Homoeologous Recombination in *Brassica napus*

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy
In the Department of Biology
University of Saskatchewan
Saskatoon

BY

Erin E. Higgins

Permission to Use

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of Department of Biology
University of Saskatchewan
112 Science Place
Saskatoon, Saskatchewan S7N 5E2 Canada

OR

Dean
College of Graduate and Postdoctoral Studies University of Saskatchewan
116 Thorvaldson Building, 110 Science Place Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Polyploidy is very common among plants but having multiple sets of chromosomes creates additional challenges for chromosome pairing and recombination during meiosis. *Brassica napus* is an allotetraploid comprised of the A and C genomes from *B. rapa* and *B. oleracea*, respectively. In adapted *B. napus* lines, the chromosomes of the A and C genomes pair almost exclusively with their true homologue during meiosis, but in newly resynthesized *B. napus* plants there is a significant increase in pairing between homoeologues, the closely related chromosomes from the other genome, i.e. A/C pairings. This interaction between homoeologues can result in an uneven distribution of chromosomes in the gametes so restricting pairing is important for overall plant fitness. However, this unequal crossing over can also serve to introduce novel variation allowing for species diversification and adaptation.

I have developed a new method for detecting homoeologous recombination events in *B. napus* using a single nucleotide polymorphism (SNP) array. Traditionally scientists have used cytology or restriction fragment length polymorphism markers but the SNP array offers much quicker data generation and a higher density of markers for more precise identification of crossover points and detection of smaller exchanges. Using this new methodology I measured recombination between the A and C genomes in natural *B. napus* lines and have shown that homoeologous exchanges continue to happen in modern *B. napus* cultivars at a relatively high frequency.

I have also used the SNP array to analyze a resynthesized *B. napus* population segregating for the level of homoeologous recombination and mapped three quantitative trait loci (QTL) controlling this phenomenon. Further analysis of the genes underlying these QTL can help to identify the mechanisms that have evolved in natural *B. napus* to control meiotic chromosome pairing and manipulation of those genes could be used to increase homoeologous recombination rates to introduce novel traits from diverse species.

Acknowledgements

First and foremost, thank you to Dr. Isobel Parkin for providing me the opportunity to undertake this project, for your patience with how long it's taken and your friendship and support during this process.

Thank you to the members of my committee, Drs Pierre Fobert, Jitao Zou, Carlos Carvalho, and Curtis Pozniak for your guidance and input.

Thank you to Drs Elaine Howell and Sue Armstrong from the University of Birmingham for the cytological work to complement the genetic marker analysis in this thesis and their perseverance to get this work published.

Thank you to the AAFC Parkin lab, Robinson group, second floor, and coffee crew. It's been a very long journey and I couldn't have done this without your support. Bioinformatics assistance from Wayne Clarke is greatly appreciated and thank you to the greenhouse staff for their excellent care of my plant material.

Thank you to my friends and family for resisting the urge to ask, "When will you be finished?". For your support and patience I will be forever grateful.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAFC	Agriculture and Agri-Food Canada
ADT	Assay design tool
ASK	Arabidopsis SKP-like
ASY	Asynaptic
At	Arabidopsis thaliana
BAC	Bacterial Artificial Chromosome
BLAT	Basic local alignment tool
Bna	Brassica napus
BnaPh1	Brassica napus Ph1
bp	basepair
BRCA	Breast cancer like
Cas9	CRISPR associated protein
CDK	Cyclin dependent kinase
CENP	Centromere protein
cM	Centimorgan
CO	Crossover
Col	Columbia
CRISPR	Clustered regularly interspaced short palindromic repeats
CTAB	cetylcetyltrimethylammonium bromide
CTF	Chromosome transmission fidelity
DFO	DSB Formation
dHj	Double Holliday junction
DH	Doubled haploid
DMC	Disruption of meiotic control
DNA	Deoxyribonucleic Acid
DSB	Double Stranded Break
dsDNA	double-stranded DNA
EME	Essential meiotic endonuclease
EMS	Ethyl methanesulfonate
ERCC	Excision Repair
FANCM	Fanconi anemia complementation group
FISH	Fluorescence in situ hybridisation
Gb	Gigabase
GBS	Genotyping-by-sequencing
GISH	Genomic in situ hybridisation

GSL	Glucosinolate
GWAS	Genome Wide Association Study
HEI	Homolgue of enhancer of cell invasion
HeR	Homoeologous Recombination
Hj	Holliday junction
HNRT	Homoeologous non-reciprocal translocation
HOP	Homologous pairing protein
KASH	Klarsicht/ANC-1/Syne-1 homology
Kb	Kilobase
KO	Knockout
LD	Linkage disequillibrium
Ler	Landsberg erecta
LOD	Likelihood of odds
Mb	Megabase
MCM	Minichromosome maintenance
MEI	Meiosis defective
MET	Methyltransferase
MI	Metaphase I
MLH	MutL homologue
MRE	Meiotic recombination gene
MRN	MRE-RAD-NBS
MSH	MutS homologue
MUS	MMS and UV sensitive
NBS	Nijmegen breakage syndrome
NCBI	National Centre for Biological Information
NCO	Non-crossover
NMD	Nonsense mediated decay
Os	Oryza sativa
PE	Paired-end
Ph1	Pairing homoeologous 1
PrBn	Pairing regulator in Brassica napus
PRD	Putative recombination initiation defective
PSS	Separation of sister chromatids
PTD	Part time dancers
QTL	Quantitative Trait Locus
RAD	RAD51 Recombinase
RCK	Rock-n-rollers
rDNA	ribosomal DNA
RFC	Replication factor C

RFLP	Restriction Fragment Length Polymorphism
RNAi	RNA interfering
RNAi	Ribonucleic acid
RNAseq	RNA sequencing
RPA	Replication Protein A
RTR	RecQ-Topoisomerase-RMI1 Complex
SC	Synaptonemal Complex
SCC	Sister chromatid cohesion
SDS	Solo dancers
SDSA	Synthesis dependent strand annealing
SEND	Structure specific endonuclease
SGO	Shugoshin
SHOC	Shortage in chiasmata
SMC	structural maintenance of chromosomes
SMRT	Single Molecule Real Time
SNP	Single Nucleotide Polymorphism
SPO	Sporulation specific protein
SPS	Synaptic Partner Switch
SPSC	Synaptic Partner Switch accompanied by at least one crossover
ssDNA	single-stranded DNA
SSR	Simple Sequence Repeat
SUN	Sad1-UNc84 homology
SV	Structural Variant
SWI	Switch protein
SYN	Synaptic
TOP	Topoisomerase
VIGS	Virus induced gene silencing
WGD	Whole Genome Duplication
WGS	Whole Genome Sequencing
wt	Wildtype
XRCC	X-ray repair cross complementing
XRS	X-ray sensitive
ZIP	Zipper protein
ZMM	ZIP-MSH-MER Complex

1. Literature Review

1.1 Brassica

1.1.1 Brassica genetics

The Brassicaceae is a large family comprised of >3,800 species which grow on six continents in every type of environment. The family is highly diverse and includes common, well known vegetables and oilseeds, weedy pests and isolated species endemic to a single island. The vegetable and oilseed Brassica varieties are of great agricultural importance across the world, in particular *B. napus* (canola or oilseed rape) which is widely grown in both the Northern and Southern hemispheres and is harvested primarily for its oil.

The genetic relationship between the six major Brassica species was originally described by U (1935). This model, called the "Triangle of U" identifies the three Brassica diploids, *B. rapa* (A genome), *B. nigra* (B genome), *B. oleracea* (C genome) and the three allotetraploids formed by the pairwise combinations, *B. napus* (AACC), *B. juncea* (AABB), and *B. carinata* (BBCC) (Figure 1-1). *Brassica napus* is a relatively young polyploid, believed to have formed within the last 10,000 years from a small number of hybridizations, likely in the Mediterranean region (CHALHOUB *et al.* 2014). The diploid progenitors, *B. rapa* and *B. oleracea,* which are grown primarily as vegetable crops have high genetic and morphological diversity including turnip, Chinese cabbage, broccoli, cauliflower, kale and many others (DIXON 2006).

Early genetic maps for *B. napus* made using Restriction Fragment Length Polymorphism (RFLP) markers identified the primary homoeologues of the A and C genomes (PARKIN *et al.* 1995; SHARPE *et al.* 1995). It was clear from these maps that some of the chromosomes have a single homoeologue with nearly identical marker order while others have multiple homoeologues and chromosomes have been rearranged or inverted relative to one another (Figure 1-2).

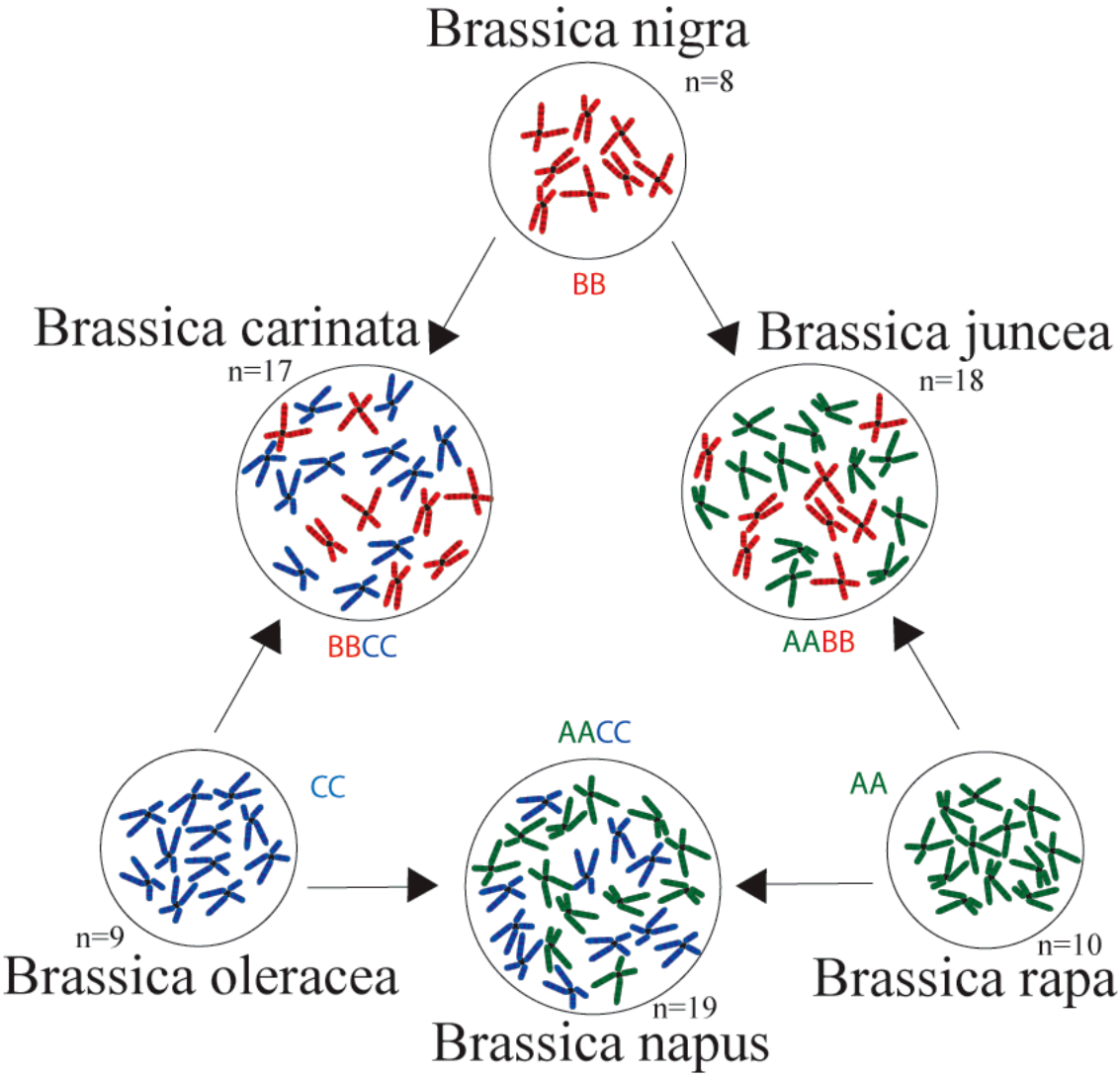


Figure 1-1: The Triangle of U
Genetic relationship between the six major Brassica species. The diploids, *B. rapa, B. nigra* and *B. oleracea* which have the A, B and C genomes, respectively, and the allotetraploids created by their hybridization, *B. juncea* (A and B genomes), *B. napus* (A and C genomes, and *B. carinata* (B and C genomes). (Image from Wikipedia)

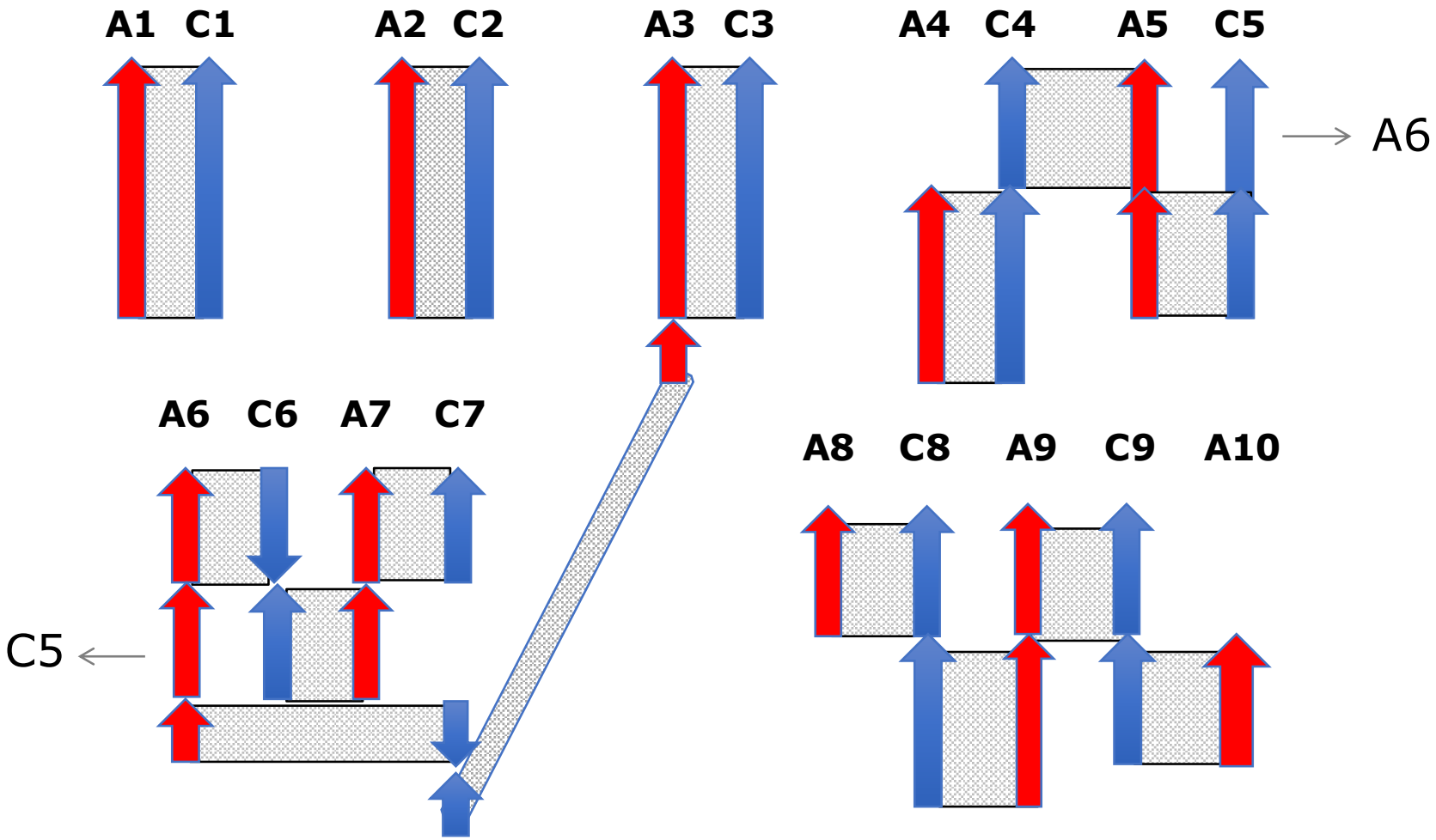


Figure 1-2: Alignment of A and C genomes of B. napus
Alignment of the homoeologous chromosomes in *B. napus,* the 10 A genome chromosomes are shown in red and the 9 C genome chromosomes are shown in blue. Some chromosome pairs are completely aligned (A1/C1) and others have multiple homoeologues and large rearrangements (A6). Directional arrows are based on RFLP marker order.

Due to its economic importance, there are extensive public and private breeding programs aimed at increasing *B. napus* productivity and quality (INIGUEZ-LUY AND FEDERICO 2011; SNOWDON AND INIGUEZ LUY 2012). However the limited number of hybridisation events that led to the formation of *B. napus* has created a very limited amount of genetic diversity within the *B. napus* genepool. The high level of diversity particularly within the diploids *B. rapa* and *B. oleracea* but also in other Brassica species would be a significant source of variation if traits of interest could be incorporated into *B. napus* varieties.

1.1.2 Brassica genomics

The development of cheap, accurate, high-throughput sequencing has created a new era of genomics in all aspects of research, including agriculture. The first major Brassica genome published was *B. rapa* in 2011 (WANG *et al.* 2011), followed by *B. oleracea* (PARKIN *et al.* 2014) and the first *B. napus* genome was published in 2014 (CHALHOUB *et al.* 2014). These genome assemblies were created using 2nd generation short read (32-125bp) sequences. Genome assembly is difficult with short-read data because it is difficult to correctly assemble low complexity regions or to uniquely assign reads to a genomic position in highly duplicated genomes, which is particularly problematic in plants which have gone through several whole genome duplication (WGD) throughout evolutionary history (MASTERSON 1994).

However, this highly accurate short read data was particularly useful for identifying single nucleotide polymorphisms (SNPs) which can be easily converted to genetic markers. These were a major improvement in genetic mapping because unlike Simple Sequence Repeats (SSRs) which are based on repeats and are therefore difficult to have in genic regions, SNPs were identified throughout the genome making them easier to closely associate with traits of interest in breeding programs. The further development of high-density arrays has made it possible to quickly and accurately identify variation within a population for genome-wide association studies (GWAS), QTL mapping, or diversity analysis (MASON *et al.* 2017). Initially, density was limited

to only 100's or 1000's of markers on each array, but today arrays up to 1 Million SNPs are routinely used for human studies.

The development of new long-read sequencing technology is once again changing plant genomics. These new 3rd generation technologies include Oxford Nanopore Sequencing which is capable of producing individual reads of >1 Mb and routinely produces reads >100 Kb, and PacBio Single Molecule Real Time (SMRT) which accurately sequences molecules up to 50 Kb. In addition, optical mapping and chromosome conformation capture technologies have greatly improved scaffolding and chromosome assembly and several Brassica genome assemblies have been improved using these new technologies (BELSER *et al.* 2018; ZHANG *et al.* 2018; PERUMAL *et al.* 2020).

1.2 Meiosis

1.2.1 Overview

Meiosis is the process by which all sexually reproducing organisms form haploid gametes from diploid progenitor cells through a clearly defined process of a single round of DNA replication followed by two reductive divisions, Meiosis I and Meiosis II. The two rounds of cell division are each divided into four distinct stages: Prophase, Metaphase, Anaphase and Telophase. Prophase I is the longest of these and is further divided into leptotene, zygotene, pachytene, diplotene and diakinesis. During Meiosis I, homologous chromosomes pair, double stranded breaks (DSBs) form, crossing over occurs linking the two homologous chromosomes, and finally the pair is separated. In meiosis II the sister chromatids are separated resulting in the formation of four daughter cells each with a haploid set of chromosomes. The faithful segregation of the replicated homologous chromosomes during Meiosis I is essential for producing viable gametes containing a complete haploid set of chromosomes.

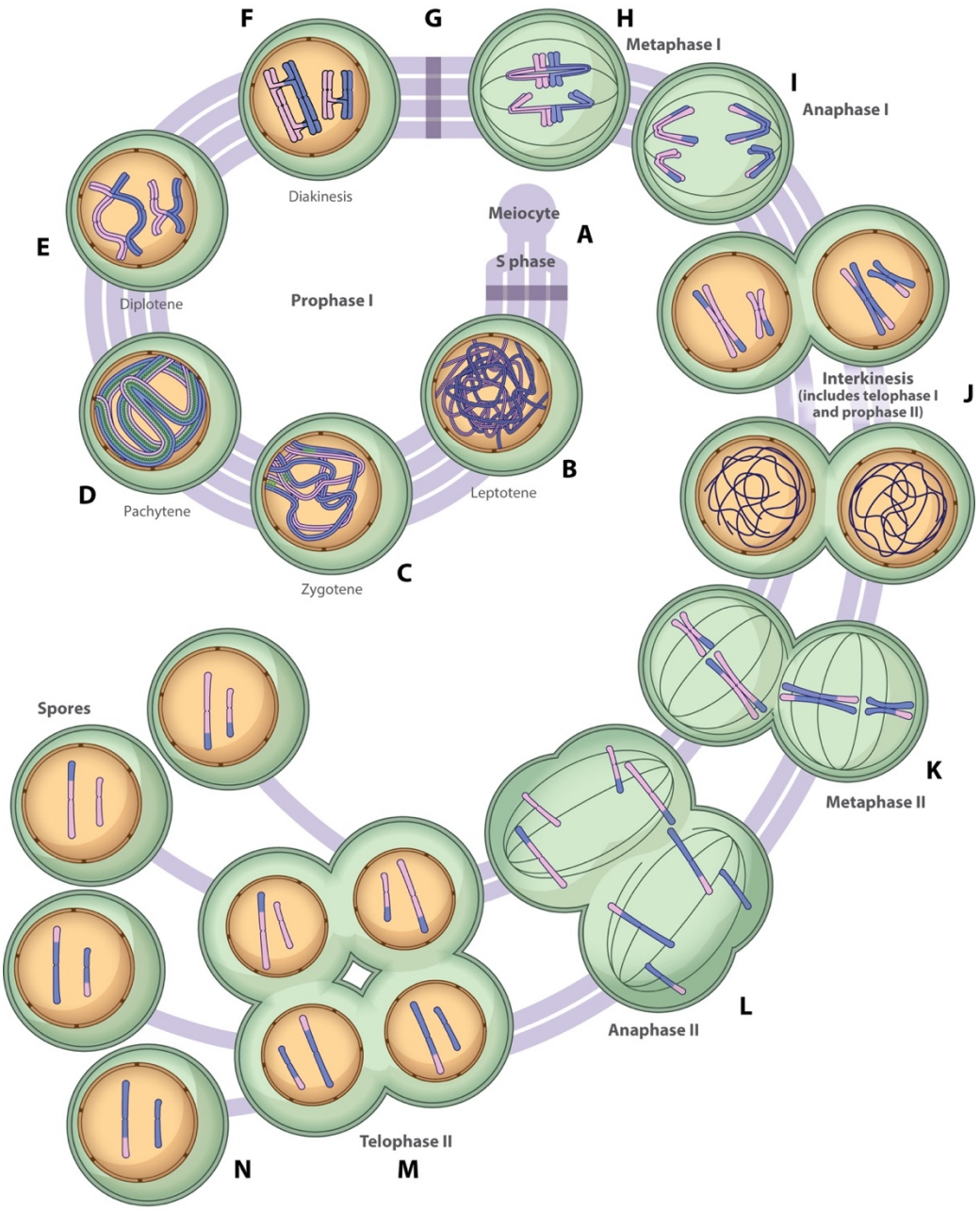


Figure 1-3: Overview of Meiosis

A) Premeiosis – meiocyte differentiation and meiotic S phase **B)** Leptotene – formation of chromosome axes and initiation of recombination **C)** Zygotene – synaptonemal complex forms and recombination progresses **D)** Pachytene – completion of synapsis and further progression of recombination **E)** Diplotene – synaptonemal complex disassembles, homologous chromosomes are now connected by crossovers **F)** Diakinesis – chromosomes condense and bivalents can be distinguished **G)** Prophase I completes and the nuclear envelope breaks down **H)** Metaphase I – bivalents align along metaphase plate **I)** Anaphase I – chromosomes migrate to two poles **J)** Interkinesis including telophase I and prophase II – two nuclei form, chromosomes briefly decondense **K)** Metaphase II – chromosomes align on metaphase plate **L)** Anaphase II – cohesion release and separation of sister chromatids **M)** Telophase II – four nuclei form **N)** Cytokinesis – haploid spores are released. Image from Mercier et al (2015).

Cytological observation of cells undergoing meiosis has provided a detailed picture of chromosome behavior at each stage of the process (Figure 1-3). During the pre-meiotic S-phase of the cell cycle, cohesion is established between sister chromatids by loading of the cohesion complex. A proteinaceous axis extends along the chromatids at the end of the G2 phase forming a structure where the chromatids are joined at the loop bases. Following this, DSBs are created along the chromosome arms and the homologous chromosomes start to align. Formation of the synaptonemal complex (SC) begins during zygotene and by pachytene the SC binds the chromosome pairs along their entire length and homologous recombination proceeds. Once recombination is complete, the homologous chromosomes begin to condense and the SC breaks down. At the end of prophase I the homologous chromosomes are linked at chiasmata, the sites where recombination has occurred.

The chromosome pairs align along the metaphase plate and attach to the spindle apparatus. During anaphase I the chromosomes separate to opposite poles but cohesion between the sister chromatids is maintained. This first meiotic division is followed by meiosis II which involves a mitotic-like division during which cohesion between sister chromatids is lost and haploid gametes are formed. The meiotic processes have been the focus of intense study for over a century and much is now known about the genes that control chromosome pairing, recombination and segregation. A summary of the meiosis-related genes identified in Arabidopsis are listed in Table 1-1 (KAUR *et al.* 2006; MA 2006; GEUTING *et al.* 2009; BAUKNECHT AND KOBBE 2014; DE *et al.* 2014; OH *et al.* 2014; LARIO *et al.* 2015; MERCIER *et al.* 2015; WRIGHT *et al.* 2015; VRIELYNCK *et al.* 2016; BOLANOS-VILLEGAS *et al.* 2018; CHAMBON *et al.* 2018; FERNANDES *et al.* 2018; ROHRIG *et al.* 2018).

