GENETIC ANALYSIS IN ADVANCED BACKCROSS POPULATIONS IN *Brassica oleracea*

by

LAN-SHUAN SHUANG

(Under the Direction of Andrew H. Paterson)

ABSTRACT

The single species *Brassica oleracea* encompasses a remarkable diversity of morphotypes, including cauliflower, broccoli, kohlrabi, marrowstem kale, cabbage, and Brussels sprouts as well as rapid-flowering morphologically simple genotypes reminiscent of the leading botanical model, *Arabidopsis thaliana*. To dissect the molecular basis of morphological diversity of *B. oleracea*, two backcross populations were developed by using inbred lines of cabbage (Badger Inbred) and cauliflower (Orange) as donor parents, and a rapid cycling line (TO1434) as the recurrent parent. Genotypes of the two populations in their BC_4F_1 generation were determined by genotyping-by-sequencing (GBS). The two populations were evaluated in the field for two seasons. Morphological traits, including flower color and 14 leaf-, stem-, and flower-traits, were segregating within the two populations, based on which we found 219 marker-trait associations. The two populations provide the foundation to construct panels of near isogenic lines covering most of the genome, and reveal QTLs for morphological traits in finer resolution and higher mapping power in the near future.

INDEX WORDS: *Brassica oleracea*, Backcross populations, Morphological traits, Genetic mapping, Genotyping-by-sequencing (GBS), Quantitative trait loci (QTL)

GENETIC ANALYSIS IN ADVANCED BACKCROSS POPULATIONS IN *Brassica oleracea*

by

LAN-SHUAN SHUANG

B.S., National Taiwan University, Taiwan, 2011

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2015

GENETIC ANALYSIS IN ADVANCED BACKCROSS POPULATIONS IN *Brassica oleracea*

by

LAN-SHUAN SHUANG

Major Professor:	Andrew Paterson
Committee:	Andrea Sweigart
	Juan Carlos Diaz-Perez
	Peng-Wah Chee

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2015

ACKNOWLEDGEMENTS

I would like to acknowledge my major professor Dr. Paterson for his guidance, support and patience and all the committee members, Dr. Sweigart, Dr. Diaz and Dr. Chee, for their valuable suggestions. I would like to thank Hugo Cuevas for starting this project and Cornelia Lemke for maintaining the populations. I would also like to thank the assistance from Crop and Soil Sciences greenhouse, Plant Biology greenhouse, and Horticulture farm, especially Crop and Soil Sciences greenhouse manager, Kurk Lance. I would like to express my gratitude to all the members in PGML who have helped me in various ways.

Special thanks to those who have been supportive along the way, my parents, Shang-Hong Chang and Ta-Ning Shuang, my dearest sister, Yiming Shuang, and my best friend, Yu-Ru Chen.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

About *Brassica oleracea*

Brassica oleracea, a species in the genus *Brassica* (RAKOW 2004), is of great economic importance for its many vegetable crops, including Brussels sprouts (var. *gemmifera*), broccoli (var. *italica*), cauliflower (var. *botrytis*), Chinese kale (var. *alboglabra*), kale and collard (var. *acephala*), kohlrabi (var. *gongylodes*), and cabbage (var. *capitata*) (DIXON 2007). These vegetables are abundant in fiber, vitamins and minerals (DIXON 2007) and regarded as anticarcinogens due to their glucosinolate content. Glucosinolates can be hydrolyzed to isothiocyanates, which are inducers of detoxification enzymes (BEECHER 1994; VERHOEVEN *et al.* 1996; VAN POPPEL *et al.* 1999; TERRY *et al.* 2001; KRISTAL and LAMPE 2002; FOWKE *et al.* 2003; DIXON 2007). According to FAO statistics, production of cabbages and other Brassicas is about 70,104,972 t, while cauliflowers and broccoli is about 21,266,789 t, which are in the fifth and 15th places of worldwide vegetable and melon production in 2012 (FAOSTAT 2014).

Remarkable diversity in morphology is another well-known characteristic of *Brassica oleracea*. Each morphotype has an enlarged edible organ, such as the lateral buds of Brussels sprouts, the inflorescence of cauliflower and broccoli, apical meristem and leaves of cabbage, bulbous lower stem epicotyl of kohlrabi, and leaves of kale

(BABULA *et al.* 2007). Also, rapid cycling lines have been selected for short generation time, self-compatibility, absence of vernalization and seed dormancy, and are considered to be a genetic standard (WILLIAMS and HILL 1986). It is suggested that the morphological variation is because the crops cross-breed with wild types in the vicinity and being domesticated and selected in different environments (DIXON 2007). Morphotypes in the species can be divided into three groups, kale, cabbage and broccoli by RFLP markers (SONG *et al.* 1988). Cauliflower is considered to be derived from broccoli (SONG *et al.* 1990). Regarding chloroplast genetic diversity, which is considered to be maternally inherited, broccoli and cauliflower have the same haplotype while cabbage, kohlrabi and Chinese kale have another (ZHANG *et al.* 2012).

B. oleracea has been considered to be a mesopolyploid (QUIROS *et al.* 1987; MCGRATH *et al.* 1990; SLOCUM *et al.* 1990; LAN *et al.* 2000), with its genome size estimated at 630 MB. Besides *B. oleracea* (2n=18, CC), there are diploid *B. rapa* (2n=20, AA), and *B. nigra* (2n=16, BB), and amphidiploid *B. napus* (2n=38, AACC), *B. juncea* (2n=36, AABB), and *B. carinata* (2n=34, BBCC). The relationship between diploid and amphidiploid Brassicas has been described by "the triangle of U" (NAGAHARU 1935). Two *B. oleracea* genomes have been published recently (YU *et al.* 2013; LIU *et al.* 2014b; PARKIN *et al.* 2014). Several shared whole genome polyploidization events, including 'gamma' triplication event and alpha and beta duplication, predated its divergence with another sequenced model organism in the same family, *Arabidopsis thaliana*, (BOWERS *et al.* 2003; JIAO *et al.* 2012). An additional whole genome triplication event is estimated to have occurred ~15.9 million years ago (MYA), before divergence of *B. oleracea* and *B. rapa* lineages (~4.6 MYA) (LYSAK *et al.* 2005; LIU *et al.* 2014b).

Genetic mapping in *B. oleracea*

Association mapping is an approach to identify genomic regions associated with traits of interest, by cosegregation with morphological, biochemical or molecular markers (COLLARD *et al.* 2005). DNA-based molecular markers are most widely used, including restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), expressed sequence tag (EST), simple sequence repeats/ microsatellites (SSR) and single nucleotide polymorphisms (SNPs). Nowadays, SNPs are of value due to their abundance and genome wide distribution (RAFALSKI 2002), which allow higher throughput of genotyping. Also, with advances in sequencing techniques, the discovery and genotyping of SNPs has experienced dramatic reductions in cost, labor and time (DESCHAMPS *et al.* 2012; HE *et al.* 2014). Therefore, SNPs are becoming more prevalent than in past decades.

Populations that often used in linkage mapping are F_2, backcross (BC), recombinant inbred (RI) and doubled haploid (DH). The former two populations are fast and easy to develop and the latter two are reproducible due to their composition of homozygous (COLLARD *et al.* 2005). Furthermore, advanced backcross populations, such as introgression lines and near isogenic lines (NILs), are also resources for genetic mapping and breeding (KOOKE *et al.* 2012).

In *B. oleracea*, several traits were studied for which mapping suggested single gene control. Morphological traits such as bulb, anthocyanin pigmentation, glossy foliage, vernalization requirement (KIANIAN and QUIROS 1992), leaf morphology (LANDRY *et al.* 1992), flower color (FARINHÓ *et al.* 2004; ZHANG *et al.* 2015), were studied. A single

gene mutation, *Or*, acting as semi-dominant, causes accumulations of beta-carotene and makes tissues become orange, is mapped and studied in cauliflower (LI *et al.* 2001b; LI and GARVIN 2003; LI *et al.* 2003b). Glucosinolate content has been considered an important objective for breeding because of its anticarcinogencity. Genes involved in its biochemical pathway were mapped (LI and QUIROS 2001; LI *et al.* 2001a; NAGAOKA *et al.* 2010). Self-incompatibility, controlled by the S-locus, involves identification by the pistil of 'self' pollen and inhibition of its germination by virtue of possessing the same S-allele on both stigma and pollen, with this locus and candidate genes well studied (OCKENDON and GATES 1975; KIANIAN and QUIROS 1992; RAMSAY *et al.* 1996; BOYES *et al.* 1997; CAMARGO *et al.* 1997; HU *et al.* 1998; SCHOPFER *et al.* 1999). Resistance to downy mildew disease was found to be controlled by a single gene (FARNHAM *et al.* 2002; GIOVANNELLI *et al.* 2002; COELHO and MONTEIRO 2003; FARINHÓ *et al.* 2004).

Genomic regions that contain genes influencing numerous quantitative traits have been mapped. Morphological traits, including stem and leaf (KENNARD *et al.* 1994; LAN and PATERSON 2001; SEBASTIAN *et al.* 2002; WALLEY *et al.* 2012), inflorescence architecture (LAN and PATERSON 2000; GAO *et al.* 2007), flowering time and its requirement for vernalization (KENNARD *et al.* 1994; CAMARGO and OSBORN 1996; BOHUON *et al.* 1998; RAE *et al.* 1999; LAN and PATERSON 2000; AXELSSON *et al.* 2001; OKAZAKI *et al.* 2007) were dissected to reveal remarkable diversity among morphotypes. Resistance to *Plasmodiophora brassicae*, which causes clubroot disease (LANDRY *et al.* 1992; FIGDORE *et al.* 1993; GRANDCLÉMENT and THOMAS 1996; VOORRIPS *et al.* 1997; MORIGUCHI *et al.* 1999; ROCHERIEUX *et al.* 2004; NAGAOKA *et al.* 2010), and *Sclerotinia sclerotiorum*, which causes Sclerotinia stem rot (MEI *et al.* 2013), were studied. QTLs

related to *Agrobacterium* transformation and plant regeneration from protoplasts were mapped to identify genetic factors that influence transformation efficiency (HANSEN *et al.* 1999; COGAN *et al.* 2002; COGAN *et al.* 2004; HOLME *et al.* 2004; SPARROW *et al.* 2004). QTLs involved in varying glucosinolate degradation (HENNIG *et al.* 2013), carotenoid content (BROWN *et al.* 2014), potassium use efficiency (WHITE *et al.* 2010), calcium and magnesium concentrations in shoots (BROADLEY *et al.* 2008), and fatty acid synthesis and modification (BARKER *et al.* 2007) were mapped. Traits related to plant growth, like circadian period (SALATHIA *et al.* 2007) and seed germination (BETTEY *et al.* 2000) were also studied.

Comparative analysis is useful to study genome evolution between related species and transfer genomic information from well-studied species to others (KACZMAREK *et al.* 2009). Comparative analysis among diploid species in Brassica showed intergenomic conserved regions, with A and C genomes showing highest homology, and suggesting the hypothesis that these genomes evolved from a smaller genome through genome duplication and reshuffling (MCGRATH and QUIROS 1991; LAGERCRANTZ and LYDIATE 1996; TRUCO *et al.* 1996). Comparative analysis also showed that nine chromosomes of *B. oleracea* are highly conserved with chromosomes 11-19 of *B. napus*, indicating that these are the chromosomes that were contributed to *B. napus* by a *B. oleracea*-like progenitor (LYDIATE *et al.* 1993; PARKIN *et al.* 1995; BOHUON *et al.* 1996; CHEUNG *et al.* 1997). The sequenced model organism, *Arabidopsis thaliana*, is in the same family, Brassicaceae, as Brassica. Comparative analysis between *A. thaliana* and *B. oleracea* showed their homology and suggested that polyploidization and extensive chromosomal rearrangement had contributed to the divergence of the two species (lineages) during evolution

(KOWALSKI *et al.* 1994; LAN *et al.* 2000; BABULA *et al.* 2003; LI *et al.* 2003a; LUKENS *et al.* 2003; TOWN *et al.* 2006; KACZMAREK *et al.* 2009). Furthermore, expansion in genome size of *B. oleracea* was considered to be contributed by transposable elements (ZHANG and WESSLER 2004; QIU *et al.* 2009).

Genetic analysis for morphological variation

QTL mapping is a way to discover the genetic control of natural variation (ALONSO-BLANCO *et al.* 2009). Morphological variation in *B. oleracea* has been dissected as leaf-, stem-, and inflorescence-related traits and mapped in various types of populations (KENNARD *et al.* 1994; LAN and PATERSON 2000; LAN and PATERSON 2001; SEBASTIAN *et al.* 2002; WALLEY *et al.* 2012). In addition, methylation polymorphism was considered to contribute to morphological variation within the species and was previously reported to correlate with leaf morphology (SALMON *et al.* 2008).

Among the traits related to morphology, phenotypes of curd formation and flowering time have been most extensively studied. Formation of curd was characterized by three changes in SADIK (1962), including curtailed leaf development, lateral buds elongated into shoots which make up the surface of curd, and shortened internodes. In Arabidopsis, double mutants in *CAULIFLOWER* (*CAL)* and *APETALA1* (*AP1*) showed a phenotype similar to cauliflower and a nonsense mutation was identified in cauliflower which suggested that inactivation of the *CAULIFLOWER* (*CAL*) gene was related to the curd phenotype (KEMPIN *et al.* 1995; PURUGGANAN *et al.* 2000). Furthermore, in a doubled haploid population developed by crossing cauliflower and broccoli, curding phenotype showed cosegregation with *BoAP1-a* on O6, and *BoCAL-a* on O3 (SMITH and

KING 2000). However, these genes were not the only contributors to curd morphology. Indeed, the *CAL* mutation was also found in broccoli and other *B. oleracea* plants that produce non-curding and wild type inflorescences (PURUGGANAN *et al.* 2000). LABATE *et al.* (2006) used *BoCAL-a*, *BoAP1-a*, and *BoGSL-ELONG*, a gene involved in glucosinolate synthesis that was found in cauliflower, to predict inflorescence type in broccoli and cauliflower and suggested that these three genes were not sufficient to explain the inflorescence phenotypes. DUCLOS and BJÖRKMAN (2008) showed that homologs of flowering formation genes in Arabidopsis were not able to explain the inflorescence formation and structure of broccoli and cauliflower.

In early QTL mapping studies, the curd was described by the number of buds on the main inflorescence (KENNARD *et al.* 1994). It was later on better depicted by measuring first-rank branching, side-branches, cluster width, curd width, apical shoot length and first branch length (LAN and PATERSON (2000)). GAO *et al.* (2007) studied a population developed by crossing broccoli and cauliflower, with curd formation classified visually as broccoli-like, intermediate and cauliflower-like, and mapped three QTLs, one of which was associated with *BoAP1-a*. Collectively, these prior QTL studies make clear that formation of curd is affected by multiple genes.

Flowering time is an important trait because of its economic significance, and it is widely studied. Models have been developed to predict flowering time based on QTLs (UPTMOOR *et al.* 2008; UPTMOOR *et al.* 2012). In several studies, QTLs for flowering time were often located on regions containing Brassica homologs of flowering genes in Arabidopsis, including *CONSTANS* (*CO*) and *FLOWERING LOCUS* C (*FLC*) (BOHUON

et al. 1998; AXELSSON *et al.* 2001; OKAZAKI *et al.* 2007). In Arabidopsis, *CO* is involved in the photoperiod response pathway and induces flowering under long days, while expression of *FLC* would repress flowering until the plant is vernalized (GREENUP *et al.* 2009). One *FLC* homolog, *BoFLC4-1*, was proposed to play a similar role in cabbage (LIN *et al.* 2005). In BOHUON *et al.* (1998), six QTLs were mapped on O2, O3, O5 and O9, three of which (on O2, O3, O9) were shown to contain homologs of *CO* in Arabidopsis. In the same population, RAZI *et al.* (2008) exclude the possibility of an *FLC* homolog being a candidate gene for flowering time. In AXELSSON *et al.* (2001), through the mapping of homologs of *CO* and *FLC* and QTLs for flowering time, a *CO* homolog did co-localize with a flowering QTL. OKAZAKI *et al.* (2007) found that one *FLC* homolog, *BoFLC2*, was located in the same region as a QTL with the largest effect in the study, and that sequence polymorphism of *BoFLC2* was well correlated with flowering time. The different candidate genes detected in studies above might due to populations derived by various parents.

B. oleracea, a fascinating species with diverse morphology and extensive genomic polyploidization, may provide a model for evolutionary genetics. Its close relationship with *A. thaliana* enables transferability of genetic information between the two species. With its genome sequenced recently (YU *et al.* 2013; PARKIN *et al.* 2014), identification of the genes responsible for trait variation within the species, among the morphotypes, could be greatly facilitated.

Reference

ALONSO-BLANCO, C., M. G. AARTS, L. BENTSINK, J. J. KEURENTJES, M. REYMOND *et al.*, 2009 What has natural variation taught us about plant development, physiology, and adaptation? The Plant Cell Online **21:** 1877-1896.

AXELSSON, T., O. SHAVORSKAYA and U. LAGERCRANTZ, 2001 Multiple flowering time QTLs within several *Brassica* species could be the result of duplicated copies of one ancestral gene. Genome **44:** 856-864.

BABULA, D., M. KACZMAREK, A. BARAKAT, M. DELSENY, C. F. QUIROS *et al.*, 2003 Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. Mol Genet Genomics **268:** 656-665.

BABULA, D., M. KACZMAREK, P. A. ZIÓŁKOWSKI and J. SADOWSKI, 2007 *Brassica oleracea*, pp. 227-285 in *Vegetables*. Springer.

BARKER, G. C., T. R. LARSON, I. A. GRAHAM, J. R. LYNN and G. J. KING, 2007 Novel insights into seed fatty acid synthesis and modification pathways from genetic diversity and quantitative trait loci analysis of the *Brassica* C genome. Plant physiology **144:** 1827-1842.

BEECHER, C. W., 1994 Cancer preventive properties of varieties of *Brassica oleracea*: a review. The American journal of clinical nutrition **59:** 1166S-1170S.

BETTEY, M., W. FINCH-SAVAGE, G. J. KING and J. LYNN, 2000 Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in *Brassica oleracea*. New Phytologist **148:** 277-286.

BOHUON, E. J. R., D. J. KEITH, I. A. P. PARKIN, A. G. SHARPE and D. J. LYDIATE, 1996 Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor Appl Genet **93:** 833-839.

BOHUON, E. J. R., L. D. RAMSAY, J. A. CRAFT, A. E. ARTHUR, D. F. MARSHALL *et al.*, 1998 The association of flowering time quantitative trait loci with duplicated fegions and candidate loci in *Brassica oleracea*. Genetics **150:** 393-401.

BOWERS, J. E., B. A. CHAPMAN, J. RONG and A. H. PATERSON, 2003 Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433-438.

BOYES, D. C., M. E. NASRALLAH, J. VREBALOV and J. B. NASRALLAH, 1997 The self-incompatibility (S) haplotypes of Brassica contain highly divergent and rearranged sequences of ancient origin. The Plant Cell Online **9:** 237-247.

BROADLEY, M. R., J. P. HAMMOND, G. J. KING, D. ASTLEY, H. C. BOWEN *et al.*, 2008 Shoot calcium and magnesium concentrations differ between subtaxa, are highly heritable, and associate with potentially pleiotropic loci in *Brassica oleracea*. Plant Physiology **146:** 1707-1720.

BROWN, A. F., G. G. YOUSEF, K. K. CHEBROLU, R. W. BYRD, K. W. EVERHART *et al.*, 2014 High-density single nucleotide polymorphism (SNP) array mapping in Brassica oleracea: identification of QTL associated with carotenoid variation in broccoli florets. Theoretical and Applied Genetics **127:** 2051-2064.

CAMARGO, L., and T. OSBORN, 1996 Mapping loci controlling flowering time in *Brassica oleracea*. Theoretical and Applied Genetics **92:** 610-616.

CAMARGO, L., L. SAVIDES, G. JUNG, J. NIENHUIS and T. OSBORN, 1997 Location of the self-incompatibility locus in an RFLP and RAPD map of *Brassica oleracea*. Journal of Heredity **88:** 57-60.

CHEUNG, W., G. CHAMPAGNE, N. HUBERT and B. LANDRY, 1997 Comparison of the genetic maps of *Brassica napus* and *Brassica oleracea*. Theoretical and Applied Genetics **94:** 569-582.

COELHO, P. S., and A. A. MONTEIRO, 2003 Inheritance of downy mildew resistance in mature broccoli plants. Euphytica **131:** 65-69.

COGAN, N., J. LYNN, G. KING, M. KEARSEY, H. NEWBURY *et al.*, 2002 Identification of genetic factors controlling the efficiency of Agrobacterium rhizogenes-mediated transformation in *Brassica oleracea* by QTL analysis. Theoretical and Applied Genetics **105:** 568-576.

COGAN, N. O., H. J. NEWBURY, A. M. OLDACRES, J. R. LYNN, M. J. KEARSEY *et al.*, 2004 Identification and characterization of QTL controlling Agrobacterium-mediated transient and stable transformation of *Brassica oleracea*. Plant biotechnology journal **2:** 59-69.

COLLARD, B., M. JAHUFER, J. BROUWER and E. PANG, 2005 An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica **142:** 169-196.

DESCHAMPS, S., V. LLACA and G. D. MAY, 2012 Genotyping-by-sequencing in plants. Biology **1:** 460-483.

DIXON, G. R., 2007 *Vegetable brassicas and related crucifers / Geoffrey R Dixon*. Wallingford, UK ; Cambridge, MA : CABI Pub., c2007.

DUCLOS, D. V., and T. BJÖRKMAN, 2008 Meristem identity gene expression during curd proliferation and flower initiation in *Brassica oleracea*. Journal of experimental botany **59:** 421-433.

FARINHÓ, M., P. COELHO, J. CARLIER, D. SVETLEVA, A. MONTEIRO *et al.*, 2004 Mapping of a locus for adult plant resistance to downy mildew in broccoli (*Brassica oleracea* convar. *italica*). Theoretical and applied genetics **109:** 1392-1398.

FARNHAM, M., M. WANG and C. THOMAS, 2002 A single dominant gene for downy mildew resistance in broccoli. Euphytica **128:** 405-407.

FIGDORE, S., M. FERREIRA, M. SLOCUM and P. WILLIAMS, 1993 Association of RFLP markers with trait loci affecting clubroot resistance and morphological characters in Brassica oleracea L. Euphytica **69:** 33-44.

FOOD, and N. AGRICULTURE ORGANIZATION OF THE UNITED, 2014 FAOSTAT, pp. FAO, Rome, Italy.

FOWKE, J. H., F.-L. CHUNG, F. JIN, D. QI, Q. CAI *et al.*, 2003 Urinary isothiocyanate levels, brassica, and human breast cancer. Cancer Research **63:** 3980-3986.

GAO, M., G. LI, B. YANG, D. QIU, M. FARNHAM *et al.*, 2007 High-density Brassica oleracea linkage map: identification of useful new linkages. Theor Appl Genet **115:** 277-287.

GIOVANNELLI, J. L., M. W. FARNHAM, M. WANG and A. E. STRAND, 2002 Development of sequence characterized amplified region markers linked to downy mildew resistance in broccoli. Journal of the American Society for Horticultural Science **127:** 597-601.

GRANDCLÉMENT, C., and G. THOMAS, 1996 Detection and analysis of QTLs based on RAPD markers for polygenic resistance to Plasmodiophora brassicae Woron in *Brassica oleracea* L. Theoretical and Applied Genetics **93:** 86-90.

GREENUP, A., W. J. PEACOCK, E. S. DENNIS and B. TREVASKIS, 2009 The molecular biology of seasonal flowering-responses in Arabidopsis and the cereals. Annals of Botany **103:** 1165-1172.

HANSEN, L. N., R. ORTIZ and S. B. ANDERSEN, 1999 Genetic analysis of protoplast regeneration ability in *Brassica oleracea*. Plant cell, tissue and organ culture **58:** 127-132.

HE, J., X. ZHAO, A. LAROCHE, Z.-X. LU, H. LIU *et al.*, 2014 Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in plant science **5**.

HENNIG, K., R. VERKERK, M. DEKKER and G. BONNEMA, 2013 Quantitative trait loci analysis of non-enzymatic glucosinolate degradation rates in *Brassica oleracea* during food processing. Theoretical and Applied Genetics **126:** 2323-2334.

HOLME, I., A. TORP, L. HANSEN and S. ANDERSEN, 2004 Quantitative trait loci affecting plant regeneration from protoplasts of *Brassica oleracea*. Theoretical and Applied genetics **108:** 1513-1520.

HU, J., J. SADOWSKI, T. OSBORN, B. LANDRY and C. QUIROS, 1998 Linkage group alignment from four independent *Brassica oleracea* RFLP maps. Genome **41:** 226-235.

JIAO, Y., J. LEEBENS-MACK, S. AYYAMPALAYAM, J. E. BOWERS, M. R. MCKAIN *et al.*, 2012 A genome triplication associated with early diversification of the core eudicots. Genome Biol **13:** R3.

KACZMAREK, M., G. KOCZYK, P. A. ZIOLKOWSKI, D. BABULA-SKOWRONSKA and J. SADOWSKI, 2009 Comparative analysis of the *Brassica oleracea* genetic map and the *Arabidopsis thaliana* genome. Genome **52:** 620-633.

KEMPIN, S. A., B. SAVIDGE and M. F. YANOFSKY, 1995 Molecular basis of the cauliflower phenotype in Arabidopsis. Science **267:** 522-525.

KENNARD, W., M. SLOCUM, S. FIGDORE and T. OSBORN, 1994 Genetic analysis of morphological variation in *Brassica oleracea* using molecular markers. Theoretical and Applied Genetics **87:** 721-732.

KIANIAN, S., and C. QUIROS, 1992 Generation of a *Brassica oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theoretical and Applied Genetics **84:** 544-554.

KOOKE, R., E. WIJNKER and J. J. KEURENTJES, 2012 Backcross populations and near isogenic lines, pp. 3-16 in *Quantitative Trait Loci (QTL)*. Springer.

KOWALSKI, S. P., T.-H. LAN, K. A. FELDMANN and A. H. PATERSON, 1994 Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics **138:** 499-510.

KRISTAL, A. R., and J. W. LAMPE, 2002 Brassica vegetables and prostate cancer risk: a review of the epidemiological evidence. Nutrition and cancer **42:** 1-9.

LABATE, J. A., L. D. ROBERTSON, A. M. BALDO and T. BJORKMAN, 2006 Inflorescence identity gene alleles are poor predictors of inflorescence type in broccoli and cauliflower. JOURNAL-AMERICAN SOCIETY FOR HORTICULTURAL SCIENCE **131:** 667.

LAGERCRANTZ, U., and D. J. LYDIATE, 1996 Comparative genome mapping in *Brassica*. Genetics **144:** 1903-1910.

LAN, T.-H., T. A. DELMONTE, K. P. REISCHMANN, J. HYMAN, S. P. KOWALSKI *et al.*, 2000 An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Research **10:** 776-788.

LAN, T.-H., and A. H. PATERSON, 2000 Comparative mapping of quantitative trait loci sculpting the curd of *Brassica oleracea*. Genetics **155:** 1927-1954.

LAN, T. H., and A. H. PATERSON, 2001 Comparative mapping of QTL determining the plant size of *Brassica oleracea*. Theor Appl Genet **103:** 383-397.

LANDRY, B. S., N. HUBERT, R. CRETE, M. S. CHANG, S. E. LINCOLN *et al.*, 1992 A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). Genome **35:** 409-420.

LI, G., M. GAO, B. YANG and C. QUIROS, 2003a Gene for gene alignment between the *Brassica* and Arabidopsis genomes by direct transcriptome mapping. Theoretical and Applied Genetics **107:** 168-180.

LI, G., and C. F. QUIROS, 2001 Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. Theoretical and Applied Genetics **103:** 455-461.

LI, G., A. RIAZ, S. GOYAL, S. ABEL and C. QUIROS, 2001a Inheritance of three major genes involved in the synthesis of aliphatic glucosinolates in *Brassica oleracea*. Journal of the American society for horticultural science **126:** 427-431.

LI, L., and D. GARVIN, 2003 Molecular mapping of Or, a gene inducing beta-carotene accumulation in cauliflower (*Brassica oleracea* L. var. *botrytis*). Genome **46:** 588-594.

LI, L., S. LU, D. M. O'HALLORAN, D. F. GARVIN and J. VREBALOV, 2003b High-resolution genetic and physical mapping of the cauliflower high-beta-carotene gene Or (Orange). Mol Genet Genomics **270:** 132-138.

LI, L., D. J. PAOLILLO, M. V. PARTHASARATHY, E. M. DIMUZIO and D. F. GARVIN, 2001b A novel gene mutation that confers abnormal patterns of β-carotene accumulation in cauliflower (*Brassica oleracea* var. *botrytis*). The Plant Journal **26:** 59-67.

LIN, S.-I., J.-G. WANG, S.-Y. POON, C.-L. SU, S.-S. WANG *et al.*, 2005 Differential regulation of FLOWERING LOCUS C expression by vernalization in cabbage and Arabidopsis. Plant physiology **137:** 1037-1048.

LIU, S., Y. LIU, X. YANG, C. TONG, D. EDWARDS *et al.*, 2014 The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nature communications **5**.

LUKENS, L., F. ZOU, D. LYDIATE, I. PARKIN and T. OSBORN, 2003 Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. Genetics **164:** 359-372.

LYDIATE, D., A. SHARPE, U. LAGERCRANTZ and I. PARKIN, 1993 Mapping the *Brassica* genome. Outlook on agriculture.

