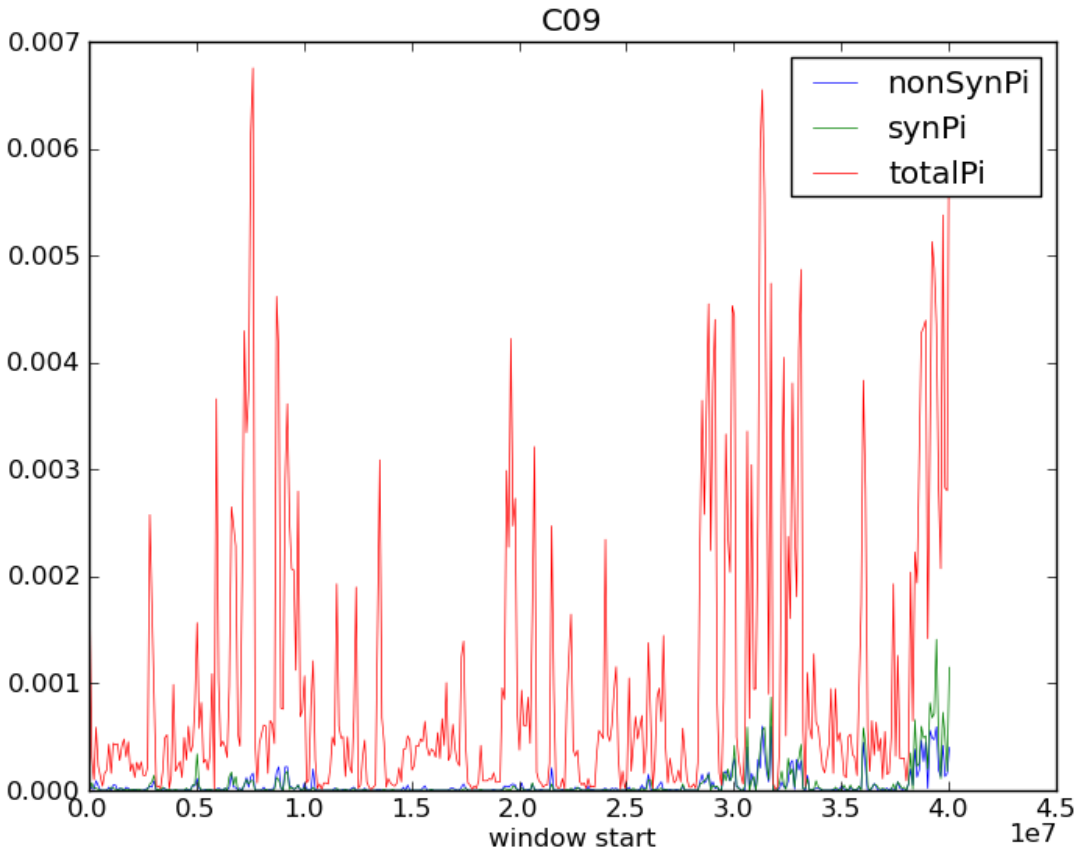
Supplementary Figure 5.19 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C06.



Supplementary Figure 5.20 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C07.



Supplementary Figure 5.21 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C08.



Supplementary Figure 5.22 Sliding window estimates of nonsynonymous (blue), synonymous (green) and total SNP density across chromosome C09.