Table 1-1: Known Meiotic Genes in Arabidopsis thaliana

AT gene code	Gene symbol	Description	Copy No. in B. napus
AT1G01690	PRD3, ATPRD3	ARABIDOPSIS THALIANA PUTATIVE RECOMBINATION INITIATION DEFECTS 3, putative recombination initiation defects 3	4
AT1G01880	AtGEN1, GEN1	ortholog of HsGEN1	2
AT1G04650	FLIP	FIDGETIN-LIKE-1 INTERACTING PROTEIN	4
AT1G05180	AXR1	AUXIN RESISTANT 1	6
AT2G32410	AXL1, AXL		2
AT1G06660	JASON	JASON	4
AT1G07060	DFO, ATDFO	DSB formation	2
AT1G07745	RAD51D, SSN1, ATRAD51D	SUPPRESOR OF SNI1, homolog of RAD51 D	2
AT1G08880	H2AXA, G-H2AX, GAMMA-H2AX, HTA5	GAMMA H2AX, histone H2A 5, gamma histone variant H2AX	5
AT1G54690	H2AXB, HTA3, G-H2AX, GAMMA-H2AX	histone H2A 3, GAMMA H2AX, gamma histone variant H2AX	0
AT1G10710	PHS1	POOR HOMOLOGOUS SYNAPSIS 1	2
AT1G10930	ATSGS1, RECQ4A, ATRECQ4A		4
AT1G60930	RECQL4B, ATRECQ4B, RECQ4B	RECQ helicase L4B, RECQ HELICASE L4B	2
AT1G10940	SRK2A, SNRK2-4, ASK1, SNRK2.4	SNF1-related protein kinase 2.4, SUCROSE NONFERMENTING 1-RELATED PROTEIN KINASE 2-4, ARABIDOPSIS SERINE/THREONINE KINASE 1	5
AT1G60940	SNRK2.10, SNRK2-10, SRK2B |	SNF1-related protein kinase 2.10, SUCROSE NONFERMENTING 1-RELATED PROTEIN KINASE 2-10, SNF1-RELATED KINASE 2B	3
AT1G11060	WAPL1, AtWAPL1	Wings apart-like protein 1	4
AT1G61030	AtWAPL2, WAPL2	Wings apart-like protein 2	0
AT1G12790	PTD	Class I CO pathway	2
AT1G13330	HOP2, AHP2	Arabidopsis Hop2 homolog, HOMOLOGOUS-PAIRING PROTEIN 2	2

AT1G14750	SDS	SOLO DANCERS	6
AT1G15570	CYCA2;3	CYCLIN A2;3	6
AT1G80370	CYCA2;4	Cyclin A2;4	6
AT1G16330	CYCB3;1	cyclin b3;1	4
AT1G22260	AtZYP1a, ZYP1, ZYP1a		2
AT1G22275	ZYP1, ZYP1b		2
AT1G27900	no symbol available	RNA helicase family protein	2
AT1G29400	AML5, ML5	MEI2-like protein 5	4
AT1G32310	SAMBA		2
AT1G34355	ATPS1, PS1	PARALLEL SPINDLE 1	6
AT1G35530	FANCM, At-FANCM	Fanconi anemia complementation group M	3
AT1G48360	FAN1	Fanconi/FANCD2 associ- ated nuclease I	2
AT1G50240	TIO, FU	TWO IN ONE, FUSED	4
AT1G53490	HEI10	homolog of human HEI10 (Enhancer of cell Invasion No.10)	5
AT1G60460	MTOPVIB	meiotic topoisomerase VIB-like	2
AT1G63990	SPO11-2	sporulation 11-2	2
AT1G66170	DUET, MMD1	DUET, MALE MEIOCYTE DEATH 1	2
AT1G67370	ATASY1, ASY1	ASYNAPTIC 1	3
AT1G75950	UIP1, ASK1, SKP1A, ATSKP1, SKP1 ARABIDOPSIS SKP1 HOMOLOGUE 1, UFO INTERACTING PROTEIN 1, S phase kinase-associated protein 1	ARABIDOPSIS SKP1 HOMOLOGUE 1, UFO INTERACTING PROTEIN 1, S phase kinase-associated protein 1	7
AT1G77320	MEI1	meiosis defective 1	1
AT1G77390	DYP, CYCA1, CYCA1;2, TAM	CYCLIN A1;2, TARDY ASYNCHRONOUS MEIOSIS, CYCLIN A1	4
AT1G77600	AtPDS5B, PDS5B		2

AT1G78790	AtMHF2, MHF2	Arabidopsis homolog of mammalian MHF2	2
AT2G06510	RPA1A, RPA70A, ATRPA70A, ATRPA1A	replication protein A 1A, RPA70-KDA SUBUNIT A, ARABIDOPSIS THALIANA RPA70-KDA SUBUNIT A, ARABIDOPSIS THALIANA REPLICATION PROTEIN A 1A	2
AT2G21800	ATEME1A, EME1A	essential meiotic endonuclease 1A	2
AT2G22140	ATEME1B, EME1B	essential meiotic endonuclease 1B	2
AT2G27170	TTN7, SMC3	TITAN7, STRUCTURAL MAINTENANCE OF CHROMOSOMES 3	4
AT2G28560	ATRAD51B, RAD51B		2
AT2G31970	ATRAD50, RAD50		2
AT2G33793	ASY4	ASYNAPTICA4	4
AT2G42890	ML2, AML2	MEI2-like 2	4
AT2G45280	ATRAD51C, RAD51C	RAS associated with diabetes protein 51C	2
AT2G46980	AtASY3, ASY3	Arabidopsis thaliana ASYNAPTIC 3, ASYNAPTIC 3	2
AT2G47980	SCC3, ATSCC3	sister-chromatid cohesion protein 3, SISTER-CHROMATID COHESION PROTEIN 3	2
AT3G02680	ATNBS1, NBS1	nijmegen breakage syndrome 1	2
AT3G02980	MCC1	MEIOTIC CONTROL OF CROSSOVERS1	1
AT3G05480	RAD9, ATRAD9		2
AT3G09660	AtMCM8, MCM8	minichromosome maintenance 8	2
AT3G10440	SGO1, AtSGO1	SHUGOSHIN 1	2
AT3G12280	ATRBR1, RBR1, RB1, RB, RBR	RETINOBLASTOMA-RELATED, RETINOBLASTOMA 1, RETINOBLASTOMA-RELATED PROTEIN 1, retinoblastoma-related 1	6
AT3G13170	ATSPO11-1		2
AT3G14190	PANS1, CMR1	COPPER MODIFIED RESISTANCE 1, PATRONUS 1	3
AT3G18524	MSH2, ATMSH2	MUTS homolog 2	2
AT3G19210	CHR25, ATRAD54, RAD54	homolog of RAD54	2
AT3G20475	ATMSH5, MSH5	MUTS-HOMOLOGUE 5, MUTS-homologue 5	2
AT3G22880	ATDMC1, ARLIM15, DMC1	DISRUPTION OF MEIOTIC CONTROL 1, ARABIDOPSIS HOMOLOG OF LILY MESSAGES INDUCED AT MEIOSIS 15, ARABIDOPSIS THALIANA DISRUPTION OF MEIOTIC CONTROL 1	2

AT3G24495	ATMSH7, MSH6-2, MSH7	MUTS homolog 7, ARABIDOPSIS THALIANA MUTS HOMOLOG 7, MUTS HOMOLOG 6-2	2
AT3G25100	CDC45	cell division cycle 45	4
AT3G26890	*no symbol available*	*Meiosis chromosome segregation family protein*	*4*
AT3G27120	FIG1/FIDGETIN-LIKE 1	AAA-ATPase	2
AT3G27730	MER3, RCK	ROCK-N-ROLLERS	2
AT3G33520	SUF3, ARP6, ESD1, ATARP6	SUPPRESSOR OF FRI 3, EARLY IN SHORT DAYS 1, actin-related protein 6	2
AT3G43210	NACK2, TES, ATNACK2	ARABIDOPSIS NPK1-ACTIVATING KINESIN 2, NPK1-ACTIVATING KINESIN 2, TETRASPORE	2
AT3G47460	CAP-E2	SMC2 homolog	4
AT3G48190	ATATM, PIG1, ATM	ARABIDOPSIS THALIANA ATAXIA-TELANGIECTASIA MUTATED, ataxia-telangiectasia mutated, pcd in male gametogenesis 1	2
AT3G48750	CDC2AAT, CDKA;1, CDK2, CDC2, CDKA1, CDC2A	cell division control 2	5
AT3G48900	AtSEND1, SEND1		2
AT3G52115	ATGR1, ATCOM1, GR1, COM1	gamma response gene 1	2
AT3G54670	TITAN8	ATSMC1, SMC1, STRUCTURAL MAINTENANCE OF CHROMOSOMES 1, TITAN8, TTN8	5
AT3G57300	INO80, ATINO80	INO80 ortholog, INO80 ORTHOLOG	4
AT3G57860	UVI4-LIKE, GIG1, OSD1	GIGAS CELL 1, UV-B-insensitive 4-like, OMISSION OF SECOND DIVISION	2
AT3G59550	ATSYN3, ATRAD21.2, SYN3	SISTER CHROMATID COHESION 1 PROTEIN 3	3
AT3G63480	AtPSSI, KIN-1	KINESIN 1	0
AT4G00020	BRCA2(IV), EDA20, BRCA2A, MEE43	BREAST CANCER 2 like 2A, EMBRYO SAC DEVELOPMENT ARREST 20, MATERNAL EFFECT EMBRYO ARREST 43	0
AT5G01630	BRCA2(V), ATBRCA2(V), BRCA2B	BRCA2-like B	3
AT4G01370	MPK4, ATMPK4, MAPK4	MAP kinase 4	3
AT4G02070	ATMSH6, MSH6-1, MSH6	MUTS HOMOLOG 6-1, ARABIDOPSIS THALIANA MUTS HOMOLOG 6, MUTS homolog 6	2

AT4G05190	ATK5, KINESIN 5	Homology to KATA	2
AT4G09140	ATMLH1, MLH1	MUTL-homologue 1, ARABIDOPSIS THALIANA MUTL-HOMOLOGUE 1	2
AT4G14180	PRD1, AtPRD1	putative recombination initiation defect 1	2
AT4G14220	RHF1A	RING-H2 group F1A	2
AT4G14970	ATFANCD2, FACND2, FANCONI ANEMIA D2	required for meiotic homologous recombination	2
AT4G17380	ATMSH4, MSH4	MUTS HOMOLOG 4, ARABIDOPSIS MUTS HOMOLOG 4, MUTS-like protein 4	2
AT4G18120	AtML3	Pseudogene? MEI2-like 3	2
AT4G20900	MS5, TDM1	MALE-STERILE 5	2
AT4G21270	KATA, KATAP, ATK1	KINESIN-LIKE PROTEIN IN ARABIDOPSIS THALIANA A, KINESIN-LIKE PROTEIN IN ARABIDOPSIS THALIANA A PROTEIN, kinesin 1	2
AT4G22910	FZR2, CCS52A1	cell cycle switch protein 52 A1, FIZZY-related 2	4
AT4G11920	CCS52A2, FZR1	FIZZY-RELATED 1, cell cycle switch protein 52 A2	4
AT4G22970	RSW4, ESP, AESP	EXTRA SPINDLE POLES, RADIALLY SWOLLEN 4, homolog of separase	4
AT4G24710	PCH2/CRC1	pachytene checkpoint protein	2
AT4G25540	MSH3, ATMSH3	homolog of DNA mismatch repair protein MSH3	2
AT4G29170	ATMND1		2
AT4G30870	MUS81, ATMUS81	ARABIDOPSIS THALIANA MMS AND UV SENSITIVE 81, MMS AND UV SENSITIVE 81	2
AT4G31400	CTF7, ECO1, AtCTF7	CHROMOSOME TRANSMISSION FIDELITY 7	4
AT4G33270	AtCDC20.1, CDC20.1	cell division cycle 20.1	6
AT4G35520	ATMLH3, MLH3	MUTL protein homolog 3	2
AT5G05490	REC8, SYN1, DIF1, AtREC8	DETERMINATE, INFERTILE 1, SYNAPTIC 1	2
AT5G07290	AML4, ML4 MEI2-like 4	MEI2-like 4	2
AT5G07660	SMC6A	structural maintenance of chromosomes 6A	4
AT5G08110	HRQ1	Homologous to RecQ helicase	2
AT5G13840	CCS52B, FZR3	FIZZY-related 3, cell cycle switch protein 52 B	2
AT5G15540	EMB2773, ATSCC2, SCC2	EMBRYO DEFECTIVE 2773, ARABIDOPSIS THALIANA SISTER-CHROMATID COHESION 2, SISTER-CHROMATID COHESION 2	4

AT5G15920	EMB2782, AtSMC5, SMC5	EMBRYO DEFECTIVE 2782, structural maintenance of chromosomes 5	2
AT5G16270	SYN4, ATRAD21.3	sister chromatid cohesion 1 protein 4, ARABIDOPSIS HOMOLOG OF RAD21 3	6
AT5G19400	SMG7		4
AT5G20850	ATRAD51, RAD51		4
AT5G21150	AGO9	ARGONAUTE 9	6
AT5G22000	RHF2A	RING-H2 group F2A	6
AT5G22010	AtRFC1, RFC1	replication factor C1, replication factor C 1	4
AT5G22750	RAD5A		2
AT5G24280	GMI1	GAMMA-IRRADIATION AND MITOMYCIN C INDUCED 1	4
AT5G25380	CYCA2;1	cyclin a2;1	3
AT5G40820	ATR, ATRAD3, ATATRA	taxia telangiectasia-mutated and RAD3-related, ARABIDOPSIS THALIANA ATAXIA TELANGIECTASIA-MUTATED AND RAD3-RELATED	2
AT5G41110	MEE6.18	Meiosis chromosome segregation family protein	6
AT5G43530	RAD5B		2
AT5G45400	RPA70C, RPA1C, ATRPA70C	Replication Protein A 1C	4
AT5G47690	PDS5A, PDS5, AtPDS5A		4
AT5G48390	ATZIP4, AtSPO22	A. thaliana homologue of yeast SPO22	2
AT5G48720	XRI1, XRI	X-RAY INDUCED TRANSCRIPT 1, X-RAY INDUCED TRANSCRIPT	6
AT5G50930	MHF1, AtMHF1	homolog of human MHF1	2
AT5G51330	DYAD, SWI1	SWITCH1	4
AT5G52290	SHOC1	SHORTAGE IN CHIASMATA 1	2
AT5G54260	MRE11, ATMRE11	ARABIDOPSIS MEIOTIC RECOMBINATION 11, MEIOTIC RECOMBINATION 11	2
AT5G57450	ATXRCC3, XRCC3	ARABIDOPSIS THALIANA HOMOLOG OF X-RAY REPAIR CROSS COMPLEMENTING 3 (XRCC3), homolog of X-ray repair cross complementing 3	2
AT5G57880	PRD2, ATPRD2, MPS1	ARABIDOPSIS THALIANA PUTATIVE RECOMBINATION INITIATION DEFECTS 2, PUTATIVE RECOMBINATION INITIATION DEFECTS 2, MULTIPOLAR SPINDLE 1	4
AT5G61460	MIM, ATRAD18, SMC6B	STRUCTURAL MAINTENANCE OF CHROMOSOMES 6B, hypersensitive to MMS, irradiation and MMC	4

AT5G62410	ATSMC4, SMC2, ATCAP-E1, TTN3	TITAN 3, structural maintenance of chromosomes 2	4
AT5G63370	CDKG1	cyclin-dependent kinase G1	2
AT5G63540	BLAP75, ATRMI1, RMI1	RECQ MEDIATED INSTABILITY 1	2
AT5G63920	AtTOP3alpha, TOP3A	topoisomerase 3alpha	2
AT5G64520	ATXRCC2, XRCC2	homolog of X-ray repair cross complementing 2	2
AT5G66130	ATRAD17, RAD17	RADIATION SENSITIVE 17	2

1.2.2 Cohesion, chromosome movement and homologue recognition

During chromosomal cohesion, a cohesion complex forms a ring which traps duplicated sister chromatids and provides a physical link holding them together. Cohesion proteins are cleaved at anaphase by a separase enzyme and the sister chromatids separate. The cohesion complex is composed of four proteins: SMC1, SMC3, SYN1 and SCC3 (NASMYTH 2002) and other accessory proteins are involved in loading and maintenance of the complex including SCC2, CTF7, SWI1, SGO, and ASK1 proteins (Figure 1-4).

Loading of the cohesion complex occurs prior to S phase and requires the SCC2 protein (SEBASTIAN *et al.* 2009). Acetylation of SMC3 residues by CTF7 closes the ring and cohesion is established (SINGH *et al.* 2013). SWI1 is known to be essential for sister chromatid cohesion because in *swi1*, there are 20 chromatids at metaphase I which segregate randomly, but the precise role of SWI1 is not known (MERCIER *et al.* 2001). Chromosome arm cohesion is dissolved by separase in anaphase I, but centromere cohesion is maintained until anaphase II. The shugoshin proteins, AtSGO1 and AtSGO2 protect centromeric SYN1 from separase until anaphase II (ZAMARIOLA *et al.* 2014a). In *ask1* mutants, cohesion distribution and function is altered, and localization of SYN1 from zygotene to anaphase I is also affected (ZHAO *et al.* 2006). From these observations it seems likely that ASK1 is required for the release of chromatin from the nuclear membrane in leptotene, thus affecting the conformation and remodeling of meiotic chromosomes (YANG *et al.* 2006).

Chromosome movement is a prominent feature of meiotic prophase I. It is mediated by proteins located on the inner nuclear membrane with a highly conserved functional SUN domain and proteins on the outer nuclear membrane that contain a KASH domain. Interaction of these proteins creates a functional bridge between the cytoplasm and nucleoplasm, linking the cytoskeleton at one end with the chromosome at the other end. This bridge transduces the cytoskeleton forces that are responsible for chromosome movement during homologue pairing and recombination in many eukaryotic species including plants (STARR AND FRIDOLFSSON 2010). In *Arabidopsis*, class I kinesin I (AtPSS1) is essential for complete synapsis and bivalent formation

based on the partial synapsis at pachytene and a mixture of univalents and bivalents at metaphase I in *Atpss1* mutants. CO distribution is also affected in *Atpss1*, to the degree that CO interference is undetectable. This is similar to the phenotype seen in SUN protein mutations, indicating that AtPSS1 acts in the same pathway to regulate CO formation (DUROC *et al.* 2014). AtPSS1 is a member of the kinesin family of proteins that have the ability to walk on microtubules in an ATP dependant manner (ENDOW *et al.* 2010), leading to speculation that AtPSS1 moves along microtubules and generates forces that are transduced through the membrane via KASH-SUN proteins and affects chromosome pairing and synapsis (DUROC *et al.* 2014). Mutants of the rice homologue, OsPSS1, also have univalents at metaphase I, suggesting that this mechanism is conserved among plants (ZHOU *et al.* 2011). Duroc et al (2014) hypothesize that AtPSS1-dependent movement could promote synapsis by enabling the search for the homologous chromosome partner which could require chromosome movement across a large distance in the nucleus (KIM *et al.* 2010); or conversely, the AtPSS1-dependent movement could be used for the disentanglement of non-homologous chromosome associations that likely occur during the searching process due the small size of the nucleus and the relatively large number of non-homologous chromosomes (STORLAZZI *et al.* 2010).

Telomeres play an important role in the process of homologue recognition. In most species they cluster together on a small region of the nuclear membrane in a bouquet formation during early zygotene and persisting throughout zygotene (HARPER *et al.* 2004). Plants do not have a typical microtubule organizing centre where the bouquet attaches to the nuclear membrane as seen in other eukaryotes, but still form a bouquet though the attachment to the membrane is unclear (ROBERTS *et al.* 2009). *Arabidopsis* is one of the exceptions to the formation of a telomere bouquet, instead they have a cluster that associates with the nucleolus in the pre-meiotic phase (ARMSTRONG *et al.* 2001). It is suggested that the role of the telomere bouquet is to bring the chromosome ends into close proximity to promote homologous pairing (HARPER *et al.* 2004), but it may also be important in facilitating the movement of chromosomes that takes place during zygotene (SHEEHAN AND PAWLOWSKI 2009). In

the maize *pam1* mutant, the telomeres attach to the nuclear membrane but are unable to form a tight bouquet. Consequently, these plants show a dramatic reduction in homologous pairing (GOLUBOVSKAYA *et al.* 2002).

The role of centromere associations in homologue recognition and synapsis is less clear than that of telomeres. Early in prophase I non-homologous centromeres are coupled together, while late in prophase I homologous centromeres are paired but the mechanism of this transition is unknown. The late centromere pairing begins during assembly of the SC and persists after the SC has been dissolved. This late, homologous pairing appears to have a direct role in chromosome segregation, but the function of the early non-homologous centromere associations are unknown (OBESO *et al.* 2014).

Chromosome movement and pairing is closely tied to synapsis. Formation of the SC between the paired homologues stabilizes the pairing interactions. The SC is a proteinaceous structure made up of two axial elements at the bases of the chromosome loops and then linked together by transverse filament proteins. SC formation is closely tied to CO formation and most species depend on COs for complete synapsis (PAGE AND HAWLEY 2004). Zip1 is a coiled-coil protein that forms the transverse filament of the SC (SYM *et al.* 1993) for which *Arabidopsis* has two homologues AtZYP1a and AtZYP1b (HIGGINS *et al.* 2005). Mutants have extensive aberrant recombination that results in the formation of multivalents, nonhomologous bivalents and univalents at metaphase I, indicating that loss of AtZYP1 results in the loss of CO control (HIGGINS *et al.* 2005). In rice, mutants of the homologous gene, ZEP1, showed an increase in chiasmata, indicating a role in controlled CO formation (WANG *et al.* 2010).

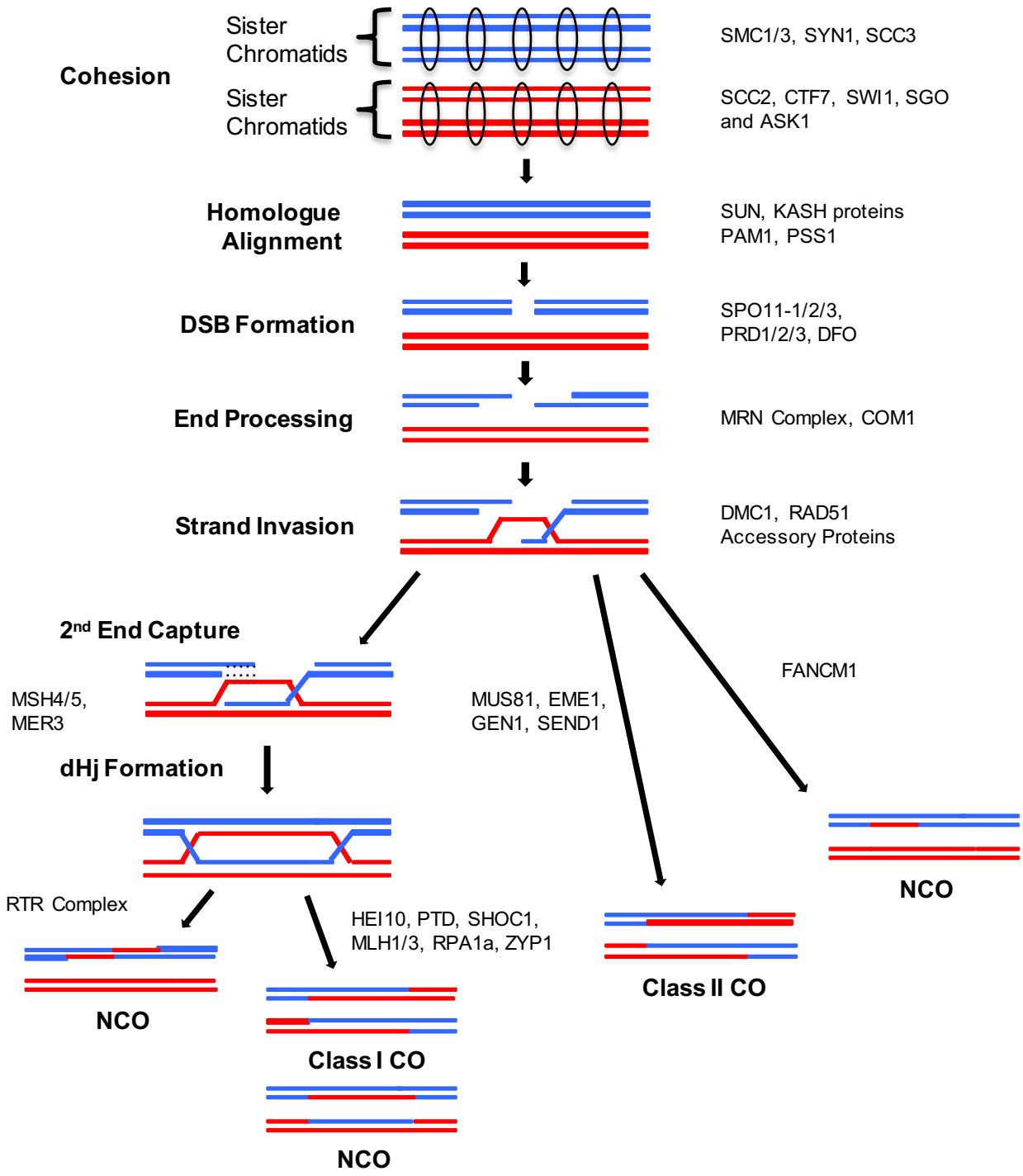


Figure 1-4: Overview of Meiotic Recombination
In pre-meiotic S-phase, sister chromatids are bound together by the cohesion complex, and then homologous chromosomes are aligned. Double-stranded breaks (DSBs) are created and processed prior to single strand invasion. If the invading strand is captured a double Holliday junction (dHj) is formed. The dHj can then be resolved as a Class I crossover (CO) or non-crossover (NCO). Processing by the RTR Complex also results in NCOs. When the second strand is not captured the single Hj can be resolved by the Class II pathway resulting in a CO or NCO

1.2.3 Double-stranded breaks and strand exchange

1.2.3.1 Double-stranded break formation

Homologous recombination leading to CO formation is initiated early in prophase I through the formation of DSBs (Figure 1-4). When these DSBs are repaired as COs, they create a physical link between homologous chromosomes, thus ensuring that the homologues are properly separated at the end of meiosis I and not randomly distributed between the resulting gametes. In yeast at least 10 proteins are required for formation of DSBs (Spo11, Mre11, Rad50, Xrs2, Ski8, Rec102, Rec104, Rec114, Mei4 and Met2) (COLE *et al.* 2010). These genes are poorly conserved in other eukaryotes, or in some cases they are conserved but are not required for meiosis, e.g. the Ski8 homologue does not have a meiotic function in *Arabidopsis* (JOLIVET *et al.* 2006). Other homologues have different meiotic functions in *Arabidopsis* than in yeast. In yeast, Mre11 and Rad50 are required for DSB formation, but AtMRE11 and AtRAD50 are necessary for 5' resection of the break rather than DSB formation (PUIZINA *et al.* 2004). Arabidopsis has six genes that are essential for DSB formation: AtSPO11-1, AtSPO11-2, AtPRD1, AtPRD2, AtPRD3, and AtDFO (DE MUYT *et al.* 2007; HARTUNG *et al.* 2007; DE MUYT *et al.* 2009; ZHANG *et al.* 2012).

Spo11 catalyses the formation of DSBs in nearly all eukaryotic systems. Spo11 is a homodimer that attacks the phosphodiester backbone of the DNA and remains attached to either side of the break (KEENEY *et al.* 1997). Arabidopsis has three SPO11 genes, two of which are required in meiosis, AtSPO11-1 and AtSPO11-2, though other eukaryotes require only one for meiotic DSB formation (KEENEY *et al.* 1997; HARTUNG *et al.* 2007). Both AtSPO11-1 and AtSPO11-2 have a number of alternative transcripts, many containing a premature stop codon. These transcripts are presumed to be targeted by the non-sense mediated decay (NMD) pathway, which could act as a means of controlling DSB formation (SPRINK AND HARTUNG 2014).

AtPRD1, AtPRD2, and AtPRD3 were identified through forward genetics screens to be required for DSB formation. Characteristic of other known DSB genes, *prd* mutants show no synapsis or chiasmata, resulting in univalents at metaphase I and random chromosome segregation at anaphase I (DE MUYT *et al.* 2007; DE MUYT *et al.*

2009). AtPRD1 and AtPRD2 show weak homology to the meiotic MEI1 and MEI4 genes which are required for DSB formation in mammals (LIBBY *et al.* 2003; KUMAR *et al.* 2010), and the AtPRD3 homologue OsPAIR1 is necessary for DSB formation in rice (NONOMURA *et al.* 2007), indicating a conservation in the genes required for the control of DSB formation across diverse species.

Lastly, AtDFO is a plant-specific DSB formation protein (ZHANG *et al.* 2012). Mutants have normal thread-like chromosomes in leptotene and early zygotene, but then they fail to form the thick strands that are seen in pachytene in the wild type. This results in ten univalents at diakinesis and subsequent random separation in anaphase I. Genetic analysis of progeny from Ler/Col wt and Ler/*Atdfo* crosses showed a severe disruption of recombination in the *Atdfo* mutant. *Atmre11* mutants had fragmented chromosomes indicating that DSBs are created but not repaired, but *Atdfo/Atmre11* mutants did not show chromosome fragmentation indicating that AtDFO is required for DSB formation rather than repair (ZHANG *et al.* 2012).