LYSAK, M. A., M. A. KOCH, A. PECINKA and I. SCHUBERT, 2005 Chromosome triplication found across the tribe Brassiceae. Genome Res **15:** 516-525.

MCGRATH, J. M., and C. QUIROS, 1991 Inheritance of isozyme and RFLP markers in *Brassica campestris* and comparison with *B. oleracea*. Theoretical and Applied Genetics **82:** 668-673.

MCGRATH, J. M., C. F. QUIROS, J. J. HARADA and B. S. LANDRY, 1990 Identification of *Brassica oleracea* monosomic alien chromosome addition lines with molecular markers reveals extensive gene duplication. Molecular and General Genetics MGG **223:** 198-204.

MEI, J., Y. DING, K. LU, D. WEI, Y. LIU *et al.*, 2013 Identification of genomic regions involved in resistance against *Sclerotinia sclerotiorum* from wild *Brassica oleracea*. Theoretical and Applied Genetics **126:** 549-556.

MORIGUCHI, K., C. KIMIZUKA-TAKAGI, K. ISHII and K. NOMURA, 1999 A genetic map based on RAPD, RFLP, isozyme, morphological markers and QTL analysis for clubroot resistance in *Brassica oleracea*. Breeding science **49:** 257-265.

NAGAHARU, U., 1935 Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jap J Bot **7:** 389-452.

NAGAOKA, T., M. DOULLAH, S. MATSUMOTO, S. KAWASAKI, T. ISHIKAWA *et al.*, 2010 Identification of QTLs that control clubroot resistance in *Brassica oleracea* and comparative analysis of clubroot resistance genes between *B. rapa* and *B. oleracea*. Theoretical and applied genetics **120:** 1335-1346.

OCKENDON, D., and P. GATES, 1975 GROWTH OF CROSS-AND SELF-POLLEN TUBES IN THE STYLES OF *BRASSICA OLERACEA*. New Phytologist **75:** 155-160.

OKAZAKI, K., K. SAKAMOTO, R. KIKUCHI, A. SAITO, E. TOGASHI *et al.*, 2007 Mapping and characterization of *FLC* homologs and QTL analysis of flowering time in *Brassica oleracea*. Theor Appl Genet **114:** 595-608.

PARKIN, I., A. SHARPE, D. KEITH and D. LYDIATE, 1995 Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome **38:** 1122-1131.

PARKIN, I. A., C. KOH, H. TANG, S. J. ROBINSON, S. KAGALE *et al.*, 2014 Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome biology **15:** R77.

PURUGGANAN, M. D., A. L. BOYLES and J. I. SUDDITH, 2000 Variation and selection at the *CAULIFLOWER* floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. Genetics **155:** 855-862.

QIU, D., M. GAO, G. LI and C. QUIROS, 2009 Comparative sequence analysis for *Brassica oleracea* with similar sequences in *B. rapa* and *Arabidopsis thaliana*. Plant cell reports **28:** 649-661.

QUIROS, C., O. OCHOA, S. KIANIAN and D. DOUCHES, 1987 Analysis of the *Brassica oleracea* genome by the generation of *B. campestris-oleracea* chromosome addition lines: characterization by isozymes and rDNA genes. Theoretical and Applied Genetics **74:** 758-766.

RAE, A., E. HOWELL and M. KEARSEY, 1999 More QTL for flowering time revealed by substitution lines in *Brassica oleracea*. Heredity **83:** 586-596.

RAFALSKI, A., 2002 Applications of single nucleotide polymorphisms in crop genetics. Current opinion in plant biology **5:** 94-100.

RAKOW, G., 2004 Species origin and economic importance of *Brassica*, pp. 3-11 in *Brassica*. Springer.

RAMSAY, L., D. JENNINGS, M. KEARSEY, D. MARSHALL, E. BOHUON *et al.*, 1996 The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome **39:** 558-567.

RAZI, H., E. HOWELL, H. NEWBURY and M. KEARSEY, 2008 Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in *Brassica oleracea*? Theoretical and Applied Genetics **116:** 179-192.

ROCHERIEUX, J., P. GLORY, A. GIBOULOT, S. BOURY, G. BARBEYRON *et al.*, 2004 Isolate-specific and broad-spectrum QTLs are involved in the control of clubroot in *Brassica oleracea*. Theoretical and applied genetics **108:** 1555-1563.

SADIK, S., 1962 Morphology of the curd of cauliflower. American Journal of Botany**:** 290-297.

SALATHIA, N., J. R. LYNN, A. J. MILLAR and G. J. KING, 2007 Detection and resolution of genetic loci affecting circadian period in *Brassica oleracea*. Theoretical and Applied Genetics **114:** 683-692.

SALMON, A., J. CLOTAULT, E. JENCZEWSKI, V. CHABLE and M. J. MANZANARES-DAULEUX, 2008 *Brassica oleracea* displays a high level of DNA methylation polymorphism. Plant Science **174:** 61-70.

SCHOPFER, C. R., M. E. NASRALLAH and J. B. NASRALLAH, 1999 The male determinant of self-incompatibility in *Brassica*. Science **286:** 1697-1700.

SEBASTIAN, R., M. KEARSEY and G. J. KING, 2002 Identification of quantitative trait loci controlling developmental characteristics of *Brassica oleracea* L. Theoretical and Applied Genetics **104:** 601-609.

SLOCUM, M., S. FIGDORE, W. KENNARD, J. SUZUKI and T. OSBORN, 1990 Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theoretical and Applied Genetics **80:** 57-64.

SMITH, L. B., and G. J. KING, 2000 The distribution of BoCAL-a alleles in *Brassica oleracea* is consistent with a genetic model for curd development and domestication of the cauliflower. Molecular Breeding **6:** 603-613.

SONG, K., T. OSBORN and P. WILLIAMS, 1988 *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). TAG Theoretical and Applied Genetics **76:** 593-600.

SONG, K., T. C. OSBORN and P. H. WILLIAMS, 1990 *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). TAG Theoretical and Applied Genetics **79:** 497-506.

SPARROW, P., T. TOWNSEND, A. ARTHUR, P. DALE and J. IRWIN, 2004 Genetic analysis of *Agrobacterium tumefaciens* susceptibility in *Brassica oleracea*. Theoretical and Applied Genetics **108:** 644-650.

TERRY, P., A. WOLK, I. PERSSON and C. MAGNUSSON, 2001 Brassica vegetables and breast cancer risk. Jama **285:** 2975-2977.

TOWN, C. D., F. CHEUNG, R. MAITI, J. CRABTREE, B. J. HAAS *et al.*, 2006 Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell **18:** 1348-1359.

TRUCO, M., J. HU, J. SADOWSKI and C. QUIROS, 1996 Inter-and intra-genomic homology of the *Brassica* genomes: implications for their origin and evolution. Theoretical and Applied Genetics **93:** 1225-1233.

UPTMOOR, R., J. LI, T. SCHRAG and H. STÜTZEL, 2012 Prediction of flowering time in *Brassica oleracea* using a quantitative trait loci-based phenology model. Plant Biology **14:** 179-189.

UPTMOOR, R., T. SCHRAG, H. STÜTZEL and E. ESCH, 2008 Crop model based QTL analysis across environments and QTL based estimation of time to floral induction and flowering in *Brassica oleracea*. Molecular Breeding **21:** 205-216.

VAN POPPEL, G., D. T. VERHOEVEN, H. VERHAGEN and R. A. GOLDBOHM, 1999 Brassica vegetables and cancer prevention, pp. 159-168 in *Advances in Nutrition and Cancer 2*. Springer.

VERHOEVEN, D. T., R. A. GOLDBOHM, G. VAN POPPEL, H. VERHAGEN and P. A. VAN DEN BRANDT, 1996 Epidemiological studies on brassica vegetables and cancer risk. Cancer Epidemiology Biomarkers & Prevention **5:** 733-748.

VOORRIPS, R., M. JONGERIUS and H. KANNE, 1997 Mapping of two genes for resistance to clubroot (Plasmodiophora brassicae) in a population of doubled haploid lines of *Brassica oleracea* by means of RFLP and AFLP markers. Theoretical and applied genetics **94:** 75-82.

WALLEY, P. G., J. CARDER, E. SKIPPER, E. MATHAS, J. LYNN *et al.*, 2012 A new broccoli× broccoli immortal mapping population and framework genetic map: tools for breeders and complex trait analysis. Theoretical and Applied Genetics **124:** 467-484.

WHITE, P., J. P. HAMMOND, G. J. KING, H. C. BOWEN, R. M. HAYDEN *et al.*, 2010 Genetic analysis of potassium use efficiency in *Brassica oleracea*. Annals of botany **105:** 1199-1210.

WILLIAMS, P. H., and C. B. HILL, 1986 Rapid-cycling populations of *Brassica*. Science **232:** 1385-1389.

YU, J., M. ZHAO, X. WANG, C. TONG, S. HUANG *et al.*, 2013 Bolbase: a comprehensive genomics database for *Brassica oleracea*. BMC genomics **14:** 664.

ZHANG, B., C. LIU, Y. WANG, X. YAO, F. WANG *et al.*, 2015 Disruption of a CAROTENOID CLEAVAGE DIOXYGENASE 4 gene converts flower colour from white to yellow in *Brassica* species. New Phytologist.

ZHANG, X., and S. R. WESSLER, 2004 Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. Proceedings of the National Academy of Sciences of the United States of America **101:** 5589-5594.

ZHANG, Y., Z. FANG, Q. WANG, Y. LIU, L. YANG *et al.*, 2012 Chloroplast Subspecies-Specific SNP Detection and Its Maternal Inheritance in Brassica oleracea L. by Using a dCAPS Marker. Journal of Heredity **103:** 606-611.

CHAPTER 2

TRANSMISSION GENETICS OF TWO *Brassica oleracea* MORPHOTYPES IN CROSSES WITH A RAPID-CYCLING GENOTYPE

Lan-Shuan Shuang, Hugo Cuevas, Cornelia Lemke, Changsoo Kim, Hui Guo and Andrew H. Paterson. To be submitted to G3: Genes| Genomes| Genetics.

Abstract

Brassica oleracea is a species with remarkable morphological diversity, perhaps due in part to its complex genome composition, but certainly contributing to its great economic importance. To dissect morphological diversity of *B. oleracea*, two backcross populations were developed by using inbred lines of cabbage (Badger Inbred) and cauliflower (Orange) as donor parents, and a rapid cycling line (TO1434) as the recurrent parent. Genotypes of the two populations in their BC_4F_1 generation were determined by genotyping-by-sequencing (GBS). Across the 75 and 89 BC_4F_1 lines from the two populations, respectively, donor introgressions collectively covered most of the genome and were over-represented in some regions. Recombination events were not detected in the pericentromeric regions, and on one arm of chromosome 6. Based on these two populations, near isogenic lines of *B. oleracea* can be constructed in the near future.

Introduction

Single nucleotide polymorphism (SNP) markers had been considered to be an informative genetic marker since the beginning of the sequencing era. Their abundance and genome wide distribution offered the potential for increased genotyping throughput and marker density. With advances in sequencing technology, reduced cost and time in sequencing now enables SNPs to be widely applied in genetic studies, either by array-based or genotyping-by-sequencing (GBS) methods. GBS generally utilizes multiplex sequence genomic libraries with reduced genome complexity by a combination of sample barcoding, restriction enzyme digestion, and a next-generation sequencing (NGS) platform (ELSHIRE *et al.* 2011). Array-based SNP genotyping generally targets SNPs

known *a priori*, while GBS scans a reduced-representation sample of a genome *de novo*, and can be applied in different species without allelic bias caused by sequence diversity (DESCHAMPS *et al.* 2012; HE *et al.* 2014). Its ease of operation, high-throughput, time-, cost- and labor-efficiency, allow GBS to be utilized in various genetic studies, i.e. genome wide association (BANDILLO *et al.* 2013; ROMAY *et al.* 2013; UITDEWILLIGEN *et al.* 2013; SONAH *et al.* 2014), genomic diversity (FU and PETERSON 2011; LU *et al.* 2013; FU *et al.* 2014), genetic mapping (BAIRD *et al.* 2008; CHUTIMANITSAKUN *et al.* 2011; SPINDEL *et al.* 2013; WARD *et al.* 2013; DEOKAR *et al.* 2014; HUANG *et al.* 2014; LAMBEL *et al.* 2014; LI *et al.* 2014b; LIU *et al.* 2014a; TALUKDER *et al.* 2014; VÍQUEZ-ZAMORA *et al.* 2014; CONSORTIUM 2015), and genomic selection (POLAND *et al.* 2012; JARQUÍN *et al.* 2014). However, there are some challenges remaining in utilizing GBS. When applying GBS in large, complex, polyploid genomes or species without reference genomes, data analysis plays a critical role (HE *et al.* 2014). In addition, filtering informative SNPs from background noise remains a major concern in GBS studies.(DAVEY *et al.* 2013).

Brassica oleracea is a species with remarkable morphological diversity, perhaps in part due to its complex genome composition, but certainly contributing to its great economic importance. Many vegetables are in this species, including Brussels sprouts (var. *gemmifera*), broccoli (var. *italica*), cauliflower (var. *botrytis*), Chinese kale (var. *alboglabra*), kale and collard (var. *acephala*), kohlrabi (var. *gongylodes*), and cabbage (var. *capitata*), which act as sources of fiber, vitamins and minerals in our daily diet (DIXON 2007). Within the species, morphological variation is striking, including 'morphotypes' with enlarged edible organs such as the lateral buds of Brussels sprouts, the inflorescence of cauliflower and broccoli, apical meristem and leaves of cabbage,

bulbous lower stem epicotyl of kohlrabi, and leaves of kale (BABULA *et al.* 2007), and rapid cycling lines which are considered to be a genetic standard in the species (WILLIAMS and HILL 1986). Nuclear and chloroplast variation among the morphotypes have been described by molecular markers (SONG *et al.* 1988; ZHANG *et al.* 2012). *B. oleracea* has been considered a mesopolyploid with several genome duplication and triplication events followed by gene loss and chromosomal reshuffling (BOWERS *et al.* 2003; LYSAK *et al.* 2005; JIAO *et al.* 2012; LIU *et al.* 2014b), and triplicated blocks are distributed across the genome (LIU *et al.* 2014b).

QTL mapping studies in *B. oleracea* have mostly used F_2, F_3 and doubled haploid (DH) populations (BABULA *et al.* 2007). Production of F_2 and F_3 populations is fast and easy in self-compatible species, represents all three possible genotypes at a locus, and provides unique recombinational information from two gametes per individual. DH is developed by chromosome doubling of haploid of F_1 haploids, a technology which varies in feasibility among species, but confers immediate homozygosity (COLLARD *et al.* 2005) and therefore is attractive for replicated studies and inbreeding species where the phenotype associated with the heterozygote is of minimal importance.

To validate and increase the resolution of QTL, advanced backcross populations could be of value. By several cycles of backcrossing, a donor genome could be dissected in the homogenous background of a recurrent parent. Furthermore, with a combination of molecular markers, we can select a set of near isogenic lines (NILs) collectively, sampling most regions of the genome.

Near isogenic lines (NILs), with single or small numbers of introgressed segments from a donor parent in a homogeneous genetic background of a recurrent parent, can serve as a good resource in both genetic mapping and breeding (KOOKE *et al.* 2012). Since there are few (ideally one) introgressed segment(s) in each NIL, phenotypes due to QTLs on the segment(s) are rendered much more discrete than in F_2 or backcross populations, often behaving as simple Mendelian factors (PARAN and ZAMIR 2003). QTL mapping based on NILs can thus increase accuracy of QTL position, and detect small effect QTLs that might otherwise be obscured by larger-effect genes in more complex populations. In addition, because of the fixed genotype of NILs, they can be replicated in different environments to test interaction between genetic and environmental factors (MONFORTE and TANKSLEY 2000). By crossing NILs to the recurrent parent, fine mapping of specific QTLs toward their cloning is facilitated.

Different from the mapping populations commonly used in *B. oleracea*, F_2, F_3, and DH, construction of NILs is labor- and time-extensive, since it required multiple cycles of backcrossing. Similarly to DH, NILs are immortal homozygous lines. In contrast with the other populations, NILs have different genetic composition. While NILs are mostly composed of a homogenous genetic background from the recurrent parent, the other types of lines have approximately equal contributions from both parents. As a result, epistasis among various loci may be much more prominent in F_2, F_3 and DH, but in NILs is limited to interactions between closely-linked genes.

As a valuable resource, NILs have been developed in many species. A population of recombinant backcross substitution lines has been developed in *B. oleracea* (RAMSAY

et al. 1996). The population was employed for QTL mapping of flowering time and results suggested that more QTL could be found than in a doubled haploid population (RAE *et al.* 1999). To maximize the identification of introgressed fragments, high marker density is preferred and can now be reached through SNP markers (FLETCHER *et al.* 2013). Using SNPs to reveal NIL genotypes has been applied in Arabidopsis (FLETCHER *et al.* 2013), maize (PEA *et al.* 2013), soybean (SEVERIN *et al.* 2010), and rice (ARBELAEZ *et al.* 2015).

To dissect cabbage and cauliflower genomes in a rapid cycling background, we developed two BC_4 populations. A rapid cycling line, with fast growth, short generation time and self-compatible characteristics, is the recurrent parent (WILLIAMS and HILL 1986). Cauliflower and cabbage, which are diverse in both morphology and molecular genetics and also sample diverse lineages within *Brassica oleracea* (SONG *et al.* 1988), are donor parents.

Materials and Methods

Plant materials and population development

Two populations were constructed using an inbred rapid cycling line (TO1434) as common recurrent parent and cabbage [*B. oleracea* ssp. *capitata* L., 'Badger Inbred' (BIL)], and cauliflower [*B. oleracea* ssp. *botrytis* L., mutant for *Orange* gene (ORG) (Lu et al., 2006)] as donor parents. TO1434 was pollinated by BIL and ORG to produce F_1s, so that the progenies would have cytoplasm from TO1434. Then, TO1434 was used as pollen plant for backcrosses to reach BC_4F_1. During population construction, plants were grown in a growth chamber with a 15.5 hour photoperiod and 23℃/20℃ day/night

temperature. Eight weeks after sowing, plants that had not formed flower buds were vernalized for 70 days with an eight hour photoperiod and constant 4 ℃ temperature.

DNA extraction

DNA of each individual was extracted from fresh leaves as reported in PATERSON *et al.* (1993), with extraction buffer replaced by 0.35M sorbitol, 0.1M Tris, 0.005M EDTA, 0.04 M Sodium bisulfite and lysis buffer replaced by 0.2M Tris at pH8.0, 0.05M EDTA, 2M NaCl, 0.05M CTAB.

Genotyping-by-sequencing (GBS)

GBS library construction followed a slightly modified MSG (ANDOLFATTO *et al.* 2011) procedure. Constructed libraries were sequenced using an Illumina Miseq. Sequencing data was analyzed using TASSEL-GBS (GLAUBITZ *et al.* 2014). In TASSEL-GBS, the first 64 bps of each reads are mapped on a reference genome to decide the position of the reads. SNP is called based on the alignment of reads. Heterozygosity at a locus is 'called' (inferred) if two alleles inferred to be present at a probability greater than that of sequencing error. Otherwise, a locus will be called as a homozygote, depending on which allele is found.

SSR markers and PCR genotyping

A total of 479 SSR markers were ordered from published literature (LOWE *et al.* 2004; PIQUEMAL *et al.* 2005; BURGESS *et al.* 2006; INIGUEZ-LUY *et al.* 2008; WANG *et al.* 2012; IZZAH *et al.* 2014; SHI *et al.* 2014). Since some marker names from SHI *et al.* (2014) conflict with those of WANG *et al.* (2012), markers ordered from the former are indicated

with .j at the end. Each marker was BLASTed to the genome to reveal its physical location (PARKIN *et al.* 2014). Polymerase chain reaction (PCR) was performed using 30 ng of DNA as template, 1 U Taq polymerase, 1 µL of 10X PCR buffer (100 mM Tris-HCl at pH 9, 500 mM KCl, and 15 mM MgCl2), 1 µL of 2 mM dNTP, 1 µL of 25 mM MgCl2, 0.5 µL of 10 µM of each primer, with final reaction volume of 10 µL. PCR reaction was denatured at 95℃ for 3 minutes, followed by 11 cycles of 95℃ for 30 seconds, 55-65℃ for 1 minute with 1℃ increases at each cycle, and 72℃ for 1 minute; then another 33 cycles of 95℃ for 30 seconds, 55℃ for 1 minute and 72℃ for 1 minute. The final cycle at 72℃ was for 5 minutes, then the samples were held at 4℃. Amplified fragments were analyzed in 10% polyacrylamide gels with silver staining.

Data analysis

Recombination frequencies in BC_4F_1 generation were estimated by R/qtl (BROMAN *et al.* 2003). Introgression was inferred by finding two consecutive SNP alleles from the donor parent. Evaluation of introgression frequencies was tested by chi-squared test. T-tests were analyzed in statistical software R (TEAM 2014).

Results

Quality control

Two BC_4F_1 populations, of cabbage and cauliflower respectively, were genotyped by genotyping-by-sequencing (GBS) to reveal introgressed segments in each line. We found 2336 and 2500 SNP markers between the parents of the respective backcross populations.

Since GBS results are prone to some sequencing error, a SNP filtration was set up for BC_4F_1 populations (Table 2.1). First, SNPs that indicated heterozygotes in either of the inbred parents were discarded, totaling 798/2336 in the cabbage population and 1134/2500 in the cauliflower population. After several cycles of backcrossing, the donor parent allele could be confounded with the recurrent parent allele. Therefore, these SNPs would provide little information regarding introgressed segments.

Second, markers for which donor parent alleles were extremely enriched were also discarded. In the BC_4F_1 generation, an average of one in 16 individuals should be heterozygous for the donor allele. Markers that called heterozygotes with frequencies of more than 50% were discarded, totaling 18/2336 and 70/2500 in cabbage and cauliflower populations, respectively. Moreover, we would only have heterozygotes for the donor parent allele in this generation. Because of this, SNPs that called homozygotes for donor parent alleles in the individual were considered to be heterozygotes in this study.

Third, SNPs that failed to call heterozygotes in the population were discarded. These markers could have either low coverage in calling heterozygotes in any individuals or be prone to sequencing error in the parents which resulted in mis-called polymorphic markers, totaling 168/2336 in cabbage population and 95/2500 in cauliflower population.

Fourth, SNPs with abnormal segregation patterns when compared to flanking markers, totaling 21/2236 in the cabbage population and 19/2500 in the cauliflower population, were discarded, since homologous sequence could cause contamination in SNP calling and heterozygotes could be called as homozygotes when the locus has low coverage. Since SNPs were distributed densely, we were able to assume that pairwise

SNPs are linked and the recombination frequency between them should be smaller than 0.5. As a result, recombination frequencies and their LOD scores were calculated by R/qtl taken as an index. If a marker was unlinked with its flanking markers, the marker was discarded.

Finally, we combined pairwise SNPs separated by distances less than 100 bp. As a result, 424/2236 SNPs in the cabbage population and 361/2500 SNPs in the cauliflower population were discarded which would provide redundant information.

After filtering, 907 and 821 SNPs were used to genotype each individual with an average of 100.8 and 91.2 SNP markers on each chromosome, in cabbage and cauliflower backcross populations, respectively (Table 2.2). In summary, we have the most markers on chromosome 3, with 143 SNPs and 128 SNPs spanning 64.8 Mb, and the least markers on chromosome 6, with 59 SNPs across 39.1 Mb and 47 SNPs spanning 38.7 Mb in cabbage and cauliflower population, respectively. SNPs were most densely distributed on chromosome 2 with 0.4 Mb/SNP in the cabbage population and chromosome 1 with 0.5 Mb/SNP in the cauliflower population. Chromosome 6 in both populations had the lowest SNP density, with 0.7 Mb/SNP and 0.8 Mb/SNP in cabbage and cauliflower population. Distribution of SNPs across each chromosome is shown in Figure 2.1 and Figure 2.2. Several regions larger than 5 Mbp were not covered by SNPs, located on chromosomes 1, 6, and 8 in the cabbage backcross population and chromosomes 5, 7, and 9 in cauliflower. These ‘gaps’ were not covered because the SNPs in these regions were filtered due to either calling heterozygotes in one of the parents or failing to call heterozygotes in individuals. Since these two possible explanations might result from

either biological factors or sequencing error, more markers might be needed in these regions to reveal the genotypes.

Here, the threshold of calling introgressed segments was to have two consecutive markers called heterozygotes for donor alleles. If SNPs called homozygotes were flanked by SNPs called heterozygotes, we could infer homozygotes by checking the coverage, i.e. the number of sequencing reads that cover each of the SNPs. Under an assumption that the genotype was actually a heterozygote, the probability of drawing one of the alleles should be 0.5 and follow a binomial distribution. The higher the coverage, the less likely for the genotype to be heterozygote. A probability of 0.025 was set as a threshold. If the cumulative coverage of homozygote SNPs were more than five times, the regions were considered homozygous.

Genotype of the cabbage and cauliflower backcross populations

In the cabbage BC_4F_1 populations, 75 lines were genotyped and a total of 193 introgressed segments were inferred. We failed to identify any introgressed segments from 11 lines, including seven lines with more than 70% missing data and four lines without introgressed segments. We found 2.8 (193/68) segments on average, with 0 to 7 introgressed segments in each line. We have 11 lines with one segment, 16 with two, 13 with three and 24 with four or more segments (Table 2.3).

In the cauliflower BC_4F_1 population, 89 individuals were genotyped and a total of 191 introgressed segments were found. We failed to identify any introgressed segment in 11 individuals, which included six individuals with more than 70% missing data and five individuals without introgressed segments. Three additional segments were found by SSR

markers. In summary, 194 segments were identified, with individual lines averaging 2.3 (194/83) introgressed segments and ranging from one to seven segments. We found 23 lines with one introgressed segment, 21 with two, 18 with three, and 16 with four or more introgressed segments (Table 2.3).

It was expected that there would be 2.5 introgressed segments in the BC_4F_1 generation, inferred from the estimation of introgressed segment size of 22 cM (HANSON 1959). Here, we identified an average of 2.8 and 2.3 introgressed segments in the respective populations, which was not significantly different from our expectation by a one sample T-test.

Genome coverage

Most of the genome was covered by introgressed segments in both populations, excepting some small gaps. The distribution of the introgressed segments across the genomes are shown in Figure 2.3. Introgressed segments covered the cabbage genome 0 to 16 times and the cauliflower genome 0 to 13 times. Some gaps that were not covered by introgressions were on one end of chromosome 7 in the cabbage population, and two regions on chromosome 4 in cauliflower (Table 2.4). In contrast, there are several regions with over-abundance of donor alleles, including four regions on chromosomes 2, 4, 7 and 9 in the cabbage population and two regions on chromosome 5 in the cauliflower population (Table 2.5). If a region was inherited without interruption by recombination, it would be shown as a plateau in the distributions. The distributions of introgressed segments in two populations were combined (Figure 2.3) to compare the introgression

pattern from the two donor genomes. Most of the non-recombinant regions shared by the two populations were located near the centromeres on all chromosomes. Additionally, a shared region was on the arm of chromosome 6.

SSR markers can be applied in a targeted manner to validate introgression. To determine if regions with high levels of introgressions were due to segregation distortion, segregation ratios of their selfed progenies were investigated by SSR markers. Regions on chromosomes 2, 4 and 9 in the cabbage population and chromosome 5 in the cauliflower population were targeted. At each region, we sampled one to five BC_4F_1 families, totaling 15 to 66 individuals. No marker showed significant deviation toward the donor allele.

The average coverage of introgressed segments was 5.29 and 4.51 in the cabbage and cauliflower population, respectively. The distribution of average coverage on each chromosome is shown in Figure 2.4. The highest coverage was on chromosome 2 in the cabbage population and chromosome 8 in the cauliflower population. The lowest coverage was on chromosome 6, covered 2.85 times, in the cabbage population and chromosome 7, covered 3.37 times, in the cauliflower population. Most chromosomes were covered 2 to 4 times on average in the cabbage population, other than chromosome 2, 4, and 9 with higher coverage. In the cauliflower population, most chromosomes were covered 3 to 6 times. To compare the coverages of introgressed segments on each chromosome in two populations, two sample t-tests were performed. Introgressions in the cauliflower population were significantly higher on chromosomes 1, 5, 6 and 8 and lower on chromosomes 2, 4 and 9 than the cabbage population.

Discussion

We found 907 and 821 SNPs for two BC_4F_1 populations by GBS, for a marker density higher than most previous Brassica studies (SLOCUM *et al.* 1990; KIANIAN and QUIROS 1992; LANDRY *et al.* 1992; BOHUON *et al.* 1996; CAMARGO and OSBORN 1996; RAMSAY *et al.* 1996; CAMARGO *et al.* 1997; CHEUNG *et al.* 1997; VOORRIPS *et al.* 1997; HU *et al.* 1998; MORIGUCHI *et al.* 1999; SEBASTIAN *et al.* 2000; FARINHÓ *et al.* 2004; ROCHERIEUX *et al.* 2004; INIGUEZ-LUY *et al.* 2009). High marker density could result in finer resolution of introgression identification and better definition of recombination break points (FLETCHER *et al.* 2013). Furthermore, with the assistance of a reference genome, the order of markers could be decided by placing them along the physical map, instead of constructing a *de novo* linkage map (POLAND and RIFE 2012).

Though finer resolution was reached, several factors in GBS could affect introgression identification. Biological factors, including presence-absence variation, polymorphic restriction sites and differential methylation, along with technical factors including sequencing error and depth, can give rise to missing data and erroneous genotypes (POLAND and RIFE 2012). Variation at restriction sites and low sequencing depth can result in mis-calling a heterozygote as a homozygote, by uneven sampling of the two alleles or missing data at a locus. There are only two genotypes in a BC_4F_1 population, including homozygotes for recurrent parent alleles and heterozygotes. This suggests that genotypes called (by TASSEL) as homozygotes for the donor parent allele may actually be heterozygotes, with too little depth of coverage to find the recurrent parent allele. We still fail to distinguish between homozygotes for the recurrent parent

allele and heterozygotes, which may lead to potential introgressions hidden in SNPs showing recurrent parent background.

To ensure the accuracy of GBS results, different thresholds have been set in various studies. In FLETCHER *et al.* (2013), 2b-RAD sequencing was used to characterize a set of near isogenic lines in *A. thaliana*. A fairly stringent threshold in both SNP calling and introgression identification was set. When genotyping by 2b-RAD sequencing, 20x sequence coverage of the parental genome and 10x coverage of all other individuals was used. Introgression was identified by having three continuous markers called donor allele. To avoid genotyping error in heterozygotes, which might be the result of low quality SNPs or incorrect alignments of reads from homologous regions (XIE *et al.* 2010), we set the threshold for inferring introgression as having at least two consecutive SNPs calling heterozygote in the individual. Some true heterozygotes might fail to be indicated by nearby markers. Nonetheless, potential genotyping error listed above can be validated by SSR markers. GBS genotyping still provides a good foundation for NIL selection.

Some chromosome segments remained intact over four cycles of backcrossing in the two populations. Most such segments were located in pericentromeric heterochromatin, considered a recombination cold spot (EGEL and LANKENAU 2007). Sequence diversity might play a role in recombination. One shared non-recombinant region was on the arm of chromosome 6 near the self-incompatibility locus (S locus). Self-incompatibility is controlled by a single locus with multiple alleles in *B. oleracea*. Within the locus, several genes are involved. The most well-known are the genes regarding self-recognition, which are S receptor kinase (SRK) in stigma(STEIN *et al.* 1991)

and S locus protein 11/ S-locus cysteine-rich protein (SP11/SCR) in pollen (SCHOPFER *et al.* 1999; SUZUKI *et al.* 1999). The combination of alleles from these genes is known as the S haplotype. When comparing two S haplotypes in *B. napus*, it was shown that the two haplotypes had different gene orientation and one of the haplotypes had retrotransposons and haplotype-specific genes in the intergenic region, which might repress recombination (Cui et al. 1999). In *B. oleracea*, the region between SRK and SP11/SCR contained retrotransposon-like sequences (Fujimoto et al. 2006). As a result, the non-recombinant region on chromosome 6 might indicate sequence variation among the parental lines.

Recombinant backcross lines were previously developed in *B. oleracea* (RAMSAY *et al.* 1996), reaching the BC_2 generation by marker assisted selection with 138 markers then selecting 77 individuals that covered 82.6% of the genome. Even with marker assisted selection, unbalanced introgression could still be observed across the genome. Excessive introgressions were on the end of chromosome 8, where we observed a slightly higher than average coverage (8-9x) in the cauliflower population. On the other hand, a region on chromosome 6 was covered by one line and the chromosome was barely covered by the selfed-progenies because it contains a locus controlling self-incompatibility. Similar with previous results, relatively low coverages were observed on chromosome 6 in both populations (RAMSAY *et al.* 1996). Uneven introgression might be due to unintentional selection of low vigor or fertility plants and genes related to reproduction, such as sterility, self-incompatibility, unilateral incongruity, gamete and zygote viability (CHETELAT and MEGLIC 2000; MONFORTE and TANKSLEY 2000; JEUKEN and LINDHOUT 2004).

The regions with over-representation of donor alleles that we have identified overlapped with previously reported segregation distortion regions (SDR). In WANG *et al.* (2012), a cabbage genetic map was constructed based on a doubled haploid (DH) population. The longest SDR that they discovered was on chromosome 2, matching our result in the cabbage population. IZZAH *et al.* (2014) found the longest SDR on chromosome 5 from a cross between two cabbage lines, containing the region which we found in the cauliflower backcross population. However, segregation of the introgressed segments in their selfed-progeny did not deviate toward the donor allele, which suggests that the retention of the introgressions are not due to defection during meiosis. Our two populations share a common recurrent parent and the regions with over-representation of donor alleles in the two populations did not overlap, which excludes the possibility of a negative effect on fertility contributed by the recurrent allele. In contrast, it is possible that donor alleles in these regions might be advantageous to fertility so that they are favored to be retained in the population.

Regions without inferred introgression could be the result of either no genomic introgression during population development, or genotyping error that can be validated by SSR markers. In our result, low introgression is often located at the end of the chromosome. This might be resulted from either the genotyping error, including the combination of erroneous genotypes and a lack of informative markers, or higher level of recombination on the telomeric region which increases the probability of purging out the introgressions. During population development, we lost some families because of the failure of seed production in backcrosses, even though we did not do any selection. Regions with low coverage of introgressions are likely due to donor alleles carrying

genes that affect sterility, fertility, lethality or vigor. Here, 'low coverage' indicated the regions that were not covered at all, an event that is inferred to be non-random by chi-squared statistics with significance threshold of 0.05.

In this study, we reported genotypes with high SNP density of two BC4F1 populations. Based on the result, we can target backcross families with small numbers of segments and select near isogenic lines of *B. oleracea* from these families.

Reference

ANDOLFATTO, P., D. DAVISON, D. EREZYILMAZ, T. T. HU, J. MAST *et al.*, 2011 Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome research **21:** 610-617.

ARBELAEZ, J. D., L. T. MORENO, N. SINGH, C.-W. TUNG, L. G. MARON *et al.*, 2015 Development and GBS-genotyping of introgression lines (ILs) using two wild species of rice, *O. meridionalis* and *O. rufipogon*, in a common recurrent parent, *O. sativa* cv. Curinga. Molecular Breeding **35:** 1-18.

BABULA, D., M. KACZMAREK, P. A. ZIÓŁKOWSKI and J. SADOWSKI, 2007 *Brassica oleracea*, pp. 227-285 in *Vegetables*. Springer.

BAIRD, N. A., P. D. ETTER, T. S. ATWOOD, M. C. CURREY, A. L. SHIVER *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one **3:** e3376.

BANDILLO, N., C. RAGHAVAN, P. A. MUYCO, M. A. L. SEVILLA, I. T. LOBINA *et al.*, 2013 Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice **6:** 1-15.

BOHUON, E. J. R., D. J. KEITH, I. A. P. PARKIN, A. G. SHARPE and D. J. LYDIATE, 1996 Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor Appl Genet **93:** 833-839.

BOWERS, J. E., B. A. CHAPMAN, J. RONG and A. H. PATERSON, 2003 Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433-438.

BROMAN, K. W., H. WU, Ś. SEN and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890.

BURGESS, B., H. MOUNTFORD, C. J. HOPKINS, C. LOVE, A. E. LING *et al.*, 2006 Identification and characterization of simple sequence repeat (SSR) markers

derived in silico from *Brassica oleracea* genome shotgun sequences. Molecular Ecology Notes **6:** 1191-1194.

CAMARGO, L., and T. OSBORN, 1996 Mapping loci controlling flowering time in *Brassica oleracea*. Theoretical and Applied Genetics **92:** 610-616.

CAMARGO, L., L. SAVIDES, G. JUNG, J. NIENHUIS and T. OSBORN, 1997 Location of the self-incompatibility locus in an RFLP and RAPD map of *Brassica oleracea*. Journal of Heredity **88:** 57-60.

CHETELAT, R., and V. MEGLIC, 2000 Molecular mapping of chromosome segments introgressed from *Solanum lycopersicoides* into cultivated tomato (*Lycopersicon esculentum*). Theoretical and Applied Genetics **100:** 232-241.

CHEUNG, W., G. CHAMPAGNE, N. HUBERT and B. LANDRY, 1997 Comparison of the genetic maps of *Brassica napus* and *Brassica oleracea*. Theoretical and Applied Genetics **94:** 569-582.

CHUTIMANITSAKUN, Y., R. W. NIPPER, A. CUESTA-MARCOS, L. CISTUÉ, A. COREY *et al.*, 2011 Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. BMC genomics **12:** 4.

COLLARD, B., M. JAHUFER, J. BROUWER and E. PANG, 2005 An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica **142:** 169-196.

CONSORTIUM, I. C. G. M., 2015 High-Resolution Linkage Map and Chromosome-Scale Genome Assembly for Cassava (*Manihot esculenta* Crantz) from 10 Populations. G3: Genes| Genomes| Genetics **5:** 133-144.

DAVEY, J. W., T. CEZARD, P. FUENTES-UTRILLA, C. ELAND, K. GHARBI *et al.*, 2013 Special features of RAD Sequencing data: implications for genotyping. Molecular Ecology **22:** 3151-3164.

DEOKAR, A. A., L. RAMSAY, A. G. SHARPE, M. DIAPARI, A. SINDHU *et al.*, 2014 Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. BMC genomics **15:** 708.

DESCHAMPS, S., V. LLACA and G. D. MAY, 2012 Genotyping-by-sequencing in plants. Biology **1:** 460-483.

DIXON, G. R., 2007 *Vegetable brassicas and related crucifers / Geoffrey R Dixon*. Wallingford, UK ; Cambridge, MA : CABI Pub., c2007.

EGEL, R., and D.-H. LANKENAU, 2007 *Recombination and meiosis: crossing-over and disjunction*. Springer Science & Business Media.

ELSHIRE, R. J., J. C. GLAUBITZ, Q. SUN, J. A. POLAND, K. KAWAMOTO *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one **6:** e19379.

FARINHÓ, M., P. COELHO, J. CARLIER, D. SVETLEVA, A. MONTEIRO *et al.*, 2004 Mapping of a locus for adult plant resistance to downy mildew in broccoli (*Brassica oleracea* convar. *italica*). Theoretical and applied genetics **109:** 1392-1398.

FLETCHER, R. S., J. L. MULLEN, S. YODER, W. L. BAUERLE, G. REUNING *et al.*, 2013 Development of a next-generation NIL library in *Arabidopsis thaliana* for dissecting complex traits. BMC genomics **14:** 655.

FU, Y.-B., B. CHENG and G. W. PETERSON, 2014 Genetic diversity analysis of yellow mustard (*Sinapis alba* L.) germplasm based on genotyping by sequencing. Genetic Resources and Crop Evolution **61:** 579-594.

FU, Y.-B., and G. W. PETERSON, 2011 Genetic diversity analysis with 454 pyrosequencing and genomic reduction confirmed the eastern and western division in the cultivated barley gene pool. The Plant Genome **4:** 226-237.

GLAUBITZ, J. C., T. M. CASSTEVENS, F. LU, J. HARRIMAN, R. J. ELSHIRE *et al.*, 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One **9:** e90346.

HANSON, W., 1959 Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. Genetics **44:** 833.

HE, J., X. ZHAO, A. LAROCHE, Z.-X. LU, H. LIU *et al.*, 2014 Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in plant science **5**.

HU, J., J. SADOWSKI, T. OSBORN, B. LANDRY and C. QUIROS, 1998 Linkage group alignment from four independent *Brassica oleracea* RFLP maps. Genome **41:** 226-235.

HUANG, Y.-F., J. A. POLAND, C. P. WIGHT, E. W. JACKSON and N. A. TINKER, 2014 Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. PloS one **9:** e102448.

INIGUEZ-LUY, F. L., L. LUKENS, M. W. FARNHAM, R. M. AMASINO and T. C. OSBORN, 2009 Development of public immortal mapping populations, molecular markers and linkage maps for rapid cycling *Brassica rapa* and *B. oleracea*. Theor Appl Genet **120:** 31-43.

INIGUEZ-LUY, F. L., A. V. VOORT and T. C. OSBORN, 2008 Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype. Theoretical and applied genetics **117:** 977-985.

IZZAH, N. K., J. LEE, M. JAYAKODI, S. PERUMAL, M. JIN *et al.*, 2014 Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. *capitata* L.) and construction of an EST-based genetic map. BMC genomics **15:** 149.

JARQUÍN, D., K. KOCAK, L. POSADAS, K. HYMA, J. JEDLICKA *et al.*, 2014 Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC genomics **15:** 740.

JEUKEN, M., and P. LINDHOUT, 2004 The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. Theoretical and Applied Genetics **109:** 394-401.

JIAO, Y., J. LEEBENS-MACK, S. AYYAMPALAYAM, J. E. BOWERS, M. R. MCKAIN *et al.*, 2012 A genome triplication associated with early diversification of the core eudicots. Genome Biol **13:** R3.

KIANIAN, S., and C. QUIROS, 1992 Generation of a *Brassica oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theoretical and Applied Genetics **84:** 544-554.

KOOKE, R., E. WIJNKER and J. J. KEURENTJES, 2012 Backcross populations and near isogenic lines, pp. 3-16 in *Quantitative Trait Loci (QTL)*. Springer.

LAMBEL, S., B. LANINI, E. VIVODA, J. FAUVE, W. P. WECHTER *et al.*, 2014 A major QTL associated with *Fusarium oxysporum* race 1 resistance identified in genetic populations derived from closely related watermelon lines using selective genotyping and genotyping-by-sequencing for SNP discovery. Theoretical and Applied Genetics **127:** 2105-2115.

LANDRY, B. S., N. HUBERT, R. CRETE, M. S. CHANG, S. E. LINCOLN *et al.*, 1992 A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). Genome **35:** 409-420.

LI, X., Y. WEI, A. ACHARYA, Q. JIANG, J. KANG *et al.*, 2014 A Saturated Genetic Linkage Map of Autotetraploid Alfalfa (*Medicago sativa* L.) Developed Using Genotyping-by-Sequencing Is Highly Syntenous with the *Medicago truncatula* Genome. G3: Genes| Genomes| Genetics **4:** 1971-1979.

LIU, H., M. BAYER, A. DRUKA, J. R. RUSSELL, C. A. HACKETT *et al.*, 2014a An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC genomics **15:** 104.

LIU, S., Y. LIU, X. YANG, C. TONG, D. EDWARDS *et al.*, 2014b The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nature communications **5**.

LOWE, A. J., C. MOULE, M. TRICK and K. J. EDWARDS, 2004 Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species. Theor Appl Genet **108:** 1103-1112.

LU, F., A. E. LIPKA, J. GLAUBITZ, R. ELSHIRE, J. H. CHERNEY *et al.*, 2013 Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS genetics **9:** e1003215.

LYSAK, M. A., M. A. KOCH, A. PECINKA and I. SCHUBERT, 2005 Chromosome triplication found across the tribe Brassiceae. Genome Res **15:** 516-525.

MONFORTE, A. J., and S. D. TANKSLEY, 2000 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.

MORIGUCHI, K., C. KIMIZUKA-TAKAGI, K. ISHII and K. NOMURA, 1999 A genetic map based on RAPD, RFLP, isozyme, morphological markers and QTL analysis for clubroot resistance in *Brassica oleracea*. Breeding science **49:** 257-265.

PARAN, I., and D. ZAMIR, 2003 Quantitative traits in plants: beyond the QTL. Trends in Genetics **19:** 303-306.

PARKIN, I. A., C. KOH, H. TANG, S. J. ROBINSON, S. KAGALE *et al.*, 2014 Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome biology **15:** R77.

PATERSON, A. H., C. L. BRUBAKER and J. F. WENDEL, 1993 A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Molecular Biology Reporter **11:** 122-127.

PEA, G., H. H. AUNG, E. FRASCAROLI, P. LANDI and M. E. PÈ, 2013 Extensive genomic characterization of a set of near-isogenic lines for heterotic QTL in maize (Zea mays L.). BMC genomics **14:** 61.

PIQUEMAL, J., E. CINQUIN, F. COUTON, C. RONDEAU, E. SEIGNORET *et al.*, 2005 Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theoretical and Applied Genetics **111:** 1514-1523.

POLAND, J., J. ENDELMAN, J. DAWSON, J. RUTKOSKI, S. WU *et al.*, 2012 Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome **5:** 103-113.

POLAND, J. A., and T. W. RIFE, 2012 Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome **5:** 92-102.

RAE, A., E. HOWELL and M. KEARSEY, 1999 More QTL for flowering time revealed by substitution lines in *Brassica oleracea*. Heredity **83:** 586-596.

RAMSAY, L., D. JENNINGS, M. KEARSEY, D. MARSHALL, E. BOHUON *et al.*, 1996 The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome **39:** 558-567.

ROCHERIEUX, J., P. GLORY, A. GIBOULOT, S. BOURY, G. BARBEYRON *et al.*, 2004 Isolate-specific and broad-spectrum QTLs are involved in the control of clubroot in *Brassica oleracea*. Theoretical and applied genetics **108:** 1555-1563.

ROMAY, M. C., M. J. MILLARD, J. C. GLAUBITZ, J. A. PEIFFER, K. L. SWARTS *et al.*, 2013 Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol **14:** R55.

SCHOPFER, C. R., M. E. NASRALLAH and J. B. NASRALLAH, 1999 The male determinant of self-incompatibility in *Brassica*. Science **286:** 1697-1700.

SEBASTIAN, R., E. HOWELL, G. KING, D. MARSHALL and M. KEARSEY, 2000 An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations. Theoretical and Applied Genetics **100:** 75-81.

SEVERIN, A. J., G. A. PEIFFER, W. W. XU, D. L. HYTEN, B. BUCCIARELLI *et al.*, 2010 An integrative approach to genomic introgression mapping. Plant physiology **154:** 3-12.

SHI, J., S. HUANG, J. ZHAN, J. YU, X. WANG *et al.*, 2014 Genome-wide microsatellite characterization and marker development in the sequenced *Brassica* crop species. DNA research **21:** 53-68.

SLOCUM, M., S. FIGDORE, W. KENNARD, J. SUZUKI and T. OSBORN, 1990 Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theoretical and Applied Genetics **80:** 57-64.

SONAH, H., L. O'DONOUGHUE, E. COBER, I. RAJCAN and F. BELZILE, 2014 Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant biotechnology journal.

SONG, K., T. OSBORN and P. WILLIAMS, 1988 *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). TAG Theoretical and Applied Genetics **76:** 593-600.

SPINDEL, J., M. WRIGHT, C. CHEN, J. COBB, J. GAGE *et al.*, 2013 Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theoretical and Applied Genetics **126:** 2699-2716.

STEIN, J. C., B. HOWLETT, D. C. BOYES, M. E. NASRALLAH and J. B. NASRALLAH, 1991 Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. Proceedings of the National Academy of Sciences **88:** 8816-8820.

SUZUKI, G., N. KAI, T. HIROSE, K. FUKUI, T. NISHIO *et al.*, 1999 Genomic organization of the S locus: identification and characterization of genes in SLG/SRK region of S9 haplotype of *Brassica campestris* (syn. *rapa*). Genetics **153:** 391-400.

TALUKDER, S. K., M. A. BABAR, K. VIJAYALAKSHMI, J. POLAND, P. V. PRASAD *et al.*, 2014 Mapping QTL for the traits associated with heat tolerance in wheat (*Triticum aestivum* L.). BMC genetics **15:** 97.

TEAM, R., 2014 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2011; ISBN: 3-900051-07-0, pp.

UITDEWILLIGEN, J. G., A.-M. A. WOLTERS, B. BJORN, T. J. BORM, R. G. VISSER *et al.*, 2013 A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One **8:** e62355.

VÍQUEZ-ZAMORA, M., M. CARO, R. FINKERS, Y. TIKUNOV, A. BOVY *et al.*, 2014 Mapping in the era of sequencing: high density genotyping and its application for mapping TYLCV resistance in Solanum pimpinellifolium. BMC genomics **15:** 1152.

VOORRIPS, R., M. JONGERIUS and H. KANNE, 1997 Mapping of two genes for resistance to clubroot (*Plasmodiophora brassicae*) in a population of doubled haploid lines of *Brassica oleracea* by means of RFLP and AFLP markers. Theoretical and applied genetics **94:** 75-82.

WANG, W., S. HUANG, Y. LIU, Z. FANG, L. YANG *et al.*, 2012 Construction and analysis of a high-density genetic linkage map in cabbage (*Brassica oleracea* L. var. *capitata*). BMC genomics **13:** 523.

WARD, J. A., J. BHANGOO, F. FERNÁNDEZ-FERNÁNDEZ, P. MOORE, J. SWANSON *et al.*, 2013 Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC genomics **14:** 2.

WILLIAMS, P. H., and C. B. HILL, 1986 Rapid-cycling populations of *Brassica*. Science **232:** 1385-1389.

XIE, W., Q. FENG, H. YU, X. HUANG, Q. ZHAO *et al.*, 2010 Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proceedings of the National Academy of Sciences **107:** 10578-10583.

ZHANG, Y., Z. FANG, Q. WANG, Y. LIU, L. YANG *et al.*, 2012 Chloroplast Subspecies-Specific SNP Detection and Its Maternal Inheritance in *Brassica oleracea* L. by Using a dCAPS Marker. Journal of Heredity **103:** 606-611.

Table 2.1. SNP algorithm.

	Cabbage	Cauliflower
Called heterozygote in the recurrent parent	263	358
Called heterozygote in the donor parent	535	776
Over-represented with donor allele	18	70
Failed to call heterozygote	168	95
Abnormal segregation	21	19
Co-segregate with pairwise SNP	424	361
Informative SNP	907	821
Total SNP from GBS	2336	2500

Table 2.2. SNP distribution on chromosomes.

	Cabbage					Cauliflower				
Ch	Start	End	Length	Number of markers	Density	Start	End	Length	Number of markers	Density
1	2,370	43,403,250	43,400,880	75	578,678	2,370	43,272,711	43,270,341	92	470,330
2	24,496	52,517,148	52,492,652	125	419,941	207,889	52,153,912	51,946,023	88	590,296
3	144,306	64,916,538	64,772,232	143	452,953	144,306	64,916,538	64,772,232	128	506,033
4	966,792	53,519,511	52,552,719	104	505,315	280,996	52,765,106	52,484,110	96	546,709
5	253,825	46,306,108	46,052,283	97	474,766	593,798	46,834,403	46,240,605	91	508,139
6	495,391	39,688,165	39,192,774	59	664,284	495,391	39,150,600	38,655,209	47	822,451
7	265,490	48,123,466	47,857,976	113	423,522	976,514	48,054,761	47,078,247	99	475,538
8	662,682	41,048,476	40,385,794	81	498,590	396,485	41,671,105	41,274,620	72	573,259
9	1,079,644	54,574,753	53,495,109	110	486,319	288,216	53,810,144	53,521,928	108	495,573

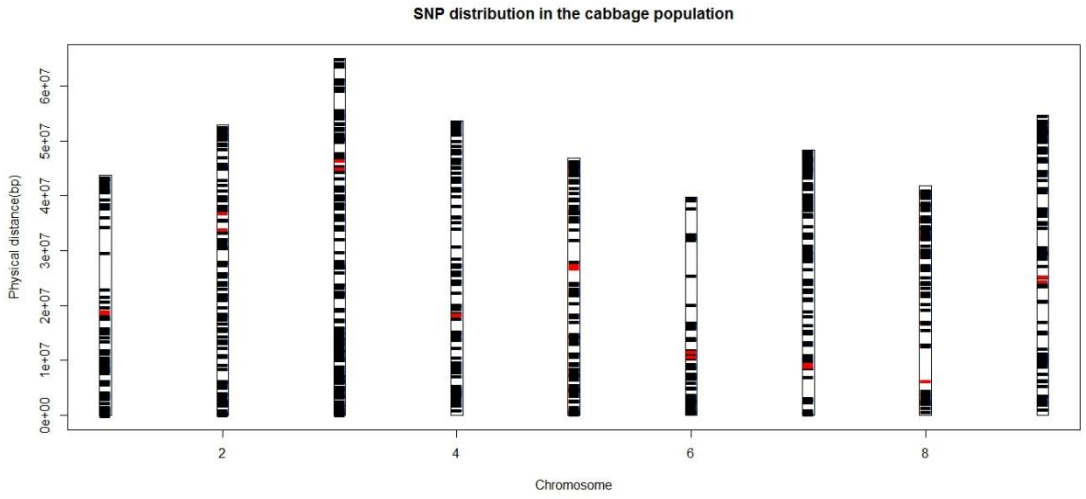


Figure 2.1. SNP distribution on chromosomes in the cabbage backcross population.

Boundaries of centromere are suggested by (PARKIN *et al.* 2014) and indicated by red bar.

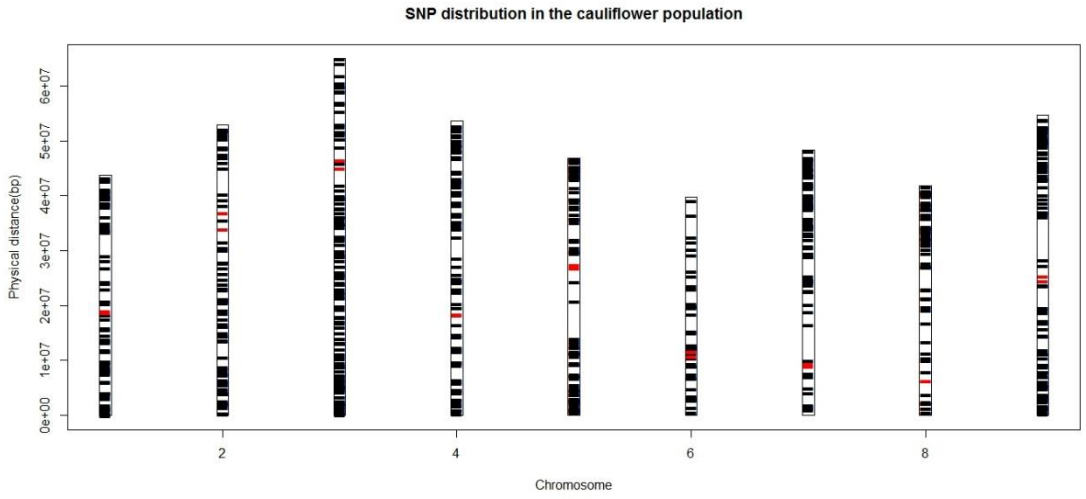


Figure 2.2. SNP distribution on chromosomes in the cauliflower backcross population.

Boundaries of centromere are suggested by (PARKIN *et al.* 2014) and indicated by red bar.

Table 2.3. Number and percentage of segments identified in two populations.

Number of segment	Cabbage	Cauliflower
Missing	7 (9.3%)	6 (6.7%)
0	4 (5.3%)	5 (5.6%)
1	11 (14.7%)	23 (25.8%)
2	16 (21.3%)	21 (23.6%)
3	13 (17.3%)	18 (20.2%)
4	13 (17.3%)	9 (10.1%)
5	8 (10.7%)	4 (4.5%)
6	2 (2.7%)	2 (2.3%)
7	1 (1.3%)	1 (1.1%)

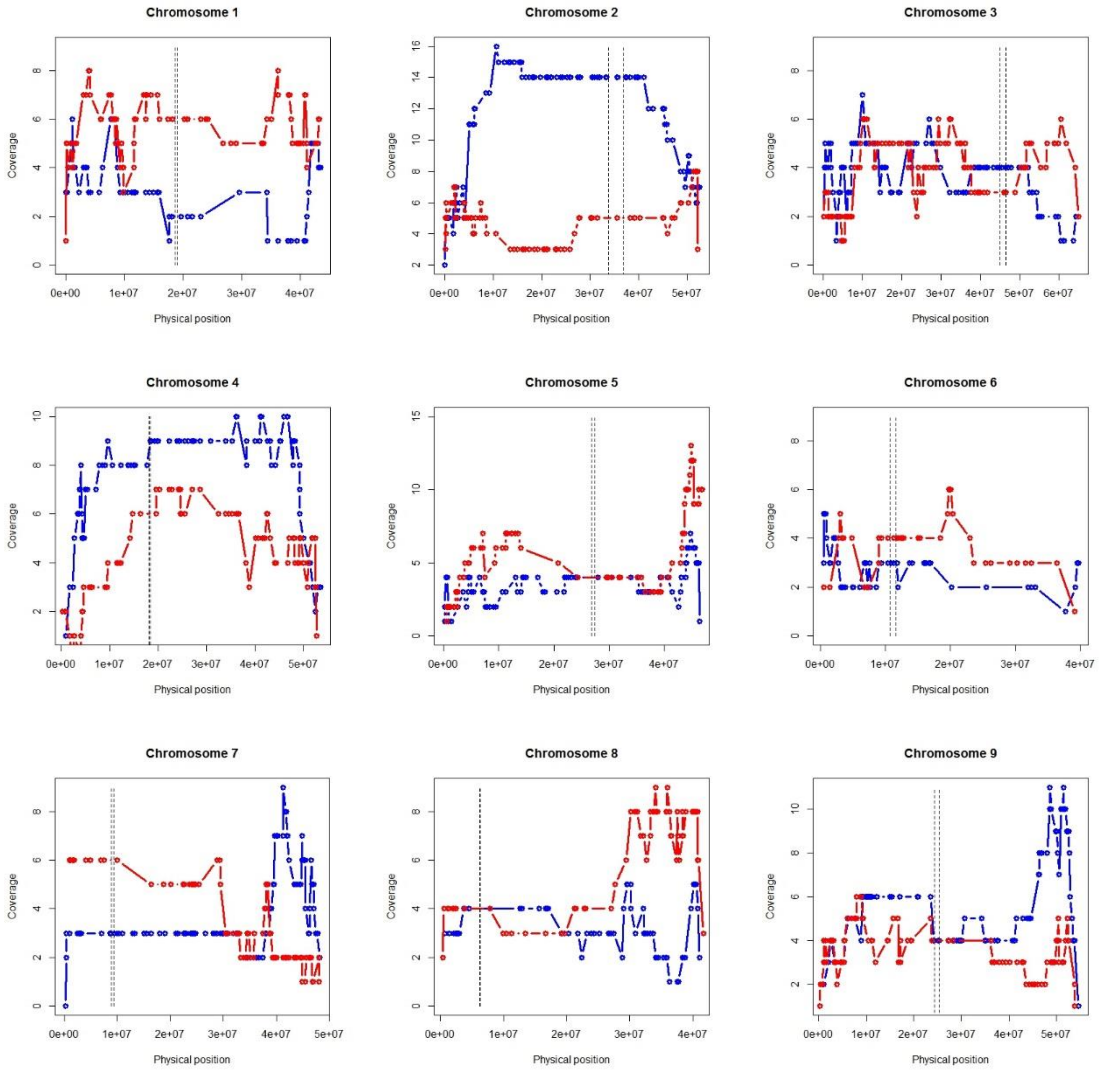


Figure 2.3. Distribution of introgressions on chromosomes in the cabbage population. X-axis represents physical distance and y-axis represents number of coverage. Boundaries of centromere are suggested by (PARKIN *et al.* 2014) and indicated by straight lines.