1.2.3.2 Processing of DSBs and strand exchange

Following DSB formation, SPO11 remains covalently bound to the DNA, and is removed by a single strand nick on either side of the break. The 5' DNA strand is then resected to leave 3' single-stranded tails. In *Arabidopsis* these processes are carried out by the complex of MRE11-RAD50-NBS1 (MRN complex) and AtCOM1. *Atmre11* and *Atrad50* mutants show lack of chromosome pairing, synapsis and DNA fragmentation indicating that they are required for the processing of DSBs rather than their formation (BLEUYARD AND WHITE 2004; PUIZINA *et al.* 2004). Within the MRN complex, MRE11 is a DNA binding protein that directly facilitates the formation of the single-stranded overhangs, RAD50 is likely to be important in maintaining the structure of the MRN complex, and NBS1 signals the presence of a DSB and regulates MRE11 activity (NEALE AND KEENEY 2006; BORDE 2007). Based on analysis of *Atcom1*, it appears to function downstream of AtSPO11 in the process of DSB repair prior to 3' strand invasion (UANSCHOU *et al.* 2007).

Resection of the DSBs leaves 3' ssDNA tails on each side of the break site. Two recombinases, Rad51 and Dmc1, form nucleoprotein filaments with the single-stranded tails. One of the filaments then invades the homologous double-stranded DNA forming a single-end invasion intermediate. The displaced DNA strand forms a D-loop, which extends along the chromosome as the invading strand polymerizes allowing 3' capture of the strand on the other side of the DSB. Ligation of the DNA ends results in two, four-way junctions, forming a structure called a double-Holliday junction (dHj) (NEALE AND KEENEY 2006) (Figure1-4).

Arabidopsis has one copy of the DMC1 and RAD51 genes (DOUTRIAUX *et al.* 1998). Mutants of AtDMC1 do not undergo synapsis and have 10 univalents at metaphase I, indicating that DSBs are being repaired (presumably by AtRAD51) but with sister chromatids rather than homologues (COUTEAU *et al.* 1999). On the other hand, *Atrad51* shows defects in chromosome pairing and synapsis but is also defective in DSB repair as indicated by chromosome fragmentation at metaphase I (LI *et al.* 2004). Together these results indicate that AtDMC1 and AtRAD51 are required for inter-homologue DSB repair but that AtRAD51 can use sister chromatids to repair DSBs independent of AtDMC1. Several additional proteins are involved in the AtDMC1/AtRAD51 mediated DSB repair and strand exchange including BRCA2, RAD51C, XRCC3, ASY1, ASY3, MND1/HOP2, SDS, MCM8, RPA1 and RFC1.

Arabidopsis has two functional homologues of BRCA2, which is well known to be involved in DNA recombination (BOULTON 2006). An *Atbrca2* knockout showed incorrect localization of AtRAD51 and AtDMC1 indicating that AtBRCA2 controls single-strand invasion through recruitment of AtRAD51 and AtDMC1 (SEELIGER *et al.* 2012).

Two RAD51 paralogues, AtRAD51C and AtXRCC3 also affect DSB repair through interaction with AtRAD51. The phenotypes of *Atrad51c* and *Atxrcc3* are very similar to *Atrad51* in that they have defects in chromosome alignment, synapsis and recombination (BLEUYARD AND WHITE 2004; LI *et al.* 2005). There are slight differences between the SC in the mutants, indicating that they are not functionally redundant but work together to promote meiotic DSB repair possibly whereby

AtRAD51 loading or activity is aided by AtRAD51C and AtXRCC3 (BLEUYARD *et al.* 2006).

Similarly, the chromosome axis proteins AtASY1 and AtASY3 are proposed to coordinate recombinase activity in favour of inter-homologue rather than sister chromatid recombination (SANCHEZ-MORAN *et al.* 2007; FERDOUS *et al.* 2012). In *Atasy1*, the association of AtDMC1 with recombination intermediates is affected, therefore synapsis and recombination are compromised but all DSBs are repaired. AtRAD51 however, appears to function independent of AtASY1 (SANCHEZ-MORAN *et al.* 2007). Studies with *Atasy3* showed that AtASY1 localization on the chromosomes is dependent on AtASY3, but in *Atasy1,* AtASY3 localization is unaffected (FERDOUS *et al.* 2012).

Another accessory factor, the AtMND1/AtHOP2 complex, has been shown to support strand exchange for DSB repair (PEZZA *et al.* 2010). In *Arabidopsis* it is necessary for DMC1-mediated inter-homologue repair but is not required for RAD51-mediated inter-sister repair (UANSCHOU *et al.* 2013). Solo Dancers (SDS) is a meiosis specific cyclin-like protein in plants that is required for DMC1 mediated, DSB inter-homologue repair (AZUMI *et al.* 2002; DE MUYT *et al.* 2009). Another protein, MCM8, appears to function alongside RAD51 to promote inter-sister repair when inter-homologue repair is compromised due to the absence of DMC1 (CRISMANI *et al.* 2013).

Replication Protein A (RPA) is a multi-subunit protein composed of RPA1, RPA2, and RPA3 that binds ssDNA. Arabidopsis has five RPA1 paralogues, two RPA2 paralogues, and three RPA3 paralogues. RPA1c is essential for DSB repair as evidenced by a chromosome fragmentation in *rpa1c* mutants (AKLILU *et al.* 2014). The RPA1a was shown to be necessary class I crossover formation (Section 2.4.2), but was not thought to be necessary for DSB repair (OSMAN *et al.* 2011). However, a recent study by Aklilu (2014) shows that RPA1a can function in DSB repair in the absence of RPA1c.

Arabidopsis rfc1 mutants show reduced fertility, multivalent formation and are defective in the formation of interference-sensitive COs, supporting the idea that RFC1 is important for dHj formation (WANG *et al.* 2012). Mutants also show RAD51

foci did not unload from zygotene to pachytene but were present until late pachytene. Wang et al (2012) propose that the loss of RFC1 function causes a failure of RAD51 and DMC1 to dissociate from the nucleoprotein filament, thus blocking dHj formation and the type I CO pathway. Instead, type II COs are formed between homologous and non-homologous chromosomes. An alternative explanation could be that the persistence of RAD51 foci could be the result of the generation of additional DSBs when fewer type I COs are formed. This would indicate the presence of a feedback mechanism to control CO formation and maintain crossover homeostasis as seen in mice and yeast systems (MARTINI *et al.* 2006).

1.2.4 Crossovers and recombination

After strand exchange, some of the recombination intermediates go on to form COs, which are necessary for complete synapsis of the homologous chromosome pairs. Most species require one obligate CO per chromosome pair to ensure accurate separation of the homologous chromosomes at anaphase I. Most COs are not randomly distributed, if there are multiple COs on the same chromosome they are spaced apart, i.e. a CO at one location makes it less likely that a CO will happen in a nearby location, a phenomenon termed CO interference (BERCHOWITZ AND COPENHAVER 2010). The mechanical stress model for CO interference proposes that a CO creates a local spot of relief from the mechanical stress that is generated by the expansion and contraction of the chromatin along the axis. This relief radiates outward along the length of the chromosome, and as the distance increases, stress builds again until another CO is created to relieve it, resulting a series of COs that are spatially separated along the chromosome (KLECKNER 2006). It has been suggested that Topoisomerase II mediates this process by adjusting the spatial relationships among DNA segments during chomatin/axis compaction (ZHANG *et al.* 2014).

In *Arabidopsis*, approximately 150 DSBs are formed during meiotic recombination (SANCHEZ-MORAN *et al.* 2007), but only a small number (~10) of recombination intermediates develop as COs (SANCHEZ-MORAN *et al.* 2002). The rest are required for pairing, chromosome alignment and synapsis initiation and are subsequently

repaired as non-crossovers (NCOs). Since there must be at least one CO per chromosome pair (JONES AND FRANKLIN 2006), selection of those DSB sites that resolve into COs can not be random. In yeast the CO/NCO decision seems to occur early, before dHj resolution and SC assembly, at or just before the formation of the single-end invasion intermediate (BISHOP AND ZICKLER 2004). This early decision model of meiotic recombination postulates that most NCOs derive from synthesis-dependant strand annealing (SDSA) and that the intermediates that go on to form dHjs are resolved as COs (ALLERS AND LICHTEN 2001). However, in *Arabidopsis* there is evidence that dHjs can be resolved as NCOs, by the RTR complex (CHELYSHEVA *et al.* 2008; HARTUNG *et al.* 2008) (Figure 1-4).

The five chromosome pairs of *Arabidopsis* have an obligate CO and CO interference is exhibited when there are multiple COs between a chromosome pair. Several proteins belonging to the Class I recombination pathway are essential for maintaining obligate COs and CO interference. There is also evidence for a class II meiotic recombination pathway, in which the COs do not exhibit CO interference and do not maintain the obligate CO on each chromosome (HIGGINS *et al.* 2004).

1.2.4.1 The class I pathway of recombination

The Class I pathway of meiotic recombination results in interference-sensitive COs, i.e. formation of a CO at one site makes formation of a CO at an adjacent site less likely. The pathway is well studied in yeast, which has identified a group of genes collectively called ZMM (Zip1, Zip2, Zip3, Zip4, Msh4, Msh5, and Mer3) that are essential for class I CO formation and SC assembly (BORNER *et al.* 2004). Homologues for most of these have been identified in Arabidopsis, along with additional plant specific genes (Figure 1-4).

It is hypothesized that MSH4 and MSH5 initially stabilize a single-end invasion by a sliding clamp mechanism that embraces the double-stranded DNA, thus enabling the conversion into a dHj and resolution as either a CO or NCO (SNOWDEN *et al.* 2004). MER3 (also called ROCK-N-ROLLERS (RCK)) is a DNA helicase that unwinds duplex DNA in a 3' to 5' direction and may promote CO formation by stimulating

extension of the DNA heteroduplex molecules (NAKAGAWA *et al.* 2001). Arabidopsis *msh4, msh5,* and *mer3* mutants all have a similar phenotype showing a dramatic reduction in chiasmata formation, and those that do form are not affected by CO interference, indicating they are formed via the Class II pathway (CHEN *et al.* 2005). The AtZYP1 proteins are homologues of the well-characterized Zip1 which is essential for the formation of the SC. Yeast Zip2, Zip3, and Zip4 function together to promote and regulate SC polymerization ensuring that it is dependent on homologous recombination (MACQUEEN AND ROEDER 2009). The Arabidopsis protein, HEI10 has been suggested as a functional homologue of Zip3, and *hei10* has the typical *zmm* phenotype with a dramatic reduction in the formation of COs indicating it is a ZMM protein (CHELYSHEVA *et al.* 2012).

Arabidopsis has two novel proteins, PARTING DANCERS (PTD) and shortage in chiasmata (SHOC1), which are specifically required for class I CO formation (MACAISNE *et al.* 2011). SHOC1 shows similarity to the Zip2 protein and has structural similarity to XPF family proteins though the sequence is not conserved (MACAISNE *et al.* 2008). PTD has sequence similarity to proteins of the ERCC1 family and has been shown to be involved in meiotic CO formation as well (WIJERATNE *et al.* 2006). XPF proteins are endonucleases that form heterodimers with non-catalytic ERCC1 proteins to recognize and process branched structures during DNA repair (CICCIA *et al.* 2008). Using a yeast two-hybrid assay, PTD and SHOC1 were shown to interact, suggesting that they could form a heterodimer, similar to an XPF-ERCC1 complex, which is required for the formation of Class I COs (MACAISNE *et al.* 2011).

In addition to its involvement in DSB processing, AtRPA1a plays a role in class I CO formation. Based on KO studies, it looks as though AtRPA1a is required after AtMSH4 and before AtMLH3 in the class I pathway. It is possible that AtRPA1a affects strand annealing and second end capture during meiosis (OSMAN *et al.* 2009). Once the dHj has formed, AtMLH1 and AtMLH3 both play a role in the resolution of the dHj, and are thought to maintain or impose a conformation that ensures it resolves as a CO (FRANKLIN *et al.* 2006). AtMLH3 contains a metal binding motif, but

AtMLH1 does not, leading to speculation that this motif catalyses the symmetrical nicking of a dHj that is required for it to resolve as a CO (OSMAN *et al.* 2011).

1.2.4.2 The class II pathway of recombination

In *Arabidopsis*, ~15% of the COs are not MSH4 dependent, are not subject to CO interference and with only an average of less than two per meiotic cell they do not ensure the obligate CO between each homologous chromosome pair (HIGGINS *et al.* 2004).

AtMUS81 and AtEME1 are homologues of the Mus81 and Eme1 genes, which form a complex that is required for all meiotic COs in fission yeast, none of which exhibit CO interference (SMITH *et al.* 2003). The MUS81 complex is an endonuclease that can cleave early recombination intermediates such as D-loops and nicked Hjs to generate COs (OSMAN *et al.* 2003). Arabidopsis has two EME11 homologues, AtEME1A and AtEME1B, both of which are capable of binding AtMUS81 and forming a functional endonuclease. These complexes are capable of cutting nicked and intact Hjs, unlike other systems (GEUTING *et al.* 2009).

Another newly discovered Hj resolution pathway in *Arabidopsis* is analogous to the E. coli Hj resolution pathway (BAUKNECHT AND KOBBE 2014). This uses the resolvase RuvC that forms a homodimer that positions two active sites at the centre of the Hj and will cleave it if a specific consensus sequence is found (SHAH *et al.* 1994; SHALEV *et al.* 1999). The analogous eukaryotic GEN1 proteins are members of the Rad2/XPG nuclease family (IP *et al.* 2008). Arabidopsis has two functional homologues of GEN1, AtGEN1 and AtSEND1. These homologues have different preferred sequences, and based on the cleavage positions it appears that both the structure and the sequence context of the Hj determine the cleavage position (BAUKNECHT AND KOBBE 2014).

1.2.4.3 Non-crossover pathways

Arabidopsis has two independent pathways for NCO repair of DSBs, one via the RTR complex and the other using FANCM-mediated DNA repair (MANNUSS *et al.* 2010). Many dHjs are resolved to NCOs via the RTR complex (Figure 1-4). This complex has

three components: 1) a recombination deficiency Q helicase, (AtRECQ4A in Arabidopsis); 2) a type 1A topoisomerase, (TOP3α in Arabidopsis); and 3) the RecQ-mediated genome instability 1 structural protein, (AtRMI1 in Arabidopsis) (ZAKHARYEVICH *et al.* 2012). The resolution process begins with the helicase catalyzing the branch migration, bringing together the crossing points of the dHj followed by the decatenation catalyzed by the topoisomerase (HARTUNG *et al.* 2007; HARTUNG *et al.* 2008). The RMI1 protein plays an essential structural role in mediating protein-protein interactions that are required for a functional RTR complex (BONNET *et al.* 2013).

The other Hj resolution pathway in Arabidopsis involves the FANCM and MLH1 proteins that have been shown as important regulators of the MUS81 dependent CO pathway (KNOLL *et al.* 2012). In fission yeast, the Mhf1 and Mhf2 proteins support the FANCM homologue, Fm11 to transform COs into NCOs via synthesis-dependent strand annealing SDSA (LORENZ *et al.* 2012). *Atfancm* mutants have 3x fewer COs compared to wild type plants (CRISMANI *et al.* 2012) and mutants of *Atmhf1*, showed connections between bivalents during diplotene and carrying through to anaphase I, resulting in unequal chromosome distribution at the end of meiosis II (DANGEL *et al.* 2014). A recent study in Brassica used EMS mutagenized populations and *fancm* mutants had 3x and 1.3x fewer COs in *B. rapa* and *B. napus,* respectively, confirming the anti-CO role for FANCM (BLARY *et al.* 2018).

1.3 Polyploidy

1.3.1 Polyploid formation and evolution

Polyploids are organisms that contain more than one set of chromosomes and may be autopolyploids (multiple copies of the same genome), or allopolyploids (two or more different genomes). Polyploidy is a natural evolutionary process that occurs in all major classes of living organisms including plants, animals and yeast. The success of a polyploid species depends on the ability to produce viable offspring i.e. to control meiosis and produce gametes that can combine to form viable embryos. It is difficult to assess the level of polyploidy among plants, but it has been shown that ~70% of

angiosperms are polyploid (MASTERSON 1994). Species like Arabidopsis and rice were historically assumed to be diploids because they have an odd number of chromosomes, but genome sequencing has shown that they are ancient polyploids (BOWERS *et al.* 2003b).

One possibility for the formation of polyploid individuals from diploid organisms is through unreduced gametes. If, by chance, two unreduced gametes formed a viable embryo, then this would create an instant polyploid. The relative rarity of unreduced gametes makes this unlikely, and a more reasonable hypothesis would be that the unreduced gamete would combine with a haploid gamete to form a triploid embryo. If this triploid could survive to maturity and produce an unreduced triploid gamete, then this could form an embryo with a haploid gamete to create a tetraploid individual (RAMSEY AND SCHEMSKE 1998). The recurrent polyploidization (formation of the same species multiple times) of many plants indicates that a greater proportion of the diversity from the diploid ancestors will be present in the polyploid gene pool than if the species had evolved only once from two diploid progenitors (SYMONDS *et al.* 2010).

All genes are functionally duplicated in the newly formed polyploid, so these duplicated genes can both remain functional, one of the genes can be lost or silenced, or the two homoeologues can diverge in function. This does not always take place over several generations, a study in cotton looked at 1400 gene pairs immediately following genome merger and found that 25% of genes already showed biased expression of one of the homoeologues (FLAGEL *et al.* 2008). Epigenetic mechanisms such as methylation, small RNA mediated silencing or histone modifications are likely to account for such an immediate response. During polyploid evolution, retention and loss of duplicated genes shows general patters indicating the process is non-random. For example, duplicate copies of genes involved in signal transduction or transcription tend to be retained, while for those genes involved in DNA repair one of the duplicates is often lost (BLANC AND WOLFE 2004; MAERE *et al.* 2005).

1.3.2 Homoeologous recombination

A unique problem for neopolyploids during meiosis is a chromosome may have more than one potential pairing partner, termed a homoeologue and cytological analysis of neopolyploids provided evidence for this occurring. In meiotic cells, bivalents, univalent and multivalents of homologous and homoeologous chromosomes can be seen and this aberrant pairing occurs more frequently in newly formed *B. napus* compared to existing lines (ATTIA AND RÖBBELEN 1986b). Recombination between homoeologues as a result of this mispairing can lead to unbalanced gametes, some of which may be inviable, and dramatic genetic changes in subsequent generations. All or part of a chromosome from one genome can be replaced with a chromosome from the other genome and if both recombinant chromosomes are not inherited there will be a loss of genetic material (Figure 1-5). Similarly, multivalent and univalent chromosomes can also lead to unbalanced gametes during anaphase I when unpaired chromosomes will move to one pole or the other and the complicated dissolution of multivalents yields unpredictable results (ZAMARIOLA *et al.* 2014b). These gametes may be inviable or produce plants which are less healthy or sterile, leading to an overall reduction in yield which would be a significant problem for crop species. Therefore, the genetic mechanisms responsible for chromosome pairing control in polyploids is of interest for maintaining plant fitness and crop productivity.

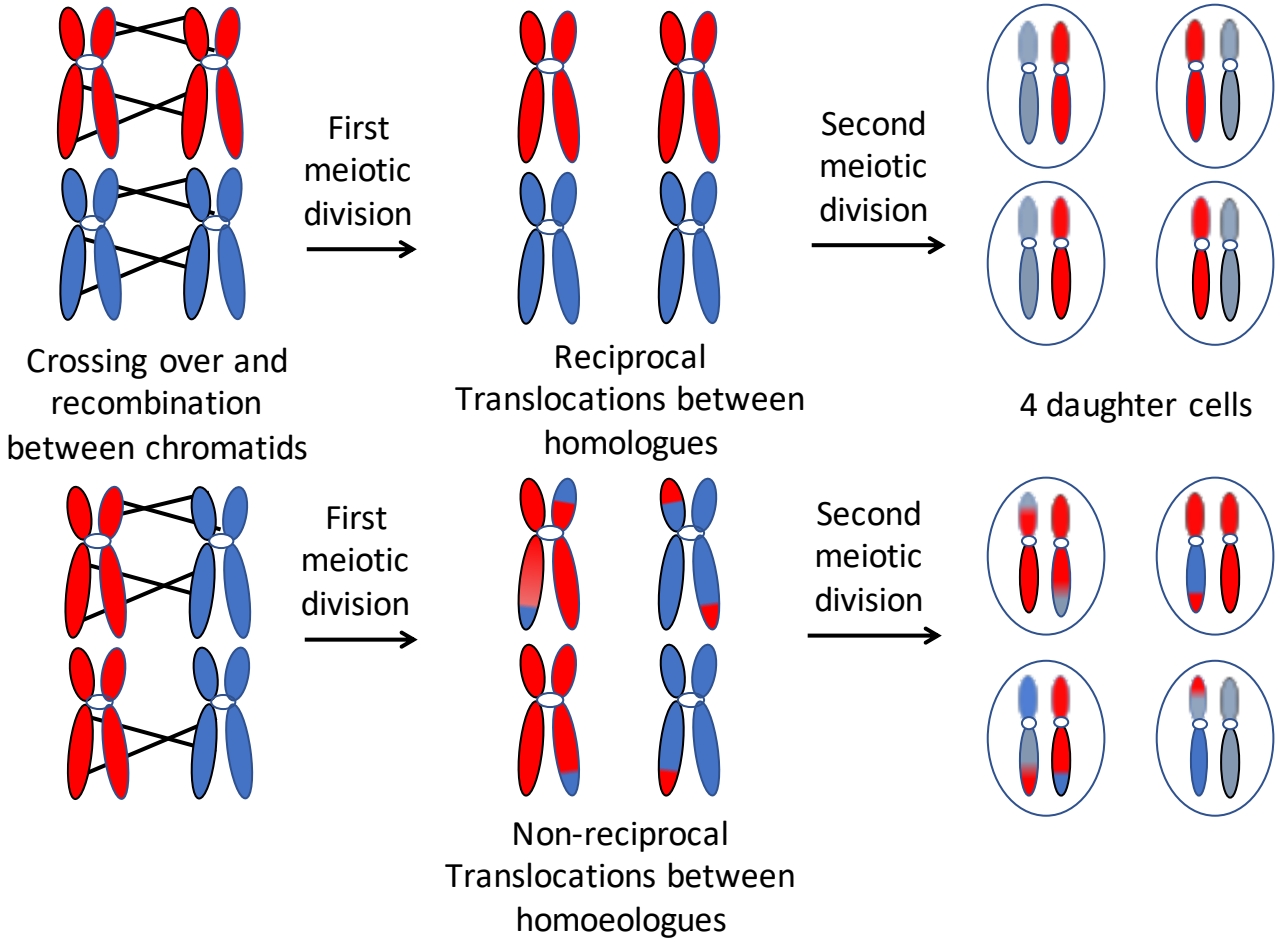


Figure 1-5: Homologous and Homoeologous Recombination
Top panel: Homologous chromosomes pair and recombine, the resulting four daughter cells each have a complete set of chromosomes. Bottom panel: Homoeologous chromosomes pair and recombine resulting in daughter cells with duplicated or missing chromosome segments. These gametes may be inviable or may form plants with decreased fitness.

1.3.3 Meiotic chromosome pairing control in polyploids

Newly formed polyploids must become genetically stable in order to ensure proper distribution of homologues/homoeologues in gamete formation. Thus they must behave as diploids and restrict pairing to true homologues. Two hypothesis could explain the diploidization of allopolyploids: 1) accumulation of structural chromosomal variation between the homoeologous genomes resulting in preferential pairing between homologues rather than homoeologues or, 2) the development of a molecular mechanism which restricts chromosome pairing to homologues. There is evidence for each of these indicating that it is most likely not one or the other but a combination of both. The identification of loci that restrict homoeologous chromosome pairing and recombination in allopolyploids indicates that divergence alone is not sufficient to restrict homoeologous chromosome associations. Two of these loci, *Ph1* in wheat and *PrBn* in *Brassica napus* have been well studied.

1.3.3.1 Wheat *Ph1*

Control of meiotic chromosome pairing has been extensively investigated in hexaploid wheat, where the Pairing homoeologous1 (*Ph1*) phenotype was first identified over 50 years ago (RILEY AND CHAPMAN 1958; SEARS AND OKAMOTO 1958). When *Ph1* is present, chromosomes pair and recombine exclusively with their homologue during meiosis, but in the absence of *Ph1* pairing occurs between both homologous and homoeologous chromosomes, resulting in chromosomal rearrangements accumulating over subsequent generations and the lines becoming infertile (MOORE 2000). The *Ph1* locus was originally identified in a deletion line and was localized to chromosome 5B (SEARS 1977). Much effort has gone into elucidating how *Ph1* functions to control chromosome pairing, but the large genome size combined with the highly repetitive nature of the wheat genome initially limited the ability to identify candidate genes (MOORE 2000).

Griffiths et al (2006) identified a 2.5 Mb region on chromosome 5B which contains a cluster of defective cyclin-dependent kinase-like (CDK) genes. CDKs control the progression of the cell cycle in eukaryotes (DEWITTE AND MURRAY 2003). The

hypothesis is that these defective CDK-like genes reduce CDK activity and this is responsible for the *Ph1* phenotype of restricting chromosome pairing between homoeologues. Use of okadaic acid, a phosphatase inhibitor that increases CDK activity, on wheat-rye hybrids induced homoeologous pairing, thus copying the *Ph1* phenotype (KNIGHT *et al.* 2010). A study determined that phosphorylation of wheat histone H1 doubled at Cdk2-like consensus sequences in the *Ph1* mutant (GREER *et al.* 2012). These results led to the hypothesis that *Ph1* delays heterochromatin decondensation which affects sister chromatid cohesion. This delay allows for repair of DSBs by sister chromatids, without the delay there is decreased specificity in the chromosome pairing process allowing non-homologous pairing to occur (GREER *et al.* 2012).

An alternative mode of action for *Ph1* was proposed by Bhullar (2014) (BHULLAR *et al.* 2014) where physical mapping of *Ph1* mutants identified a ~2.5 Mb region on chromosome 5B (GILL *et al.* 1993) different to that of Griffiths et al (2006), and comparative mapping identified a region in *Oryza sativa* which contained 91 genes (SIDHU *et al.* 2008). Silencing of the wheat orthologue of Os9g30320 through VIGS and RNAi resulted in varying levels of multivalents at metaphase I, chromosome clustering, and misalignment along the metaphase I plate (BHULLAR *et al.* 2014). The degree of multivalents, clustering and misalignment was dependent on the level of reduction in gene expression, with a 44% reduction providing a phenotype most similar to that seen in *Ph1* deletion lines. The authors proposed that this candidate *Ph1* gene functions by regulating centromere-microtubule interactions, lack of which results in the characteristic P*h1* phenotype. This hypothesis is based on similar expression patterns to CENP-E, which associates with the kinetochore to facilitate interaction with microtubules that is required for chromosome movement and alignment along the metaphase plate (YAO *et al.* 1997).

However, Rey et al (REY *et al.* 2017) disputed this, arguing that Os9g30320 had previously been characterized as a tapetal cell gene (JEON *et al.* 1999) and mutation or deletion of the gene results in chromosome clumping due to stressed meiocytes. Those authors offer evidence that a *ZIP4* paralogue within the 5B locus is the elusive

Ph1 gene, based on the phenotype and analysis of a wheat line with a mutation in the *ZIP4* gene from an EMS mutagenized population (REY *et al.* 2017).
Other loci with smaller effects have been identified including *Ph2*, which has been localized to the short arm of wheat chromosome 3D (MELLO-SAMPAYO 1971). These genes are not as well studied as *Ph1* because the suppression of homoeologous pairing is not as strong as Ph1, and their effect is difficult to measure in the presence of *Ph1* (MELLO-SAMPAYO 1971). Positional cloning of the *Ph2* locus on chromosome 3D identified MSH7 which is one of the *MutS* homologues that are part of the mismatch repair system as the *Ph2* gene (SERRA *et al.* 2021).

1.3.3.2 Brassica *PrBn*

The control of pairing in Brassica species has not been as well studied as in wheat, but work has been done to discover the mechanism by which chromosomes find their pairing partners during meiosis. Much of the research on pairing control in Brassica has focused on haploid or triploid populations. Using cytogenetic techniques to directly monitor pairing and recombination in pollen mother cells, the Pairing regulator in *B. napus* (*PrBn*), locus was identified and mapped to chromosome C9 (JENCZEWSKI *et al.* 2003). Additional QTL that had smaller but independent effects from *PrBn* were also been identified in these haploids (LIU *et al.* 2006) indicating that homologous pairing in Brassica is controlled by a suite of genes rather than a single master switch. Blary et al (2016) performed RNASeq on meiocytes isolated from the two lines used to map the *PrBn* locus, Darmor bzh and Yudal, to look at expression levels for the genes underlying the QTL but did not identify any candidate meiotic genes that showed differential expression.

It is unlikely that *Ph1* and *PrBn* control pairing by a similar mechanism (JENCZEWSKI *et al.* 2003) *PrBn* was identified using natural polymorphism that exists in *Brassica napus*, but no natural polymorphism has been identified for *Ph1*. *PrBn* was identified in haploids, but the diploids from which the haploids were generated all exhibited

disomic inheritance indicating that *PrBn* is not required for chromosome stability and fertility unlike *Ph1*.

2. Thesis Overview

Brassica napus is a recent allopolyploid and contains the A genome from *B. rapa* and the C genome from *B. oleracea*. The success of a polyploid species depends on the ability to faithfully maintain chromosome content and produce viable gametes through meiosis. Genetic control of strict homologous chromosome pairing is well studied in wheat, though relatively little is known of the process in Brassica species. The ability to determine if there is a genetic control of chromosome pairing in *B. napus* depends on: a) having a high-density genotyping platform which comprehensively covers the two constituent genomes of *B. napus* with genetic markers; b) a method for identifying allelic gain and loss at homoeologous loci and; c) use of this method in a *B. napus* mapping population which segregates for the control of meiotic chromosome pairing and recombination to identify loci important for controlling this trait. This thesis is based on three publications which address each of these challenges.

One of the barriers to comprehensively measure recombination was the cumbersome nature of RFLP and SSR markers but the development of high-density SNP arrays has made it possible to quickly and relatively cheaply genotype large numbers of samples at tens of thousands of loci. A Brassica Illumina Infinium SNP array containing approximately 60,000 SNP assays was developed. The design and testing of this array is described in **Clarke WE, Higgins EE, Pileske J, Wieseke R, Sidebottom C, et al. A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single locus markers in the allotetraploid genome. Theoretical and Applied Genetics 2016 129:1887-1899; doi:10.1007/s-00122-0016-2746-7 or Chapter 3.**

My contribution to the work detailed in the paper was two-fold. First, I analyzed a diverse set of ~400 *B. napus* lines to screen all SNPs on the array and identify those with cluster patterns indicative of non-specific hybridization so they could be either removed from analysis completely or adjusted so the genotyping software could accurately detect segregating SNPs. This information was used to create a custom cluster file that is available to all users of the Brassica 60K array to filter out poorly performing SNPs from their own data. Secondly, after applying the cluster file, I used the remaining set of SNP markers to identify those polymorphic in a resynthesized *B. napus* population and created a genetic map to confirm the position of the markers in the *B. napus* genome. I examined SNP markers which did not map to their expected position based on physical alignment of the probe sequence to the *B. napus* genome to determine if they mapped to the predicted homoeologous location in the genome or elsewhere.

Once a set of high quality, single locus SNP markers was established, development of a method to identify homoeologous recombination needed to be established. To do this, I chose ten *B. napus* cultivars from a worldwide collection of spring-type oilseed rape lines and analysed progeny from these lines with the Brassica 60K SNP array. The resultant data and analyses are presented in **Higgins EE, Clarke WE, Howell EC, Armstrong SJ and Parkin IAP. Detecting *de novo* homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. G3: Genes|Genomes|Genetics August 2018 8:2673-2683; doi: 10.1534/g3.118.200118** or **Chapter 4**.

I used the expected position of the SNP probe sequences in the two genomes of *B. napus* to align the chromosomes as homoeologous pairs. I crossed the each of the ten cultivars to another *B. napus* line and the progeny from these crosses were run on the SNP array. I used the SNP array data to identify reciprocal gain and loss at homoeologous loci to detect changes in allelic ratios in the progeny from the 10 chosen *B. napus* lines. Finally, I validated the accuracy of the SNP array in identifying homoeologous recombination events through analysis of whole genome sequencing

data of one of the parental lines. As the first author on this paper I carried out the analyses, interpreted the data and drafted the manuscript.

The doubled-haploid (DH) *B. napus* population that was used for initial genetic mapping of the Brassica 60K SNP array was created by crossing a DH *B. napus* line with a resynthesized *B. napus* line created by crossing a *B. rapa* and *B. oleracea* lines. Newly resynthesized lines are known to have higher rates of homoeologous recombination than established *B. napus* so the resulting DH population segregated for homoeologous recombination rate. Now with two key pieces in place, the development of a fast and affordable comprehensive genotyping system and development of an assay capable of measuring homoeologous recombination using this platform, it was possible to combine these tools and map loci important for controlling meiotic chromosome pairing. This work is detailed in **Higgins EE, Howell EC, Armstrong SJ and Parkin IAP. A major quantitative trait locus on chromosome A9, *BnaPh1*, controls homoeologous recombination in *Brassica napus*. New Phytologist. doi: 10.1111/nph.16986 or Chapter 5**.
I created testcross populations for 48 lines from the DH population, assayed the testcross individuals with the SNP array and analysed the data for homoeologous recombination events. I then used the recombination rate for each of the 48 lines for QTL mapping. This work was done in parallel with cytologists at the University of Birmingham who analysed meiotic chromosome pairing in lines from the same DH population. I used their cytogenetic data in the QTL mapping as a corroboration of my SNP array work. I created RNASeq libraries for leaf and developing bud tissue from the two parents of the DH population and combined this with meiocyte RNASeq data from the collaborators in Birmingham to look at expression of meiotic genes underlying the three identified QTLs.

3. A high density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single locus markers in the allotetraploid genome

Citation:

Wayne E. Clarke, **Erin E. Higgins**, Joerg Plieske, Ralf Wieseke, Christine Sidebottom, Yogendra Khedikar, Jacqui Batley, Dave Edwards, Jinling Meng, Ruiyuan Li , Cynthia Taylor Lawley, Jérôme Pauquet, Benjamin Laga, Wing Cheung, Federico Iniguez-Luy, Emmanuelle Dyrszka, Stephen Rae, Benjamin Stich, Andrew G. Sharpe, Martin W. Ganal, Isobel A. P. Parkin. A high density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single locus markers in the allotetraploid genome. **Theoretical and Applied Genetics 2016 129:1887-1899;** doi:10.1007/s-00122-0016-2746-7.

*All supplementary data, figures and tables in this chapter are available through the Theoretical and Applied Genetics open access article. https://rdcu.be/cgdqj

3.1 Abstract

A high-density single nucleotide polymorphism (SNP) Illumina Infinium array, containing 52,157 markers, was developed for the allotetraploid *Brassica napus*. A stringent selection process employing the short probe sequence for each SNP assay was used to limit the majority of the selected markers to those represented a minimum number of times across the highly replicated genome. As a result approximately 60% of the SNP assays display genome-specificity, resolving as three clearly separated clusters (AA, AB, BB) when tested with a diverse range of *B. napus* material. This genome-specificity was supported by analysis of the diploid ancestors

of *B. napus* whereby 26,504 markers and 29,720 were scorable in *B. oleracea* and *B. rapa*, respectively. Forty-four percent of the assayed loci on the array were genetically mapped in a single doubled-haploid *B. napus* population allowing alignment of their physical and genetic coordinates. Although strong conservation of the two positions was shown, at least 3% of the loci were genetically mapped to a homoeologous position compared to their presumed physical position in the respective genome, underlying the importance of genetic corroboration of locus identity. In addition, the alignments identified multiple rearrangements between the diploid and tetraploid Brassica genomes. Although mostly attributed to genome assembly errors some are likely evidence of rearrangements that occurred since the hybridisation of the progenitor genomes in the *B. napus* nucleus. Based on estimates for linkage disequilbrium decay, the array is a valuable tool for genetic fine mapping and genome-wide association studies in *B. napus* and its progenitor genomes.

3.2 Introduction

Brassica napus is an economically important oilseed crop that is primarily grown to extract the healthy edible oil from the seed, but it is now also grown as a renewable feedstock for biodiesel. In addition, there are vegetable types of the species that have been bred for both human and animal consumption. It is a temperate crop widely grown in both the Northern and Southern hemispheres due to available genotypic variation for flowering time and response to photoperiod. It is believed that *B. napus* emerged from a small number of hybridisation events between the diploid progenitors *Brassica oleracea* (C genome) and *Brassica rapa* (A genome) (U 1935) that probably occurred in the southern Mediterranean and possibly regions of Asia around 7,000-10,000 years ago (CHALHOUB *et al.* 2014). The progenitors, *B. oleracea* and *B. rapa,* are also important predominantly vegetable crop species that each display a wide range of genetic and morphological diversity (DIXON 2006).

There are extensive worldwide breeding efforts in *B. napus* and its diploid relatives in both the public and private domains that contribute to developing higher value

crops with improved yields (INIGUEZ-LUY AND FEDERICO 2011; SNOWDON AND INIGUEZ LUY 2012). Such breeding efforts are benefiting from access to a burgeoning collection of genetic and genomic resources for the *Brassica* species culminating in the recent release of the diploid and amphidiploid genomes that complete one axis of U's triangle and define the *B. napus* genome (WANG *et al.* 2011; CHALHOUB *et al.* 2014; PARKIN *et al.* 2014). The now available genome sequences can be exploited to identify candidate genes for traits of interest but their primary utility in breeding is in the development of genetic markers for marker assisted selection and more recently genomic selection. Genomic selection or predictive breeding is showing potential for application in crop species, where traits can be controlled by multiple small effect QTLs, as more sophisticated algorithms have been developed to overcome the statistical challenges of working with disproportionately larger numbers of marker loci than samples tested (JANNINK *et al.* 2010).

The availability of genome sequences and access to relatively economical next generation sequencing technologies has provided the impetus to identify extensive nucleotide variation among different plant species. The abundance of single nucleotide polymorphisms (SNPs) across plant genomes has made them highly desirable for marker development (GANAL *et al.* 2009; GANAL *et al.* 2012). High throughput (tens of thousands or higher) SNP screening can be achieved effectively by either, genotyping-by-sequencing (GBS) or high-density SNP arrays. GBS requires no former knowledge of available SNPs within a species but is heavily reliant on bioinformatics capacity, and although common SNP will be found across experiments, the SNP profile identified is dependent on the genotypes queried (DESCHAMPS *et al.* 2012). In comparison, high-density SNP arrays provide a common platform that can be continuously used and replicated across multiple labs with minimal computational requirement (GANAL *et al.* 2012). However, such SNP genotyping arrays involve significant development costs to identify sufficient numbers of robust, informative loci that fulfill assay design criteria. Identifying high quality SNP loci for array design requires sequence data from sufficient numbers of genotypes to be able to

assess polymorphism levels and associated allele ratios across the diversity of a species in order to minimise ascertainment bias. In addition, genome duplication in polyploid genomes such as *B. napus* confounds the design of SNP assays, since nucleotide variation among closely related orthologous or paralogous sequences is often misinterpreted as allelic variation (PARKIN *et al.* 2010). Further, since the SNPs are evaluated through hybridisation, multiple homologous and homoeologous loci may hybridise to a single SNP oligonucleotide probe leading to highly compressed and often irresolvable SNP patterns.

The current manuscript describes the development of a high density (>50,000) Illumina Infinium® SNP array designed for genotyping in *B. napus*, that can also be applied to the diploids, *B. oleracea* and *B. rapa*. Next generation sequence data from both genomic and transcriptome sources were utilised to identify millions of preliminary SNP loci across the *B. napus* genome. Extensive filtering of these data led to the development of a highly effective tool for Brassica breeding with the majority of the SNP assays targeting single loci within the amphidiploid genome. The efficacy of the array was tested through the generation of cluster files, which define common allele clusters across a range of genotypes in all three species, and a high-density genetic map for *B. napus*.

3.3 Materials and Methods

3.3.1 Reference mapping and variant calling

Pseudo-genome sequences of the diploid A and C genomes (283.8 Mb and 488.6 Mb respectively) were combined into a single reference sequence set for mapping. Sequence reads from each genotype were aligned independently using the CLC Genomics Server v3.6. Default parameters for the mapping algorithm were used except for the mapping identity parameter which was increased to 98% in order to facilitate resolution of homoeologous sequence reads. Mapped reads were interrogated for sequence variation using the CLC Genomics Server v3.6 variant

discovery algorithm. A minimum depth of coverage of 3x for 454 and 8x for Illumina data was required for SNP calling. Mapping data and variant calls were exported from CLC in the form of SAM alignment files and tab-delimited text files, respectively. Data from these files were combined using a custom Perl script in order to determine a missing, reference, or variant call in each genotype at each covered position of the genome.

3.3.2 SNP filtering

Combined SNP results were filtered using custom Perl scripts and eliminated based on the following criteria: 1) SNP positions without suitable flanking sequence (50bp on at least one side of the SNP with no variation); 2) SNP positions with more than two variations within the surveyed genotypes; 3) SNP positions with high levels of heterozygous calls, biased allele ratio, or missing data; 4) Illumina Assay Design Tool (ADT) score less than 0.6; 5) SNP positions where the variation was the result of a transversion.

3.3.3 Probe Matching and SNP selection

Probe sequences for all filtered SNPs were obtained from Illumina and then aligned to the reference sequences using the open source alignment tool BLAT with default parameters (KENT 2002). These alignments were parsed using a custom Perl script to determine the number of times the probe sequence from a particular SNP matched to the reference sequence set. A probe alignment was considered to be matched if 35 consecutive base pairs of the probe were fully aligned. SNPs were ranked based on the number of times their probe sequence matched the reference sequence set and SNPs with fewer probe matches preferentially selected.

3.3.4 Experimental SNP data collection

The cluster file for *B. napus* was generated at AAFC through analysis of 437 genotypes and at TraitGenetics through the analysis of 432 genotypes. The cluster files for *B. oleracea* and *B. rapa* were generated with 129 and 121 samples,

respectively. In both laboratories, DNA was extracted from young leaf tissue of greenhouse grown plants using a cetyltrimethylammonium bromide (CTAB) based method (MURRAY AND THOMPSON 1980). DNA was quantified and 200ng were hybridised to the Brassica 60K Infinium array as described in the manufacturer's protocol (Illumina Inc., San Diego, CA). The arrays were scanned using an Illumina HiScan or BeadArray Reader and SNP data were analysed using the Genotyping module of the GenomeStudio software package with the setting for the No Call threshold set to 0.05.

3.3.5 Generation of the genetic map

DNA from 124 lines of a doubled haploid (DH) population (derived from a cross between DH12075 and PSA12 and named SG DH, Parkin, unpublished) was hybridised to the Brassica 60K Infinium array and allele calls were made using the newly generated cluster file. The genetic linkage map was generated using the MSTmap software package (WU *et al.* 2008). The map order was checked manually to ensure the optimal placement of the SNP loci, and a bin map was generated. Final map distances were calculated using the Kosambi mapping function and the Mapmaker v3 software (LANDER *et al.* 1987).

3.4 Results

3.4.1 Array Design

A set of 54,866 SNP assays, previously identified and tested on the Illumina platform were provided from a number of different sources (BUS *et al.* 2012; DALTON-MORGAN *et al.* 2014) (Cheung, Dryszka, Laga, Pauquet, Rae, unpublished data). The remainder of the SNP assays that were used in the array design were processed using a single pipeline (Supplementary Figure 1). Next generation sequencing data were collated from two previously published datasets described in HARPER *et al.* (2012), which contributed RNASeq data from 42 different *B. napus* genotypes, and CLARKE *et al.* (2013), which contributed Illumina and Roche 454 sequence capture data from

nine *B. napus* genotypes. In addition, Roche 454 (1.16 Gb) data from genomic material and Illumina HiSeq (417.85 Gb) data from both genomic and transcriptome sources were generated for an additional 13 *B. napus*, four *B. oleracea* and three *B. rapa* genotypes (Supplementary Table 1).

The array was designed prior to the release of the *B. napus* genome sequence (CHALHOUB *et al.* 2014). Thus high quality sequence reads were reference-mapped using CLC Genomics Server v3.6 to a pseudo *B. napus* genome derived from concatenating the genome sequences of *B. rapa* (WANG *et al.* 2011) and *B. oleracea* (PARKIN *et al.* 2014). Considering only uniquely matching reads, over 570 Gb of sequence data were aligned to the pseudo-genome providing an estimated 738x depth coverage, although it should be noted the inclusion of transcriptome data can bias the overall distribution with over-representation of some genic regions. SNP calling was completed using the SNP Discovery algorithm of the CLC Genomics Server and all relevant data were exported for further filtering. Custom Perl scripts were used to generate an output file that included the SNP id, reference id and position, flanking sequence where available, the reference allele, and for each individual surveyed, the SNP call, depth and frequency data. These data were then filtered in three steps. First, SNPs were excluded if there was insufficient SNP-free flanking sequence (50 bp on at least one side). This step removed the largest number of the identified SNP loci (76%) (Table 3-1). In the second step, SNPs were excluded if they were multi-allelic (more than 2 alleles) since these cannot be efficiently assayed using the Illumina platform. The final step identified high confidence SNP loci, SNPs were excluded when the frequency of individuals with missing data was greater than 70%, the frequency of individuals that showed heterozygous calls was greater than 40%, and finally if the allele frequency was higher than 0.8 or lower than 0.2. Table 3-1 shows the attrition at each filtering step. A final set of 180,398 SNP loci consisting of filtered and previously tested SNPs were submitted to the Illumina Assay Design Tool, which returned 161,917 SNPs with a minimum recommended score at or above 0.6.

Table 3-1: SNPs and filtering steps used for array design

Filter Step	SNPs Excluded	SNP Count
None	0	24,528,374
Flanking Sequence	18,619,172	5,909,202
Multi-Allele SNP	7,671	5,901,531
Confidence[1]	5,742,443	159,088
Illumina ADT Score (<0.6)	33,556	125,532
Transversions	1,318	124,214

[1] SNP positions were filtered for high levels of heterozygous calls, biased allele ratio, or missing data as described in the Results section

The specificity of each Illumina SNP assay is reliant on a single 50 bp probe sequence flanking one side of the SNP, the length of which can lead to ambiguous matching across genomes with any level of redundancy. In order to filter potential designs to reduce the impact of high copy probe sequences, the 50 bp probe for each possible assay design were matched using BLAT to the pseudo-genome. More than half (74%) of the filtered SNP assays had probe sequences that mapped to multiple regions of the Brassica genome. The final SNP list submitted for Illumina bead design contained 15,141 previously tested SNP loci, 32,294 newly designed SNP loci that matched the pseudo-genome uniquely and 11,029 SNP loci that matched twice. Once synthesized, 52,157 SNP markers on the Brassica 60K array passed bead representation and decoding quality metrics, including 1,213 A/T or C/G SNPs, which are represented by Infinium I bead types that require two beads per assay.

3.4.2 Cluster file generation for reliable scoring of the SNP markers in *Brassica napus* and its diploid ancestors

The most efficient high-throughput application of an Illumina array can be achieved with the development of a robust cluster file that defines the expected intensity level of the three genotype classes (AA, AB, BB) for each SNP locus. The cluster file is

applied to intensity data to automatically call the genotypes for experimental samples, thus allowing easy comparison of data across labs (Figure 3-1a). At AAFC Saskatoon the first dataset included 327 *B. napus* genotypes of both annual and biennial type, from diverse origins, ten F_1 lines, and a subset of lines from two DH mapping populations. Independently at TraitGenetics, a second dataset was generated that consisted of 432 mostly winter-type *B. napus* genotypes, including 67 hybrids, 88 F_2, and 20 resynthesized *B. napus* lines. The two datasets were analysed independently and the resultant cluster files compared. After filtering out 173 SNP from the cluster file that displayed low intensity across the majority of the samples, 51,984 SNP remained. Based on the genotypes tested at AAFC 1678 loci were monomorphic and for the genotypes tested at TraitGenetics, 2444 markers were monomorphic. Due to the strong sequence similarity between the A and C genomes of *B. napus* it was anticipated that some of the SNP loci would display cluster patterns reflecting co-hybridisation of homoeologous loci. In such instances, when both homoeologous loci are polymorphic the resultant SNP patterns are not automatically resolvable, generally these result in 4-5 clusters and the Illumina software will identify exceptionally high numbers of heterozygotes (Figure 3-1b). However, when one of the homoeologous loci are monomorphic the genotype cluster intensities are shifted to one side of the theta space (actual genotype would be for example, AAAA, AAAB and AABB), leading to false cluster assignment with the routine analysis tools (Figure 3-1c), yet the cluster definition of such a SNP locus can be optimised manually to reflect the correct genotype positions, rendering it perfectly scorable (Figure 3-1d). The two labs independently assayed for such loci, and manually adjusted the cluster assignments where necessary. By assessing the number of polymorphic markers that showed a cluster pattern that was indicative of a single copy locus, with three possible allelic states (homozygous allele AA, heterozygous AB, homozygous allele BB) distributed over the entire theta space (difference between mean AA θ and mean BB $\theta > 0.6$), between 34,248 (TraitGenetics) and 37,536 (AAFC Saskatoon) loci were determined to be effectively genome specific. For a small number of markers, fluorescence was observed for only one allele, which

could be informative in certain populations but would be unable to detect heterozygous individuals (Figure 3-1e). In total, 47,304 markers were defined as scorable and are retained in the current *B. napus* cluster file.

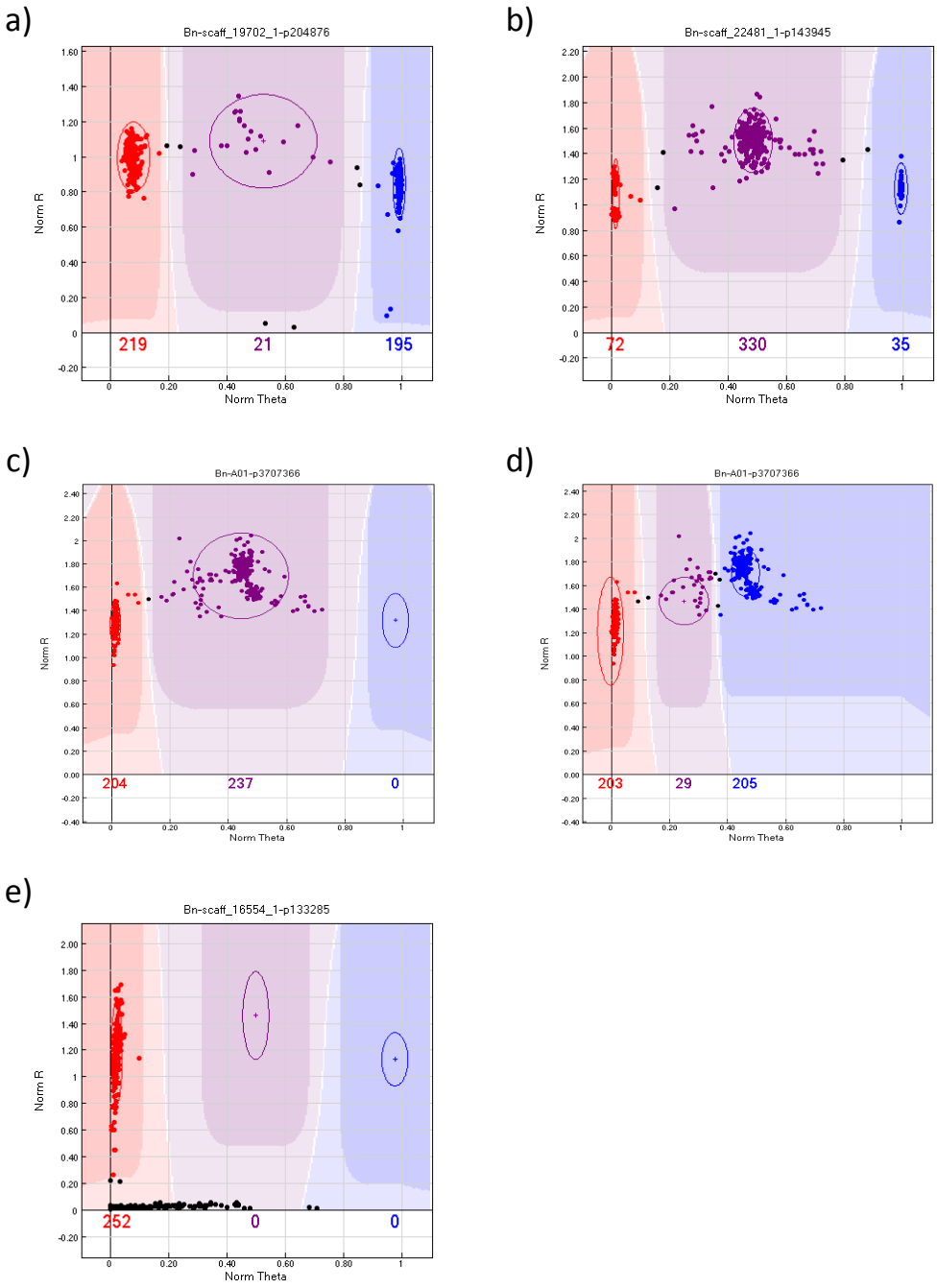
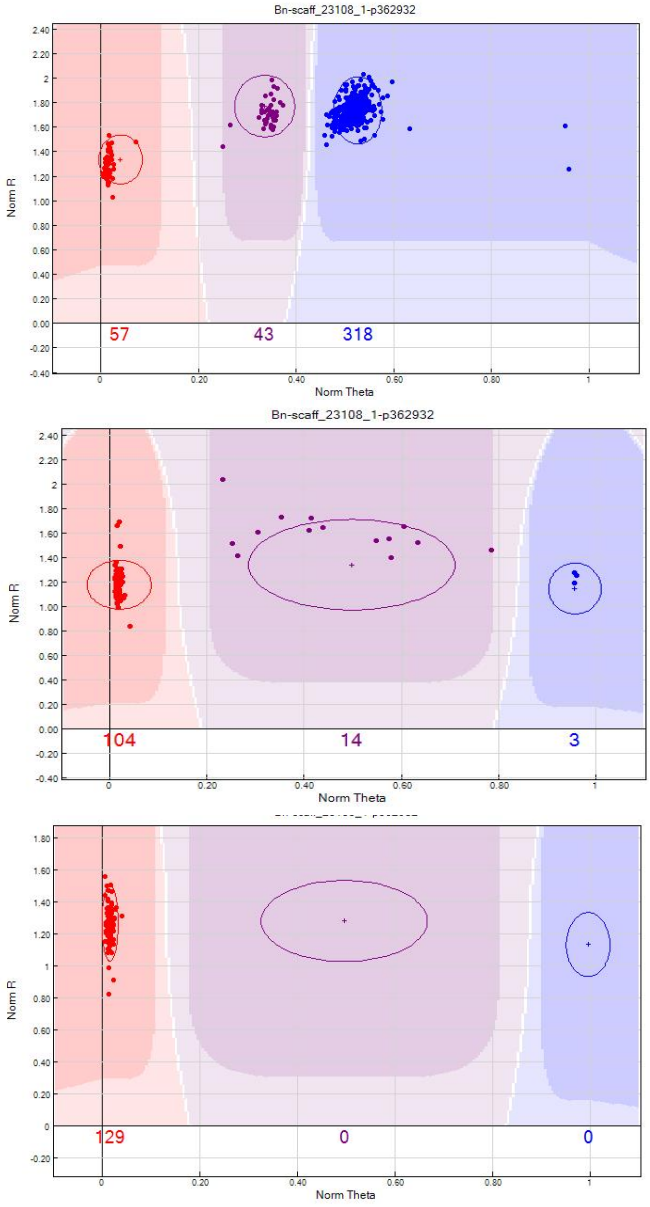


Figure 3-1: GenomeStudio images showing representative SNP cluster patterns across B. napus

One cluster representing one parental allele is coloured in red (AA), the second in blue (BB), and heterozygote genotypes in purple. **a** Shows a genome-specific SNP marker in *B. napus,* almost 60% of the SNP loci show this clear separation of the expected three genotypes. **b** SNP locuslikely resulting from hybridization of two segregating homoeologous loci reveals five clusters and an excess of heterozygotes. **c** and **d** Show a SNP locus called automatically by the software and after manual adjustment of the cluster profile, respectively. **e** SNP locus where one parental allele shows no hybridization (presence/absence marker).