Table 2.4. Regions not covered by introgression.

	Chromosome	Start	End
Cabbage	7	265,490	441,607
Cauliflower	4	1,799,207	2,720,242
	4	2,815,887	4,149,259

Table 2.5. Regions with ‘excessive’ introgressions as defined by chi-squared statistics at significance threshold of 5%.

	chromosome	Start	End
Cabbage	2	5,048,939	47,045,929
	4	9,695,149	48,349,695
	7	41,414,251	41,888,667
	9	48,517,732	50,071,532
	9	50,929,968	52,518,556
Cauliflower	5	43,874,291	45,287,319
	5	46,268,854	46,834,403

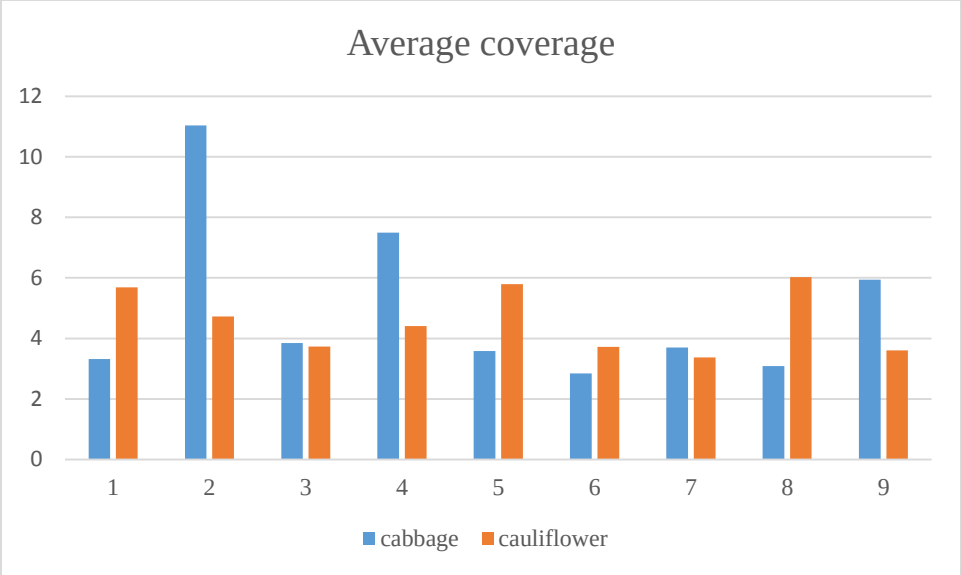


Figure 2.4. Average coverage of each chromosome by introgression in two populations. Blue represents the cabbage population and orange represents the cauliflower population. X-axis represents chromosome and y-axis represents depth of coverage.

CHAPTER 3

INITIAL QTL ANALYSIS OF MORPHOLOGICAL TRAITS

Lan-Shuan Shuang, Hugo Cuevas, Cornelia Lemke, Changsoo Kim, Hui Guo and Andrew H. Paterson. To be submitted to Theoretical and Applied Genetics.

Abstract

The single species *Brassica oleracea* encompasses a remarkable diversity of morphotypes, including cauliflower, broccoli, kohlrabi, marrowstem kale, cabbage, and Brussels sprouts as well as rapid-flowering morphologically simple genotypes reminiscent of *Arabidopsis thaliana*. To reveal the genetic control of the phenotypic variation among morphotypes, two advanced backcross populations, with inbred lines of cabbage (Badger inbred) and cauliflower (Orange) as the donor parents and a rapid cycling line (TO1434) as the recurrent parent, were planted in the field for two seasons. Phenotypic variation was observed between the two seasons that indicated the environmental sensitivity of the traits. Flower color and 14 leaf-, stem-, and flower-traits were segregating within the two populations, based on which we found 219 marker-trait associations. More replications of genotypes in multiple environments will enable us to reach higher mapping power and finer resolution in the future.

Introduction

QTL mapping can build the foundation for studying genetic control of natural variation (ALONSO-BLANCO *et al.* 2009). The diverse morphology within the single species *B. oleracea* has long been of interest, and genetic mapping in regard to its morphological traits has been studied since 1992. Traits like annual habit, glossy foliage, and leaf morphology were studied as simple Mendelian traits (KIANIAN and QUIROS 1992; LANDRY *et al.* 1992). A more detailed morphological trait analysis, describing leaf-,stem- and flower-morphology by 22 traits, was done in an F_2 population (KENNARD *et al.* 1994). To investigate intra-species variation, three F_2 populations were developed using one

common parent, from which 47 QTLs and 86 QTLs were mapped for plant size and inflorescence architecture, respectively, and putative Arabidopsis mutants for traits were inferred though comparative mapping (LAN and PATERSON 2000; LAN and PATERSON 2001). Two immortal doubled haploid (DH) populations were also used to studied developmental traits (SEBASTIAN *et al.* 2002; WALLEY *et al.* 2012). In addition, because of the economic importance of flowering time, its genetic control was extensively studied (KENNARD *et al.* 1994; CAMARGO and OSBORN 1996; BOHUON *et al.* 1998; RAE *et al.* 1999; LAN and PATERSON 2000; AXELSSON *et al.* 2001; OKAZAKI *et al.* 2007).

As an out-crossing crop, populations used for genetic mapping in *B. oleracea* were mostly F_2, F_3 and doubled haploid (DH). QTL mapping in these population types often resulted in large confidence intervals and lack sensitivity for identifying small-effect QTL, which tend to be masked by large-effect QTLs and genetic interactions from the background. To improve on these limitations, QTL mapping in near isogenic lines (NILs) was first proposed in tomato (ESHED and ZAMIR 1995). Near isogenic lines (NILs), also known as backcross inbred lines (BILs) and introgression lines (ILs), have single or small numbers of introgressed segments from a donor parent in a homogeneous genetic background of a recurrent parent. NILs are developed by repeated backcrossing and can be utilized in fine mapping, QTL mapping and breeding (PATERSON *et al.* 1990; LIPPMAN *et al.* 2007; KOOKE *et al.* 2012). Since there are few introgressed segments in each near isogenic line, QTLs on the segment(s) are rendered much more discrete than in F2 or backcross populations, often behaving as simple Mendelian factors (PARAN and ZAMIR 2003; KOOKE *et al.* 2012). QTL mapping based on NILs can thus increase accuracy of QTL position, and detect small effect QTLs that might otherwise be obscured by larger-

effect genes in more complex populations. In addition, because of the fixed genotype of NILs, they can be replicated in different environments to test interaction between genetic and environmental factors (MONFORTE and TANKSLEY 2000). By crossing NILs to the recurrent parent, fine mapping of specific QTLs toward their cloning is facilitated.

QTL mapping in NILs has been well demonstrated in tomato. Introgression lines were developed by dissecting a wild tomato genome, *Solanum pennellii*, in a domesticated tomato genetic background, *Solanum lycopersicum.* QTL mapping and gene cloning for numerous traits were performed. In addition, the introgressions which related to heterosis, were applied in tomato breeding (LIPPMAN *et al.* 2007). In *B. oleracea*, recombinant backcross substitution lines resembling NILs have been developed and employed for QTL mapping of flowering time (RAE *et al.* 1999). It was suggested that in this population more QTLs could be mapped than in a doubled haploid population based on the same advantages that accrue to NILs. Beyond the species mentioned above, QTL mapping in NILs was also applied in many other species, such as Arabidopsis (KEURENTJES *et al.* 2007), rice (MATSUBARA *et al.* 2008), rapeseed (BUTRUILLE *et al.* 1999), corn (SZALMA *et al.* 2007; LI *et al.* 2014a) and wheat (LIU *et al.* 2006). As an immortal population, the mapping power of NILs was compared with recombinant inbred lines (RILs). It was previously proposed that mapping power is higher in RILs than NILs (KAEPPLER 1997). QTL mapping in RILs could detect large-effect QTL and genetic interaction. On the other hand, mapping in NILs could detect small-effect QTLs that failed to be detected in RILs because NILs have a relatively homogeneous genetic background (KEURENTJES *et al.* 2007).

In this study, we developed two backcross populations. TO1434, a 'rapid cycling' genotype with short generation time and self-compatibility, was a common recurrent parent (WILLIAMS and HILL 1986). Cabbage ('Badger Inbred', var. *capitata*) and cauliflower ('ORG' inbred, var. *botrytis*), with orange curd, were donor parents. Genotypes of BC_4F_1 lines were revealed by genotyping-by-sequencing (GBS). Their selfed-progenies, BC_4F_2 and BC_4F_3 families, were planted in the field in spring and fall, respectively. The phenotypes of the backcross families were dissected into leaf-, stem-, and inflorescence- related traits and their associated SSR markers were reported.

Materials and methods

Plant materials and population development

Two populations were constructed using a rapid cycling line (TO1434) as common recurrent parent and cabbage [*B. oleracea* ssp. *capitata* L., 'Badger Inbred' (BIL)], and cauliflower [*B. oleracea* ssp. *botrytis* L., mutant for *Orange* gene (Lu et al., 2006)] as donor parents. TO1434 was pollinated by BIL and ORG to produce F_1 plants, so that the progenies would have TO1434 cytoplasm. Then, TO1434 was used as pollen parent for backcrosses. During population construction, plants were grown in a growth chamber with a 15.5 hour photoperiod and 23℃/20℃ day/night temperature. After eight weeks from sowing, plants that have not formed flower buds were vernalized for 70 days with an eight hour photoperiod and constant 4 ℃ temperature.

DNA extraction

DNA of each individual was extracted from fresh leaves as reported in PATERSON *et al.* (1993), with extraction buffer replaced by 0.35M sorbitol, 0.1M Tris, 0.005M EDTA, 0.04 M Sodium bisulfite and lysis buffer replaced by 0.2M Tris at pH8.0, 0.05M EDTA, 2M NaCl, 0.05M CTAB.

SSR markers and PCR genotyping

A total of 479 SSR markers were ordered from published literature (LOWE *et al.* 2004; PIQUEMAL *et al.* 2005; BURGESS *et al.* 2006; INIGUEZ-LUY *et al.* 2008; WANG *et al.* 2012; IZZAH *et al.* 2014; SHI *et al.* 2014). Since some markers' name from SHI *et al.* (2014) conflict with those of WANG *et al.* (2012), markers ordered from the former have .j at the end. Each marker was BLASTed to the published genome (PARKIN *et al.* 2014) to reveal its location. The polymerase chain reaction (PCR) was performed using 30 ng of DNA as template, 1 U Taq polymerase, 1 µL of 10X PCR buffer (100 mM Tris-HCl at pH 9, 500 mM KCl, and 15 mM MgCl2), 1 µL of 2 mM dNTP, 1 µL of 25 mM MgCl2, 0.5 µL of 10 µM of each primer, with final reaction volume of 10 µL. PCR reactions were denatured at 95℃ for 3 minutes, followed by 11 cycles of 95℃ for 30 seconds, 55-65℃ for 1 minute with 1℃ increase at each cycle, and 72℃ for 1 minute; then another 33 cycles of 95℃ for 30 seconds, 55℃ for 1 minute and 72℃ for 1 minute. The final cycle at 72℃ was for 5 minutes, then the sample was held at 4℃. Amplified fragments were analyzed in 10% polyacrylamide gels with silver staining.

Phenotype evaluation

BC_4F_2 and BC_4F_3 families were planted in the field in 2014 spring and fall, respectively. Seedlings were transplanted to the field when they were 3 to 4 weeks old. In

spring, each BC_4F_2 family was represented by 5 individuals. In fall, each family was represented by 2 to 5 individuals. Phenotypes which are listed below were recorded on the day that the first flower opened.

1. Lamina length: length of lamina of the largest leaf.
2. Lamina width: width of lamina of the largest leaf.
3. Blade shape: ratio between lamina width and lamina length.
4. Petiole length: length of petiole of the largest leaf.
5. Node number: number of nodes along main stem.
6. Plant height: length from ground to apex of plant
7. Internode distance: divide plant height by node number.
8. Stem width: widest width of stem
9. Bud number: number of buds on the first cluster.
10. Cluster width: widest width of the first cluster.
11. Curd with: widest width of the main curd.
12. Flower color: color of petal.
13. Budding time: days from transplanting to budding.
14. Flowering time: days from transplanting to flowering.

Statistical analysis

Pearson's correlation coefficients were calculated. Associations between markers and traits were tested by two comparisons, between introgressed individuals and (a) the recurrent parent and (b) non-introgressed individuals. Marker genotype was taken as the independent variable and phenotypic value was the dependent variable. Each

combination of marker and phenotype was examined by analysis of variance (ANOVA) with two assumptions, which were (1) whether the residuals of the model were normally distributed, which was tested by the Shapiro-Wilk test, and (2) variance within each group was homogenous, which was tested by Bartlett's test. If a marker-trait combination met the assumptions, then it would be analyzed by ANOVA. For phenotypes that were not normally distributed, Box-Cox transformation was applied. Phenotypes that failed to meet normality test but had homogenous variance among groups, were analyzed by Kruskal-Wallis rank sum test for markers with three genotypic classes and two-sample Wilcoxon test for markers with two genotypic classes. When the combination of marker and phenotype failed to meet both assumptions, two-sample t-test was applied when the phenotype was normally distributed. The significance threshold was set as p-value of 0.001, to mitigate the multiple-comparison problem that is fundamental to genome-wide QTL mapping studies. R^2 of multilocus models were reported by taking all the significant markers as independent variables and phenotypic value as dependent variable. If several markers on the same introgressed chromosome segment showed significant association with phenotype, the most significant one was reported. Additive effects were estimated by half of the difference of phenotypic values between the homozygous donor and recurrent parent alleles. Dominance effects were estimated by the difference of phenotypic values between heterozygotes and the average of the two homozygote genotypes. To reach a finer resolution of markers detected on overlapped introgressions, individuals with the same introgressions were pooled together and compared to the recurrent parent. Phenotypic values of the introgressed families were reported as the mean of the phenotypes within each family.

Results

Genotypes

This study included two seasons of field data. BC_4F_2 populations were planted in 2014 spring, including 32 cabbage and 29 cauliflower BC_4F_2 families. We were able to genotype 89% (80/90) and 88% (77/88) of introgressed segments within cabbage and cauliflower BC_4F_2 families each by 80 and 96 SSR markers. In fall, three and 29 BC_4F_3 families, which derived from one and 15 cabbage and cauliflower BC_4F_1 families, respectively, were planted and genotyped by 65 SSR markers.

Phenotypes

We observed phenotypic variation for almost all traits, except plant height, between the two seasons (Figure 3.1). The BC_4F_3 families that were planted in fall reached larger plant size, which indicated larger leaves, thicker stems, more nodes and buds, shorter internode distance, later budding and flowering. In addition, plants formed more prominent curds in fall than spring.

2014 spring

The phenotypic distributions of cabbage and cauliflower BC_4F_2 population are shown in Figure 3.1, Figure 3.2 and Table 3.1. The two populations shared similar distributions of phenotypes. The cabbage BC4F2 families had higher frequencies of plants with larger lamina length and width, leaf length, more nodes and buds and later budding and flowering. Most of the distributions were asymmetric except plant height in the cauliflower population. Distributions of budding time and flowering time were

bimodal while others were unimodal. Modes of each phenotype were close to the mean of the recurrent parent, which was consistent with the similarities between recurrent parent and backcross progenies.

Correlation coefficients between traits were shown in Figure 3.4 and Figure 3.5 for the cabbage and cauliflower populations respectively. Higher correlations were shown among the traits related to plant size and flowering time.

2014 fall

Phenotypic distributions for cabbage and cauliflower BC_4F_3 population are shown in Figure 3.3 and Table 3.1. Distributions of the two populations overlapped. The cabbage BC_4F_3 populations had higher frequencies of plants with longer leaf length and thicker stem. The means of recurrent parent values were close to the population mean for all traits.

Correlations for all traits are listed in Figure 3.6. Consistent with spring phenotypes, leaf traits were highly correlated with each other and stem width. Budding time was correlated with flowering time.

QTL analysis

Flower color

Flower color was observed in two populations and segregation was detected in the BC_4F_3 generation. Four BC_4F_3 families derived from a single BC_4F_1 family were planted, each with three BC_4F_3 individuals (Figure 3.7). Flower color segregated within this BC_4F_1 family, so that we were able to map the locus in the physical region between

nucleotides 55,802,077 (BoESSR073) and 60,314,017 (BoSF2423) The putative region include *CAROTENOID CLEAVAGE DIOXYGENASE 4*, located at 566015961-56607751 which was recently identified as the gene controlling flower color in *B. oleracea* and its amphidiploid relative *B. napus* and *B. carinata* (Zhang et al. 2015).

Lamina length

Fourteen markers were significantly associated with lamina length in BC_4F_2 populations in spring (Table 3.2). In the cabbage population, nine loci were found on chromosomes 1, 2, 3, and 4, collectively explaining 60.73% of phenotypic variance. Among six markers that were detected on chromosome 2, families with introgression at the interval between BoSF1304.j and BoSF239 had significantly larger lamina length than the recurrent parent (Figure 3.9). In the cauliflower population, five loci were discovered on chromosomes 5 and 9, explaining 44.7% of phenotypic variance. Even though detected markers on chromosome 9 were on different introgressed segments, no finer resolution could be obtained since there is no replicate genotypes within other introgressed families.

Lamina width

Seventeen loci were significantly associated with lamina width in BC_4F_2 populations in spring (Table 3.3). Twelve loci were detected on chromosomes 1, 2, 3, 4, 5, 8 and 9 in the cabbage population, explaining 70.7% of phenotypic variation. Four associated markers were on chromosome 2. Families with introgression at the interval between BoSF1304.j and BoSF239 showed significantly greater lamina width than the recurrent parent (Figure 3.9). Furthermore, BoSF2425, Na10-A08, FITO546 and

BoSF2564 detected introgressions in single BC_4F_2 families, which also possess introgression at other associated markers. As a result, there might be false-positive detection among these markers. In the cauliflower population, five loci were detected on chromosomes 5 and 9, explaining 59.8% of phenotypic variance.

Leaf length

Nineteen loci were found across two seasons (Table 3.4). In the cabbage BC_4F_2 population, 11 loci were detected on chromosomes 1, 2, 3, 4, 5, and 9, explaining 65.8% of phenotypic variance. Among the detected markers on chromosome 2, families with introgression at the interval between BoSF1304.j and BoSF239 showed significantly longer leaf length than the recurrent parent (Figure 3.9). Among the cauliflower BC_4F_2 families, seven loci were found on chromosomes 3, 5 and 9, explaining 50.4% of phenotypic variance. In the BC_4F_3 population, an associated marker was on chromosome 3, explaining 0.98% of phenotypic variance.

Blade shape

Six loci were detected for blade shape in the cauliflower BC_4F_2 and BC_4F_3 populations (Table 3.5). In the BC_4F_3 population, three associated loci were on chromosomes 2 and 3 when compared with non-introgressed individuals, explaining 8.6% of phenotypic variance. On chromosome 3, two associated markers were found. Introgressed families at Ol11-B05 showed significant phenotypic difference from the recurrent parent, which suggested closer linkage between the QTL and the marker (Figure 3.17). In the BC_4F_2 population, three loci were detected on chromosome 9 explaining 22.3% of phenotypic variance. However, only one introgressed family showed

greater blade shape than the recurrent parent, which might suggest linkage between the putative QTL and CB10103 (Figure 3.16).

Petiole length

Twenty-one loci were associated significantly with petiole length (Table 3.6). In the cabbage BC_4F_2 population, nine loci were found on chromosomes 1, 2, 3, 4, and 9, explaining 48.5% of phenotypic variance. On chromosome 2, introgressed families at the interval between BoSF1304.j and BoSF239 showed significantly longer petiole length than the recurrent parent (Figure 3.9). Two markers were detected on chromosome 3 and two BC_4F_2 families with introgressions at BoESSR492 both showed phenotypic differences from the recurrent parent, which suggested closer linkage to this marker (Figure 3.10). In addition, FITO523 and BoSF2423 detect introgression in the same BC_4F_2 family, which might lead to false-positive detection by these markers. In the cauliflower BC_4F_2 population, 11 loci were observed on chromosomes 1, 4, 5, 8 and 9, explaining 46.1% of phenotypic variance. Families with introgression at BoSF2436.j also had introgression at three markers on chromosome 9, BoSF2564, Ol10-D08 and BoSF2389.j, which might lead to false-positive detection. Even though four associated markers were on chromosomes 4 and 5, no finer resolution could be provided since families with introgression did not necessarily show phenotypic differences (Figure 3.14 and Figure 3.15). In addition, in the cauliflower BC_4F_3 population, a locus on chromosome 5 showed association and explained 15.9% of phenotypic variance.

Node number

A total of 23 loci were found to be associated significantly with node number across two seasons (Table 3.7). In the cabbage BC_4F_2 population, 15 loci on chromosomes 1, 2, 3, 4, 5, 7, 8 and 9 were found when compared with the recurrent parent, explaining 74.6% of phenotypic variance. Four associated markers were on chromosome 2. By comparing introgressed families with the recurrent parent, two putative QTL regions were suggested, with one link to BoSF1304.j and another in the interval between BoSF2532 and BoSF0439.j (Figure 3.9). While on chromosome 9, the putative interval might fall between PBCGSSRBo34 and BoSF258 (Figure 3.13). Three markers, Na10-A08, FITO546 and BoSF2564 observed introgression in the same family which might lead to false-positive detection. In the cauliflower BC_4F_2 population, seven loci were also detected on chromosomes 1 and 9, explaining 63.7% of phenotypic variance. Introgressed families at BoSF2436.j, on chromosome 1, also had introgression on chromosome 9 which might lead to false-positive detection. On chromosome 9, six markers were detected. In comparison with the recurrent parent, families with introgression at the interval between Ol10-D08 and CB10103 had significantly higher node number (Figure 3.16). In the cauliflower BC_4F_3 population, a locus was found on chromosome 1, explaining 2.6% of phenotypic variance.

Plant height

Two loci were found to be significantly associated with plant height (Table 3.8). Respectively, a locus on chromosome 3 was identified among BC_4F_3 families, explaining 17% of phenotypic variance, and a locus on chromosome 2 was detected in the cabbage BC_4F_2 population by comparing with non-introgressed individuals, explaining 8.4% of phenotypic variance

Internode distance

Ten loci were found to be associated with internode distance across two seasons (Table 3.9). In the cabbage BC_4F_2 population, four significant loci were on chromosomes 2 and 7, explaining 44.8% of phenotypic variance. Among the associated markers on chromosome 2, families with introgression at the interval between BoSF2294 and BoSF1304.j had shorter internode distance than the recurrent parent which suggests this might be the putative QTL region (Figure 3.9). In the cauliflower BC_4F_2 population, four associated loci were on chromosome 9, explaining 31.1% of phenotypic variance. Among the associated markers on chromosome 9, only one introgressed family showed significance when compared with the recurrent parent which suggests linkage between the causal QTL and CB10103 (Figure 3.16). In the BC_4F_3 population, two loci were on chromosomes 3 and 9, explaining 12.5% of phenotypic variance. Because two markers on different chromosomes identified introgression in the same family, one of which might be false-positive.

Stem width

Thirty-one loci were found for stem width in BC_4F_2 populations in spring (Table 3.10). In the cabbage population, 21 loci on chromosomes 1, 2, 3, 4, 5, 7, 8, and 9 were significant and could explain 85.4% of phenotypic variance. Two markers, Na10-A08 and FITO546 detected introgression in the same family, which might lead to false-positive detection. Multiple associated markers were on several chromosomes. Eight associated markers were on chromosome 2, with families with introgression at the interval between BoSF1304.j and BoSF239 having significantly thicker stem width than

the recurrent parent which suggest this interval might be the putative QTL region (Figure 3.9). Two markers were detected on chromosome 4, with most of the families with introgression at BrSF0537 showing significance in comparison to the recurrent parent except family # 95 (Figure 3.11). More evidence is needed to discern the putative QTL interval. Two markers on chromosome 5 were on independent introgressions. Three associated markers were on chromosome 9, with families with introgression at the interval between PBCGSSRBo34 and BoSF258 showing significantly thicker stem width than the recurrent parent, suggesting the location of putative QTLs (Figure 3.13). In the cauliflower population, 10 loci on chromosomes 1, 3, 5 and 9 were found to be associated with stem width, explaining 71.8% of phenotypic variance. Two markers were detected on chromosome 5, introgressed families at BoSF2878 showing significant difference from the recurrent parent (Figure 3.15). Among the associated markers on chromosome 9, only introgressed families at BoSF0211.j showed significance in comparison to the recurrent parent (Figure 3.16). Introgressed families at other markers could not provide information useful to this analysis. A common locus on chromosome 5 was detected in two BC_4F_2 populations

Bud number

Thirteen loci were found to be associated with bud number across two seasons (Table 3.11). In the cabbage BC_4F_2 population, five significant loci were on chromosomes 1, 2, 4, and 9, explaining 28.3% of phenotypic variance. Two markers detected on chromosome 2 were on independent introgressed segments on each end of the chromosome. In the cauliflower BC_4F_2 population, four loci were detected, explaining 28.3% of phenotypic variance. Families carrying introgression in the interval between

Ol10-D08 and BoSF2389.j had more buds than the recurrent parent. In the BC_4F_3 population, four loci were detected on chromosomes 2 and 4 explaining 24.9% of phenotypic variance. Two BC_4F_3 families showed introgressions at detected markers on chromosome 4 (Figure 3.18). BC_4F_3 family #57-10 had introgression at BoESSR333, while #9-12 had introgression at three markers. Since #57-10 did not show significance when compared to the recurrent parent, we were able to exclude BoESSR333 from the putative regions. Higher coverages by introgressions would be needed to identify the causal QTL on chromosome 4. In addition, FITO377 also detected introgressions in family #9-12 which might lead to false-positive detection by either of these markers.

Cluster width

Twenty-six loci were associated with cluster width in the BC_4F_2 populations in spring (Table 3.12). Sixteen loci on chromosomes 1, 2, 4, 5, 6, 7, 8 and 9 were found in the cabbage BC_4F_2 population and could explain 43.7% of phenotypic variance. Again, multiple markers were detected on several chromosomes. On chromosome 2, four markers were detected. Comparisons between introgressed families and the recurrent parent suggested two putative regions on chromosome 2, one between BoSF2294 and BoSF1304.j, with closer linkage to BoSF2294, and another on the other end of the chromosome, in the interval between BoSF2532 and BoSF0439.j (Figure 3.9). On chromosome 5, comparison between introgressed families and the recurrent parent suggested that causal QTL might share closer linkage with BoESSR945 (Figure 3.12). Two associated markers on chromosome 7 were on independent introgressions. On chromosome 9, introgressed families in the interval between PBCGSSRBo34 and BoSF258 had larger cluster width than the recurrent parent (Figure 3.13). Several

markers, including BoSF2425, Na10-A08, FITO546 and Ol10-D08, each detect introgression in different single families, while these families also showed introgression at other associated markers which might lead to false-positive detection.

In the cauliflower population, ten loci were identified on chromosomes 1, 5 and 9, explaining 44.7% of phenotypic variance. Three associated markers were on chromosome 5, with comparison between introgressed families and the recurrent parent suggesting that BoSF2878 and CB10027 would be the putative loci (Figure 3.15). Among six associated markers on chromosome 9, no evidence could be provided about the region between Ol10-D08 and CB10103, since only one introgressed family was comparable. On the other hand, comparisons suggested closer linkage between the causal QTL and BoSF0211.j (Figure 3.16). A locus on chromosome 9, Ol10-D08, was detected in both populations.

Curd width

Curd width was only scored in fall. Two loci were found on chromosomes 1 and 3, explaining 26.4% of phenotypic variance (Table 3.13). A locus on chromosome 1, which detected introgression in a single family, was found by comparing the introgressed individuals with the recurrent parent. A locus on chromosome 3 showed association when compared with non-introgressed BC_4F_3 individuals.

Budding time

Fifteen loci were found to be associated with budding time in the two BC_4F_2 populations (Table 3.14). Eight loci were found in the cabbage BC_4F_2 population that were on chromosomes 1, 2, 3, 4, 7 and 9, explaining 45.7% of phenotypic variance. In

comparison with the recurrent parent, introgressed families at the interval between BoSF1304.j and BoSF239 on chromosome 2 had significantly longer budding time (Figure 3.9). In the cauliflower population, seven loci on chromosomes 3 and 9 were detected, explaining 49.9% of phenotypic variance. Among markers detected on chromosome 9, introgressed families at the interval between BoSF2364.j to CB10103 had significantly longer budding time than the recurrent parent (Figure 3.16).