Diploid samples included in the initial analysis indicated that a subset of the loci could not be scored accurately in the diploids using the *B. napus* cluster file (Figure 3-2a-c). In order to facilitate the use of the Brassica array for scoring in the diploid ancestors, 129 *B. oleracea* lines, hybrids and representative segregating material, and 121 *B. rapa* lines, F_1s and representative segregating material were analysed with the array. Based on these results, cluster files for the two ancestral species were generated, mainly based on modified cluster positions for those markers that were not genome specific. The final cluster file for *B. oleracea* contained 26,504 scorable markers of which 21,113 were polymorphic in the investigated material and the *B. rapa* cluster file contained 29,720 scorable markers of which 22,695 were polymorphic in the investigated material (Supplementary Table 2).



a) *B. napus*

b) *B. rapa*

c) *B. oleracea*

Figure 3-2: GenomeStudio images showing representative SNP cluster patterns in the different Brassica species.
One parental allele is coloured in red (AA), the second in blue (BB), and heterozygotes in purple. The SNP marker is polymorphic, but not genome-specific in *B. napus,* **a** resulting in condensed clusters due to the detection of the homoeologous locus on the other genome. In the diploid *B. rapa,* **b** this marker is polymorphic and shows widely distributed clusters (no second homoeologous locus detected, typical diploid pattern). **c** In *B. oleracea* material, this marker is monomorphic.

3.4.3 Physical and genetic position of SNPs on *Brassica* genomes

The physical positions of the assayed loci in the diploid A and C genomes were determined during the SNP calling process, based on unique read mapping to the reference sequence set, consisting of the A genome of *B. rapa* (WANG *et al.* 2011) and the C genome of *B. oleracea* (PARKIN *et al.* 2014). The physical position were also determined in two recently completed *B. napus* genomes, one a winter type (CHALHOUB *et al.* 2014) and the second a spring type (Parkin et al, unpublished), by aligning the DNA flanking each of the SNP loci to each genome using BLAT (KENT 2002). The best hit and associated percent identity of the match for each genome were then extracted from the BLAT results. Based on a percent identity of at least 85%, 50,255 SNPs were positioned in the spring type genome sequence and 49,794 were positioned in the winter type genome sequence. Taking both genome sequences together, a total of 51,172 SNPs could be matched to one or both *B. napus* genomes (Figure 3-3). It was recognized that the length of the query sequence could lead to ambiguities or erroneous matches due to the highly redundant nature of each genome. The latter would be particularly true for matches to the *B. napus* genome, where in addition to the strong homology between the two constituent genomes there are also regions of effective identity resulting from homoeologous exchanges between the A and C genomes (CHALHOUB *et al.* 2014). Based on the BLAT scores 22,258 and 23,191 SNPs could be unambiguously positioned on the A and C genomes, respectively, while 2,138 were placed on either the A or C with equal probability (Supplementary Table 3). Additionally, 4,570 SNPs could not be positioned on the pseudochromosomes as a result of either missing data in *B. napus* or the alignment of the SNP sequence to an unanchored scaffold in one or both *B. napus* genomes. The SNP loci were largely found in non-coding regions, although 17,955 lay within annotated gene sequences, only 8,681 of which were positioned within an exon (Supplementary Table 4).

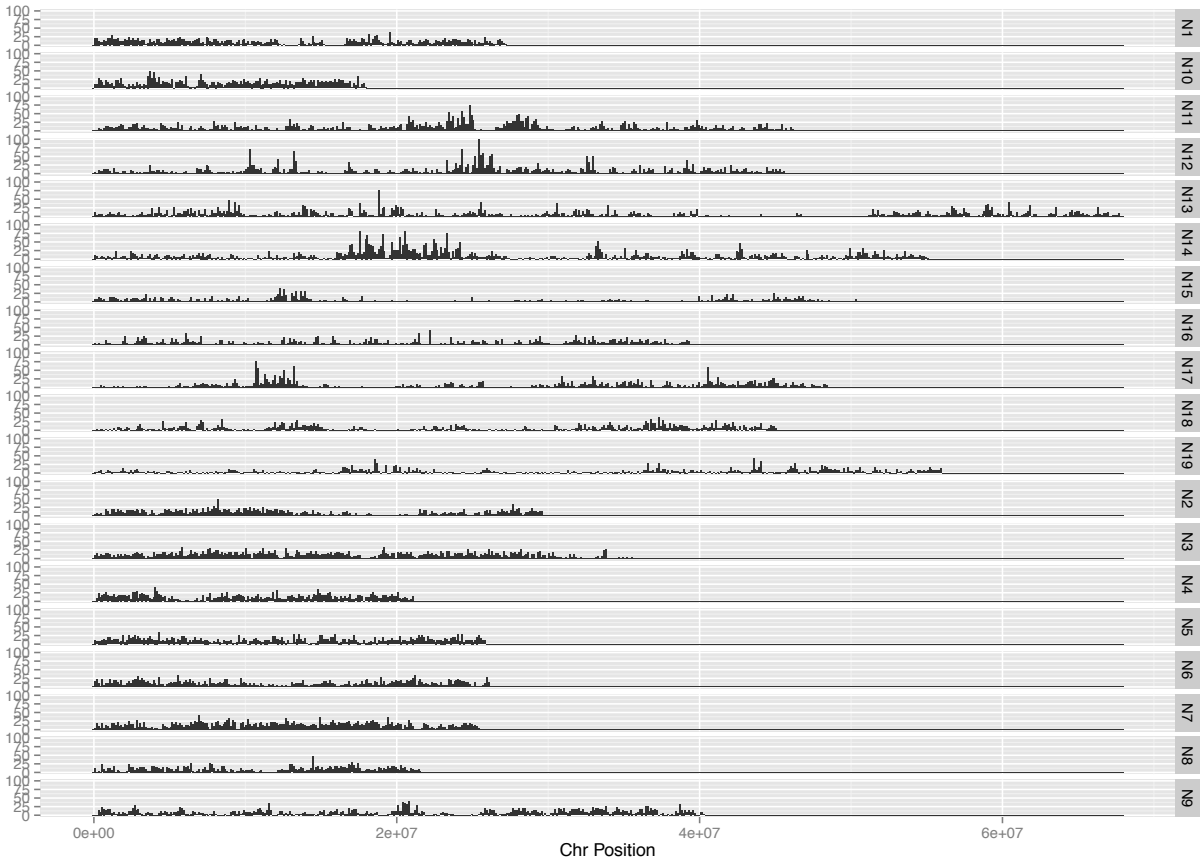


Figure 3-3: Physical distribution of SNP loci across the B. napus genome
The SNP loci were aligned to the genome of spring-type DH12075 based on BLAT scores, with the numbers of SNP loci per 125 Kb window indicated on the y-axis for each chromosome

In order to genetically position 21,766 (46%) of the SNP loci, the highly polymorphic SG DH population derived from a cross between a resynthesized *B. napus* and an established *B. napus* line was used (Supplementary Tables 5 and 6). Based on informative recombination events, these loci were placed in 1,310 bins across the 19 linkage groups and covered a length of 1,815 cM (Table 3-2). The loci were distributed with on average one marker every 0.15 cM or less (Table 3-2). There were a number of genetically defined bins with a higher than average density of markers, which tended to cluster together and were associated with regions of low recombination, predominantly found in the vicinity of the presumed centromeric regions (Figure 3-3).

Table 3-2: Distribution of genetically and physically positioned SNP loci in the B. napus genome

Linkage Group	Number of genetically defined bins[1]	Number of physically positioned SNP loci[2]	Physical length of pseudo-molecule (Kb)	Number of mapped SNP loci	Map Distance (cM)	Markers/cM	cM/marker	Kb/marker
A1	60	2672	27,105	913	75.9	12.03	0.08	10.19
A2	74	2425	29,627	946	98.2	9.63	0.10	12.22
A3	94	3185	35,753	1330	114.3	11.64	0.09	11.23
A4	50	2112	21,080	1085	57.6	18.84	0.05	9.98
A5	78	2332	25,706	1121	99.6	11.26	0.09	11.06
A6	89	2302	26,146	1019	100.1	10.18	0.10	11.32
A7	46	2529	25,458	1333	60	22.52	0.04	10.07
A8	55	1863	21,685	953	85	11.21	0.09	11.64
A9	89	2452	40,546	1279	127.5	10.03	0.10	16.43
A10	63	2053	17,911	841	75.6	11.12	0.09	8.72
C1	46	3418	45,604	1882	71.3	26.40	0.04	13.65
C2	47	3743	47,311	1241	69.8	17.78	0.06	12.20
C3	118	3870	67,777	1804	165.3	10.91	0.09	17.51
C4	88	4399	55,069	1443	136.9	10.54	0.09	12.55
C5	65	1600	48,717	651	124.4	5.23	0.19	31.46
C6	40	1982	40,797	980	42.6	23.00	0.04	19.88
C7	82	2784	48,823	1360	101.7	13.37	0.07	17.54
C8	68	2151	44,716	742	104	7.13	0.14	20.97
C9	58	1782	55,995	843	106	7.95	0.13	31.42
Total	1310	49,744	725,833	21766	1814.9	11.99	0.08	14.59

[1] The genetic map position is based on mapping data from the SG DH population

[2] The physical position is based on the original reference mapped position in the diploid genome sequence

Only markers that were positioned both genetically and physically could be definitively positioned on the *B. napus* genome. In general, there was good correspondence between the two, with 20,138 of the 21,766 (93%) SNP loci genetically mapping to the position expected based on sequence alignment. Additionally, 3% of loci that were physically mapped to one position were genetically mapped to the homoeologous region of the genome (Figure 3-4). The remaining 4% did not match either the expected physical position or the homoeologous position. Some of these loci were genetically mapped to unanchored *B. napus* scaffolds, but others mapped to a different position in *B. napus* than expected based on the diploid genome from which the SNP probe was designed, suggesting inconsistencies with the original assemblies. This was particularly true for the *B. rapa* genome where genomic regions of varying sizes (0.1 – 1.1 Mb) were found to be anchored to the wrong chromosome relative to the two *B. napus* genome sequences (Supplementary Table 3). This is likely due to the relatively low marker density that was used to anchor the *B. rapa* genome assembly (WANG *et al.* 2011). It is possible that some differences between the genomes could reflect true chromosomal rearrangements; indeed the relatively large inversion found to differentiate both ends of A10 from N10, could be evidence of such an event, since it appears common to both *B. napus* genotypes (Figure 3-5). However, a similar but smaller inversion at the top of A7/N7 is specific to only the Darmor bzh *B. napus* genotype (Supplementary Table 3). The physical and genetic positions of the SNP loci have been imported into a web tool that visualizes these alignments (http://aafc-aac.usask.ca/Bn60). Particular regions of the genome can be selected to identify potentially useful SNP loci in any region of interest.

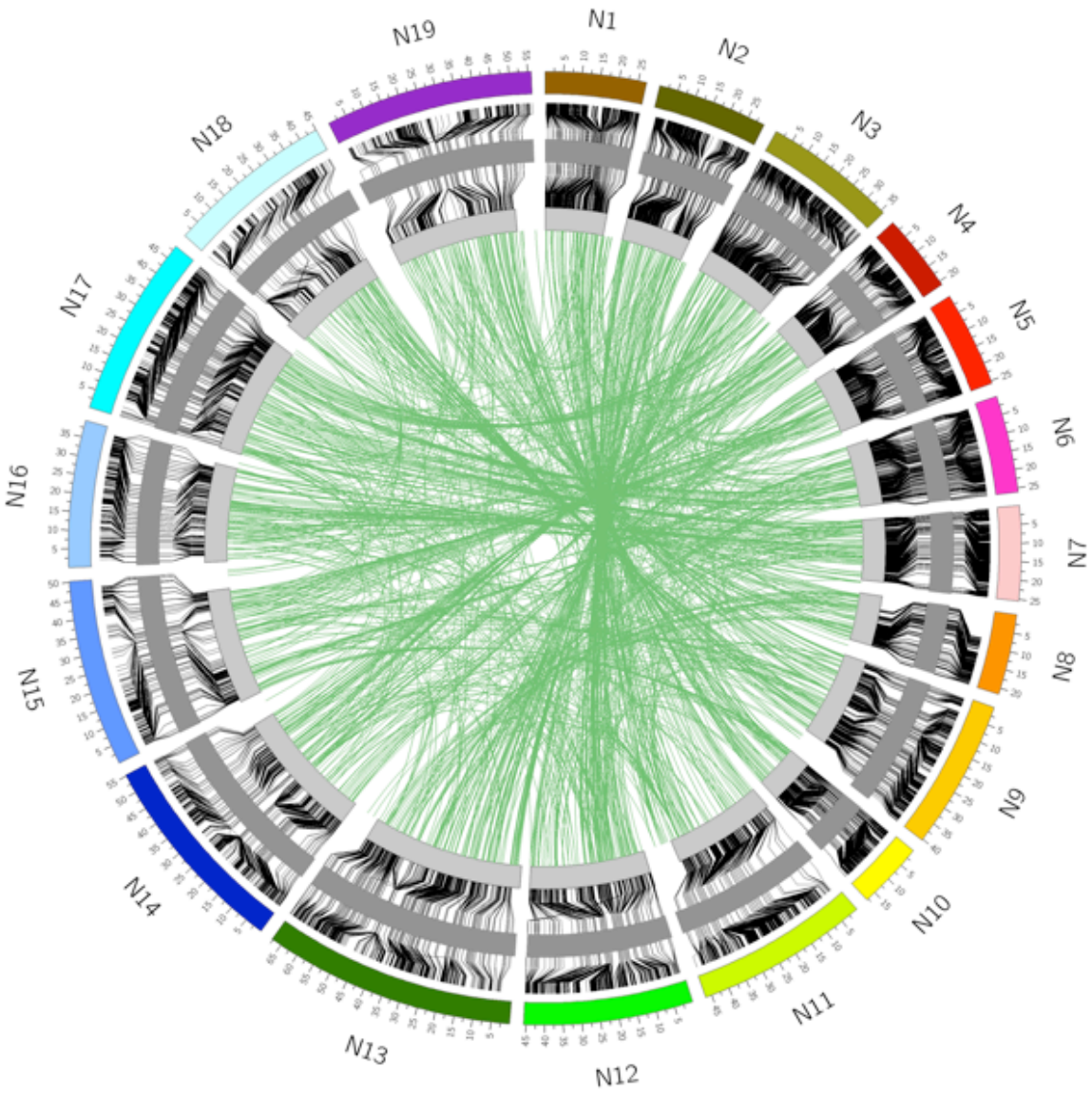


Figure 3-4: Relationship between the physical and genetic positions of the SNP loci in B. napus

The inner circle represents the genetic map which is flanked to the outside by the physical position in the spring-type DH12075 and to the inside by the physical position in the winter-type Darmor bzh. The green lines connecting across the centre of the circle represent those loci that are genetically positioned to an alternate (mostly homoeologous) position compared to their physical coordinates in the genome sequence.

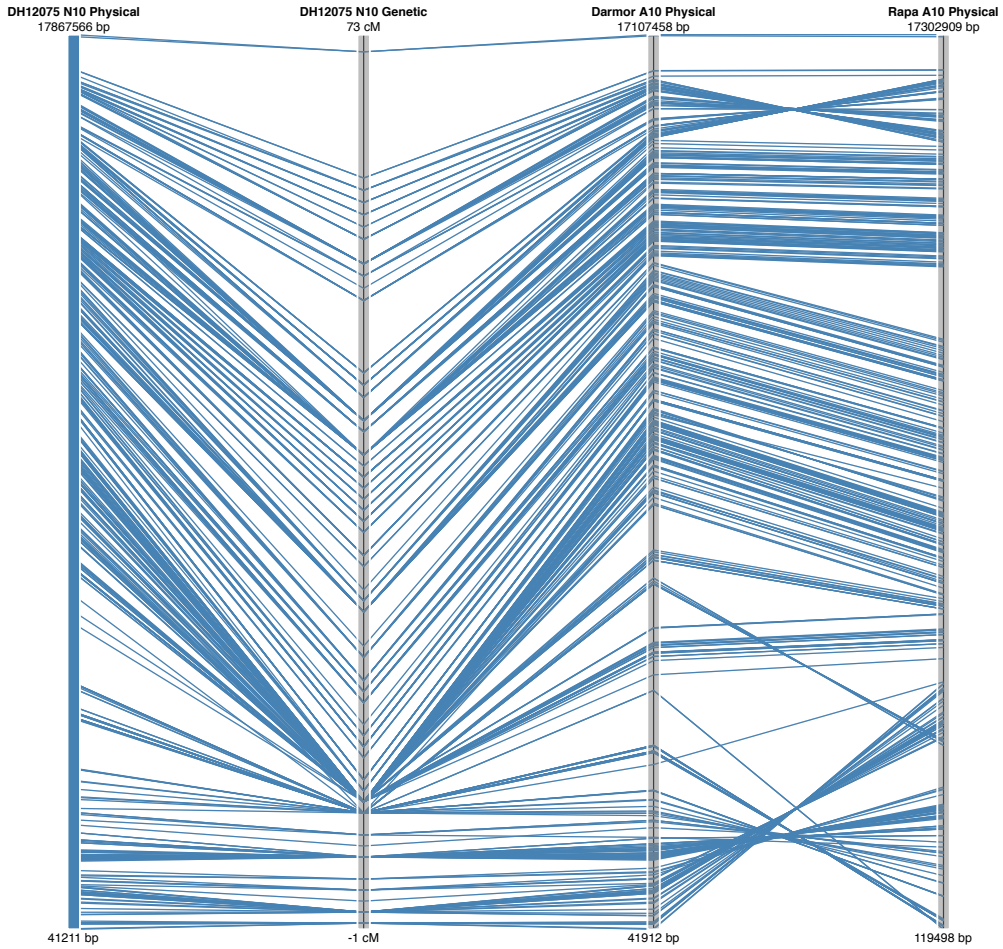


Figure 3-5: Alignment of the genetic map for linkage group N10 of B. napus
Alignment of the spring-type *B. napus* DH12075 with the genetic position in the SGDH population and the physical position in winter-type *B. napus* Darmor and *B. rapa*

3.5 Discussion

The recent release of three Brassica crop genome sequences has provided opportunities for the development and application of new breeding tools. The complexity of the *B. napus* genome becomes evident when carrying out genetic mapping even for relatively simple quality based traits, where multiple loci control their expression. In order to dissect and follow genetically complex traits within Brassica breeding programs robust sequence based markers are required. The high throughput genotyping array described here offers an excellent platform for

facilitating such analyses and allows ready access to a set of well-characterised markers.

The SNP pipeline used to develop the majority (approximately 74%) of the assays on the Brassica array included a number of steps to limit the impact of genome duplication and allopolyploidy on the resultant design. Reference mapping of short read sequences to the constituent diploid genomes was optimised to prevent matching to homoeologous regions, sequence variation was avoided in the immediate proximity of the target SNP, and regions where multiple alleles were identified were excluded, all limiting the calling of paralogous SNPs. The final step of remapping the oligonucleotide probe for each SNP assay back to the reference genomes and selecting those with the lowest copy number further facilitated the design of the array, since 81% of the SNPs that matched only once in the *B. napus* genome produced genome-specific three cluster patterns. Approximately 58% of all assays produced clear genome-specific genotype calls for a wide range of *B. napus* genotypes including F_1 individuals. This is a marked improvement over similar arrays designed for hexaploid and tetraploid wheat where not more than 25.5% of the SNP assays produced clear genome-specific cluster patterns (WANG *et al.* 2014). However, the stringency of the design pipeline limited the variation available to be employed for the array design, effectively reducing the number of initial SNPs by 99%. This high level of attrition could cause some marker selection bias. Thus, in order to achieve an optimal density of markers across the genome, pre-validated SNP assays, which matched two or more regions of the reference genome were also included.

Upon testing *B. napus* cultivars originating from multiple continents and covering the range of annual and biennial types, the SNP assays proved to be highly polymorphic, with only 3.5% monomorphic loci identified indicating the value of the selected SNPs. This was also confirmed independently in a recently published work that used the array to assess diversity within a collection of predominantly Asian *B. napus* lines (QIAN *et al.* 2014). The array was tested through the generation of a dense

SNP map for *B. napus* with 21,766 (46%) of the loci mapped in one DH population. Due to the design process, each of the SNP loci was physically anchored to a specific base pair position in one or other of the diploid progenitor genomes (Supplementary Table 3). On extending this analysis to *B. napus* however, the alignment of only the isolated short SNP regions to the genome sequence introduced a level of ambiguity, with some SNPs being equally likely to align, or in the case of the assays, hybridise to the A or C genome of *B. napus*. In addition, due to the prevalent homoeologous recombination events which have occurred during the evolution of *B. napus* (CHALHOUB *et al.* 2014) it was expected that some SNPs may map to alternate orthologous positions in different *B. napus* genotypes. Comparing the genetic and physical position for each of the SNP loci, it was found that 3% of the loci were genetically mapped to a homoeologous region relative to their physically defined coordinates (Figure 3-4). Further studies with the array are likely to uncover additional ambiguities that should be considered when utilising the array for analyses, especially when the loci cannot be genetically anchored in the population or specific genotypes being queried.

Although not specifically designed for this purpose, assessment of the array using DNA from the diploid genomes of *B. oleracea* and *B. rapa* demonstrated its value for genetic analysis of these two important vegetable crop species. Although, the genome-specificity of many of the markers could be seen as a disadvantage since those specific for the other genome result in failed assays, there are still 26,504 and 29,720 clearly scorable markers for *B. oleracea* and *B. rapa*, respectively. In addition, despite the analysed sample number being lower, 21,113 and 22,695 markers were shown to be polymorphic in *B. oleracea* and *B. rapa,* respectively, demonstrating the utility of the array for these two diploid species. Indeed this was confirmed through the recent generation of a genetic map for *B. oleracea* using the array (BROWN *et al.* 2014).

Although the array offers relatively good coverage of the *B. napus* genome with the SNP loci physically distributed across each of the chromosomes at an approximate

density of 1 marker every 15 Kb based on the diploid genome length, there was significant sub-genome bias observed with a higher density in the A compared to the C genome (one marker every 11 Kb cf. 19 Kb) (Table 3-2). The genetic map was based on a highly polymorphic cross allowing almost 50% of the SNP loci to be positioned on the *B. napus* genome. The genetic map showed only small gaps with five ≥ 9 cM. However, when considering the physical distribution of the mapped loci, each chromosome apart from N3 and N5 had at least one interval greater than 500 Kb, these larger intervals were also biased to the C genome with 89% (140/158) of such intervals being localized to the C genome. This could reflect differing levels of genetic variation between the sub-genomes of *B. napus*, as observed by others (DELOURME *et al.* 2013; QIAN *et al.* 2014) or may suggest further optimisation of the array should focus on selection of C genome loci. Nevertheless any bias in distribution of loci should be considered in downstream applications using the array. Once aligned to *B. napus*, this distribution did not change markedly and no large physical gaps were observed based on the overall marker selection (Figure 3-3). The saturated coverage is partly a reflection of the genome organisation, with extensive blocks of repetitive elements largely limited to the pericentromeric regions. In addition, although 34.4% (17,955) of the SNP loci fall within annotated genes the array design did not focus on functional SNPs, which can bias the marker distribution. In maize, a similar high-density array was developed that targeted genic regions and even bearing in mind the greater genome size there were significant gaps in the physical SNP coverage on many chromosomes of at least 1 Mb per chromosome (GANAL *et al.* 2011). Alignment of the physical and genetic maps for *B. napus* showed good collinearity; however, a number of rearrangements were noted on comparison with the diploid genomes. Although, some of these can be attributed to artifacts of the genome assembly process in each species, the larger rearrangements that are common to the two *B. napus* genomes may indicate chromosomal changes that have occurred since the fusion of the two progenitor genomes in the *B. napus* nucleus (Figure 3-5).

The Brassica 60K Infinium array provides a robust and efficient tool for genetic studies in *B. napus*. Due to the natural and breeding bottlenecks created in modern *Brassica* germplasm much emphasis is now placed on capturing the wider allelic diversity within the species gene pool (BUS *et al.* 2011). Genome wide association studies (GWAS) in a number of other crop species have suggested the value of such analyses for exploiting untapped variation to identify causative loci for key economic traits (ZHAO *et al.* 2011; COOK *et al.* 2012). An essential prerequisite for GWAS is the ability to query genome-wide polymorphisms that are spaced such that the analyses are not limited by the observable linkage disequilibrium (LD) in the species of interest. More recent estimates for *B. napus* suggest LD breakdown across the genome, ranging from 0.3-1.7 cM (DELOURME *et al.* 2013) and 0.25-2.5 Mb (QIAN *et al.* 2014) with LD decaying more rapidly in the A genome. The distribution of SNP loci across the genome, which lie well within current LD estimates, should facilitate the use of the array for GWAS or QTL mapping to identify genes underlying traits of interest. The demonstration of the utility of the array within *B. napus* as well as its diploid ancestors *B. oleracea* and *B. rapa* indicates that the developed array can be used in the entire crossing range of these three species, providing a valuable tool for Brassica breeding applications.

4. Detecting *de novo* homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array

Citation:

Erin E. Higgins, Wayne E. Clarke, Elaine C. Howell, Susan J. Armstrong, Isobel A. P. Parkin. Detecting *de novo* homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array**. G3: Genes|Genomes|Genetics August 2018 8:2673-2683**; doi: 10.1534/g3.118.200118

*All supplementary data, figures and tables are available through the Open Access article on the G3 website. https://doi.org/10.25387/g3.6207530

4.1 Abstract

The heavy selection pressure due to intensive breeding of *Brassica napus* has created a narrow gene pool, limiting the ability to produce improved varieties through crosses between *B. napus* cultivars. One mechanism that has contributed to the adaptation of important agronomic traits in the allotetraploid *B. napus* has been chromosomal rearrangements resulting from homoeologous recombination between the constituent A and C diploid genomes. Determining the rate and distribution of such events in natural *B. napus* will assist efforts to understand and potentially manipulate this phenomenon. The Brassica high-density 60K SNP array, which provides genome-wide coverage for assessment of recombination events, was used to assay 254 individuals derived from 11 diverse cultivated spring type *B. napus*. These analyses identified reciprocal allele gain and loss between the A and C genomes and allowed visualization of *de novo* homoeologous recombination events across the *B. napus*

genome. The events ranged from loss/gain of 0.09 Mb to entire chromosomes, with almost 5% aneuploidy observed across all gametes. There was a bias towards sub-telomeric exchanges leading to genome homogenisation at chromosome termini. The A genome replaced the C genome in 66% of events, and also featured more dominantly in gain of whole chromosomes. These analyses indicate *de novo* homoeologous recombination is a continuous source of variation in established *Brassica napus* and the rate of observed events appears to vary with genetic background. The Brassica 60K SNP array will be a useful tool in further study and manipulation of this phenomenon.

4.2 Introduction

The genomic relationship between the major Brassica species was first described by U (1935) and is defined by three diploid species: *B. rapa* (A genome), *B. nigra* (B genome) and *B. oleracea* (C genome); and three allotetraploids created from each pair-wise hybridisation of these genomes: *B. juncea* (A and B genomes), *B. napus* (A and C genomes) and *B. carinata* (B and C genomes). Of these Brassica species, *B. napus* (canola or oilseed rape) is the most economically important and is believed to have been formed in the last 10,000 years centered around Mediterranean Europe (CHALHOUB *et al.* 2014) and is now grown on all continents, harvested predominantly for its oil. The Brassica species provide an excellent platform for the study of genome evolution in polyploids since they encompass multiple ancient genome duplication events (CHALHOUB *et al.* 2014). These events include the gamma triplication event common to most eudicots, the α and β whole genome duplication common to all Brassicaceae, a Brassica lineage specific whole genome triplication that led to the formation of the Brassica diploids (or mesopolyploids), and most recently whole genome hybridisation resulting in the three allopolyploid (or neopolyploid) species (MASTERSON 1994; BOWERS *et al.* 2003a; LYSAK *et al.* 2005; SCHRANZ *et al.* 2006).