Flowering time

Nineteen loci were found across two seasons (Table 3.15). Twelve loci were found in the cabbage BC_4F_2 population on chromosomes 1, 2, 4, 8 and 9, explaining 55.2% of phenotypic variance. In comparison with the recurrent parent, two putative regions were on chromosome 2, BoSF1304.j to BoSF239 and BoSF2532 to BoSF0439.j, while on chromosome 9, the putative interval is between PBCGSSRBo34 and BoSF258 (Figure 3.9 and Figure 3.13). Four loci were found in the cauliflower BC_4F_2 population on chromosomes 5 and 9, explaining 33.6% of phenotypic variance. No finer resolution on chromosome 9 could be provided yet, until more recombinants are obtained. In fall, three loci were detected, two of which were on chromosomes 2 and 9 in the cabbage BC_4F_3 population, explain 5.4% of phenotypic variance. One marker on chromosome 9 was found in the cauliflower BC_4F_3 population, explaining 36.8% of phenotypic variance.

Discussion

To dissect the morphological variation in *B. oleracea*, cabbage and cauliflower – derived backcross populations in a rapid cycling background were evaluated in the field for two seasons. Phenotypic variations were observed between and within the two

seasons and two populations (Table 3.1). Based on the phenotypic variation between introgressed individuals and (a) the recurrent parent and/or (b) non-introgressed individuals, we were able to map genetic loci for one qualitative trait and 14 quantitative traits. From the two populations, finer resolution of the genomic regions could be obtained for some QTLs than by using conventional QTL mapping, based on the overlapping regions from various introgressed chromosome segments.

Flowering time was segregating between and within two seasons and might influence development of plant size. Striking phenotypic variation was observed between two seasons with individuals planted in fall reaching larger plant size and flowering later. Within the BC_4F_2 populations planted in spring, QTL regions detected for flowering time on chromosome 2 in the cabbage population and chromosome 9 in both populations showed significant associations with many other traits. BC_4F_2 families with introgressions on these chromosomal regions showed great phenotypic variation from the recurrent parent. In addition, flowering time was highly correlated with traits related to plant size in spring but the correlations became not significant in fall. This might indicate that longer time in fall vegetative development had permitted plants to fully develop and reach larger plant size. A similar result was also suggested by RAE *et al.* (1999), in which positive correlations of plant height, node number and flowering were observed. These suggest the environmental sensitivity of these genotypes, especially in flowering time.

Molecular mechanisms for flowering time have not been discovered in *B. oleracea*, but are well studied in its close relative, Arabidopsis. Four pathways controlling flowering time were characterized in Arabidopsis, including vernalization, photoperiod, autonomous and gibberellic acid (GA) with the former two related to environmental

signals (BENTLEY *et al.* 2013). Temperature is another environmental factor that affects flowering time with lower temperature delaying flowering time in Arabidopsis (GREENUP *et al.* 2009). While the recurrent parent was a rapid cycling genotype, donor parents required vernalization to flower. However, plants in the field formed buds in two seasons. As a result, vernalization requirement was not observed in field evaluations. The variation in flowering time between two seasons might be due to differences in photoperiod and temperature. *B. oleracea* includes both photoperiod-sensitive genotypes stimulated to flower by long days, and day-neutral genotypes (GÓMEZ-CAMPO 1999) and higher temperature will accelerate flowering (UPTMOOR *et al.* 2012).

We detected 19 DNA markers associated with flowering time, 15 for budding time and eight for both traits. From previous studies, it was suggested that *CONSTANS* (*CO*) and *FLOWERING LOCUS C* (*FLC*) could be candidate genes for flowering time by localization of QTLs to regions homologous to their locations on Arabidopsis chromosomes (BOHUON *et al.* 1998; AXELSSON *et al.* 2001; OKAZAKI *et al.* 2007; RAZI *et al.* 2008). In Arabidopsis, *CO* is involved in the photoperiod response pathway and induces flowering under long days, while expression of *FLC* would repress flowering until the plant is vernalized (GREENUP *et al.* 2009). Among our significant associated regions, chromosomes 2 and 9 included the homologs of *CO* and *FLC*. Additionally, there were BC_4F_2 introgressed individuals in the cabbage population showing environment-insensitive phenotypes, which flowered at approximately the same number of days after planting in both spring and fall. Based on these individuals, we might be able to infer that the markers associated with flowering time in the cabbage BC_4F_2 population were related to environmental signal. However, since genetic control of

flowering in *B. oleracea* remains unknown, higher resolution of genetic mapping will be required before inferring the genes responsible for flowering.

It was previously known that flower color is under single gene control, with white being dominant to yellow, and the locus was mapped on O3 (RAMSAY *et al.* 1996; HOWELL *et al.* 2002; FARINHÓ *et al.* 2004; SHARMA *et al.* 2012). Here, flower color was segregating in two backcross populations, since the recurrent parent had white flowers while donor parents had yellow flowers. Backcross progenies with yellow flowers were first observed in the BC_4F_2 generation, which confirmed the dominance of white flowers. However, only two plants, each in a BC_4F_2 population, were observed in the field, which failed to provide sufficient mapping information. We were able to map the locus in an interval between 55,802,077 and 60,314,017 on chromosome 3 in the BC_4F_3 generation, which includes the *CAROTENOID CLEAVAGE DIOXYGENASE 4* (*CCD4*) gene, located at 566015961-56607751 (PARKIN *et al.* 2014). *CCD4*, which catalyzes oxidative cleavage of carotenoids, was identified as the gene responsible for flower color in *B. napus* and the result was validated in *B. carinata* (BBCC) and *B. oleracea*. The study indicated that yellow flowers were controlled by loss-of-function alleles, with carotenoids accumulating in the petal (ZHANG *et al.* 2015).

Co-localization of QTLs detected for different traits in similar genomic regions is often inferred to represent either close linkage of multiple genes or pleiotropic effects of single genes. It was previously reported that QTLs detected for multiple leaf-related traits were located in a proximal region and thus suggested that the QTLs might actually be responsible for leaf size instead of independent traits (KENNARD *et al.* 1994; LAN and PATERSON 2001; SEBASTIAN *et al.* 2002). In this study, 12 markers were detected for

lamina width along with either lamina length or petiole length, each of which might be involved in controlling leaf size. In addition to leaf traits, co-localization of QTLs for stem width and node number was also discovered (LAN and PATERSON 2001). In this study, 19 markers were detected for both stem width and node number which might be involved in stem development.

Formation of curd has been described as involving curtailed leaf development, lateral buds elongated into shoots which make up the curd surface, and shortened internodes (SADIK 1962). In this study, curd was defined as clusters of flower buds at the same surface. The curd phenotype was not prominent in BC_4F_2 (spring) populations compared to BC_4F_3 (fall) populations, which might be due to early spring flowering and shorter budding to flowering interval. In addition, it was suggested by SMITH and KING (2000), that non-curding brassicas have alleles that could complement the curding phenotypes by observing F_1 and F_2 progenies derived from the cross between cauliflower and Chinese white kale, with wild type inflorescence. The progenies showed no curding phenotypes and only one intermediate type. To describe the flower architecture of two BC_4F_2 populations in spring, cluster width instead of curd width was recorded. In addition, to better characterize cluster size, we also determined bud number of the first formed cluster. The correlation between bud number and cluster width was 0.74 in spring and 0.37 in fall. Furthermore, only six markers were detected for both traits. This suggested that in addition to cluster size, density of cluster or curd should be considered to describe curd mass, as indicated in LAN and PATERSON (2000)

For each marker-trait association analysis, in addition to comparisons between introgressed individuals and the recurrent parent, we also compared introgressed and non-

introgressed individuals. More significant marker-trait associations were observed in comparisons between the recurrent parent and introgressed individuals in BC_4F_2 population than in BC_4F_3 population, since the mean of the recurrent parent was lower than those of BC_4F_2 populations and closer to those of the BC_4F_3 populations. Through comparisons between introgressed and non-introgressed individuals, we were able to identify the introgressions that were responsible for larger phenotypic differences from the recurrent parent. Additionally, with the presence of multiple introgressed segments, the comparisons between introgressed individuals and the recurrent parents might be more prone to false positive results.

Phenotypic differences between backcross progeny and the recurrent parent are attributed to introgressions from donor parent to progeny. Families with introgressions at detected markers were compared to the recurrent parent. If introgressed families at a detected marker showed a significant difference from the recurrent parent, we could infer close linkage with causal QTLs. However, chance co-segregation of unlinked introgressions within a family might lead to false-positive detection by some markers. False-positive association may be reduced somewhat by segregation in multiple lines within a population. Lacking replication of genotypes, phenotypic variation also currently decreases mapping power (KEURENTJES *et al.* 2007). More evidence, such as more replicated selfed-progeny with fewer introgressions, will provide important validation of causal QTLs.

From BC_4F_2 and BC_4F_3 families developed from crosses of cabbage and cauliflower and a common rapid cycling parent, we were able to dissect morphological variation within the species and discover associated SSR markers. Phenotypic variation

between two seasons revealed their sensitivity to environment. With these populations and the aid of genome sequence, we hope to discover the genetic control of morphological traits. However, to reach higher mapping power and finer resolution, more introgressed families of a more advanced generation and more replicates will be required.

Reference

ALONSO-BLANCO, C., M. G. AARTS, L. BENTSINK, J. J. KEURENTJES, M. REYMOND *et al.*, 2009 What has natural variation taught us about plant development, physiology, and adaptation? The Plant Cell Online **21:** 1877-1896.

ANDOLFATTO, P., D. DAVISON, D. EREZYILMAZ, T. T. HU, J. MAST *et al.*, 2011 Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome research **21:** 610-617.

ARBELAEZ, J. D., L. T. MORENO, N. SINGH, C.-W. TUNG, L. G. MARON *et al.*, 2015 Development and GBS-genotyping of introgression lines (ILs) using two wild species of rice, *O. meridionalis* and *O. rufipogon*, in a common recurrent parent, *O. sativa* cv. Curinga. Molecular Breeding **35:** 1-18.

AXELSSON, T., O. SHAVORSKAYA and U. LAGERCRANTZ, 2001 Multiple flowering time QTLs within several *Brassica* species could be the result of duplicated copies of one ancestral gene. Genome **44:** 856-864.

BABULA, D., M. KACZMAREK, A. BARAKAT, M. DELSENY, C. F. QUIROS *et al.*, 2003 Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. Mol Genet Genomics **268:** 656-665.

BABULA, D., M. KACZMAREK, P. A. ZIÓŁKOWSKI and J. SADOWSKI, 2007 Brassica oleracea, pp. 227-285 in *Vegetables*. Springer.

BAIRD, N. A., P. D. ETTER, T. S. ATWOOD, M. C. CURREY, A. L. SHIVER *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one **3:** e3376.

BANDILLO, N., C. RAGHAVAN, P. A. MUYCO, M. A. L. SEVILLA, I. T. LOBINA *et al.*, 2013 Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice **6:** 1-15.

BARKER, G. C., T. R. LARSON, I. A. GRAHAM, J. R. LYNN and G. J. KING, 2007 Novel insights into seed fatty acid synthesis and modification pathways from genetic diversity and quantitative trait loci analysis of the *Brassica* C genome. Plant physiology **144:** 1827-1842.

BEECHER, C. W., 1994 Cancer preventive properties of varieties of *Brassica oleracea*: a review. The American journal of clinical nutrition **59:** 1166S-1170S.

BENTLEY, A., E. F. JENSEN, I. MACKAY, H. HÖNICKA, M. FLADUNG *et al.*, 2013 Flowering time, pp. 1-66 in *Genomics and breeding for climate-resilient crops*. Springer.

BETTEY, M., W. FINCH- SAVAGE, G. J. KING and J. LYNN, 2000 Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in *Brassica oleracea*. New Phytologist **148:** 277-286.

BOHUON, E. J. R., D. J. KEITH, I. A. P. PARKIN, A. G. SHARPE and D. J. LYDIATE, 1996 Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor Appl Genet **93:** 833-839.

BOHUON, E. J. R., L. D. RAMSAY, J. A. CRAFT, A. E. ARTHUR, D. F. MARSHALL *et al.*, 1998 The association of flowering time quantitative trait loci with duplicated fegions and candidate loci in *Brassica oleracea*. Genetics **150:** 393-401.

BOWERS, J. E., B. A. CHAPMAN, J. RONG and A. H. PATERSON, 2003 Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433-438.

BOYES, D. C., M. E. NASRALLAH, J. VREBALOV and J. B. NASRALLAH, 1997 The self-incompatibility (S) haplotypes of *Brassica* contain highly divergent and rearranged sequences of ancient origin. The Plant Cell Online **9:** 237-247.

BROADLEY, M. R., J. P. HAMMOND, G. J. KING, D. ASTLEY, H. C. BOWEN *et al.*, 2008 Shoot calcium and magnesium concentrations differ between subtaxa, are highly heritable, and associate with potentially pleiotropic loci in *Brassica oleracea*. Plant Physiology **146:** 1707-1720.

BROMAN, K. W., H. WU, Ś. SEN and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890.

BROWN, A. F., G. G. YOUSEF, K. K. CHEBROLU, R. W. BYRD, K. W. EVERHART *et al.*, 2014 High-density single nucleotide polymorphism (SNP) array mapping in *Brassica oleracea*: identification of QTL associated with carotenoid variation in broccoli florets. Theoretical and Applied Genetics **127:** 2051-2064.

BURGESS, B., H. MOUNTFORD, C. J. HOPKINS, C. LOVE, A. E. LING *et al.*, 2006 Identification and characterization of simple sequence repeat (SSR) markers derived in silico from *Brassica oleracea* genome shotgun sequences. Molecular Ecology Notes **6:** 1191-1194.

BUTRUILLE, D. V., R. P. GURIES and T. C. OSBORN, 1999 Linkage analysis of molecular markers and quantitative trait loci in populations of inbred backcross lines of *Brassica napus* L. Genetics **153:** 949-964.

CAMARGO, L., and T. OSBORN, 1996 Mapping loci controlling flowering time in *Brassica oleracea*. Theoretical and Applied Genetics **92:** 610-616.

CAMARGO, L., L. SAVIDES, G. JUNG, J. NIENHUIS and T. OSBORN, 1997 Location of the self-incompatibility locus in an RFLP and RAPD map of *Brassica oleracea*. Journal of Heredity **88:** 57-60.

CHETELAT, R., and V. MEGLIC, 2000 Molecular mapping of chromosome segments introgressed from *Solanum lycopersicoides* into cultivated tomato (*Lycopersicon esculentum*). Theoretical and Applied Genetics **100:** 232-241.

CHEUNG, W., G. CHAMPAGNE, N. HUBERT and B. LANDRY, 1997 Comparison of the genetic maps of *Brassica napus* and *Brassica oleracea*. Theoretical and Applied Genetics **94:** 569-582.

CHUTIMANITSAKUN, Y., R. W. NIPPER, A. CUESTA-MARCOS, L. CISTUÉ, A. COREY *et al.*, 2011 Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. BMC genomics **12:** 4.

COELHO, P. S., and A. A. MONTEIRO, 2003 Inheritance of downy mildew resistance in mature broccoli plants. Euphytica **131:** 65-69.

COGAN, N., J. LYNN, G. KING, M. KEARSEY, H. NEWBURY *et al.*, 2002 Identification of genetic factors controlling the efficiency of *Agrobacterium rhizogenes*-mediated transformation in *Brassica oleracea* by QTL analysis. Theoretical and Applied Genetics **105:** 568-576.

COGAN, N. O., H. J. NEWBURY, A. M. OLDACRES, J. R. LYNN, M. J. KEARSEY *et al.*, 2004 Identification and characterization of QTL controlling Agrobacterium-mediated transient and stable transformation of *Brassica oleracea*. Plant biotechnology journal **2:** 59-69.

COLLARD, B., M. JAHUFER, J. BROUWER and E. PANG, 2005 An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica **142:** 169-196.

CONSORTIUM, I. C. G. M., 2015 High-Resolution Linkage Map and Chromosome-Scale Genome Assembly for Cassava (*Manihot esculenta* Crantz) from 10 Populations. G3: Genes| Genomes| Genetics **5:** 133-144.

DAVEY, J. W., T. CEZARD, P. FUENTES-UTRILLA, C. ELAND, K. GHARBI *et al.*, 2013 Special features of RAD Sequencing data: implications for genotyping. Molecular Ecology **22:** 3151-3164.

DEOKAR, A. A., L. RAMSAY, A. G. SHARPE, M. DIAPARI, A. SINDHU *et al.*, 2014 Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. BMC genomics **15:** 708.

DESCHAMPS, S., V. LLACA and G. D. MAY, 2012 Genotyping-by-sequencing in plants. Biology **1:** 460-483.

DIXON, G. R., 2007 *Vegetable brassicas and related crucifers / Geoffrey R Dixon*. Wallingford, UK ; Cambridge, MA : CABI Pub., c2007.

DUCLOS, D. V., and T. BJÖRKMAN, 2008 Meristem identity gene expression during curd proliferation and flower initiation in *Brassica oleracea*. Journal of experimental botany **59:** 421-433.

EGEL, R., and D.-H. LANKENAU, 2007 *Recombination and meiosis: crossing-over and disjunction*. Springer Science & Business Media.

ELSHIRE, R. J., J. C. GLAUBITZ, Q. SUN, J. A. POLAND, K. KAWAMOTO *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one **6:** e19379.

ESHED, Y., and D. ZAMIR, 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.

FAOSTAT, F., 2014 Food and Agricultural Organization of the United Nations. 2014, pp.

FARINHÓ, M., P. COELHO, J. CARLIER, D. SVETLEVA, A. MONTEIRO *et al.*, 2004 Mapping of a locus for adult plant resistance to downy mildew in broccoli (*Brassica oleracea* convar. *italica*). Theoretical and applied genetics **109:** 1392-1398.

FARNHAM, M., M. WANG and C. THOMAS, 2002 A single dominant gene for downy mildew resistance in broccoli. Euphytica **128:** 405-407.

FIGDORE, S., M. FERREIRA, M. SLOCUM and P. WILLIAMS, 1993 Association of RFLP markers with trait loci affecting clubroot resistance and morphological characters in *Brassica oleracea* L. Euphytica **69:** 33-44.

FLETCHER, R. S., J. L. MULLEN, S. YODER, W. L. BAUERLE, G. REUNING *et al.*, 2013 Development of a next-generation NIL library in *Arabidopsis thaliana* for dissecting complex traits. BMC genomics **14:** 655.

FOWKE, J. H., F.-L. CHUNG, F. JIN, D. QI, Q. CAI *et al.*, 2003 Urinary isothiocyanate levels, brassica, and human breast cancer. Cancer Research **63:** 3980-3986.

FU, Y.-B., B. CHENG and G. W. PETERSON, 2014 Genetic diversity analysis of yellow mustard (Sinapis alba L.) germplasm based on genotyping by sequencing. Genetic Resources and Crop Evolution **61:** 579-594.

FU, Y.-B., and G. W. PETERSON, 2011 Genetic diversity analysis with 454 pyrosequencing and genomic reduction confirmed the eastern and western division in the cultivated barley gene pool. The Plant Genome **4:** 226-237.

GÓMEZ-CAMPO, C., 1999 *Biology of Brassica coenospecies / edited by César Gómez-Campo*. Amsterdam ; New York : Elsevier, 1999.

GAO, M., G. LI, B. YANG, D. QIU, M. FARNHAM *et al.*, 2007 High-density *Brassica oleracea* linkage map: identification of useful new linkages. Theor Appl Genet **115:** 277-287.

GIOVANNELLI, J. L., M. W. FARNHAM, M. WANG and A. E. STRAND, 2002 Development of sequence characterized amplified region markers linked to downy mildew

resistance in broccoli. Journal of the American Society for Horticultural Science **127:** 597-601.

GLAUBITZ, J. C., T. M. CASSTEVENS, F. LU, J. HARRIMAN, R. J. ELSHIRE *et al.*, 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One **9:** e90346.

GRANDCLÉMENT, C., and G. THOMAS, 1996 Detection and analysis of QTLs based on RAPD markers for polygenic resistance to *Plasmodiophora brassicae* Woron in *Brassica oleracea* L. Theoretical and Applied Genetics **93:** 86-90.

GREENUP, A., W. J. PEACOCK, E. S. DENNIS and B. TREVASKIS, 2009 The molecular biology of seasonal flowering-responses in Arabidopsis and the cereals. Annals of Botany **103:** 1165-1172.

HANSEN, L. N., R. ORTIZ and S. B. ANDERSEN, 1999 Genetic analysis of protoplast regeneration ability in *Brassica oleracea*. Plant cell, tissue and organ culture **58:** 127-132.

HANSON, W., 1959 Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. Genetics **44:** 833.

HE, J., X. ZHAO, A. LAROCHE, Z.-X. LU, H. LIU *et al.*, 2014 Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in plant science **5**.

HENNIG, K., R. VERKERK, M. DEKKER and G. BONNEMA, 2013 Quantitative trait loci analysis of non-enzymatic glucosinolate degradation rates in *Brassica oleracea* during food processing. Theoretical and Applied Genetics **126:** 2323-2334.

HOLME, I., A. TORP, L. HANSEN and S. ANDERSEN, 2004 Quantitative trait loci affecting plant regeneration from protoplasts of *Brassica oleracea*. Theoretical and Applied genetics **108:** 1513-1520.

HOWELL, E. C., G. C. BARKER, G. H. JONES, M. J. KEARSEY, G. J. KING *et al.*, 2002 Integration of the cytogenetic and genetic linkage maps of *Brassica oleracea*. Genetics **161:** 1225-1234.

HU, J., J. SADOWSKI, T. OSBORN, B. LANDRY and C. QUIROS, 1998 Linkage group alignment from four independent *Brassica oleracea* RFLP maps. Genome **41:** 226-235.

HUANG, Y.-F., J. A. POLAND, C. P. WIGHT, E. W. JACKSON and N. A. TINKER, 2014 Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. PloS one **9:** e102448.

INIGUEZ-LUY, F. L., L. LUKENS, M. W. FARNHAM, R. M. AMASINO and T. C. OSBORN, 2009 Development of public immortal mapping populations, molecular markers and linkage maps for rapid cycling *Brassica rapa* and *B. oleracea*. Theor Appl Genet **120:** 31-43.

INIGUEZ-LUY, F. L., A. V. VOORT and T. C. OSBORN, 2008 Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype. Theoretical and applied genetics **117:** 977-985.

IZZAH, N. K., J. LEE, M. JAYAKODI, S. PERUMAL, M. JIN *et al.*, 2014 Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. *capitata* L.) and construction of an EST-based genetic map. BMC genomics **15:** 149.

JARQUÍN, D., K. KOCAK, L. POSADAS, K. HYMA, J. JEDLICKA *et al.*, 2014 Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC genomics **15:** 740.

JEUKEN, M., and P. LINDHOUT, 2004 The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. Theoretical and Applied Genetics **109:** 394-401.

JIAO, Y., J. LEEBENS-MACK, S. AYYAMPALAYAM, J. E. BOWERS, M. R. MCKAIN *et al.*, 2012 A genome triplication associated with early diversification of the core eudicots. Genome Biol **13:** R3.

KACZMAREK, M., G. KOCZYK, P. A. ZIOLKOWSKI, D. BABULA-SKOWRONSKA and J. SADOWSKI, 2009 Comparative analysis of the Brassica oleracea genetic map and the *Arabidopsis thaliana* genome. Genome **52:** 620-633.

KAEPPLER, S., 1997 Quantitative trait locus mapping using sets of near-isogenic lines: relative power comparisons and technical considerations. Theoretical and applied genetics **95:** 384-392.

KEMPIN, S. A., B. SAVIDGE and M. F. YANOFSKY, 1995 Molecular basis of the cauliflower phenotype in Arabidopsis. Science **267:** 522-525.

KENNARD, W., M. SLOCUM, S. FIGDORE and T. OSBORN, 1994 Genetic analysis of morphological variation in *Brassica oleracea* using molecular markers. Theoretical and Applied Genetics **87:** 721-732.

KEURENTJES, J. J., L. BENTSINK, C. ALONSO-BLANCO, C. J. HANHART, H. BLANKESTIJN-DE VRIES *et al.*, 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.

KIANIAN, S., and C. QUIROS, 1992 Generation of a Brassica oleracea composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theoretical and Applied Genetics **84:** 544-554.

KOOKE, R., E. WIJNKER and J. J. KEURENTJES, 2012 Backcross populations and near isogenic lines, pp. 3-16 in *Quantitative Trait Loci (QTL)*. Springer.

KOWALSKI, S. P., T.-H. LAN, K. A. FELDMANN and A. H. PATERSON, 1994 Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics **138:** 499-510.

KRISTAL, A. R., and J. W. LAMPE, 2002 Brassica vegetables and prostate cancer risk: a review of the epidemiological evidence. Nutrition and cancer **42:** 1-9.

LABATE, J. A., L. D. ROBERTSON, A. M. BALDO and T. BJORKMAN, 2006 Inflorescence identity gene alleles are poor predictors of inflorescence type in broccoli and cauliflower. JOURNAL-AMERICAN SOCIETY FOR HORTICULTURAL SCIENCE **131:** 667.

LAGERCRANTZ, U., and D. J. LYDIATE, 1996 Comparative genome mapping in *Brassica*. Genetics **144:** 1903-1910.

LAMBEL, S., B. LANINI, E. VIVODA, J. FAUVE, W. P. WECHTER *et al.*, 2014 A major QTL associated with *Fusarium oxysporum* race 1 resistance identified in genetic populations derived from closely related watermelon lines using selective genotyping and genotyping-by-sequencing for SNP discovery. Theoretical and Applied Genetics **127:** 2105-2115.

LAN, T.-H., T. A. DELMONTE, K. P. REISCHMANN, J. HYMAN, S. P. KOWALSKI *et al.*, 2000 An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Research **10:** 776-788.

LAN, T.-H., and A. H. PATERSON, 2000 Comparative mapping of quantitative trait loci sculpting the curd of *Brassica oleracea*. Genetics **155:** 1927-1954.

LAN, T. H., and A. H. PATERSON, 2001 Comparative mapping of QTL determining the plant size of *Brassica oleracea*. Theor Appl Genet **103:** 383-397.

LANDRY, B. S., N. HUBERT, R. CRETE, M. S. CHANG, S. E. LINCOLN *et al.*, 1992 A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). Genome **35:** 409-420.

LI, F., H. JIA, L. LIU, C. ZHANG, Z. LIU *et al.*, 2014a Quantitative trait loci mapping for kernel row number using chromosome segment substitution lines in maize. Genetics and Molecular Research **13:** 1707-1716.

LI, G., M. GAO, B. YANG and C. QUIROS, 2003a Gene for gene alignment between the Brassica and Arabidopsis genomes by direct transcriptome mapping. Theoretical and Applied Genetics **107:** 168-180.

LI, G., and C. F. QUIROS, 2001 Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. Theoretical and Applied Genetics **103:** 455-461.

LI, G., A. RIAZ, S. GOYAL, S. ABEL and C. QUIROS, 2001a Inheritance of three major genes involved in the synthesis of aliphatic glucosinolates in *Brassica oleracea*. Journal of the American society for horticultural science **126:** 427-431.

LI, L., and D. GARVIN, 2003 Molecular mapping of Or, a gene inducing beta-carotene accumulation in cauliflower (*Brassica oleracea* L. var. *botrytis*). Genome **46:** 588-594.

LI, L., S. LU, D. M. O'HALLORAN, D. F. GARVIN and J. VREBALOV, 2003b High-resolution genetic and physical mapping of the cauliflower high-beta-carotene gene Or (Orange). Mol Genet Genomics **270:** 132-138.

LI, L., D. J. PAOLILLO, M. V. PARTHASARATHY, E. M. DIMUZIO and D. F. GARVIN, 2001b A novel gene mutation that confers abnormal patterns of β-carotene accumulation in cauliflower (*Brassica oleracea* var. *botrytis*). The Plant Journal **26:** 59-67.

LI, X., Y. WEI, A. ACHARYA, Q. JIANG, J. KANG *et al.*, 2014b A Saturated Genetic Linkage Map of Autotetraploid Alfalfa (*Medicago sativa* L.) Developed Using Genotyping-by-Sequencing Is Highly Syntenous with the *Medicago truncatula* Genome. G3: Genes| Genomes| Genetics **4:** 1971-1979.

LIN, S.-I., J.-G. WANG, S.-Y. POON, C.-L. SU, S.-S. WANG *et al.*, 2005 Differential regulation of *FLOWERING LOCUS C* expression by vernalization in cabbage and Arabidopsis. Plant physiology **137:** 1037-1048.

LIPPMAN, Z. B., Y. SEMEL and D. ZAMIR, 2007 An integrated view of quantitative trait variation using tomato interspecific introgression lines. Current opinion in genetics & development **17:** 545-552.

LIU, H., M. BAYER, A. DRUKA, J. R. RUSSELL, C. A. HACKETT *et al.*, 2014a An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC genomics **15:** 104.

LIU, S., Y. LIU, X. YANG, C. TONG, D. EDWARDS *et al.*, 2014b The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nature communications **5**.

LIU, S., R. ZHOU, Y. DONG, P. LI and J. JIA, 2006 Development, utilization of introgression lines using a synthetic wheat as donor. Theoretical and Applied Genetics **112:** 1360-1373.

LOWE, A. J., C. MOULE, M. TRICK and K. J. EDWARDS, 2004 Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species. Theor Appl Genet **108:** 1103-1112.

LU, F., A. E. LIPKA, J. GLAUBITZ, R. ELSHIRE, J. H. CHERNEY *et al.*, 2013 Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS genetics **9:** e1003215.