Polyploid formation leads to a phase of genomic shock in response to the duplication of all genes, with concomitant gene balance and regulatory issues (WENDEL 2000;

DOYLE *et al.* 2008). During meiosis, homologous chromosomes pair and the ensuing recombination facilitates the production of viable gametes, each with a complete set of chromosomes. However, in neopolyploids formed from closely related species, a chromosome may have more than one potential pairing partner, termed a homoeologue. Cytological analysis of meiotic cells in neopolyploids shows formation of bivalents, unpaired univalents and multivalents of homologous and homoeologous chromosomes (ATTIA AND RÖBBELEN 1986b). Recombination between homoeologous chromosomes where all or part of a chromosome from one genome is replaced with the homologous regions from the second genome can result in inheritance of either the recombined segments from both homoeologues, with no apparent loss of genetic material, or only one of the recombined segments, leading to gain and loss of genetic material. The latter events have often been termed homoeologous non-reciprocal translocations (HNRT), although by their nature they are derived from reciprocal exchange, to prevent confusion such events will be referred to as homoeologous recombination (HeR) events or exchanges. Multivalent and univalent formation can also lead to unbalanced gametes. During anaphase I unpaired chromosomes either move to one pole or are split between the poles by the spindle apparatus. The complicated dissolution of multivalents leads to unpredictable separation of chromosomes, most of which will result in unbalanced gametes (ZAMARIOLA *et al.* 2014b). Such gametes can be nonviable or the resultant embryos may produce plants that are sterile or unfit for their current environment resulting in an overall decrease in yield (JENCZEWSKI AND ALIX 2004). The mechanisms responsible for the genetic stabilisation (or diploidisation) of neopolyploids are of interest for maintaining fitness in crops, limiting gene flow to native plants, and exploiting the diploid gene pools of polyploid progenitors for novel traits.

Polyploidy is very common in plants, including several important crop species such as wheat, cotton, canola, coffee and peanut, and evidence exists for at least some level of genetic control of chromosome pairing in all of these species (CIFUENTES *et al.* 2010; LASHERMES *et al.* 2016; NGUEPJOP *et al.* 2016). The most well characterized of these

is wheat where chromosome pairing control has been studied since the 1950's after discovery of the *Pairing homoeologous1 (Ph1)* locus that had a major effect on the control of homoeologue pairing and recombination (RILEY AND CHAPMAN 1958). Using cytology it was observed that in plants lacking the *Ph1* locus there were more univalents and multivalents at metaphase I of meiotic cells rather than the typical prevalence of homologous bivalents, but a precise mechanism for this phenotype continues to be investigated (see (GREER *et al.* 2012; BHULLAR *et al.* 2014; MARTÍN *et al.* 2017; REY *et al.* 2017) for recent work). Though fewer studies have focused on the genetic control of pairing in Brassica, it is an excellent system for studying pairing control and homoeologous recombination because the allotetraploid species of the triangle of U (U 1935) can be recreated through crossing and subsequent chromosome doubling of the two constituent diploid species (SNOWDON 2007). Researchers have successfully used sequential fluorescence in situ hybridization (FISH) and genomic in situ hybridization (GISH) to distinguish the A and C genome chromosomes in *B. napus* (HOWELL *et al.* 2008), and a novel chromosome painting technique was used to identify all of the chromosomes from *B. rapa, B. napus* and *B. oleracea* (XIONG AND PIRES 2011). This makes it possible to not only identify homoeologous bivalents and multivalents but to pinpoint the chromosomes preferentially pairing in meiotic cells. Cytological analysis previously identified a major quantitative trait locus (QTL) that contributed to variation in homoeologous chromosome pairing in allohaploid *B. napus* plants of two genotypes (JENCZEWSKI *et al.* 2003). In addition, this locus appeared to impact homologous recombination; however, it did not seem to contribute to variable homoeologous pairing in allotetraploids (diploids) of the same *B. napus* lines (NICOLAS *et al.* 2009).

Molecular markers have previously been used to identify homoeologous recombination events in *B. napus* (PARKIN *et al.* 1995; SHARPE *et al.* 1995; UDALL *et al.* 2005; ROUSSEAU-GUEUTIN *et al.* 2017; STEIN *et al.* 2017). By visualizing both A and C genome loci with restriction fragment length polymorphism (RFLP) markers it was possible to resolve homoeologous recombination events by the gain of an allele at one

locus coupled with loss of an allele at the homoeologous locus (PARKIN *et al.* 1995). This simultaneous gain and loss of alleles at genetically linked loci on homoeologous chromosomes provided evidence of HeR events. Such analyses of a population derived from a cross between a newly resynthesized *B. napus* (created by crossing a *B. rapa* and *B. oleracea* line followed by chromosome doubling to produce an allotetraploid) and an established *B. napus* parent line showed a significant increase in homoeologous recombination between the A and C genomes relative to a population derived from a cross between two adapted *B. napus* parent lines (PARKIN *et al.* 1995; SHARPE *et al.* 1995). While highly reproducible the laborious nature of RFLP markers makes them difficult to assess for a large number of lines across the whole genome. Simple sequence repeat (SSR) markers have been used to show reciprocal gain and loss of A1 and C1 loci in progeny of a resynthesised *B. napus* (SZADKOWSKI *et al.* 2010) but SSRs offer only a marginal advantage in assay time compared to RFLP markers. The development of the Brassica 60K Infinium single nucleotide polymorphism (SNP) array (CLARKE *et al.* 2016) provides a high-density genome-wide platform to assess homoeologous recombination. SNP arrays allow genotyping of hundreds of lines at thousands of loci in a matter of days and have been successfully used for genetic mapping in bi-parental populations (LIU *et al.* 2013; WANG *et al.* 2015; YANG *et al.* 2017), differentiating between the different Brassica species of U's triangle (MASON *et al.* 2015), identification of parental alleles in interspecific Brassica species (MASON *et al.* 2014), and genome wide association studies (GWAS) on diverse sets of *B. napus* germplasm (HATZIG *et al.* 2015; KÖRBER *et al.* 2015). Use of the Brassica 60K array to identify segmental deletions in resynthesized *B. napus* has been combined with cytological analysis to identify translocations caused by homoeologous recombination in resynthesized *B. napus* individuals (ROUSSEAU-GUEUTIN *et al.* 2017). Similarly, the lack of amplification at physically linked SNP loci was used in conjunction with re-sequencing data to reveal homoeologous exchanges underlying QTL for *B. napus* seed quality traits (STEIN *et al.* 2017).

This paper describes use of the Brassica 60K SNP array to identify *de novo* homoeologous recombination events in allotetraploid *B. napus*. The high-density coverage provided by the array allows for genome-wide detection of recombination events at a greater depth and higher resolution than previous marker-based assays. The efficacy of this method was tested by assaying levels of *de novo* homoeologous recombination in 10 testcross populations derived from established *B. napus* lines. These data provide a range of expected levels for such events in *B. napus* and define genomic regions more prone to homoeologous recombination.

4.3 Materials and Methods

4.3.1 Testcross Population Development

Ten *B. napus* lines were chosen from a collection of spring-type cultivars (ACSRsyn1, Bronowski, Daichousen (fuku), Maris Haplona, PAK85912, Surpass 400, Svalof's Gulle, Topas, Tribune, Zhongyou 821) based on diverse geographical distribution (Canada, Poland, Korea, United Kingdom, Pakistan, Australia, Sweden, Canada, Australia, China, respectively) and where available molecular information (BUS *et al.* 2011). Formation of ACSRsyn1 was created by crossing a *B. napus*/*B. oleracea* triploid with a *B. napus*/*B. rapa* triploid (DTN-1/*B. alboglabra* 89-5402//DTN-1/*B. rapa* Parkland) and selecting for an individual with a complete AACC genome followed by selfing for several generations (provided by Sally Vail, Agriculture and Agri-Food Canada, Saskatoon). Plants were grown in a greenhouse at 18˚C with a 16/8 hour photoperiod (day/night). Hand pollinations were used to cross the *B. napus* individuals with the Australian *B. napus* cultivar "Rainbow" to produce a testcross population for each line. Young leaf tissue of sixteen individuals for each of the populations was harvested and freeze-dried for DNA extraction. Three of the testcross populations, PAK85912, Zhongyou821 and Maris Haplona were expanded to 48 individuals each, though two of the PAK85912 progeny were determined to be selfs and were disregarded from the data set.

4.3.2 Brassica SNP Array

High quality DNA was extracted from freeze-dried leaf tissue using a cetyltrimethylammonium bromide (CTAB) based method (MURRAY AND THOMPSON 1980). DNA was quantified with the Quant-it Picogreen dsDNA assay kit (Life Technologies Inc., Burlington ON, Canada) and 200 ng was hybridized to the Brassica 60K Infinium array (CLARKE *et al.* 2016) as described in the manufacturer's protocol (Illumina Inc., San Diego, CA). The arrays were scanned using an Illumina HiScan and SNP data was analysed using the genotyping module of the GenomeStudio software package (Illumina Inc.) using default settings with the exception of the no-call threshold, which was set at 0.05 and a custom cluster file was applied (CLARKE *et al.* 2016). The software creates a two-dimensional image for each SNP marker where the graphical position of each individual is determined by the fluorescent intensity (R value - y axis) and the ratio of the two allele-specific fluorophores (θ value – x axis). Individuals are assigned a genotype based on their position in the graph. The software is designed for diploid species with two alleles at each locus (AA/BB) so in a typical cross between two homozygous parents, a classic three cluster profile is produced, the AA and BB clusters would reflect the genotype of each parent and the AB cluster would represent heterozygous progeny. Single copy SNP markers were pre-selected by aligning the flanking sequence provided in the manifest file for the Brassica 60K array to the *B. rapa* and *B. oleracea* genome assemblies (WANG *et al.* 2011; PARKIN *et al.* 2014) using BLAT (KENT 2002) and selecting those markers with >90% identity in one diploid genome and <90% identity in the other diploid genome. This resulted in a set of 38,970 markers that were then filtered for polymorphism between the parents of each population. The level of SNP polymorphism for each testcross population is given in Table 4-1 for each linkage group and the GenomeStudio exported SNP data for each population is provided in Supplementary Tables S1-S10. Three or more consecutive missing or duplicated SNPs were used to identify affected regions; however, the majority of the events (91.5%) were defined by 10 or more physically linked SNP loci (Supplementary Table S11). HeR events were

identified by analysis of the homoeologous regions in the testcross individuals as described in the RESULTS.

Table 4-1: Summary of distribution of polymorphic SNP markers per chromosome for each testcross population

Linkage Group	ACSRsyn1	Bronowski	Daichousen	Maris Haplona	PAK85912	Surpass 400	Svalof's Gulle	Topas	Tribune	Zhouyou 821
A1 (21.7 Mb)	631	653	572	715	641	741	537	511	711	560
A2 (29.6 Mb)	380	274	443	441	441	331	344	432	264	455
A3 (35.8 Mb)	875	999	799	761	1012	941	837	849	776	1090
A4 (21.1 Mb)	653	670	565	755	804	655	720	746	623	618
A5 (25.7 Mb)	864	680	552	868	700	552	878	884	371	490
A6 (26.1 Mb)	787	552	665	779	648	790	728	759	692	772
A7 (25.5 Mb)	738	424	647	715	630	771	690	672	684	722
A8 (21.7 Mb)	576	383	280	599	617	605	388	575	523	441
A9 (40.5 Mb)	578	745	865	581	647	535	651	505	495	1051
A10 (17.9 Mb)	359	428	520	360	454	461	481	377	463	533
C1 (45.6 Mb)	1793	1803	1631	1914	820	1008	825	1842	1976	1683
C2 (47.3 Mb)	1145	2685	2087	1239	1857	1312	1162	1504	1316	1807
C3 (67.8 Mb)	1603	2098	1985	1620	1567	2060	1609	1631	1645	1492
C4 (55.1 Mb)	3062	1304	3315	1528	2434	3206	1273	3072	1837	3305
C5 (48.7 Mb)	970	974	572	969	828	627	915	898	706	916
C6 (40.8 Mb)	736	872	828	746	792	747	798	762	800	830
C7 (48.8 Mb)	939	802	1631	991	1204	627	1103	915	1236	1275
C8 (44.7 Mb)	1217	1196	1432	729	1109	942	1042	1120	1309	1379
C9 (56.0 Mb)	757	780	1126	786	753	730	872	690	766	966
TOTAL	18663	18322	20515	17096	17958	17641	15853	18744	17193	20385

4.3.3 Identification of homoeologous regions from the SNP array

The results of the BLAT alignment from aligning the flanking sequences from the SNP loci to the *B. rapa* and *B. oleracea* genome assemblies was also used to identify

the top hit in each of the A and C genomes for each SNP probe. Probes with at least 50% identity in both diploid genomes were selected, resulting in 28,334 SNP markers mapped to the A and C genome (Supplementary Table S12) that could be used to determine the homoeologous alignment of the A and C genomes.

4.3.4 Detection of inherited HeR events using whole genome shotgun (WGS) data

DNA from Zhongyou821 was extracted from nuclei according to Parkin et al (2014). A short insert (350 bp) Illumina DNA sequencing library was constructed according to the manufacturer's instructions (Illumina, Inc.) and 125 bp paired end (PE) data was generated on the HiSeq2000 platform, providing in total 111 million (M) PE reads (estimated 23x coverage of 1200 Mb genome). Trimmomatic v0.32 (BOLGER *et al.* 2014) with the following parameters, LEADING:15 TRAILING:15 SLIDINGWINDOW:4:15 MINLEN:55, was used to remove low quality reads, short inserts and adapter sequences, resulting in 103 M high quality PE reads. A combined pseudo-reference genome was generated from concatenating the *B. rapa* and *B. oleracea* genome assemblies (WANG *et al.* 2011; PARKIN *et al.* 2014). Bowtie2 v2.3.3.1 (LANGMEAD AND SALZBERG 2012) was used to align PE data to the pseudo-reference using the following parameters --local --sensitive --phred33 --minins 0 --maxins 1000 --no-mixed --no-discordant --no-unal --k 20 --dovetail. A custom perl script was used to retain the best alignment for each read as long as the next hit was significantly lower in call stringency. The overall alignment rate was 90.05%, with 92.8 M mapped PE reads. The resultant alignment file was analysed using the R scripts described in Samans et al (2017), which normalise the read depth across the length of each chromosome, identify regions of the genome where the read depth significantly differs from the chromosomal mean (at 1.5 SD), and finally compare homoeologous regions to define potential fixed HeR events.

4.3.5 Data Availability Statement

All supplementary data is available at https://gsajournals.figshare.com/. Table S1-10 contain the SNP marker data for all testcross populations. Table S11 lists all HeR, duplication and deletion events found in the testcross populations. Table S12 provides details of the SNP markers used for homoeologous alignment of A and C genomes. Table S13 shows the compressed SNP data which summarises *de novo* chromosome gain and loss in each testcross individual. Table S14 lists HeR events in Zhongyou821 identified through whole genome sequencing. The WGS data for Zhongyou821 has been uploaded to the NCBI short read data archive (https://trace.ncbi.nlm.nih.gov/) under BioProject ID PRJNA454160.

4.4 Results

4.4.1 Dissecting SNP marker patterns to identify homoeologous recombination events

Recombination rates for an individual can be determined by studying the products of meiosis or the genotypes of resulting progeny in the subsequent generation. Ten test populations were derived by crossing ten spring-type *B. napus* lines with the Australian *B. napus* cultivar Rainbow. In total, 256 individuals were assayed with the Brassica 60K SNP array, two lines were identified as self progeny. Initially 16 individuals for each of the 10 *B. napus* testcross populations were assessed and three populations with differing levels of observed events, PAK85912, Maris Haplona and Zhongyou821, were expanded to 46-48 individuals each. For each individual, the meiosis of both the *B. napus* line and the testcross parent Rainbow could be assessed, meaning the products of 508 meioses were evaluated.

Although the nature of polyploid genomes can impact the use of hybridisation based tools such as Infinium arrays, the Brassica 60K SNP array was designed such that ~58% of the loci on the array were estimated to amplify a single genome (CLARKE *et al.* 2016). However, conversely as much as 32% of the SNPs on the array identify both

an A and C genome locus. Although intuitively these loci might appear useful in the analyses of homoeologous recombination, as detailed in Mason et al (2017) since both loci use the same two fluorophores for the A and B alleles, they cannot readily be distinguished, thus such loci were eliminated from the analyses.

Since homoeologous recombination events are relatively rare, affecting only a small part of the genome, the most common pattern observed for single locus assays, approximately 90% of the time, in all populations was the typical three cluster pattern, AA/AB/BB, expected for an F_1 with no gain or loss of alleles in the testcross individuals. The two parents were found in the AA and BB clusters respectively and all the progeny were found in the heterozygote cluster, suggesting normal homologous recombination and segregation had occurred during meiosis in the testcross parents (Figure4-1A).

The gain and loss of alleles due to homoeologous recombination created more complicated cluster patterns, polymorphic SNPs expected to amplify a single genome (A or C), produced patterns with the expected three, but also one, four and five distinct clusters and were used to score the testcross populations. In each population, between 4-19% of the polymorphic markers showed these aberrant cluster patterns. These loci had the hallmarks of single copy SNPs, with parental alleles optimally separated (AA θ value <0.15, BB θ value >0.85) and heterozygote genotypes falling equi-distant between the two, yet additional clusters were observed across the horizontal plane. Those showing four distinct clusters had testcross individuals falling into three different groups: 1) with either one of the two parents; 2) in the expected AB cluster; or c) in a new cluster between the AB group and one of the parents (Figure4-1B). These patterns can be explained by the gain or loss of alleles due to homoeologous recombination in one of the parents. Segregation of testcross individuals with either parent indicates they are missing an allele that should have been inherited from the other parent. Since these lines only carry an allele from one parent they were designated as genotype A0 (or B0). Individuals in the expected

heterozygote cluster are presumed to have inherited one allele from each parent, and are therefore genotype AB. Based on the position of the new cluster between one of the parents and the AB group, those individuals are presumed to have inherited one allele from the first parent and two copies of the allele from the second parent and are therefore genotype ABB (or AAB). In other cases individuals were observed on both sides of the heterozygote cluster resulting in five clusters on the SNP image, indicating there were homoeologous exchanges occurring in both Rainbow and the other *B. napus* parent, which were inherited in some of the individuals, so all five genotypes, A0, AAB, AB, ABB and B0 are represented (Figure 4-1C). Lastly, some SNPs had a single cluster containing one parent and all of the testcross individuals while the other parent showed no amplification, indicating the absence of the SNP locus in one of the parents (Figure 4-1D).

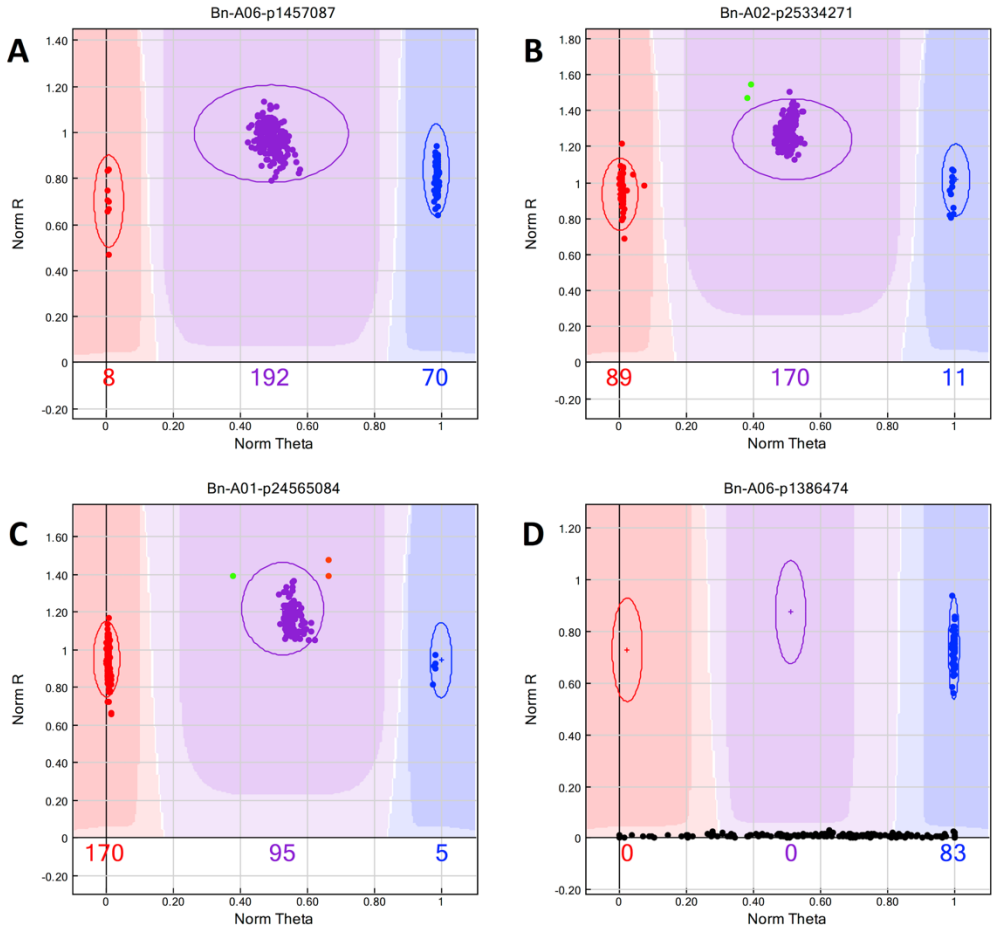


Figure 4-1: Type of observed SNP patterns in B. napus testcross individuals
A GenomeStudio image of a typical three cluster SNP with Parent 1 in red (genotype AA), Parent 2 in blue (genotype BB) and all testcross individuals of polymorphic populations in purple (genotype AB); **B** GenomeStudio image of four cluster SNP with two AAB individuals shown in green; **C** GenomeStudio image of five cluster SNP with one AAB individuals shown in green and two ABB individuals shown in Orange; **D** GenomeStudio image of a SNP with no amplification in Parent 1 therefore all testcross individuals in polymorphic crosses have genotype B0 and are in the Parent 2 BB cluster.

4.4.2 Defining the genomic position of the homoeologous recombination (HeR) events

Due to the genome specificity of the SNP assays and filtering of the SNP loci, the A and C genome loci were assessed independently, thus the gain/loss of homoeologous loci was not captured simultaneously by any one SNP assay. Homoeologous recombination (HeR) events by definition result in the exchange of chromosomal material between syntenic regions of the A and C genome within *B. napus*. Because only a single A or C genome locus can be scored for each SNP, it was necessary to first identify the syntenic regions between the genomes that can be identified by the SNP loci on the array, and then look for reciprocal allele gain and loss at SNP loci in those regions, thus determining the level and distribution of HeR events. The flanking sequences of the SNPs on the Brassica 60K array were aligned against the genomic sequence of *B. rapa* (A genome) and *B. oleracea* (C genome) which produced a clear alignment of homoeologues with comprehensive genome coverage that was in concordance with previous analyses (PARKIN *et al.* 2014) (Figure 4-2). In some cases two chromosomes were entirely aligned along their length, for example A1 and C1, while others had two or more homoeologous partners, such as A9 and A10, which were aligned with the top and bottom of C9, respectively.

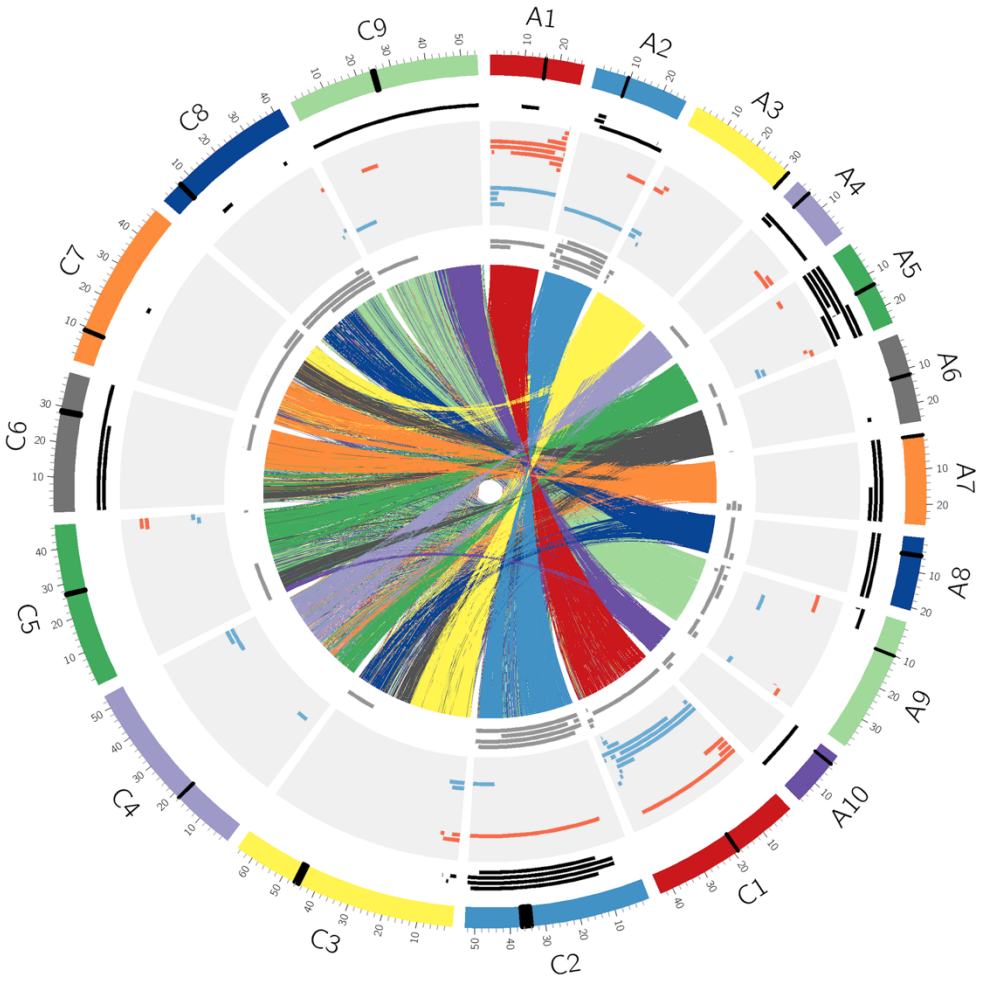


Figure 4-2 Circos Plot depicting alignment of the B. rapa and B. oleracea genomes and summary of de novo homoeologous recombination events

The *B. rapa* chromosomes A1-A10 and *B. oleracea* chromosomes C1-C9 are shown in the outer ring. Homoeologous regions between the chromosomes as identified by the sequence homology of SNP probes are shown as coloured links drawn in the centre of the image. The three rings internal to the chromosomes show from outer to inner: duplication events as black tiles; chromosomal gain and loss due to *de novo* HeR as red and blue tiles, respectively; and deletions as gray tiles. The vertical black line on each chromosome represents the approximate centromere position (WANG *et al.* 2011; PARKIN *et al.* 2014).

Genome-wide coverage and density of SNP markers along the chromosomes is important to ensure there is sufficient polymorphism between Rainbow and the 10 *B. napus* testcross parents because allele gain and loss can only be visualised when the two parents have opposing alleles. The large number of usable markers on the Brassica 60K array maximized the ability to identify homoeologous regions and determine if an event involved the two homoeologues or if only one genome was duplicated or deleted. The number of informative polymorphic markers in each testcross population is summarized in Table 4-1 and complete genotype information for all SNP markers is shown in Supplementary Tables S1-S10.

For each of the testcross individuals the genotypes for each A genome SNP locus were aligned with the genotypes for the SNPs from the homoeologous region of the C genome to allow ready visualization of the reciprocal allele gain and loss. For a given individual, when the SNP loci from say the C genome showed an extra copy of the C genome allele from one parent (scored as ABB or AAB), the allele from the corresponding region in the A genome was missing (scored as A0 or B0) (Supplementary Table S13). This gain/loss pattern was observed along both homoeologues for at least three physically linked SNP markers in order to confidently determine a HeR exchange had occurred. The number of physically linked loci identifying HeR affected regions ranged from 3-1879, with only four events defined by the minimum number of loci, and the majority of the rearrangements (91%; 151/165 affected regions) were identified by 10 or more physically linked loci (Supplementary Table S11).

4.4.3 Confirmation of fixed HeR events detected with the SNP array

Recent studies using whole genome (CHALHOUB *et al.* 2014; SAMANS *et al.* 2017) and transcriptome sequencing (HE *et al.* 2017; LLOYD *et al.* 2017) show evidence of historical HeR events which have become fixed in modern *B. napus* varieties. These types of events could be seen in the current analyses, where all individuals of a testcross family are not in the expected central heterozygous cluster for a set of

homoeologous SNP loci, but are biased toward one parent at SNP loci for one genome while at the homoeologous loci, all individuals are in the other parental homozygous cluster, indicating a duplication and deletion, respectively. Detection of such an event is shown in Figure 4-3 for a fixed HeR exchange between the A9/C9 chromosomes in the Zhongyou821 population. Whole genome shotgun sequencing of Zhongyou821 verified the presence of this event, as evidenced by a significant increase and decrease in normalised read depth as determined through sequence alignment to the A9 and C9 chromosomes, respectively (Figure 4-3). Of the eight such HeR events that could be resolved using the sequence analyses, five were detected in the SNP array data, ranging in size from one to nine Mb (Supplementary Table S14). These data provided confirmatory evidence that SNP array analysis of F_1 populations can be used to identify historical HeR fixed events in *B. napus* lines.

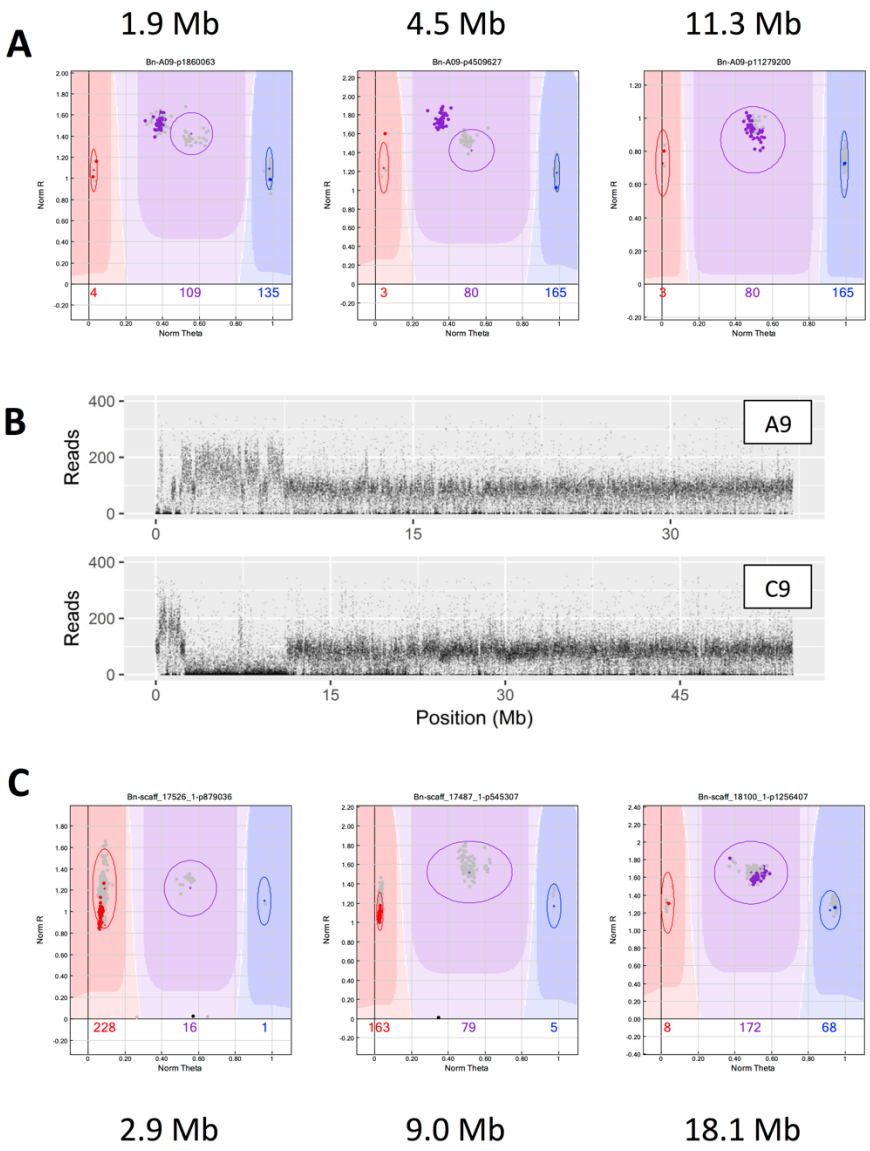


Figure 4-3: Identification of a fixed HeR event in Zhongyou821 using SNP array data and confirmed through whole genome re-sequencing

A GenomeStudio images from A9; the SNP loci sampled at 1.9 Mb and 4.5 Mb show all testcross individuals (purple) in the AAB cluster biased toward the Zhongyou 821 parent (red) and away from Rainbow (blue). Individuals in gray in the expected AB cluster are from other testcross populations without the genome rearrangement. Below the rearrangement (example at 11.3 Mb) normal SNP patterns were observed. **B** Plot of normalized sequenced read depth along the A9/C9 chromosomes from whole genome shotgun sequencing data of Zhongyou821. On A9 the first Mb is deleted and 2-8 Mb is duplicated, as opposed to C9 where the first Mb is duplicated and 2-11 Mb is deleted. **C** GenomeStudio images from C9 showing no amplification of loci in Zhongyou821 (black) at 2.9 Mb and 9.0 Mb, while all testcross individuals (red) cluster with the Rainbow parent. At 18.1 Mb the parents are in the AA and BB clusters and the testcross lines (purple) are AB. The lone individual to the left is Zhongyou821 testcross line #26, which has an extra C9 chromosome and therefore still has the genotype AAB.

4.4.4 De novo homoeologous recombination in *B. napus*

The focus of the study was to identify *de novo* rearrangements and all further discussion and calculations are based on new HeR exchanges. Such events are seen only in a single individual in a given testcross population and are presumed to have happened during the meiosis that produced the F_1 gamete. Similarly in cases where the same duplication or deletion event was seen in more than one individual within a population the event was assumed to have been present but not yet fixed (or functionally heterozygous) in the parental line and thus only inherited in a subset of the population and was not considered in subsequent analysis. Generally these events seemed to occur at similar levels to *de novo* events and testcross populations with high levels of *de novo* HeR also exhibited high levels of these segregating events (Tables S1-S10).

The products of 508 meioses were examined and 36 *de novo* events were found where both the deletion and duplication of SNP loci in the corresponding homoeologous regions could be detected (Figure 4-2, Table 4-2). In 54 instances only deletion of linked loci was observed and similarly in 39 cases only duplicated linked loci were evident (Figure 4-2, Table 4-2). These unpaired duplications and deletions were not unexpected since cytological observations of pairing control in newly resynthesized *B. napus* has frequently shown bivalents, multivalents and unpaired univalent chromosomes (ATTIA AND RÖBBELEN 1986b; SZADKOWSKI *et al.* 2010; ROUSSEAU-GUEUTIN *et al.* 2017), so it was expected there could be allele gain and loss due to the complicated dissolution of multivalents, and the segregation of unpaired chromosomes. Although such events indicate aberrant pairing only the 28% of observed events with the reciprocal gain and loss of multiple physically linked loci could be reliably attributed to homoeologous recombination in the testcross parent(s). Of the 93 events for which only the gain or loss was visible in the marker data, 27% could be attributed to aneuploid individuals with an additional (13) or missing (12) chromosome (Table 4-2). As identified in previous studies (ROUSSEAU-GUEUTIN *et al.* 2017; STEIN *et al.* 2017) deletion events are easily visible in the genotype data output

from GenomeStudio, but duplication events less so. The current assay identifies individual testcross lines with a duplication based on their θ value separation from the rest of the population caused by a difference in relative fluorescence of the two SNP alleles (Figure 1B and C). However, since the software can only call AA/AB/BB genotypes these individuals are either automatically called as AB or as a missing value. In the genotype output, multiple linked markers with mismatches and missing calls can indicate a potential duplication, but this must be validated by studying the individual SNP images making it more difficult to identify duplications, particularly very small ones. The smallest HeR event observed was 0.09 Mb, the smallest deletion was 0.04 Mb and the smallest duplication was 0.28 Mb (Table S11). The largest HeR was 40.6 Mb, the largest deletion was 28.6 Mb, and the largest duplication was 35.3 Mb. The size and position of all *de novo* HeR events, deletions and duplications observed in the testcross populations is summarized in Supplementary Table S11 and visualised in Figure 4-2.

Table 4-2: Summary of recombination events in B. napus testcross populations

Line	Number of Individuals	Duplication	Deletion	HeR	Aneuploid	TOTAL
ACSRsyn1	16	2	2	0	0	4
Bronowski	16	0	1	0	0	1
Daichousen(fuku)	16	0	0	1	1	2
MarisHaplona	48	1	8	0	1	10
Surpass400	16	0	0	0	0	0
Svalof'sGulle	16	1	0	0	1	2
Tribune	16	2	2	0	1	5
Topas	16	0	1	0	1	2
PAK85912	46	5	4	5	3	17
Zhongyou821	48	3	2	0	4	9
Rainbow	254	12	22	30	13	77

As expected chromosome exchanges at the ends of chromosomes that require only one recombination (61 events) were more common than internal exchanges that require a second recombination (43 events). Many of the HeR events (36%) as seen for A3/C3, A4/C4, A5/C5, and A9/C8 were effectively terminal, involving gain/loss of the chromosome ends (Figure 4-2). It was also noted that 12 events spanned the

centromere position (excluding aneuploids) and 25 events appeared to have breakpoints localised to the centromere position (Figure 4-2). Though centromeric breakpoints are not likely due to meiotic recombination, they are presumably a consequence of pairing between homoeologues and subsequent breakage during anaphase I and are therefore still indicative of meiotic abnormalities. In total, including aneuploids, A genome chromosomes or regions were gained 50 times and lost 44 times, while the C genome was duplicated 25 times and lost 46 times. This discrepancy is largely due to the fact that in 66% of the HeR events the A genome chromosome was duplicated and the C genome homoeologue was lost (Table 4-3).

Table 4-3: Summary of recombination events on each chromosome from the 10 B. napus testcross populations

Chromosome	Duplication	Deletion	HeR Gain	HeR Loss	Aneuploid	TOTAL
A1	3	1	13	4	1	22
A2	3	6	1	1	4	15
A3	0	3	2	3	0	8
A4	1	1	2	0	1	5
A5	6	1	3	2	2	14
A6	1	2	0	0	0	3
A7	1	3	0	0	2	6
A8	0	1	0	0	3	4
A9	2	7	3	2	0	14
A10	0	2	0	0	1	3
C1	0	3	4	13	1	21
C2	2	2	1	1	5	11
C3	3	1	3	2	0	9
C4	0	0	0	3	0	3
C5	0	2	2	2	0	6
C6	1	2	0	0	1	4
C7	1	2	0	0	1	4
C8	2	2	1	2	2	9
C9	0	1	1	1	1	4
TOTAL	26	42	36	36	25	165

Seventeen of the HeR events with concomitant gain and loss of genetic material were between A1 and C1 (47%) though unpaired deletion and duplication events for A1 and C1 were not disproportional (5% and 4%, respectively) compared to other

chromosomes (Table 4-3). The other 53% of HeR events were not distributed evenly across the chromosomes. A9 was involved in five events, three with C8 and two with C9, and though C9 did not recombine with it's other potential pairing partner A10, this association was previously observed in allohaploid *B. napus* so pairing of these chromosomes is certainly possible (GRANDONT *et al.* 2014). All chromosomes had at least one event, though C4 had no unpaired duplications or deletions but had three gain/loss HeR events and A6, A7, A8, A10, C6 and C7 had no gain/loss HeR. Curiously one Maris Haplona individual (#19) had inherited an extra copy of both A7 and C6 from Rainbow. Similar to the bias observed for the HeR events, missing or extra copies of A1/C1 and A2/C2 represented almost half (11/25) of the 25 aneuploid lines, for the reminder of the genome there were more, 9 compared to 5 aneuploid lines involving the A genome, of which 3 lines had an extra A8 chromosome.

Though the *B. napus* lines used are established breeding lines they showed varying rates of *de novo* homoeologous recombinaton. Rainbow and PAK85912 had the highest rates with 22% of individuals having at least one deletion, duplication or gain/loss HeR and Surpass400 had the lowest rate with no events in the 16 individuals tested. The ACSRsyn1 line was produced through interspecific crossing (see materials and methods for pedigree information) and was expected to have a high rate of homoeologous recombination, but only four events were observed in the 16 testcross lines, which is equivalent to the rates observed in Rainbow and PAK85912 and presumably results from stability selected over multiple generations.

One of the advantages of using the SNP array to detect HeR events is the relatively dense marker coverage as compared to RFLP or SSR markers. Smaller exchanges can be identified and the marker density allows for more precise physical positioning of the breakpoints. Calculating the distance from the identified recombination event to the nearest segregating marker gives an indication of how the density of the markers is an improvement over older marker technologies. In this data set the smallest

interval to which the recombination could be positioned was estimated to be 172 bp, while the largest was 8.7 Mb with an average size of 0.4 Mb (Table S11).

4.5 Discussion

Though *B. napus* is an excellent system for studying homoeologous recombination, relatively few studies have analysed the prevalence of *de novo* events, in part because of the limits of molecular marker technology. However, the high-density array formats for SNP markers make them ideal for quantifying homologous and homoeologous recombination rates, and the depth of coverage of markers on the array helps to overcome the limitations of studying each genome independently.

In the current study, *de novo* HeR was assayed in 508 meioses derived from 11 established *B. napus* lines, of these PAK85912 and Rainbow proved the most unstable. For each line, between 21-22% of the testcross individuals showed evidence of *de novo* segmental chromosome duplication or loss (excluding aneuploidy) presumed to be the result of homoeologous recombination or homoeologous associations. Previous studies of AC amphihaploids have suggested genotype contributes to the observed rates of HeR in *Brassica* species (ATTIA AND RÖBBELEN 1986a; ATTIA AND RÖBBELEN 1986b; JENCZEWSKI *et al.* 2003). The study by Udall et al (2005) examined three populations from natural *B. napus* crosses and found the rate of homoeologous recombination in one population was more than twice that of the other two (1.09%, 0.49% and 0.43% of all recombination events were homoeologous). In the 46 PAK85912 testcross lines there were five confirmed HeR events as well as five deletions and 4 duplications that are likely caused by chromosome mispairing, recombination and segregation. Rainbow had 30 HeR exchanges, 22 deletions and 12 duplications in the 254 individuals studied. In contrast, Zhongyou821 only had five duplication/deletion events with no HeR observed in 48 testcross lines. Considering all populations and the increased marker depth, the observed levels are comparable to those seen in a study using RFLP markers by Sharpe et al (1995), where HeR were identified in four of 174 individuals

from a *B. napus* DH population derived from a Canadian spring-type and a European winter line. Further studies will be required to assess the true range in absolute rates of *de novo* HeR in established *B. napus;* however, the cumulative evidence that genotype contributes to variation in HeR in established *B. napus* suggests the evolution of a genetic mechanism(s) to control levels of homoeologous recombination, possibly inherited from the progenitor diploids, and the variation between genotypes implies the quantitative nature of this trait.

As observed in other studies of HeR, including fixed and *de novo* events, the genome of origin and chromosome position contribute significantly to prevalence and directionality of the exchange (XIONG *et al.* 2011; CHALHOUB *et al.* 2014; HE *et al.* 2017; ROUSSEAU-GUEUTIN *et al.* 2017; SAMANS *et al.* 2017). Although all chromosomes showed evidence of *de novo* events, 38.8% of all 129 *de novo* events were contributed by the two pairs of homoeologous chromosomes (A1/C1 and A2/C2), which are syntenic along their entire length, indicating the importance of homology in determining efficiency of chromosome exchange. The gain and loss of sub-telomeric regions of homoeologous chromosomes (A3/C3, A4/C4, A5/C5 and A9/C8) was also more widely prevalent, a phenomenon previously referred to as "homogenisation" (SHARPE *et al.* 1995) and corroborates previous work that showed an increase in events with distance from the centromere (NICOLAS *et al.* 2012). However, it was also noted that 25 events were co-positioned with centromere locations (Figure 4-2). Centromeres tend to form breaks in ancestral karyotype blocks (PARKIN *et al.* 2005) suggesting they may act as evolutionary breakpoints and chromosome fusion and fission has almost certainly played a significant role in shaping Brassica genomes (FRIEBE *et al.* 2005; SCHRANZ *et al.* 2006).

The A genome dominated HeR events such that it replaced the C genome in 66% of events, this bias held with segmental duplications, with 17 of the 26 duplications involving A genome chromosomes. In contrast to Samans et al (2017), *de novo* segmental deletions were found to be higher in the A genome, 27 of the 42 deletions

were from the A genome, though the chromosomal material lost was 120 Mb as compared to the C genome, which lost 149 Mb through deletions. This could represent ascertainment bias from the SNP array since proportionally there are a lower number of informative markers per Mb derived from the C genome making it more difficult to detect small deletions (CLARKE *et al.* 2016). It was proposed in Samans et al (2017) that since the A genome is ~25% proportionally smaller than the C genome that the bias towards overall loss of the C genome supports observations of genome size reduction in neo-polyploid evolution, and this was reflected in cumulative amounts of genetic material lost and gained between the sub-genomes (Figure 4-4). He et al (2017) also evoked the known prevalence of interspecific crossing with *B. rapa* in *B. napus* breeding strategies, which could have led to biased capture of A genome regions by lone C genome chromosomes forming aberrant pairing structures in AAC triploids. One intriguing possibility for the apparent instability of the C genome in *B. napus* is the suggestion that paternal genomes are disproportionally affected in neopolyploids (LIM *et al.* 2007), which would align with suggestions that a relative of *B. rapa* was the possible maternal progenitor of *B. napus* (ALLENDER AND KING 2010).

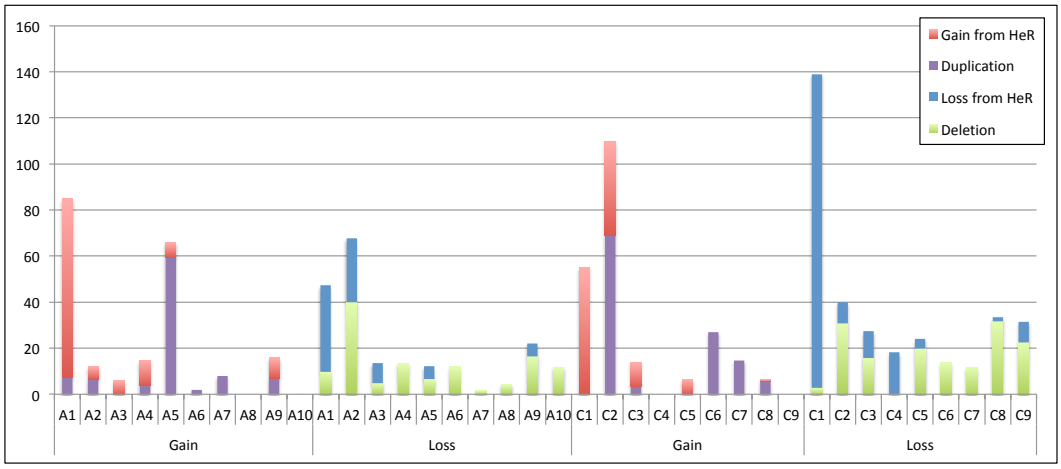


Figure 4-4: Summary of genomic gain and loss due to deletion, duplication and HeR events for each chromosome
Gain and loss of genetic material (in Mb) due to HeR is shown in red and blue, respectively. Segmental deletions and duplications are shown in green and purple, respectively.

Twenty-five aneuploid events were found among the 508 meioses, interestingly the A genome once more featured more dominantly than the C genome, with 14 of the events involving seven of the ten A genome chromosomes, and nine of which resulted in chromosome additions. As with other events A1/C1 and A2/C2 accounted for almost 50% of the events; such aneuploids were prevalent in early generations of resynthesized *B. napus* and chromosome compensation between the highly syntenic chromosomes in the polyploid nucleus is proposed to provide genome balance in the monosomic lines (XIONG *et al.* 2011). Although aneuploids are commonly produced in neopolyploids (COMAI 2005) and allopolyploids are known for their genome plasticity (LEITCH AND LEITCH 2008) the level observed (~5% across all gametes) in established *B. napus* might appear high. However, none of the observed events including the aneuploids are fixed in the testcross lines, and it would be expected that continual selection, for example in the field, for highly fertile euploids would invariably negate the presence of aneuploids and other rearrangements unless they led to a selective advantage. It would be interesting to study the effect of different parental genotypes in the development of hybrids, since it might be expected that high levels of meiotic instability would impact hybrid vigour.

The importance of homoeologous recombination in the developmental history of modern *B. napus* and its establishment as a major oilseed crop became apparent through sequence analyses of the *B. napus* genome that showed a number of important traits for Brassica oil and meal quality were derived from homoeologous recombination events (CHALHOUB *et al.* 2014). An indication of continuing genome evolution in *B. napus* was shown when a homoeologous recombination event was found to be segregating in replicate lines from the variety used to generate the *B. napus* genome assembly (LLOYD *et al.* 2017). The current study has shown that these events continue to happen frequently in natural *B. napus*, offering the opportunity to generate novel variation that could be exploited for crop improvement. The use of high-density SNP arrays has become the standard method for genetic mapping in

segregating populations, genotyping of elite lines, studies of genotypic diversity, and identification of chromosomal deletions (LIU *et al.* 2013; HATZIG *et al.* 2015; KÖRBER *et al.* 2015; MASON *et al.* 2015; ROUSSEAU-GUEUTIN *et al.* 2017; STEIN *et al.* 2017; YANG *et al.* 2017). The method described in this paper extends the use of SNP array to comprehensively measure the inheritance of homoeologous recombination events. The coverage of the Brassica 60K SNP array makes selection across the whole of the *B. napus* genome possible; in combination with cytological analysis this will present a complete picture of chromosome pairing at meiosis and the resulting homoeologous recombination that affects gamete viability, phenotypic variation and plant fitness. The SNP coverage in concert with the available *B. napus* genome sequence (CHALHOUB *et al.* 2014) also allows for a precise identification of recombination hotspots and regions of recombination repression.

5. A major quantitative trait locus on chromosome A9, *BnaPh1*, controls homoeologous recombination in *Brassica napus*

Citation: Erin E. Higgins, Elaine C. Howell, Susan J. Armstrong, Isobel A. P. Parkin. A major quantitative trait locus on chromosome A9, *BnaPh1*, controls homoeologous recombination in *Brassica napus*. **New Phytologist. 2018 229: 3281-3293;** doi: 10.1111/nph.16986

*All supplementary data, tables and figures referred to in this chapter are available through New Phytologist within the open access article. nph16986-sup-0001-FigS1-MethodsS1.pdf; nph16986-sup-0002-TablesS1-S7.xlsx

5.1 Summary

- Ensuring faithful homologous recombination in allopolyploids is essential to maintain optimal fertility of the species. Variation for the ability to control aberrant pairing between homoeologous chromosomes in *Brassica napus* has been identified. The current study exploited the extremes of such variation to identify genetic factors differentiating newly resynthesized *B. napus*, which is inherently unstable, and established *B. napus*, which has adapted to largely control homoeologous recombination.
- A segregating *B. napus* mapping population was analysed utilising both cytogenetic observations and high throughput genotyping to quantify the levels of homoeologous recombination.
- Three quantitative trait loci (QTL) were identified that contribute to the control of homoeologous recombination in the important oilseed crop *B. napus*. One major QTL on BnaA9 contributed between 32 and 58% of the observed

variation. This is the first study to assess homoeologous recombination and map associated QTLs resulting from deviations in normal pairing in allotetraploid *B. napus*.

- The identified QTL regions suggest candidate meiotic genes that could be manipulated in order to control this important trait and further allow the development of molecular markers to utilise this trait to exploit homoeologous recombination in a crop.

5.2 Introduction

The pervasiveness of polyploidy throughout the plant kingdom provides a clue as to the importance of genome duplication, hybridization and rearrangement in evolution. Over 70% of angiosperms are characterised as polyploids, though all plants are believed to have gone through whole genome duplication (WGD) at some point in their evolutionary history (MASTERSON 1994; ALIX *et al.* 2017). Plants can be broadly divided into three categories: diploids such as rice (*Oryza sativa*) where subsequent to ancient WGD events chromosomes have fused and undergone further rearrangements to the extent that they appear unique and behave independently during meiosis; autopolyploids, such as alfalfa (*Medicago sativa*) where all chromosomes have been duplicated at least once within a single nucleus; and allopolyploids, where genomes from related species have hybridised within a single nucleus, including the important monocot and dicot crop species, wheat (*Triticum aestivum*) and canola (*Brassica napus*), respectively. Like all organisms, survival of a polyploid species depends on the ability to produce viable offspring. In the allotetraploid *B. napus*, each chromosome has an identical partner (homologue) and one or more closely related partners (homoeologue) making the process of pairing, recombination and separation more difficult. *Brassica napus* was formed from the fusion of a Brassica A (n = 10) and C (n = 9) genome diploid progenitor and marker and sequence analyses have highlighted the close similarity between these genomes; however, their homoeologous relationship is not straightforward (CHALHOUB *et al.* 2014). Chromosomes such as BnaA1 and BnaC1 are aligned along their entire length

while others have more than one potential pairing partner, for example BnaA9 shares homology with both BnaC8 and BnaC9. A summary of the primary homoeologous regions of *B. napus* is shown in Figure **S1**.

Meiosis is conserved amongst sexually reproducing organisms and is tightly controlled. Recombination during meiosis ensures reduction to a haploid set of chromosomes in each cell and creates genetic diversity (MERCIER *et al.* 2015). Briefly, in most plants during prophase I of meiosis I, an axis forms along each replicated chromosome and double-strand breaks (DSB) occur. Through a process that is not fully understood homologues find each other, and the synaptonemal complex that holds them together is constructed while DSBs are repaired. If the repair of a DSB involves a non-sister chromatid, the result is a crossover (CO) or a non-crossover (NCO). A CO involves a reciprocal exchange between two homologous non-sister chromatids. The length of chromatid exchanged can be considerable, from the CO to the telomere or to the next CO, depending on which chromatids are involved. Subsequently chromosomes condense and the synaptonemal complex is dismantled but homologues are held together as bivalents at COs, seen cytogenetically as chiasmata. A minimum of one chiasma per bivalent is essential for correct chromosome alignment at metaphase I (MI). Cohesion is then lost between chromatid arms, allowing the homologues to move to opposite poles using the spindle apparatus. During meiosis II, chromatids separate and four haploid cells are produced. Because this paper focuses on COs, the term 'recombination' is used in the context of COs solely, not NCOs.

Aberrant recombination between homoeologues can cause aneuploidy as homoeologous bivalents and multivalents and unpaired univalents may not align correctly at MI. Additionally, the exchange can lead to gain/loss of homoeologous genome segments when chromatids segregate. Such rearrangements can be stably inherited and these changes can alter the balance of gene expression, impact chromosome conformation and facilitate further homoeologous mispairing during

subsequent meioses (GAETA *et al.* 2007; GAETA AND PIRES 2010). Thus, for long-term genomic stability and offspring fitness, neopolyploids must derive mechanisms to allow them to behave genetically as diploids during meiosis. In some cases these rearrangements can be advantageous in creating novel phenotypes, for instance, a deleterious allele may be replaced with its functional homoeologue or vice versa. Some of the first genetic maps provided evidence of fixed and *de novo* homoeologous recombination events in *B. napus* (PARKIN *et al.* 1995; SHARPE *et al.* 1995) and sequencing of the genome revealed several historical homoeologous non-reciprocal exchanges, including one responsible for the low glucosinolate seed content that was instrumental for the development of *B. napus* as a major oilseed crop worldwide (CHALHOUB *et al.* 2014). Current molecular techniques have uncovered relatively high rates of *de novo* homoeologous recombination events in some modern cultivars (HIGGINS *et al.* 2018) indicating that it is still an important mechanism in *B. napus* evolution.