LUKENS, L., F. ZOU, D. LYDIATE, I. PARKIN and T. OSBORN, 2003 Comparison of a Brassica oleracea genetic map with the genome of *Arabidopsis thaliana*. Genetics **164:** 359-372.

LYDIATE, D., A. SHARPE, U. LAGERCRANTZ and I. PARKIN, 1993 Mapping the *Brassica* genome. Outlook on agriculture.

LYSAK, M. A., M. A. KOCH, A. PECINKA and I. SCHUBERT, 2005 Chromosome triplication found across the tribe Brassiceae. Genome Res **15:** 516-525.

MATSUBARA, K., I. KONO, K. HORI, Y. NONOUE, N. ONO *et al.*, 2008 Novel QTLs for photoperiodic flowering revealed by using reciprocal backcross inbred lines from crosses between japonica rice cultivars. Theoretical and Applied Genetics **117:** 935-945.

MCGRATH, J. M., and C. QUIROS, 1991 Inheritance of isozyme and RFLP markers in *Brassica campestris* and comparison with *B. oleracea*. Theoretical and Applied Genetics **82:** 668-673.

MCGRATH, J. M., C. F. QUIROS, J. J. HARADA and B. S. LANDRY, 1990 Identification of Brassica oleracea monosomic alien chromosome addition lines with molecular markers reveals extensive gene duplication. Molecular and General Genetics MGG **223:** 198-204.

MEI, J., Y. DING, K. LU, D. WEI, Y. LIU *et al.*, 2013 Identification of genomic regions involved in resistance against Sclerotinia sclerotiorum from wild *Brassica oleracea*. Theoretical and Applied Genetics **126:** 549-556.

MONFORTE, A. J., and S. D. TANKSLEY, 2000 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.

MORIGUCHI, K., C. KIMIZUKA-TAKAGI, K. ISHII and K. NOMURA, 1999 A genetic map based on RAPD, RFLP, isozyme, morphological markers and QTL analysis for clubroot resistance in *Brassica oleracea*. Breeding science **49:** 257-265.

NAGAHARU, U., 1935 Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jap J Bot **7:** 389-452.

NAGAOKA, T., M. DOULLAH, S. MATSUMOTO, S. KAWASAKI, T. ISHIKAWA *et al.*, 2010 Identification of QTLs that control clubroot resistance in *Brassica oleracea* and comparative analysis of clubroot resistance genes between *B. rapa* and *B. oleracea*. Theoretical and applied genetics **120:** 1335-1346.

OCKENDON, D., and P. GATES, 1975 GROWTH OF CROSS-AND SELF-POLLEN TUBES IN THE STYLES OF *BRASSICA OLERACEA*. New Phytologist **75:** 155-160.

OKAZAKI, K., K. SAKAMOTO, R. KIKUCHI, A. SAITO, E. TOGASHI *et al.*, 2007 Mapping and characterization of *FLC* homologs and QTL analysis of flowering time in *Brassica oleracea*. Theor Appl Genet **114:** 595-608.

PARAN, I., and D. ZAMIR, 2003 Quantitative traits in plants: beyond the QTL. Trends in Genetics **19:** 303-306.

PARKIN, I., A. SHARPE, D. KEITH and D. LYDIATE, 1995 Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome **38:** 1122-1131.

PARKIN, I. A., C. KOH, H. TANG, S. J. ROBINSON, S. KAGALE *et al.*, 2014 Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome biology **15:** R77.

PATERSON, A. H., C. L. BRUBAKER and J. F. WENDEL, 1993 A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Molecular Biology Reporter **11:** 122-127.

PATERSON, A. H., J. W. DEVERNA, B. LANINI and S. D. TANKSLEY, 1990 Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics **124:** 735-742.

PEA, G., H. H. AUNG, E. FRASCAROLI, P. LANDI and M. E. PÈ, 2013 Extensive genomic characterization of a set of near-isogenic lines for heterotic QTL in maize (*Zea mays* L.). BMC genomics **14:** 61.

PIQUEMAL, J., E. CINQUIN, F. COUTON, C. RONDEAU, E. SEIGNORET *et al.*, 2005 Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theoretical and Applied Genetics **111:** 1514-1523.

POLAND, J., J. ENDELMAN, J. DAWSON, J. RUTKOSKI, S. WU *et al.*, 2012 Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome **5:** 103-113.

POLAND, J. A., and T. W. RIFE, 2012 Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome **5:** 92-102.

PURUGGANAN, M. D., A. L. BOYLES and J. I. SUDDITH, 2000 Variation and selection at the *CAULIFLOWER* floral homeotic gene accompanying the evolution of domesticated Brassica oleracea. Genetics **155:** 855-862.

QIU, D., M. GAO, G. LI and C. QUIROS, 2009 Comparative sequence analysis for Brassica oleracea with similar sequences in *B. rapa* and *Arabidopsis thaliana*. Plant cell reports **28:** 649-661.

QUIROS, C., O. OCHOA, S. KIANIAN and D. DOUCHES, 1987 Analysis of the *Brassica oleracea* genome by the generation of *B. campestris-oleracea* chromosome addition lines: characterization by isozymes and rDNA genes. Theoretical and Applied Genetics **74:** 758-766.

RAE, A., E. HOWELL and M. KEARSEY, 1999 More QTL for flowering time revealed by substitution lines in *Brassica oleracea*. Heredity **83:** 586-596.

RAFALSKI, A., 2002 Applications of single nucleotide polymorphisms in crop genetics. Current opinion in plant biology **5:** 94-100.

RAKOW, G., 2004 Species origin and economic importance of *Brassica*, pp. 3-11 in *Brassica*. Springer.

RAMSAY, L., D. JENNINGS, M. KEARSEY, D. MARSHALL, E. BOHUON *et al.*, 1996 The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. Genome **39:** 558-567.

RAZI, H., E. HOWELL, H. NEWBURY and M. KEARSEY, 2008 Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in *Brassica oleracea*? Theoretical and Applied Genetics **116:** 179-192.

ROCHERIEUX, J., P. GLORY, A. GIBOULOT, S. BOURY, G. BARBEYRON *et al.*, 2004 Isolate-specific and broad-spectrum QTLs are involved in the control of clubroot in *Brassica oleracea*. Theoretical and applied genetics **108:** 1555-1563.

ROMAY, M. C., M. J. MILLARD, J. C. GLAUBITZ, J. A. PEIFFER, K. L. SWARTS *et al.*, 2013 Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol **14:** R55.

SADIK, S., 1962 Morphology of the curd of cauliflower. American Journal of Botany**:** 290-297.

SALATHIA, N., J. R. LYNN, A. J. MILLAR and G. J. KING, 2007 Detection and resolution of genetic loci affecting circadian period in *Brassica oleracea*. Theoretical and Applied Genetics **114:** 683-692.

SALMON, A., J. CLOTAULT, E. JENCZEWSKI, V. CHABLE and M. J. MANZANARES-DAULEUX, 2008 *Brassica oleracea* displays a high level of DNA methylation polymorphism. Plant Science **174:** 61-70.

SCHOPFER, C. R., M. E. NASRALLAH and J. B. NASRALLAH, 1999 The male determinant of self-incompatibility in *Brassica*. Science **286:** 1697-1700.

SEBASTIAN, R., E. HOWELL, G. KING, D. MARSHALL and M. KEARSEY, 2000 An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations. Theoretical and Applied Genetics **100:** 75-81.

SEBASTIAN, R., M. KEARSEY and G. J. KING, 2002 Identification of quantitative trait loci controlling developmental characteristics of *Brassica oleracea* L. Theoretical and Applied Genetics **104:** 601-609.

SEVERIN, A. J., G. A. PEIFFER, W. W. XU, D. L. HYTEN, B. BUCCIARELLI *et al.*, 2010 An integrative approach to genomic introgression mapping. Plant physiology **154:** 3-12.

SHARMA, B., T. SHINADA, Y. KIFUJI, H. KITASHIBA and T. NISHIO, 2012 Molecular mapping of a male fertility restorer locus of *Brassica oleracea* using expressed sequence tag-based single nucleotide polymorphism markers and analysis of a syntenic region in *Arabidopsis thaliana* for identification of genes encoding pentatricopeptide repeat proteins. Molecular breeding **30:** 1781-1792.

SHI, J., S. HUANG, J. ZHAN, J. YU, X. WANG *et al.*, 2014 Genome-wide microsatellite characterization and marker development in the sequenced *Brassica* crop species. DNA research **21:** 53-68.

SLOCUM, M., S. FIGDORE, W. KENNARD, J. SUZUKI and T. OSBORN, 1990 Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theoretical and Applied Genetics **80:** 57-64.

SMITH, L. B., and G. J. KING, 2000 The distribution of BoCAL-a alleles in *Brassica oleracea* is consistent with a genetic model for curd development and domestication of the cauliflower. Molecular Breeding **6:** 603-613.

SONAH, H., L. O'DONOUGHUE, E. COBER, I. RAJCAN and F. BELZILE, 2014 Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant biotechnology journal.

SONG, K., T. OSBORN and P. WILLIAMS, 1988 *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). TAG Theoretical and Applied Genetics **76:** 593-600.

SONG, K., T. C. OSBORN and P. H. WILLIAMS, 1990 *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). TAG Theoretical and Applied Genetics **79:** 497-506.

SPARROW, P., T. TOWNSEND, A. ARTHUR, P. DALE and J. IRWIN, 2004 Genetic analysis of Agrobacterium tumefaciens susceptibility in *Brassica oleracea*. Theoretical and Applied Genetics **108:** 644-650.

SPINDEL, J., M. WRIGHT, C. CHEN, J. COBB, J. GAGE *et al.*, 2013 Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theoretical and Applied Genetics **126:** 2699-2716.

STEIN, J. C., B. HOWLETT, D. C. BOYES, M. E. NASRALLAH and J. B. NASRALLAH, 1991 Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. Proceedings of the National Academy of Sciences **88:** 8816-8820.

SUZUKI, G., N. KAI, T. HIROSE, K. FUKUI, T. NISHIO *et al.*, 1999 Genomic organization of the S locus: identification and characterization of genes in SLG/SRK region of S9 haplotype of *Brassica campestris* (syn. *rapa*). Genetics **153:** 391-400.

SZALMA, S., B. HOSTERT, J. LEDEAUX, C. STUBER and J. HOLLAND, 2007 QTL mapping with near-isogenic lines in maize. Theoretical and Applied Genetics **114:** 1211-1228.

TALUKDER, S. K., M. A. BABAR, K. VIJAYALAKSHMI, J. POLAND, P. V. PRASAD *et al.*, 2014 Mapping QTL for the traits associated with heat tolerance in wheat (*Triticum aestivum* L.). BMC genetics **15:** 97.

TEAM, R., 2014 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2011; ISBN: 3-900051-07-0, pp.

TERRY, P., A. WOLK, I. PERSSON and C. MAGNUSSON, 2001 Brassica vegetables and breast cancer risk. Jama **285:** 2975-2977.

TOWN, C. D., F. CHEUNG, R. MAITI, J. CRABTREE, B. J. HAAS *et al.*, 2006 Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell **18:** 1348-1359.

TRUCO, M., J. HU, J. SADOWSKI and C. QUIROS, 1996 Inter-and intra-genomic homology of the *Brassica* genomes: implications for their origin and evolution. Theoretical and Applied Genetics **93:** 1225-1233.

UITDEWILLIGEN, J. G., A.-M. A. WOLTERS, B. BJORN, T. J. BORM, R. G. VISSER *et al.*, 2013 A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One **8:** e62355.

UPTMOOR, R., J. LI, T. SCHRAG and H. STÜTZEL, 2012 Prediction of flowering time in *Brassica oleracea* using a quantitative trait loci-based phenology model. Plant Biology **14:** 179-189.

UPTMOOR, R., T. SCHRAG, H. STÜTZEL and E. ESCH, 2008 Crop model based QTL analysis across environments and QTL based estimation of time to floral induction and flowering in *Brassica oleracea*. Molecular Breeding **21:** 205-216.

VÍQUEZ-ZAMORA, M., M. CARO, R. FINKERS, Y. TIKUNOV, A. BOVY *et al.*, 2014 Mapping in the era of sequencing: high density genotyping and its application for mapping TYLCV resistance in Solanum pimpinellifolium. BMC genomics **15:** 1152.

VAN POPPEL, G., D. T. VERHOEVEN, H. VERHAGEN and R. A. GOLDBOHM, 1999 Brassica vegetables and cancer prevention, pp. 159-168 in *Advances in Nutrition and Cancer 2*. Springer.

VERHOEVEN, D. T., R. A. GOLDBOHM, G. VAN POPPEL, H. VERHAGEN and P. A. VAN DEN BRANDT, 1996 Epidemiological studies on brassica vegetables and cancer risk. Cancer Epidemiology Biomarkers & Prevention **5:** 733-748.

VOORRIPS, R., M. JONGERIUS and H. KANNE, 1997 Mapping of two genes for resistance to clubroot (*Plasmodiophora brassicae*) in a population of doubled haploid lines of *Brassica oleracea* by means of RFLP and AFLP markers. Theoretical and applied genetics **94:** 75-82.

WALLEY, P. G., J. CARDER, E. SKIPPER, E. MATHAS, J. LYNN *et al.*, 2012 A new broccoli× broccoli immortal mapping population and framework genetic map: tools for breeders and complex trait analysis. Theoretical and Applied Genetics **124:** 467-484.

WANG, W., S. HUANG, Y. LIU, Z. FANG, L. YANG *et al.*, 2012 Construction and analysis of a high-density genetic linkage map in cabbage (*Brassica oleracea* L. var. *capitata*). BMC genomics **13:** 523.

WARD, J. A., J. BHANGOO, F. FERNÁNDEZ-FERNÁNDEZ, P. MOORE, J. SWANSON *et al.*, 2013 Saturated linkage map construction in Rubus idaeus using genotyping by sequencing and genome-independent imputation. BMC genomics **14:** 2.

WHITE, P., J. P. HAMMOND, G. J. KING, H. C. BOWEN, R. M. HAYDEN *et al.*, 2010 Genetic analysis of potassium use efficiency in *Brassica oleracea*. Annals of botany **105:** 1199-1210.

WILLIAMS, P. H., and C. B. HILL, 1986 Rapid-cycling populations of *Brassica*. Science **232:** 1385-1389.

XIE, W., Q. FENG, H. YU, X. HUANG, Q. ZHAO *et al.*, 2010 Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proceedings of the National Academy of Sciences **107:** 10578-10583.

YU, J., M. ZHAO, X. WANG, C. TONG, S. HUANG *et al.*, 2013 Bolbase: a comprehensive genomics database for *Brassica oleracea*. BMC genomics **14:** 664.

ZHANG, B., C. LIU, Y. WANG, X. YAO, F. WANG *et al.*, 2015 Disruption of a *CAROTENOID CLEAVAGE DIOXYGENASE 4* gene converts flower colour from white to yellow in *Brassica* species. New Phytologist.

ZHANG, X., and S. R. WESSLER, 2004 Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. Proceedings of the National Academy of Sciences of the United States of America **101:** 5589-5594.

ZHANG, Y., Z. FANG, Q. WANG, Y. LIU, L. YANG *et al.*, 2012 Chloroplast Subspecies-Specific SNP Detection and Its Maternal Inheritance in *Brassica oleracea* L. by Using a dCAPS Marker. Journal of Heredity **103:** 606-611.

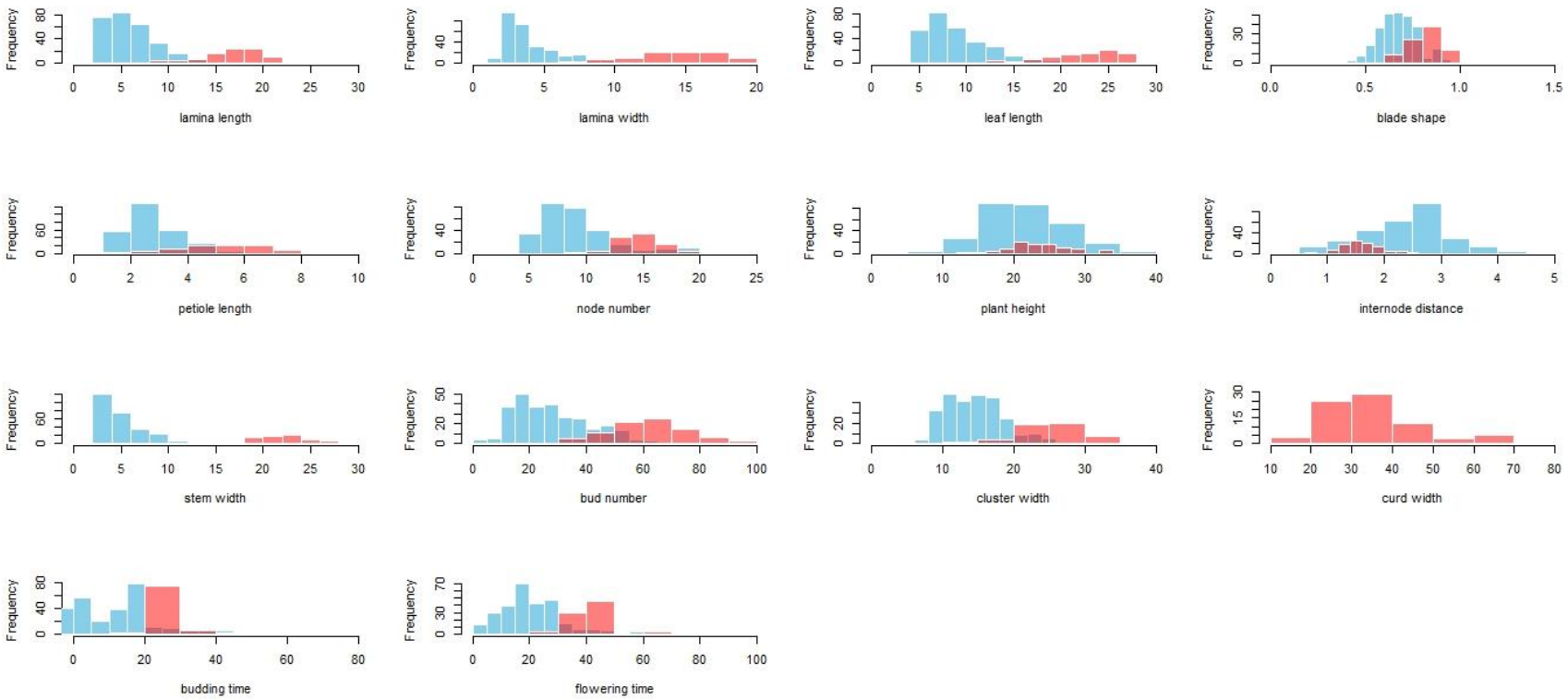


Figure 3.1. Histogram of phenotypes of two seasons, with blue represents the BC_4F_2 population planted in spring and red represents BC_4F_3 population planted in fall.

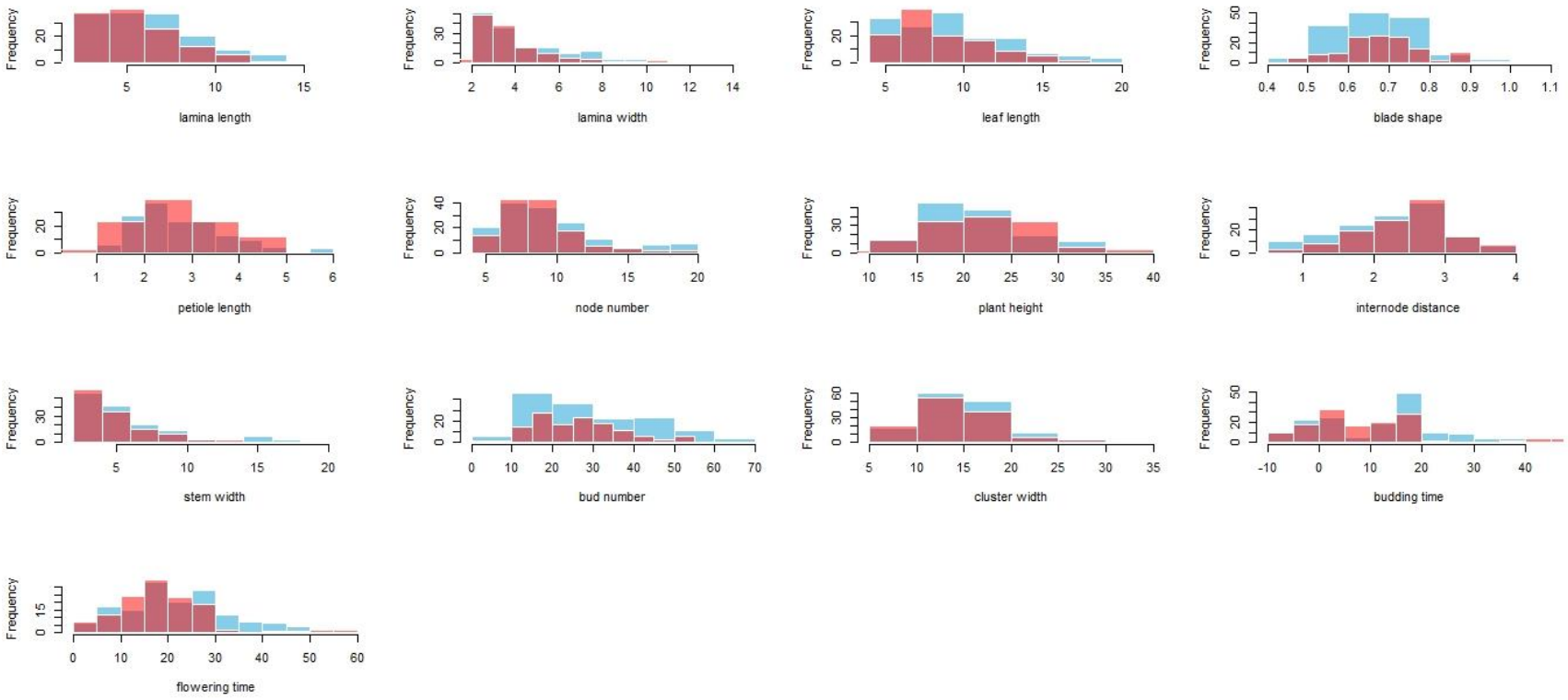


Figure 3.2. Histograms of phenotypes of two BC_4F_2 populations, with blue represents the cabbage population and red represents the cauliflower population.

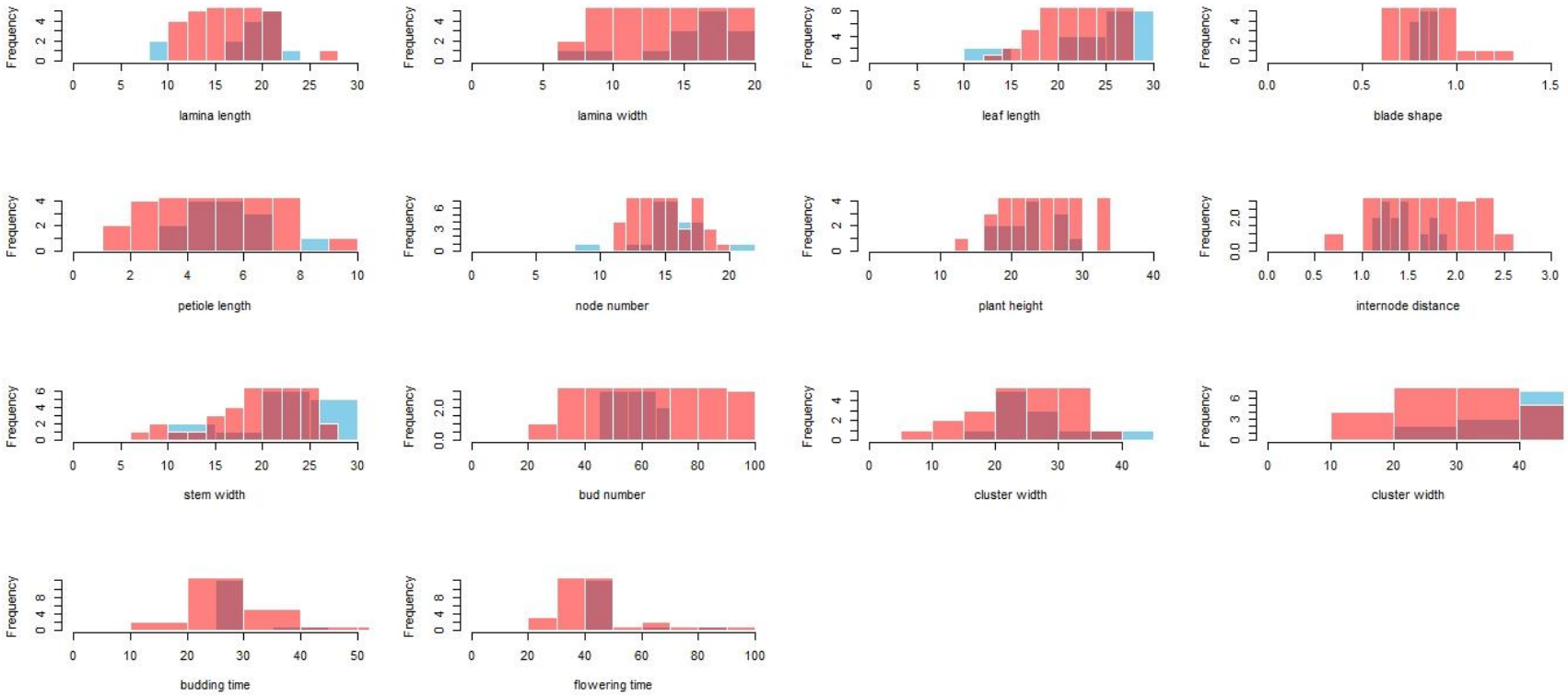


Figure 3.3. Histograms of phenotypes of two BC_4F_3 populations, with blue represents the cabbage population and red represents the cauliflower population.

Table 3.1. Descriptive statistic of parents and backcross families. NA indicates missing data.

	Spring										Fall					
	TO1434 (S6)		BIL (cabbage)		ORG (cauliflower)		cabbage BC_4F_2 families		cauliflower BC_4F_2 families		TO1434 (S6)		ORG (cauliflower)		BC_4F_3 families	
	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	Sd
lamina length (cm)	4.0	1.6	16.1	14.3	18.96	6.6	6.5	2.8	5.6	2.3	18.1	1.7	23.6	1.2	17.4	3.1
lamina width (cm)	2.9	0.9	15.2	29.5	11.78	3.7	4.5	2.2	3.9	1.9	13.5	1.4	17.0	2.7	14.5	3.1
leaf length (cm)	6.0	2.0	16.1	14.3	29.01	9.2	9.4	3.6	8.4	3.1	23.8	2.2	37.1	4.3	22.9	3.8
blade shape (cm)	0.7	0.1	0.9	0.0	0.63	0.1	0.7	0.1	0.7	0.1	0.7	0.1	0.7	0.1	0.8	0.1
petiole length (cm)	2.0	0.5	0.0	0.0	10.05	2.8	2.8	1.0	2.8	1.0	5.7	1.4	13.5	3.5	5.5	1.5
node number	7.3	1.4	17.3	1.3	17.88	2.4	10.2	3.8	9.1	2.7	15.0	2.1	22.8	1.6	15.2	2.1
plant height (cm)	20.4	7.8	11.3	17.3	6.88	2.5	21.6	5.3	22.0	5.8	27.7	3.5	11.2	0.9	23.9	4.0
internode distance (cm)	2.8	0.8	0.7	0.0	0.38	0.1	2.3	0.8	2.5	0.7	1.9	0.4	0.5	0.1	1.6	0.3
stem width (mm)	3.1	1.1	12.2	47.0	13	5.4	6.1	3.8	5.0	2.6	22.6	2.3	17.2	3.0	20.8	4.2

Table 3.1. Cont.

	Spring										Fall					
	TO1434 (S6)		BIL (cabbage)		ORG (cauliflower)		cabbage BC_4F_2 families		cauliflower BC_4F_2 families		TO1434 (S6)		ORG (cauliflower)		BC_4F_3 families	
	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	Sd
bud number	18.0	10.6	NA	NA	NA	NA	29.6	15.0	26.9	11.1	62.4	7.6	NA	NA	62.9	14.5
cluster width (mm)	10.5	2.1	NA	NA	NA	NA	15.1	4.7	14.2	4.2	30.0	3.2	22.4	2.7	25.4	5.5
curd width (mm)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	32.9	3.0	98.8	39.0	36.2	12.6
budding time (day)	1.4	8.8	NA	NA	NA	NA	12.2	11.5	10.4	12.5	27.2	1.1	NA	NA	28.3	7.1
flowering time (day)	12.2	7.6	NA	NA	NA	NA	22.7	11.1	19.6	9.9	42.4	1.8	NA	NA	43.2	11.2

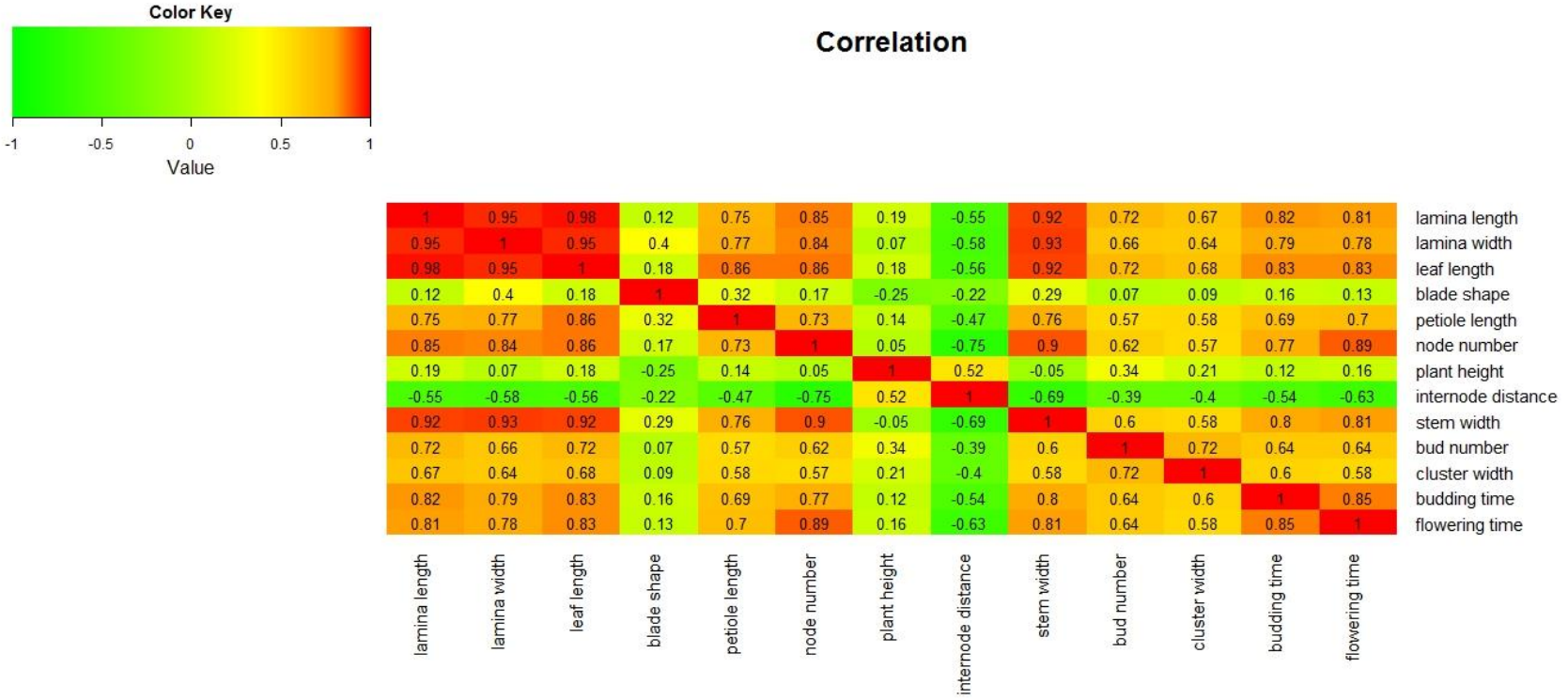


Figure 3.4. Correlation coefficient between traits among the cabbage BC_4F_2 families.

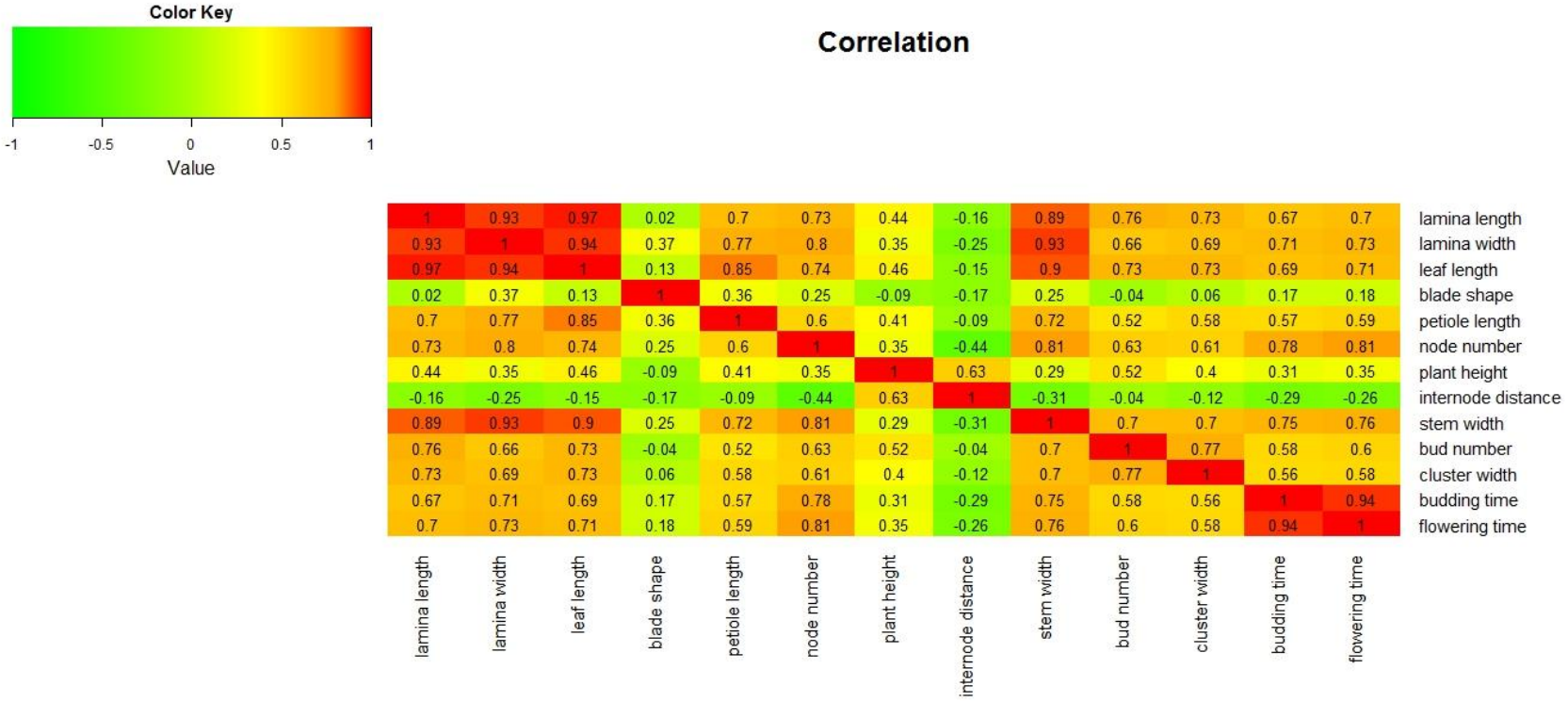


Figure 3.5. Correlation coefficient between traits among the cauliflower BC_4F_2 families.

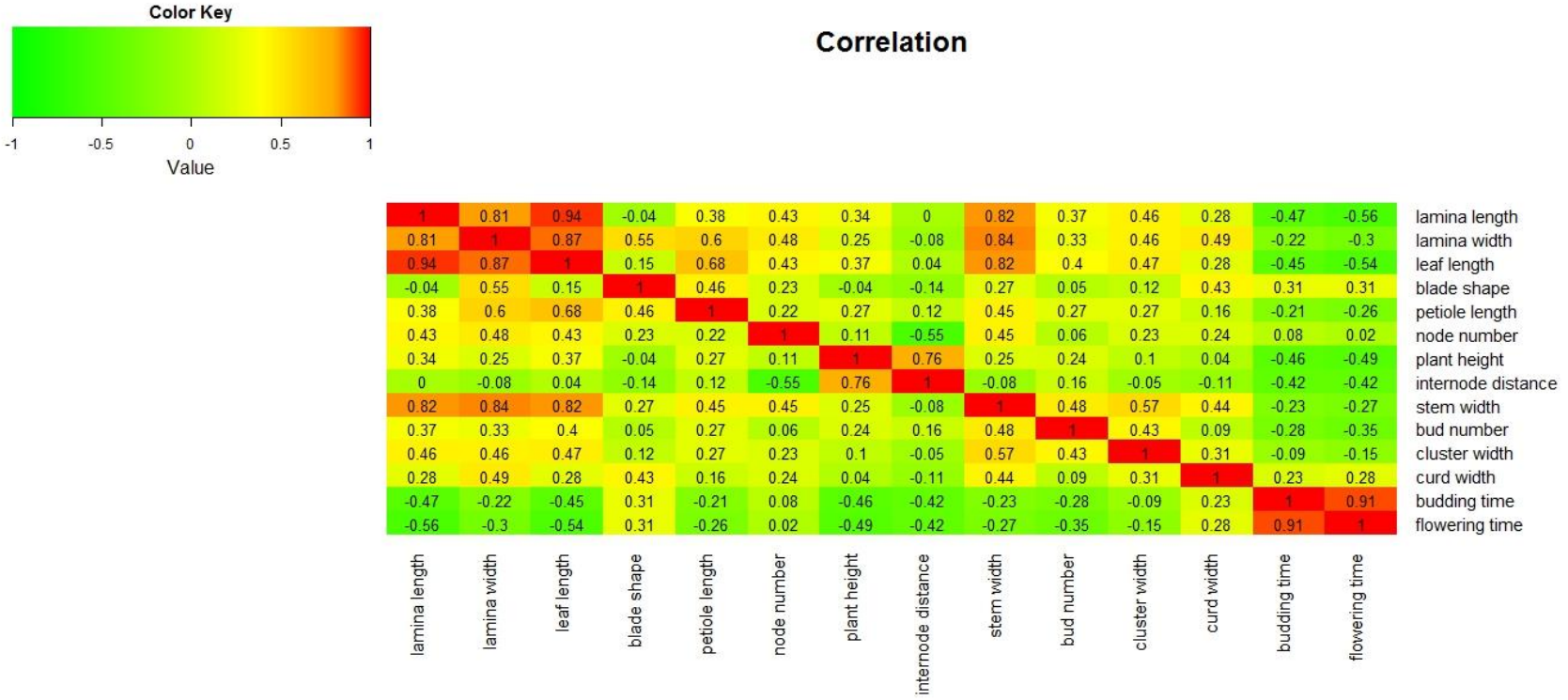
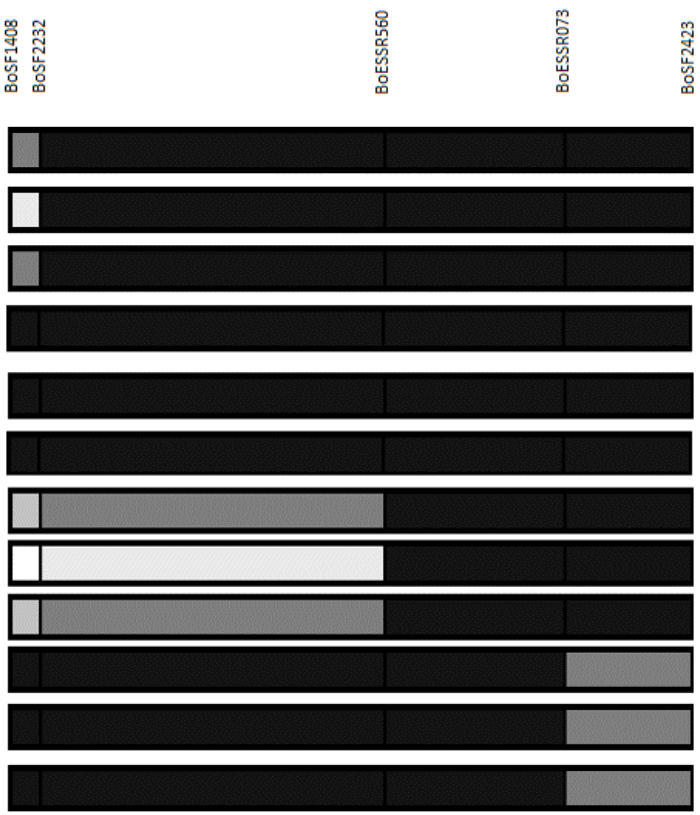


Figure 3.6. Correlation coefficient between traits among the BC_4F_3 families.



Family	Flower color
96-7-1	Yellow
96-7-2	Yellow
96-7-3	Yellow
96-8-1	Yellow
96-8-2	Yellow
96-8-3	Yellow
96-10-1	Yellow
96-10-2	Yellow
96-10-3	Yellow
96-11-1	White
96-11-2	White
96-11-3	White

Figure 3.7. Genotypes and phenotypes of the BC4F3 families segregating for flower color. Grey indicates recombination and black indicates homozygous introgression.

Table 3.2. Markers associated with lamina length. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Spring	Cabbage	BRAS067	1	42,644,937		0.000505	60.7%	-0.1	5.15	5.05
Spring	Cabbage	BoSF2294	2	204,357	0.000431	0.000572		4.01	1.28	5.28
Spring	Cabbage	BoSF1304.j	2	2,283,403	1.50E-10	3.32E-05		4.31	2.14	6.45
Spring	Cabbage	BoSF239	2	4,026,934		0.00023		2.29	2.46	4.75
Spring	Cabbage	BoSF2379.j	2	12,254,157		0.000256		1.75	1.44	3.18
Spring	Cabbage	BrSF556	2	36,950,034		0.000946		1.19	1.36	2.55
Spring	Cabbage	BoSF0439.j	2	52,314,408	0.000272	0.00025		2.58	2.56	5.15
Spring	Cabbage	BoESSR492	3	27,443,339		0.000264		NA	NA	5.12
Spring	Cabbage	BoSF103	4	42,840,233		6.47E-06		2.27	1.38	3.65
Spring	Cauliflower	BoSF2878	5	2,844,916		0.000382	44.7%	1.61	1.23	2.83
Spring	Cauliflower	Ol10-D08	9	4,242,853		0.000907		3.46	1.92	5.38
Spring	Cauliflower	BoSF2364.j	9	7,963,701		5.95E-05		3.66	2.57	6.23
Spring	Cauliflower	BoSF2389.j	9	9,838,674		6.33E-05		3.66	2.04	5.7
Spring	Cauliflower	CB10103	9	11,079,571		0.00043		NA	NA	5.00

Table 3.3. Markers detected for lamina width. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Spring	Cabbage	BoSF1961	1	15,988,128		0.000483	70.7%	2.32	1.54	3.86
Spring	Cabbage	BoSF2425	1	29,261,676		0.00082		2.32	0.87	3.19
Spring	Cabbage	BoSF2294	2	204,357	0.000213	4.61E-06		2.85	1.13	3.98
Spring	Cabbage	BoSF1304.j	2	2,283,403	1.48E-10	1.30E-10		3.36	1.56	4.92
Spring	Cabbage	BoSF239	2	4,026,934		0.00037		1.84	1.17	3.01
Spring	Cabbage	BoSF0439.j	2	52,314,408		6.42E-05		2.13	1.45	3.58
Spring	Cabbage	BoESSR492	3	27,443,339		0.000225		NA	NA	3.64
Spring	Cabbage	BoSF084	4	46,011,369		9.34E-05		1.69	0.69	2.38
Spring	Cabbage	Na10-A08	5	16,170,271		0.000973		1.8	NA	NA
Spring	Cabbage	FITO546	8	15,538,466		0.000136		1.82	1.77	3.59
Spring	Cabbage	BoSF2564	9	3,764,017		0.000665		2.1	-0.81	1.29
Spring	Cabbage	BoSF258	9	49,972,858		0.000373		-0.06	2.97	2.91
Spring	Cauliflower	BoSF2878	5	2,844,916		0.000861	59.8%	1.27	0.43	1.7
Spring	Cauliflower	Ol10-D08	9	4,242,853		7.95E-05		2.54	1.84	4.37
Spring	Cauliflower	BoSF2364.j	9	7,963,701	9.14E-06	1.89E-07		3.87	0.97	4.84
Spring	Cauliflower	BoSF2389.j	9	9,838,674		1.77E-06		3.87	0.43	4.3
Spring	Cauliflower	CB10103	9	11,079,571		0.00004		NA	NA	5.01

Table 3.4. Markers detected for leaf length. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Fall	Cauliflower	Na10-D03	3	22,677,461	1.81E-09		0.97%	-1.65	NA	NA
Spring	Cabbage	BRAS067	1	42,644,937		0.00020563	65.8%	-0.32	7.28	6.96
Spring	Cabbage	BoSF2294	2	204,357	0.000305232	8.34E-07		5.2	1.83	7.03
Spring	Cabbage	BoSF1304.j	2	2,283,403	8.01E-11	2.22E-05		5.59	3	8.59
Spring	Cabbage	BoSF239	2	4,026,934		0.00030437		2.95	3.24	6.19
Spring	Cabbage	BoSF2379.j	2	12,254,157		0.00034157		2.23	1.89	4.12
Spring	Cabbage	BoESSR151	2	27,464,074		0.00077691		1.52	2.03	3.55
Spring	Cabbage	BoSF0439.j	2	52,314,408	0.000415539	0.00017793		3.34	3.35	6.7
Spring	Cabbage	BoESSR492	3	27,443,339		5.04E-05		NA	NA	7.26
Spring	Cabbage	BoSF103	4	42,840,233		7.31E-06		2.72	2.02	4.74
Spring	Cabbage	BoSF2878	5	2,844,916		0.00025883		1.81	1.46	3.26
Spring	Cabbage	BoSF052	9	36,858,822		0.00046373		1.81	3.85	5.66

Table 3.4. Cont.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Spring	Cauliflower	BoSF2983	3	20,512,580		2.74E-05	50.4%	3.16	NA	NA
Spring	Cauliflower	Na14-H12	5	32,568,566		0.00094962		2.43	2.43	4.86
Spring	Cauliflower	Ol10-D08	9	4,242,853		0.00047755		4.95	2.68	7.63
Spring	Cauliflower	BoSF2364.j	9	7,963,701		1.01E-05		4.96	NA	NA
Spring	Cauliflower	BoSF2389.j	9	9,838,674	1.25E-07	3.72E-05		5.03	3.81	8.84
Spring	Cauliflower	CB10103	9	11,079,571	0.000708168	0.00030		NA	NA	6.79
Spring	Cauliflower	BoSF0211.j	9	52,770,497		0.00025360		NA	NA	4.16

Table 3.5. Markers detected for blade shape. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Fall	Cauliflower	FITO377	2	45,755,762	3.53E-06		8.6%	0.06	0.07	0.13
Fall	Cauliflower	BoSF2051	3	9,166,933	3.87E-07			0.09	NA	NA
Fall	Cauliflower	Ol11-B05	3	10,424,336	1.77E-05			0.1	NA	NA
Spring	Cauliflower	BoSF2364.j	9	7,963,701	2.58E-06		22.3%	0.1	-0.06	0.05
Spring	Cauliflower	BoSF2389.j	9	9,838,674	8.10E-06			0.1	-0.08	0.03
Spring	Cauliflower	CB10103	9	11,079,571	7.79E-07			NA	NA	0.14

Table 3.6. Markers detected for petiole length. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Fall	Cauliflower	BoSF2390	5	46,098,797	0.000742		15.9%	-0.76	-3.44	-4.2
Spring	Cabbage	FITO523	1	40,563,380		0.000304	48.5%	NA	NA	1.59
Spring	Cabbage	BoSF2294	2	204,357	0.000801	3.12E-05		1.2	0.56	1.76
Spring	Cabbage	BoSF1304.j	2	2,283,403	1.27E-12	6.46E-05		1.29	0.87	2.15
Spring	Cabbage	BoSF2532	2	48,056,892		0.000521		0.3	0.75	1.05
Spring	Cabbage	BoSF0439.j	2	52,314,408	0.000606	9.09E-06		0.77	0.79	1.56
Spring	Cabbage	BoESSR492	3	27,443,339		1.11E-05		NA	NA	2.15
Spring	Cabbage	BoSF2423	3	60,314,017		0.000357		0.88	0.85	1.73
Spring	Cabbage	BoSF084	4	46,011,369		0.000211		0.5	0.51	1.02
Spring	Cabbage	BoSF258	9	49,972,858		0.000729		0.33	1.13	1.46

Table 3.6. Cont.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Spring	Cauliflower	BoSF2436.j	1	3,509,864		0.005281	46.1%	NA	NA	0.91
Spring	Cauliflower	BrBAC030	4	20,213,586		5.10E-05		0.25	1.14	1.39
Spring	Cauliflower	BoESSR333	4	26,289,058		0.000734		0.28	0.92	1.2
Spring	Cauliflower	BoSF2878	5	2,844,916		0.000435		1.03	-0.18	0.85
Spring	Cauliflower	CB10027	5	17,732,720		0.000132		1.26	-0.3	0.96
Spring	Cauliflower	BoE836	8	41,048,747	0.000108			0.39	-0.53	-0.13
Spring	Cauliflower	BoSF2564	9	3,764,017		3.44E-05		0.81	0.86	1.66
Spring	Cauliflower	Ol10-D08	9	4,242,853		0.000293		1.5	0.76	2.26
Spring	Cauliflower	BoSF2389.j	9	9,838,674		1.10E-05		1.38	0.79	2.17
Spring	Cauliflower	CB10103	9	11,079,571		0.00038		NA	NA	1.8
Spring	Cauliflower	BoSF0211.j	9	52,770,497		0.000211		NA	NA	1.27

Table 3.7. Markers detected for node number. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Fall	Cauliflower	BoESSR632	1	38,689,105	1.45E-15		2.6%	-1	NA	NA
Spring	Cabbage	Ol12-F11	1	11,528,823		0.000612	74.6%	2.52	4.93	7.45
Spring	Cabbage	BRAS067	1	42,644,937		0.000927		-0.15	4.71	4.56
Spring	Cabbage	BoSF2294	2	204,357	0.000225	6.03E-06		5.35	2.06	7.41
Spring	Cabbage	BoSF1304.j	2	2,283,403	3.33E-11	4.00E-12		6.21	2.65	8.85
Spring	Cabbage	BoSF2532	2	48,056,892		0.000109		2.65	0.8	3.45
Spring	Cabbage	BoSF0439.j	2	52,314,408	0.000277	2.02E-05		3.56	2.77	6.34
Spring	Cabbage	BoESSR492	3	27,443,339		0.000538		NA	NA	5.41
Spring	Cabbage	BoSF103	4	42,840,233		5.22E-05		2.53	1.25	3.78
Spring	Cabbage	Na10-A08	5	16,170,271		0.000603		2.6	NA	NA
Spring	Cabbage	CB10204	7	35,417,201		0.000132		2.35	1.35	3.7
Spring	Cabbage	FITO546	8	15,538,466		0.00011		2.35	2.85	5.2
Spring	Cabbage	BoSF2564	9	3,764,017		0.000493		2.35	1.35	3.7
Spring	Cabbage	BoSF052	9	36,858,822		0.000137		1.52	3.18	4.7
Spring	Cabbage	PBCGSSRBo34	9	43,632,105		0.000147		1.35	3.46	4.81
Spring	Cabbage	BoSF258	9	49,972,858		0.000736		1.6	2.81	4.41

Table 3.7. cont.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Spring	Cauliflower	BoSF2436.j	1	3,509,864		0.000346	63.7%	NA	NA	6.7
Spring	Cauliflower	Ol10-D08	9	4,242,853	2.24E-05	2.42E-06		3.85	2.68	6.53
Spring	Cauliflower	BoSF2364.j	9	7,963,701	4.51E-07	2.31E-08		5.52	1.43	6.95
Spring	Cauliflower	BoSF2389.j	9	9,838,674	4.27E-05	1.46E-06		5.52	0.33	5.84
Spring	Cauliflower	CB10103	9	11,079,571		0.00003		NA	NA	8.13
Spring	Cauliflower	BoSF0347.j	9	50,141,636		0.000227		4.1	0.8	4.9
Spring	Cauliflower	BoSF0654.j	9	51,494,032		6.60E-05		4.1	-1.9	2.2

Table 3.8. Markers detected for plant height. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Fall	Cauliflower	BoSF1408	3	36058441	0.000449		17%	1.07	-3.89	-2.83
Spring	Cabbage	BoSF1304.j	2	2283403	1.00E-05		8.4%	-1.76	-0.46	-2.23

Table 3.9. Markers detected for internode distance. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Seasons	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Fall	Cauliflower	BrBAC289	3	24,201,112	1.31E-05		12.5%	-0.42	-0.38	-0.8
Fall	Cauliflower	BoSF0654.j	9	51,494,032	1.31E-05			-0.4	NA	NA
Spring	Cabbage	BoSF2294	2	204,357	5.63E-12	0.000571	44.8%	-0.84	-0.54	-1.39
Spring	Cabbage	BoSF1304.j	2	2,283,403	4.58E-31	5.05E-11		-0.95	-0.67	-1.62
Spring	Cabbage	BoSF239	2	4,026,934	0.000526			-0.77	-0.23	-1
Spring	Cabbage	BoSF2033	7	38,982,064	9.69E-15			NA	NA	-1.04
Spring	Cauliflower	Ol10-D08	9	4,242,853	0.000449		31.1%	-0.54	-0.81	-1.35
Spring	Cauliflower	BoSF2364.j	9	7,963,701	5.50E-06	0.000476		-0.86	-0.35	-1.21
Spring	Cauliflower	CB10103	9	11,079,571	1.98E-05			NA	NA	-1.35
Spring	Cauliflower	BoSF0347.j	9	50,141,636	0.000273			-0.8	-0.29	-1.09

Table 3.10. Markers detected for stem width. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Spring	Cabbage	Ol12-F11	1	11,528,823		0.000425	85.4%	2.78	4.91	7.69
Spring	Cabbage	BoSF1961	1	15,988,128		0.000139		4.03	1.95	5.98
Spring	Cabbage	BRAS067	1	42,644,937		0.000299		0.09	4.88	4.97
Spring	Cabbage	BoSF2294	2	204,357	1.46E-05	1.16E-07		5.16	2.53	7.69
Spring	Cabbage	BoSF1304.j	2	2,283,403	1.60E-11	3.51E-14		6.89	2.37	9.25
Spring	Cabbage	BoSF239	2	4,026,934	0.000521	0.000365		4.07	2.46	6.53
Spring	Cabbage	BoESSR122	2	4,394,678		0.000249		2	3.32	5.32
Spring	Cabbage	BoSF2379.j	2	12,254,157		0.000494		2.03	2.06	4.09
Spring	Cabbage	BoESSR151	2	27,464,074		0.000893		1.36	1.57	2.93
Spring	Cabbage	BoSF2532	2	48,056,892		0.000131		1.42	1.63	3.05
Spring	Cabbage	BoSF0439.j	2	52,314,408		1.90E-05		3.8	2.16	5.96
Spring	Cabbage	BoESSR492	3	27,443,339		0.00038		NA	NA	5.77
Spring	Cabbage	BoSF2370	4	13,824,817		0.000486		2.8	0.44	3.24
Spring	Cabbage	BrSF537	4	44,064,505		8.18E-06		2.16	2.27	4.43
Spring	Cabbage	BoSF2878	5	2,844,916		0.000355		0.93	0.88	1.81
Spring	Cabbage	Na10-A08	5	16,170,271		0.000974		2.58	NA	NA

Table 3.10. Cont.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Spring	Cabbage	CB10204	7	35,417,201		0.000921		1.63	0.87	2.5
Spring	Cabbage	FITO546	8	15,538,466		8.23E-06		3.07	2.1	5.17
Spring	Cabbage	BoSF052	9	36,858,822		0.000601		1.59	2.13	3.72
Spring	Cabbage	PBCGSSRBo34	9	43,632,105		0.000652		1.58	2.44	4.02
Spring	Cabbage	BoSF258	9	49,972,858		9.43E-05		0.48	3.81	4.29
Spring	Cauliflower	BoSF2436.j	1	3,509,864		0.000369	71.8%	NA	NA	4.69
Spring	Cauliflower	BoSF2983	3	20,512,580		0.000641		2.54	NA	NA
Spring	Cauliflower	BoSF2878	5	2,844,916		4.33E-05		1.54	1.55	3.1
Spring	Cauliflower	CB10027	5	17,732,720		0.00063		2.14	-0.32	1.83
Spring	Cauliflower	Ol10.D08	9	4,242,853	2.49E-05	1.49E-08		3.83	2.51	6.34
Spring	Cauliflower	BoSF2364.j	9	7,963,701	4.61E-07	0.000235		5.66	0.95	6.61
Spring	Cauliflower	BoSF2389.j	9	9,838,674	8.94E-06	8.59E-08		5.66	0.6	6.26
Spring	Cauliflower	CB10103	9	11,079,571	3.13E-06	0.00000		NA	NA	7.97
Spring	Cauliflower	BoSF0347.j	9	50,141,636		1.72E-05		3.72	-0.09	3.63
Spring	Cauliflower	BoSF0211.j	9	52,770,497		0.000739		NA	NA	2.9

Table 3.11. Markers detected for bud number. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Fall	Cauliflower	FITO377	2	45,755,762	0.000151		24.9%	16.8	7.8	24.6
Fall	Cauliflower	BoSF1568	4	11,867,081	6.30E-06	0.00022		15.43	NA	NA
Fall	Cauliflower	BoSF2863	4	34,595,747	3.95E-06	0.000913		12.72	NA	NA
Fall	Cauliflower	BoESSR763	4	45,178,599	6.30E-06	0.00022		15.47	15.13	30.6
Spring	Cabbage	Ol12-F11	1	11,528,823		0.000912	28.3%	8.17	26.83	35
Spring	Cabbage	BoSF1304.j	2	2,283,403	0.000755	9.85E-05		6.93	20.99	27.92
Spring	Cabbage	BoSF0439.j	2	52,314,408		0.000703		9.79	14.67	24.45
Spring	Cabbage	BoSF103	4	42,840,233		0.000656		12.59	4.83	17.42
Spring	Cabbage	PBCGSSRBo34	9	43,632,105		0.000302		6.33	19.22	25.56
Spring	Cauliflower	BoSF2878	5	2,844,916		0.000719	28.3%	11.5	5.61	17.11
Spring	Cauliflower	Ol10-D08	9	4,242,853	0.000672			13.25	21.25	34.5
Spring	Cauliflower	BoSF2364.j	9	7,963,701	6.93E-07	0.000147		10.17	21.43	31.6
Spring	Cauliflower	BoSF2389.j	9	9,838,674	3.66E-05	0.000421		10.17	17	27.17

Table 3.12. Markers detected for cluster width. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Spring	cabbage	BoSF2425	1	29,261,676		0.000319	43.7%	5.06	5.96	11.02
Spring	cabbage	BoSF2294	2	204,357	4.32E-05	3.19E-07		7.06	5.01	12.07
Spring	cabbage	BoSF239	2	4,026,934		0.000204		2.53	9.08	11.61
Spring	cabbage	BoSF2532	2	48,056,892		0.000831		-0.77	7.6	6.83
Spring	cabbage	BoSF0439.j	2	52,314,408	0.000169	3.47E-05		4.77	4.14	8.92
Spring	cabbage	BoSF084	4	46,011,369		0.000524		2.44	3.59	6.03
Spring	cabbage	BoESSR945	5	7,204,711		0.000129		5.62	-0.93	4.7
Spring	cabbage	Na10.A08	5	16,170,271		0.000165		5.62	NA	NA
Spring	cabbage	BoSF2292	6	19,446,298		0.000722		3.35	3.08	6.43
Spring	cabbage	CB10204	7	35,417,201		0.000227		1.33	4.78	6.1
Spring	cabbage	BoESSR758	7	48,170,105		0.000886		3.24	2.63	5.87
Spring	cabbage	FITO546	8	15,538,466		0.000961		2.2	9.05	11.25

Table 3.12. Cont.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Spring	Cabbage	Ol10-D08	9	4,242,853		0.000694		NA	NA	6.89
Spring	Cabbage	BoESSR484	9	16,943,553		0.000103		4.27	2.43	6.7
Spring	Cabbage	PBCGSSRBo34	9	43,632,105		0.000851		2.93	3.72	6.65
Spring	Cabbage	BoSF258	9	49,972,858		0.000269		2.53	4.73	7.26
Spring	Cauliflower	BoSF2436.j	1	3,509,864		0.000511	44.7%	NA	NA	3.89
Spring	Cauliflower	BoSF2878	5	2,844,916		2.63E-05		4.6	1.53	6.13
Spring	Cauliflower	CB10027	5	17,732,720		1.62E-05		5.65	0.55	6.19
Spring	Cauliflower	Na14-H12	5	32,568,566		0.000419		3.95	3.7	7.64
Spring	Cauliflower	Ol10-D08	9	4,242,853	0.00027	1.15E-05		5.52	8.47	13.99
Spring	Cauliflower	BoSF2364.j	9	7,963,701	0.000137	2.30E-05		5.81	6.24	12.05
Spring	Cauliflower	BoSF2389.j	9	9,838,674	5.22E-07	0.000201		5.81	4.11	9.93
Spring	Cauliflower	CB10103	9	11,079,571	0.000126	0.00060		NA	NA	9.88
Spring	Cauliflower	BoSF0654.j	9	51,494,032		0.00058		4	0.75	4.74
Spring	Cauliflower	BoSF0211.j	9	52,770,497		0.00022		NA	NA	6.78

Table 3.13. Markers detected for curd width. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Fall	Cauliflower	BoSF2345	1	5,136,732		0.000182	26.4%	22.08	10.13	32.22
Fall	Cauliflower	BoESSR560	3	49,453,737	0.000565			-2.49	NA	NA

Table 3.14. Markers detected for budding time. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

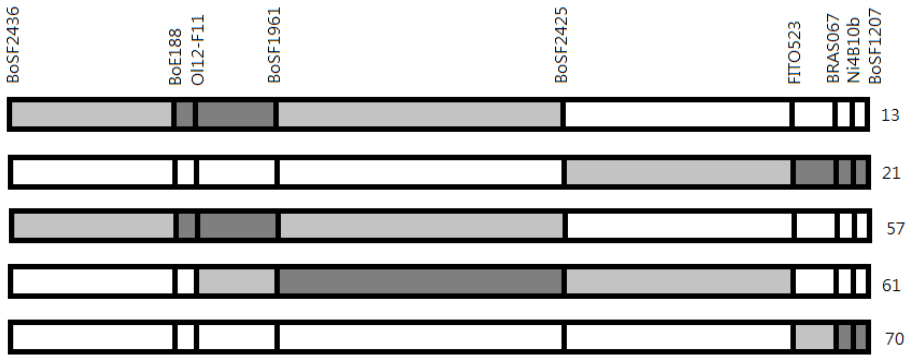
Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Spring	Cabbage	BoSF1961	1	15,988,128		0.000362	45.7%	11.47	6.47	17.93
Spring	Cabbage	BoSF2294	2	204,357	3.10E-05	1.58E-05		14.13	7.75	21.89
Spring	Cabbage	BoSF1304.j	2	2,283,403		9.60E-13		17.37	7.61	24.98
Spring	Cabbage	BoSF0439.j	2	52,314,408		0.00017		11.01	5.4	16.42
Spring	Cabbage	BoESSR492	3	27,443,339		0.000704		NA	NA	18.03
Spring	Cabbage	BoSF084	4	46,011,369		0.000274		9.8	3.8	13.6
Spring	Cabbage	BoSF2033	7	38,982,064		0.000313		NA	NA	18.6
Spring	Cabbage	PBCGSSRBo34	9	43,632,105		0.000532		9.43	4.4	13.82
Spring	Cauliflower	BoSF2983	3	20,512,580		0.00058	49.9%	8.3	NA	NA
Spring	Cauliflower	Ol10.D08	9	4,242,853	4.16E-05	2.08E-05		14.55	20.62	35.17
Spring	Cauliflower	BoSF2364.j	9	7,963,701	1.13E-05	1.14E-05		19.63	7.97	27.6
Spring	Cauliflower	BoSF2389.j	9	9,838,674	0.000314	6.51E-05		19.63	1.3	20.93
Spring	Cauliflower	CB10103	9	11,079,571		0.00020		NA	NA	27.78
Spring	Cauliflower	BoSF0347.j	9	50,141,636		0.000832		16.55	8.85	25.4
Spring	Cauliflower	BoSF0654.j	9	51,494,032		0.000978		16.55	-1.95	14.6

Table 3.15. Markers detected for flowering time. *A* represents additive effect and *D* represents dominance effect. *NA* indicates missing data. R^2 was estimated by taking all the significant markers as independent variables and phenotypic value as dependent variable.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A*+*D*
Fall	Cauliflower	BoESSR901	9	42,420,935		5.74E-05	36.8%	NA	NA	44.6
Fall	Cabbage	BrSF556	2	36,950,034	0.000353		5.4%	2.59	NA	NA
Fall	Cabbage	BoSF2717	9	52,831,044	0.000985			0.3	9.01	9.31
Spring	Cabbage	BoSF1961	1	15,988,128		0.000561	55.2%	11.23	4.9	16.13
Spring	Cabbage	BoSF2294	2	204,357	0.000177	6.45E-06		13.73	8.35	22.09
Spring	Cabbage	BoSF1304.j	2	2,283,403	7.87E-18	6.94E-15		17.04	8.14	25.18
Spring	Cabbage	BoSF239	2	4,026,934	0.000824			11.9	6.97	18.87
Spring	Cabbage	BoSF2532	2	48,056,892		0.000355		7.8	4.13	11.93
Spring	Cabbage	BoSF0439.j	2	52,314,408	0.000355	0.00012		10.04	7.21	17.25
Spring	Cabbage	BoSF103	4	42,840,233		0.000174		8.35	5.03	13.38
Spring	Cabbage	FITO546	8	15,538,466	2.13E-10	7.54E-05		7.9	8.4	16.3
Spring	Cabbage	Ol10.D08	9	4,242,853	8.61E-06	4.45E-05		NA	NA	16.55
Spring	Cabbage	BoESSR484	9	16,943,553	1.82E-09	1.34E-05		8.15	9.65	17.8
Spring	Cabbage	PBCGSSRBo34	9	43,632,105		0.000279		7.07	8.51	15.58
Spring	Cabbage	BoSF258	9	49,972,858		0.000273		5.65	10.72	16.37

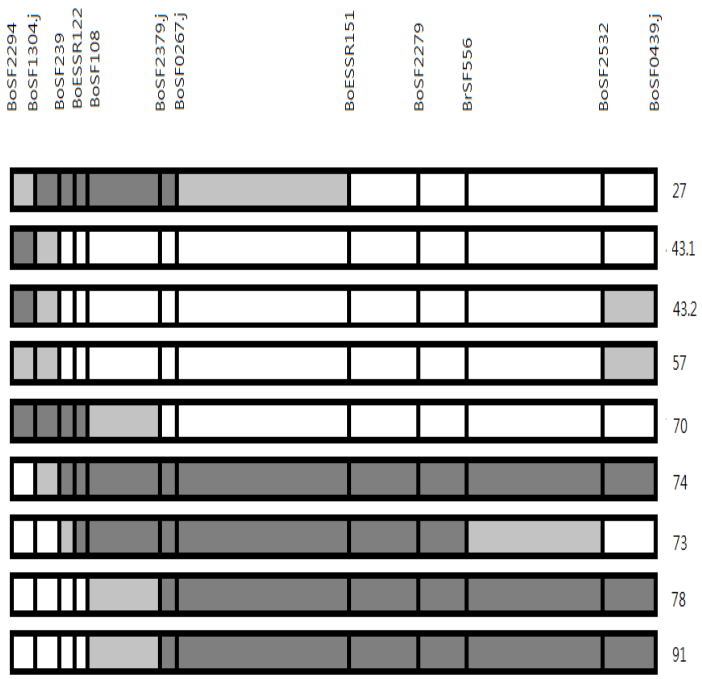
Table 3.15. Cont.

Season	Population	Marker	Chr.	Position	Non-introgressed individuals (*p*-value)	Recurrent parent (*p*-value)	R^2	*A*	*D*	*A+D*
Spring	Cauliflower	BoSF2878	5	2,844,916		0.000855	33.6%	6.9	4.23	11.13
Spring	Cauliflower	BoSF2364.j	9	7,963,701	2.95E-05	2.85E-05		18.07	4.33	22.4
Spring	Cauliflower	BoSF2389.j	9	9,838,674	5.98E-05	3.57E-05		18.07	2.07	20.13
Spring	Cauliflower	CB10103	9	11,079,571	0.00054	0.00017		NA	NA	25.63



Family	Lamina width (cm)	Stem width (mm)
13	0.11	1.91
21	1.08*	3.35*
57	4.76 (NA)	9.6 (NA)
61	2.54*	5.96 (NA)
70	1.79 (NA)	5.16 (NA)

Figure 3.8. Introgressions on chromosome 1 of cabbage BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.



Family	Lamina length (cm)	Lamina width (cm)	Leaf length (cm)	Petiole length (cm)	Node number
27	7.96*	5.49*	11.21*	3.26*	13.2*
43.1	7.76*	5.99*	10.26*	2.51*	10.2*
43.2	7.21*	6.09*	10.16*	2.96*	12.7*
57	9.74**	8.41 (NA)	12.49*	2.76*	11.7*
70	6.18**	3.94*	7.73*	1.56*	7.03*
74	3.37**	1.72*	4.32*	0.96**	2.3*
73	0.16	-0.21	0.86	0.71	-0.3
78	1.51	0.39	2.06	0.56	4.7 (NA)
91	6.76*	4.94*	8.36 (NA)	1.61*	9.2*

Figure 3.9. Introgressions on chromosome 2 of cabbage BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.

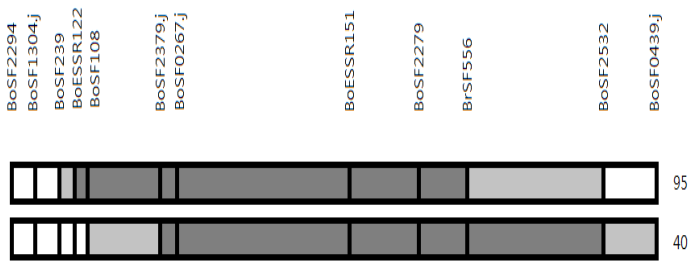
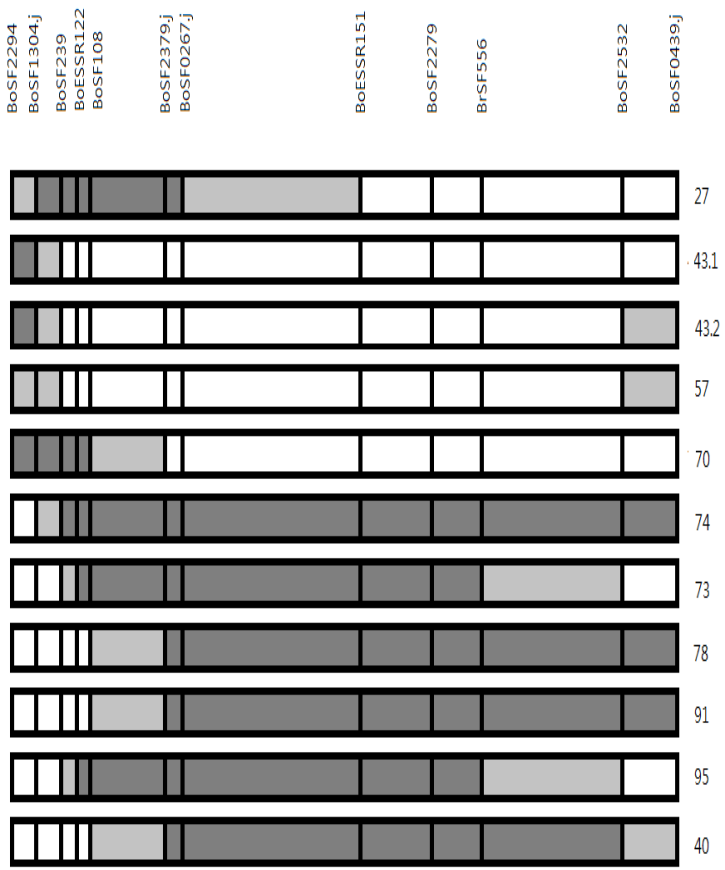


Figure 3.9. Cont.

Family	Lamina length (cm)	Lamina width (cm)	Leaf length (cm)	Petiole length (cm)	Node number
95	0.91	0.67	0.89	-0.01	0.7
40	2.24*	0.51	2.73	0.49	3.37**

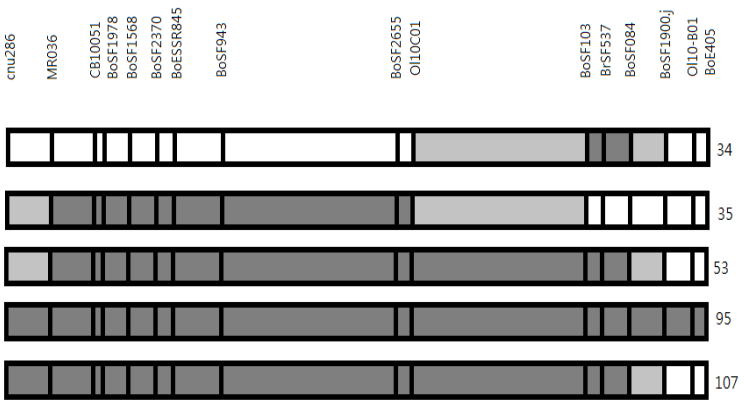


Family	Internode distance (cm)	Stem width (mm)	Bud number	Cluster width (mm)	Budding time (day)	Flowering time (day)
27	-1.92**	12.87*	8	1.31	33.6*	35.8*
43.1	-1.58*	9.06 (NA)	22	13.99*	24.6**	26.8**
43.2	-1.83**	13.4*	21	15.75*	32.6*	30.3***
57	-1.72***	13.62 (NA)	39*	11.5 (NA)	31.27*	30.47***
70	-1.57***	7.61**	23	12.72*	23.93***	24.8***
74	-0.17	2.58**	11.2*	7.49***	11.2*	9.2*
73	-0.21	0.02	4 (NA)	0.9	0.1	-3.7
78	-0.84	1.62	21.5	4.63*	17.6	13.8*
91	-0.62	7.26*	36*	NA	9.1	22.8**
95	-0.62	1.53	-1.67	3.31	4.93	6.8
40	-0.9	1.6*	11.67	4.87**	10.6	13.47*

Figure 3.9. cont.

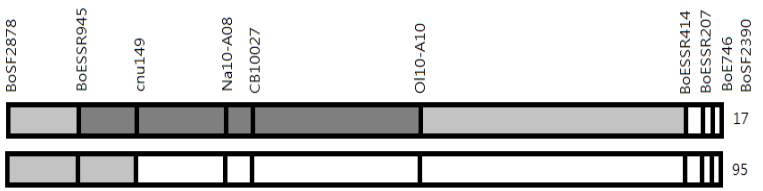
Family	Petiole length (cm)
21	1.74**
43	3.01*

Figure 3.10. Introgressions on chromosome 3 of cabbage BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.



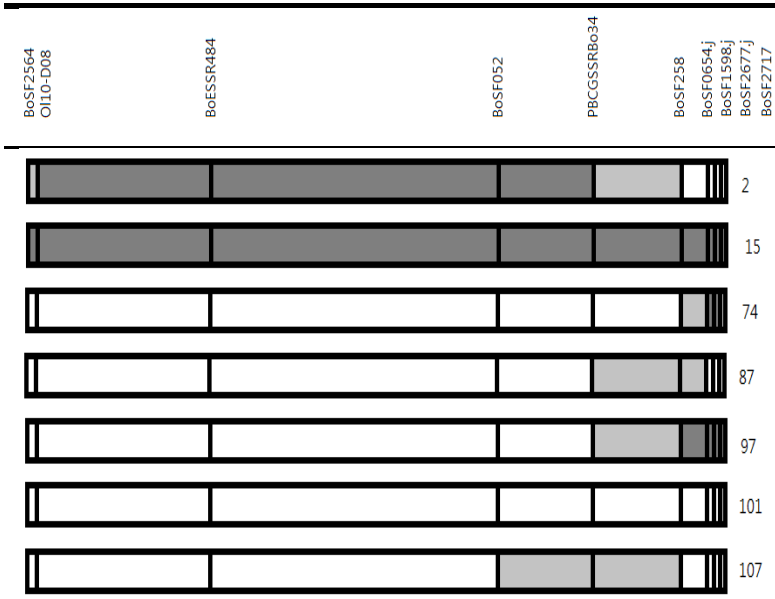
Family	Stem width (mm)
34	3.59*
35	-0.2
53	3.82*
95	2.19
107	2.4**

Figure 3.11. Introgressions on chromosome 4 of cabbage BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.



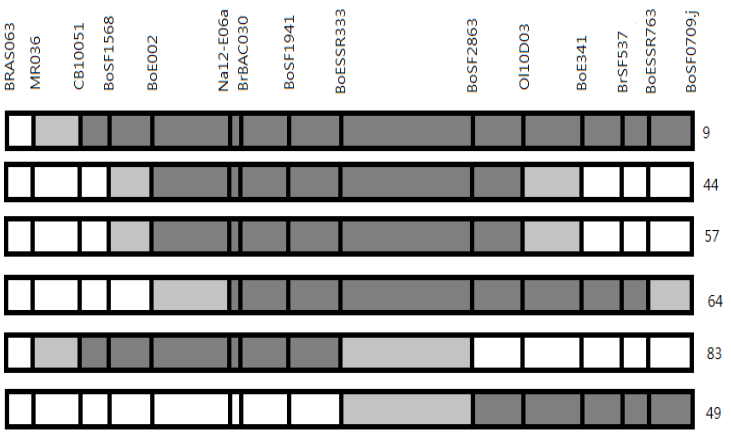
Family	Cluster width (mm)
17	11.25*
95	4.76*

Figure 3.12. Introgressions on chromosome 5 of cabbage BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.



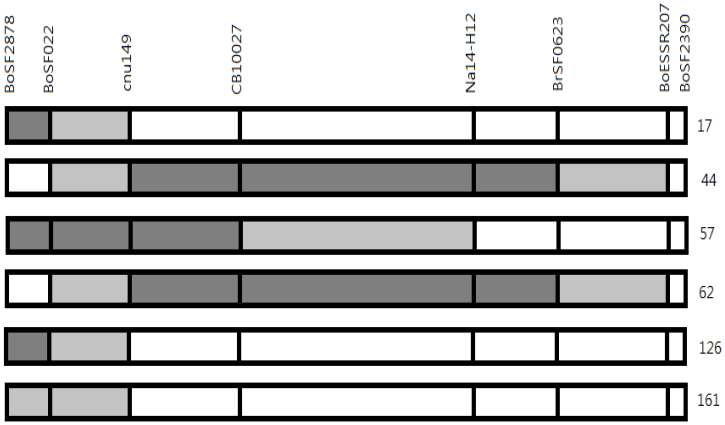
Family	Lamina width (cm)	Node number	Stem width (mm)	Cluster width (mm)	Flowering time (day)
2	1.99	4.7**	3.76	7.1**	17.8**
15	3.37*	4.03**	4.13*	7.08**	15.8***
74	1.74*	2.45*	2.57*	7.33**	9.55*
87	2.14	2.2	3.44*	4.92*	12.8***
97	-0.11	3.2*	0.96	5.07*	11.3
101	-0.11	-0.8	-0.02 (NA)	0.2	-1.2
107	0.79	2.2*	2.4**	3.91*	9.3

Figure 3.13. Introgressions on chromosome 9 of cabbage BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.



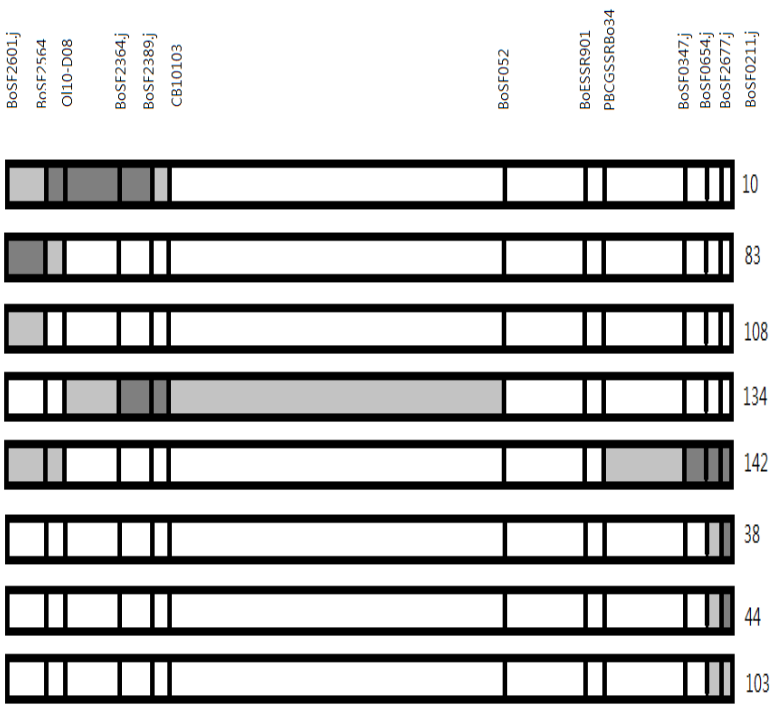
Family	Petiole length (cm)
9	0.66
44	1.66*
57	0.43
64	1.81*
83	0.96*
49	0.09

Figure 3.14. Introgressions on chromosome 4 of cauliflower BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.



Family	Petiole length (cm)	Stem width (mm)	Cluster width (mm)
17	1.71*	5.14*	7.69**
44	2.01	2.89*	9.79*
57	0.89	2.32*	7.46**
62	1.56*	2.49	5.79*
126	0.76*	3.19*	5.12**
161	0.31	0.1	0.09

Figure 3.15. Introgressions on chromosome 5 of cauliflower BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.



Family	Lamina length (cm)	Lamina width (cm)	Leaf length (cm)	Blade shape	Petiole length (cm)	Node number
10	8.94**	6.07**	12.56*	-0.05	3.63 (NA)	5.7*
83	0.81	0.51	1.89	0.77	1.13*	1.03
108	1.81	0.84	2.19	-0.11	0.39	3.37*
134	6.64**	7.02**	9.29**	0.19*	2.66**	9.95**
142	1.91	0.54	2.81	-0.16	0.91	2.7 (NA)
38	4.21*	2.44 (NA)	5.26*	-0.11	1.06	2.37*
44	2.81*	1.74	4.81*	-0.08	2.01*	0.7
103	0.21	-0.03	0.53	-0.06	0.33	1.03

Figure 3.16. Introgressions on chromosome 9 of cauliflower BC_4F_2 families. The comparisons were between introgressed individuals and the recurrent parent. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.

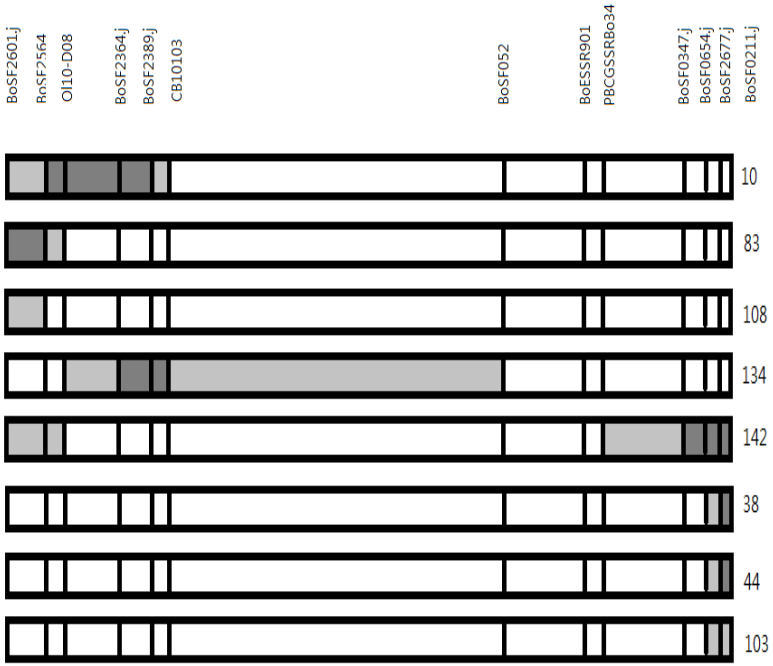
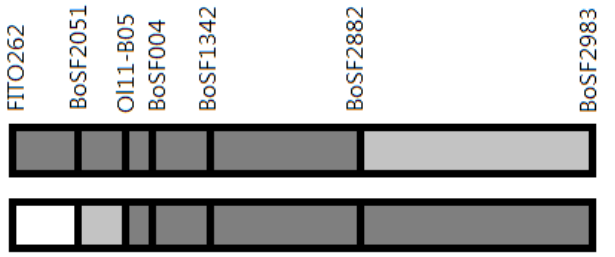


Figure 3.16.Cont.

Family	Internode distance (cm)	Stem width (mm)	Bud number	Cluster width (mm)	Budding time (day)	Flowering time (day)
10	-0.52	7.32**	29.33*	11.72*	25.6*	25.13*
83	0.28	0.99	4.67	4.44*	1.6	1.47
108	-0.15	1.12	17	4.12*	10.93	9.8
134	-1.49**	9.92 (NA)	24.25*	11.27 (NA)	34.1**	31.8***
142	-0.01	1.89	13	4.64*	14.6 (NA)	9.3
38	-0.04	2.06*	12.33	6.02**	8.93	8.47
44	-0.23	2.89*	17.5	9.79***	8.1	6.8
103	-0.05	0.49	2	3.22	1.6	2.13



Family	Blade shape
111-11	0.18*
17-8	0.21*

Figure 3.17. Introgressions on chromosome 3 of cauliflower BC_4F_3 families. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.

Family	Bud number
57-10	14.6
9-12	30.93**

Figure 3.18. Introgressions on chromosome 4 of cauliflower BC_4F_3 families. Effect of introgression was reported by the difference between the mean of introgressed individuals and the recurrent parent. Significance was reported by *, with p-value<0.05 as *, <0.01 as **, <0.001 as ***. Grey indicates recombination region and darker grey indicates introgression.

CHAPTER 4

CONCLUSION

To dissect the genomes of two *B. oleracea* morphotypes, two backcross populations were developed using a rapid-cycling genotype as a common recurrent parent, and their genotypes were determined in the BC_4F_1 generation by genotyping-by-sequencing (GBS). To filter informative markers from the background noise accompanying GBS, a SNP algorithm was devised. Based on the SNP markers, we were able to identify introgressions from the donor parents and reveal the genome-wide introgression pattern in the two populations. Introgressions covered most of the genome and were over-represented in some regions, which might because of an advantageous effect toward plant fertility carried by the introgressions. Some introgressions that remained intact over four cycles of backcrossing were on the pericentromeric regions of several chromosomes, and on one arm of chromosome 6. Among these regions, the former had been considered a recombination 'cold spot' while the latter might be due to its rich sequence divergence. Based on the genotypes, we could select near isogenic lines from BC_4F_1 families with small numbers of introgressions.

The morphologies of the two backcross populations were evaluated in the field for two seasons. Flower color and 14 leaf-, stem-, and flower-traits were segregating in the two populations, based on which we found 219 marker-trait associations. Striking phenotypic variance was observed between the two seasons, which suggested the

environmental sensitivity of the genotypes, especially regarding flowering time. Two comparisons, between introgressed individuals and (a) the recurrent parent and (b) non-introgressed individuals, were used in finding the marker-trait associations. The former comparison might be more prone to false-positive results with the presence of multiple introgressed segments. To reach higher mapping power and finer resolution, more introgressed families of a more advanced generation and more replicates will be required.