Knowledge of the mechanisms that regulate chromosome pairing, synapsis and homoeologous recombination is important to enable their manipulation to improve crop diversity and productivity of allopolyploids. The species studied most extensively is hexaploid wheat (*Triticum aestivum*) since a line with a large deletion on chromosome 5B showed aberrant chromosome pairing at meiosis (RILEY AND CHAPMAN 1958; SEARS AND OKAMOTO 1958). The locus controlling this phenotype, *Ph1* (Pairing homoeologous 1), continues to be the subject of research (see MARTÍN *et al.* (2017); REY *et al.* (2017) for recent work). Minor effect loci have also been identified, including *Ph2* (MELLO-SAMPAYO 1971) and loci implicated in either suppressing homoeologous or promoting homologous chromosome pairing and recombination (reviewed in JENCZEWSKI AND ALIX (2004)). Less is known for *B. napus,* but one locus, *Pairing regulator in B. napus* (*PrBn*), that is involved in the control of homoeologous chromosome pairing in *B. napus* haploids (AC), was discovered by exploiting natural variation for high and low allo-syndetic pairing in haploids of two *B. napus* lines (JENCZEWSKI *et al.* 2003). This locus was mapped to chromosome BnaC9, and several

minor loci with additive or epistatic effects were also identified (LIU *et al.* 2006). However, both AACC parental lines had regular bivalent pairing (JENCZEWSKI *et al.* 2003), making the role that this locus plays in homoeologous recombination in the allopolyploid unclear.

Resynthesized *B. napus* lines, created by crossing *B. oleracea* with *B. rapa* and doubling the chromosome complement, have much higher rates of homoeologous recombination compared to established lines (PARKIN *et al.* 1995; SHARPE *et al.* 1995; UDALL *et al.* 2005; GAETA *et al.* 2007; SAMANS *et al.* 2017). Similarly, a high frequency of homoeologous bivalents and multivalents, identified by labeling the C genome with a BAC probe, was observed in resynthesized lines (SZADKOWSKI *et al.* 2010). The current study exploits the difference between established and resynthesized lines of *B. napus* to map quantitative trait loci (QTL) controlling homoeologous chromosome recombination in a segregating doubled-haploid population. Using both cytogenetic assessment of homoeologous events and quantification of homoeologous recombination events with high density SNP genotyping, one major and two minor QTL loci were genetically mapped. Candidate meiosis-specific genes underlying these QTL were identified by searches of genome annotation and supplementary gene expression data.

5.3 Materials and Methods

5.3.1 Plant Material

All experiments used individuals from a *B. napus* doubled-haploid (DH) population known to segregate for homoeologous recombination, hereafter referred to as the SGDH population (CLARKE *et al.* 2016). This population was created from two F_1 lines derived by reciprocal crossing of DH12075, a DH spring-type *B. napus* derived from a cross between Westar and Cresor, and PSA12, a resynthesized *B. napus* derived from a cross between *B. oleracea* line A12 and *B. rapa* line PS270. Individuals from this population were selfed for three or four generations prior to use for the

cytogenetic analysis or testcross population development, respectively. Details of growth conditions, crossing strategy, tissue harvest and sample preparation of plant material is provided in the Supplementary Methods.

5.3.2 Cytogenetic Analysis, SNP Array Analysis and QTL Mapping

DH12075 and 43 SGDH lines were examined cytogenetically using fluorescence *in situ* hybridization (FISH) followed by genomic *in situ* hybridization (GISH) on slides prepared from anthers with meiocytes at late diakinesis/MI (Supplementary methods) (Figure 5-1). The majority of lines were represented by one plant except three lines (SG-201, SG-309, SG-324) which had two plants examined and two lines (SG-152, SG-159) which had three plants examined for a total of 50 SGDH plants. A total of 3125 meiocytes was assessed, averaging 61 per plant (Table S1). Using FISH with 45s rDNA and a BAC probe (BoB061G14) the A and C genome chromosomes were divided into three and four groups, respectively, and within these groups, some individual chromosomes could be identified by the size and position of the 45S rDNA signal (Table 5-1). Linkage groups were assigned by comparison to published karyotypes (HOWELL *et al.* 2002; HOWELL *et al.* 2008; XIONG AND PIRES 2011). Homoeologous bivalents, trivalents, quadrivalents and higher order multivalents identified by FISH and GISH were recorded for each meiocyte. The score for each configuration is effectively the number of chromosome pairs involved minus one, e.g. a quadrivalent of two homoeologous chromosome pairs is given a score of '1' for 'one synaptic partner switch (SPS) accompanied by at least one CO (SPSC)' and a multivalent involving three pairs of chromosomes is scored as '2', etc. This is a conservative estimate of homoeologous recombination because accurate counting of chiasmata was not possible, so the number of chiasmata above one per SPS is not assessed. A detailed description of the method is provided in the supplementary methods and illustrated in Figure 5-2. The number of SPSC in each meiocyte was calculated and the total for each plant divided by the number of meiocytes scored gave a mean number of SPSC per meiocyte for each plant (Table S1). This mean (or the

cumulative mean for the five lines where more than one plant was analysed) was used as the trait value for QTL mapping.

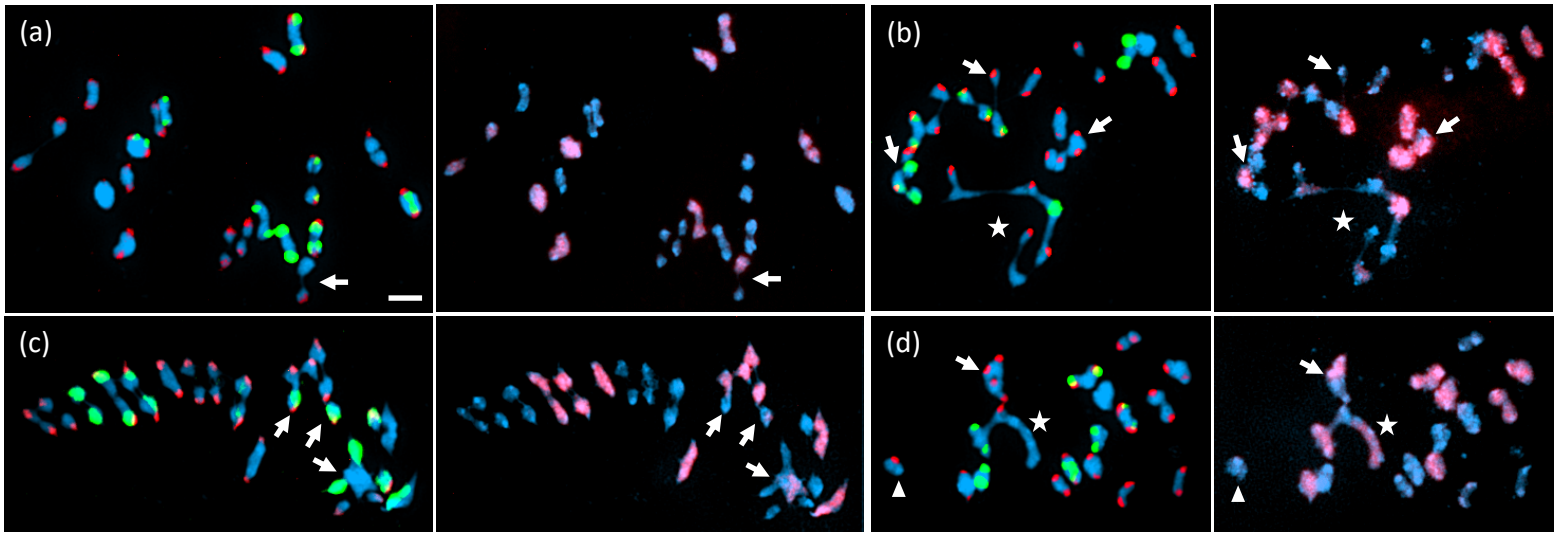


Figure 5-1: Examples of chromosome spreads from B. napus SGDH lines
Chromosome spreads of meiocytes at late diakinesis/MI examined by FISH (left image of each pair) with 45S rDNA (green) and BoB061G14 (red), and GISH (right image of each pair) with labeled C genome (red), DAPI-stained A genome (blue). Homoeologous bivalents and quadrivalents (arrows), multivalents (stars) and a univalent whose partner is in a multivalent (triangles) are highlighted. (a) SG-261, a single C9 in a bivalent with an extra A10; (b) SG-5, two AG CGS bivalents, one AG CG bivalent, one multivalent (A10 C9 A9 C8 C8 A9 C9 A10); (c) SG-235, two A1 C1 homoeologous bivalents orientated with both A1 facing the same pole and both C1 facing the other, one A3 A3 C3 C3 quadrivalent; (d) SG-25, one AG AG CG CG quadrivalent, one multivalent (C8 C8 A9 C9 C9) with univalent A9. The 45S rDNA signal on A9 in (b), (c), and (d) is not visible at this exposure. Bar 5µm.

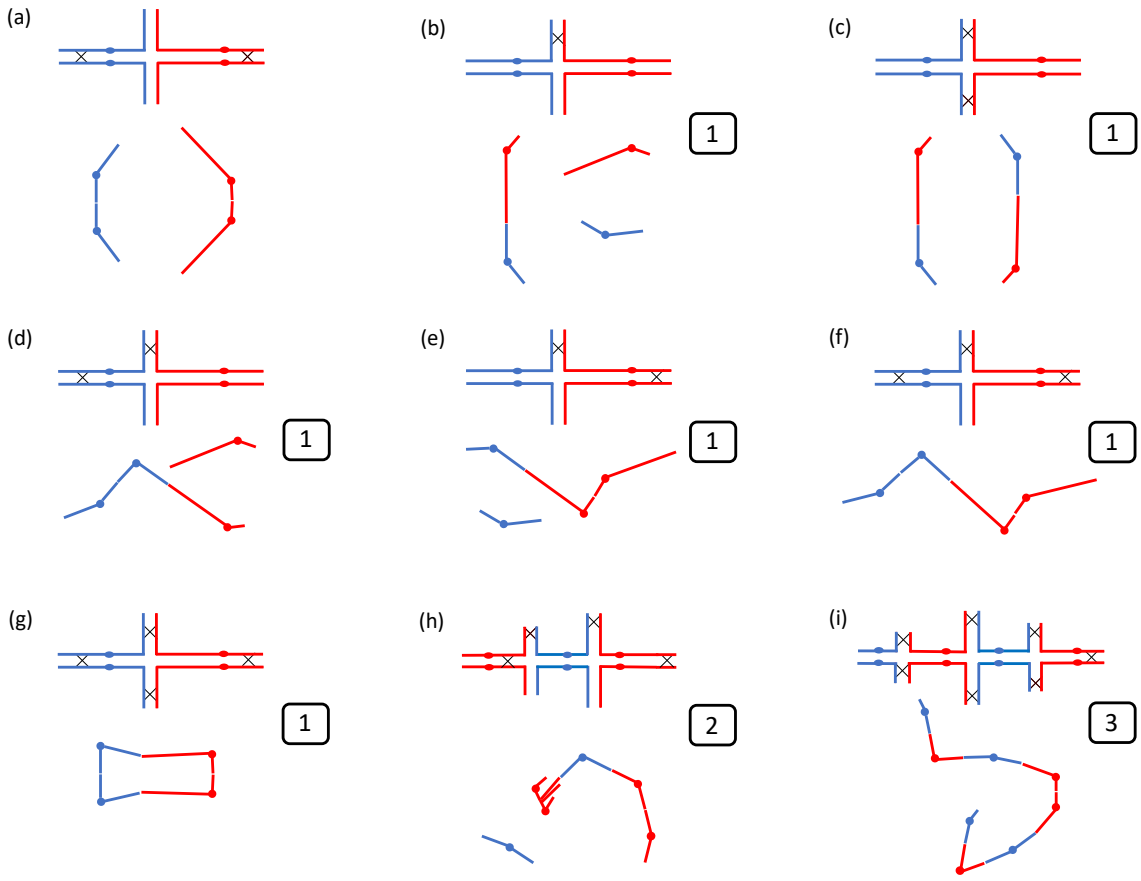


Figure 5-2: Diagrams illustrating meiotic configurations resulting from synaptic partner switches from homologues to homoeologues in B. napus SGDH lines.
Synaptic partner switches (SPS) (upper diagram), MI configuration (lower diagram). A genome chromosomes (blue), C genome chromosomes (red), centromeres (solid circles). Chromosomes consist of two chromatids (not shown). The minimum number of crossovers (CO) (black crosses) required in each synapsed region is shown. The allocated score is in the adjacent text box. (**a**) to (g) Two pairs of chromosomes with homoeology in one arm and one SPS. (a) One SPS but no CO in a region of synapsis between homoeologues results in two homologous bivalents and this has no score. (**b**) to (**g**) One SPS accompanied by one CO (one SPSC) with the outcome at MI depending on the presence/absence of COs in the other switched/not switched regions. (**b**) AC bivalent with A and C univalents. (**c**) two AC bivalents. (**d**) AAC trivalent and C univalent. (**e**) ACC trivalent and A univalent. (**f**) chain quadrivalent AACC. (**g**) ring quadrivalent AACC. (**h**) An example of two SPSC between three homoeologous chromosome pairs with additional COs resulting in one multivalent and a univalent as in Figure 1d. (**i**) An example of three SPSC with additional COs between four chromosome pairs with homoeology as in Figure 1b.

Table 5-1: Categorization of the chromosomes of Brassica napus, B. rapa and B. oleracea by the distribution of 45S rDNA (S) and BoB061G14 (G) FISH signals

Probe Signal	Brassica A genome		Brassica C genome	
	Group	Chromosome	Group	Chromosome
45S rDNA (S)	AS	A3 S large, terminal A5 S near centromere	CS	C8 S terminal
45S rDNA (S) and BoB061G14 (G)	AGS	A1 S large, near centromere A6 S near centromere A9[+] S near centromere	CGS	C7 S terminal C4* S near centromere
BoB061G14 (G)	AG	A2 A4 A7 A8 A10	CG	C1 C2 C4* C5 C6
No signals (N)	-		CN	C3 C9

[*]C4 - S is present in A12 but absent in *B. napus* DH12075
[+]A9 - S is medium strength in *B. napus* DH12075 but very faint in *B. rapa* PS270

For SNP array analysis, testcross F_1 populations were created by crossing lines from the *B. napus* SGDH population with the adapted *B. napus* line, Rainbow. Initially, 31 SG lines were selected at random from the 124 genotyped lines in the SGDH population and following preliminary QTL analysis an additional 17 lines with crossovers near putative QTL regions were added. DH12075, PSA12 and the 48 SGDH lines for the testcross populations and all testcross F_1 lines were genotyped using the Brassica 60K Illumina Infinium array (CLARKE *et al.* 2016). These data confirmed that SGDH lines were replicates of the individuals used to generate the original genetic map but identified small exchanges in DH12075 and PSA12 compared to the individuals used as parents to generate the SGDH population. The testcross F_1 lines were scored for gain and loss of alleles using the method described in HIGGINS *et al.* (2018) whereby a change in fluorescence ratio of the two fluorophores was used to determine changes in copy number at a particular SNP locus. Physical positions of the SNP loci were then used to determine the start and endpoints of chromosome gain or loss and to determine if an event was reciprocal. In total 803 F_1 lines representing 48 SGDH testcross families (between 13 and 24 individuals per

family) were genotyped and 11992 SNP loci polymorphic between the SGDH parental lines and Rainbow were used for analysis (Table S2). These markers were condensed to a subset of 826 SNPs (500 A genome markers and 326 C genome markers) representing unique co-segregating bins and used to score the F_1 lines for gain and loss of alleles. Chromosomal rearrangements were designated as homoeologous recombination (HeR) events if the reciprocal gain and loss of A and C genome loci could be detected and these events were annotated separately from those where only the duplication or deletion of chromosome segments could be identified. Although chromosomal duplications and deletions could have arisen through homoeologous chromosome pairing it is possible that other forms of aberrant pairing or intra-chromosomal associations created such anomalies, thus such events were counted separately. The average number of HeR events in the testcrosses for each SGDH line was used for QTL mapping (Table S3).

QTL Mapping was carried out using the WinQTL Cartographer software v.2.5 (BASTEN *et al.* 1999). Summaries of all methods as they apply to this data set and parameters for QTL analysis are also provided in the Supplementary Methods.

5.3.3 RNASeq

Replicate RNASeq libraries were prepared from meiocytes and leaf tissue of the natural *B. napus* parent of the SGDH population, DH12075. The sequence data has been submitted to NCBI short read archive under the accession number PRJNA664521. Resulting reads were aligned to the *B. napus* spring-type DH12075 genome sequence (I.A.P. Parkin, unpublished; available at http://Cruciferseq.ca) using STAR (DOBIN *et al.* 2013) and gene read counts were generated using FeatureCounts from the Rsubread package (https://bioconductor.org/biocLite.R). The number of aligned reads per library and number of reads per gene are detailed in Table S4. Plant growth conditions, sample preparation and analysis parameters are provided in the Supplementary Methods.

5.4 Results

5.4.1 Estimating level of homoeologous events using cytogenetics

Chromosome spreads at meiotic MI, obtained from pollen mother cells of DH12075 and 43 of the 48 SGDH lines used in the SNP Array analysis, were used to assess the variation in the level of homoeologous events in the SGDH population. A total of 64 male meiocytes of DH12075 were assessed to provide an estimate of homoeologous recombination events in the natural *B. napus* parent of the SGDH population. There was one ring quadrivalent involving A and C chromosome pairs in 14 meiocytes, giving a mean number of SPSC per meiocyte of 0.22 (14 SPSC in 64 meiocytes). It was clear that these were not all formed from the same chromosome pairs as seven of the 12 possible combinations between the A and C groups (Table 5-1) were recorded (Table S1). Their infrequent occurrence, the variety of combinations seen and the fact that the plant was DH indicates that these were newly formed homoeologous quadrivalents. No homoeologous bivalents were seen. In four other meiocytes it was noted that one pair of chromosomes had not formed a bivalent and it was not the same pair each time. The absence of the 'obligate chiasma' between homologues is of interest because it is unusual in plants and can cause aneuploidy, but it does not contribute to the SPSC score of this plant.

Of the 50 SGDH plants (representing 43 lines) examined, 36 had 38 chromosomes but four of these had numbers of A or C chromosomes differing from the expected 20A and 18C (Table S1). The other 14 plants had 35 to 40 chromosomes, with the complement of A ranging from 17 to 22 and C from 15 to 21. Of the five SGDH lines represented by two (3 lines) or three (2 lines) plants, three had plants with atypical numbers: SG-324 gained BnaA1, SG-324-a gained BnaA1 and lost BnaC1; SG-152 lost BnaC8, SG-152-a1 and a2 were normal; SG159-1 gained BnaA1 and lost BnaC1, SG159-2 was normal, SG159-3 lost BnaA1 (Table **S1**).

Example chromosome spreads of meiocytes from SGDH lines are shown in Figure 5-**1**. The most frequent configurations between A and C genome chromosomes were

quadrivalents. As far as the FISH signals allowed, the two A chromosomes appeared to be homologues, as were the two C chromosomes, as expected. Of the 12 possible combinations of the groups in Table 5-**1**, all but one, BnaA3 or BnaA5 with BnaC8, contain at least one set of A and C chromosomes with known homoeologous regions (Figure **S1**, Table **S1**). The BnaA3 or BnaA5 with BnaC8 combination was observed only six times in 3125 chromosome spreads. Mean values of SPSC per meiocyte for each of the 12 observable combinations were calculated for each plant as well as the overall mean per plant (Table **S1**).

One advantage of cytogenetics is its ability to detect chromosome configurations that may produce non-viable gametes due to mis-segregation. In some cases the orientation of a pair of homoeologous bivalents (Figure 5-**1c**) or a quadrivalent at MI was such that both A centromeres faced one pole and both C centromeres faced the other. The resulting pollen grains will have either two A chromosomes or two C chromosomes (those chromatids involved in the COs would be mainly A or C but have a section of the other genome). In some plants multivalents involving more than two pairs (Figure 5-**1b, d**) were seen, and for those in which specific chromosomes could be identified, the chromosomes that had chiasmata with two different chromosomes had known homoeology with both. The SG-261 plant (Figure 5-**1a**) had an extra BnaA10 but only one BnaC9 and in 15 of the 31 meiocytes scored these formed a homoeologous bivalent that contributed to the total score. The plants with the highest means (SG-271, 3.7 and SG-5, 3.0), had high scores for at least two combinations and both had atypical numbers of A and C chromosomes. The scores of individual meiocytes from the SG-5 plant ranged from 0 (one cell) to 6 (one cell). The mean of SG-230 at 0.11 was similar to DH12075 (0.22) but even this plant was not normal, as one chromosome arm of a C chromosome was shorter than its homologue. The maximum score for a meiocyte in SG-230 was 1 and the heteromorphic pair was not involved in any of the homoeologous quadrivalents. There was some variation between the means of the plants in the five lines represented by two or three plants, partly due to differences in chromosome numbers.

5.4.2 Estimating level and extent of homoeologous events using SNP markers

Testcross lines generated from crosses between DH12075 and PSA12, and *B. napus* line Rainbow indicated as expected that the resynthesized line PSA12 had higher levels of HeR and unpaired duplications/deletions. Forty-four testcross individuals from DH12075 were analyzed, only one HeR event was observed (0.02 HeR events per individual). The first 2 Mb of chromosome BnaC2 was missing from all DH12075 derived testcross lines and two unique duplication and six unique deletion events were identified (on average 1.18 duplication/deletion events per individual). Chromosome BnaC2 was completely deleted in two of the DH12075 testcross lines, and one individual had an extra BnaA9 chromosome. In contrast, PSA12 testcross individuals had on average 3.85 reciprocal events per line and 5.13 deletion/duplication events. Of the unpaired events, five were whole chromosome deletions and in three cases an extra copy of an entire chromosome was present. Analysis of the testcross populations indicated the PSA12 parent carried both fixed and heterozygous HeR events (Table **S5**). For example, the top 2 Mb of BnaC1 was missing in all PSA12 testcross lines, suggesting a fixed event, while an HeR event between BnaA1 and BnaC1 segregated in the F_1 (present in 20 of 40 lines) though the size of this event varied from a 2 Mb exchange to reciprocal deletion of BnaA1 and duplication of BnaC1 in their entirety. This variation in length of event is due to the heterozygous HeR region being further modified in the meiosis which produced the testcross individuals, demonstrating that homoeologous chromosome pairing and recombination continues with every generation and is further evidence of the instability of the A and C genomes in newly resynthesized *B. napus* lines. The parental and SGDH lines had been selfed for multiple generations prior to creating the testcross F_1 populations and it was clear from the SNP data in some of the lines that homoeologous exchanges had occurred in previous generations. Though it can not be known what effect a preexisting homoeologous exchange has on recombination, calculation of the number of HeR per testcross individual would include both

inherited and new events since both are caused by meiotic instability. Informative SNP array data for all testcross lines is given in Table **S2** and a summary of all quantified events is shown in Table **S3**.

Only one of the testcross families, SG-230, did not show evidence of a reciprocal HeR event, and all families had at least one event where only the duplication or deletion of alleles was observed. The highest level of HeR was seen in the SG-5 family with an average of 9.9 events per individual. Similar to observations from the PSA12 F_1 lines, the SG-5 F_1 individuals had fixed events which had been inherited from the SG-5 parent and further modified in the testcrosses. This "hyper-recombination" led to extreme values particularly in the SG-5 and SG-235 testcross families that skewed the distribution of homoeologous recombination frequency (Figure 5-**3**). The quantification of homoeologous events measured using the SNP array and cytogenetic analysis had a strong correlation (r=0.721) and the lines with extreme phenotypes such as SG-5 (high HeR) and SG-230 (low HeR) were the same in both analyses thus validating the complementarity of the two methods (Figure 5-**3**).

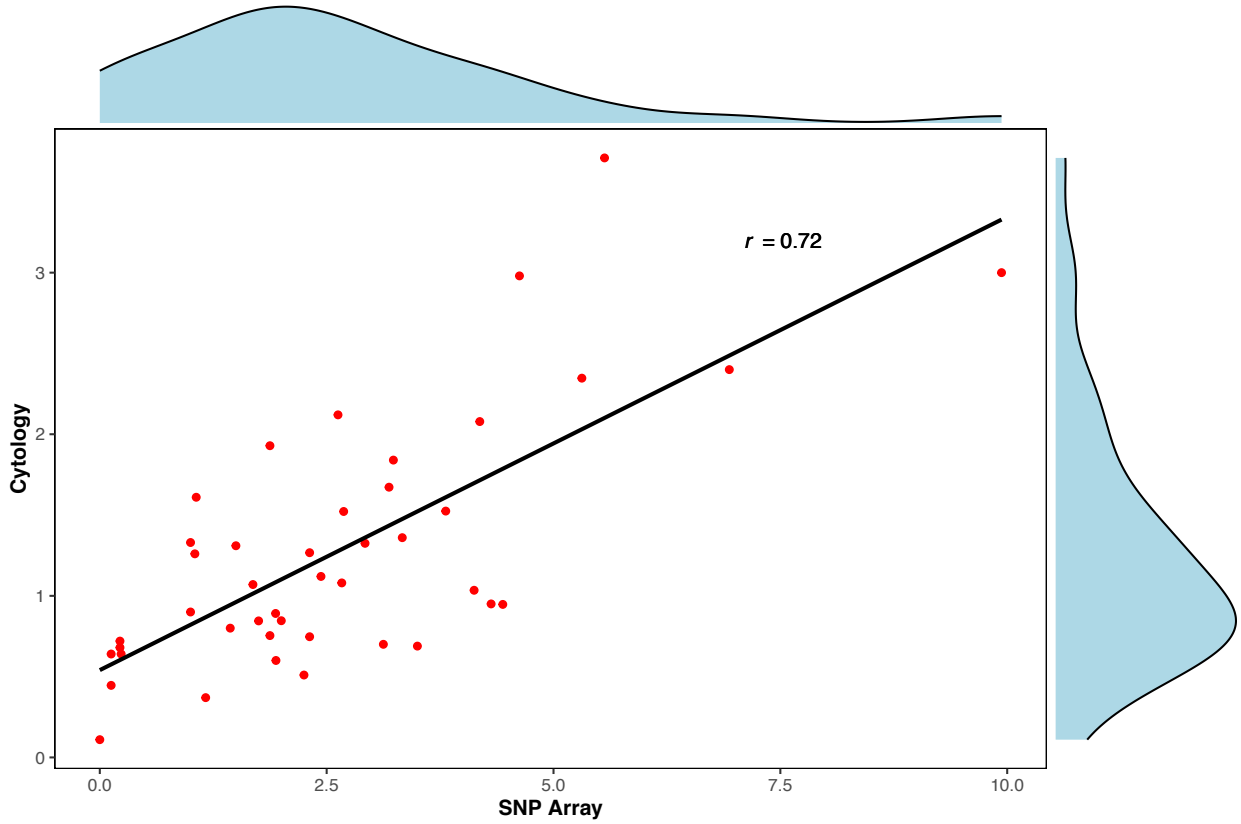


Figure 5-3: Correlation of Cytogenetic and SNP array measurements of homoeologous recombination.
The scatterplot shows the correlation of the rate of HeR measured using cytogenetics and reciprocal SNP marker gain/loss. The distribution of scores for each analysis is shown in the graphs above and to the right of the scatterplot (x-axis SNP scoring, y-axis Cytogenetic scoring) and the correlation statistic (r) is indicated.

In total, there were 2095 HeR events in the 803 lines analysed, with the most common exchanges being between the highly collinear chromosomes BnaA1 and BnaC1, which had 31% of all HeR events (Figure 5-4, Table **S3**). Chromosome BnaA6 had the fewest events, only five in total (0.2%), two with BnaC5 and three exchanges with BnaC7. Chromosome BnaA5 had a clear preference for recombination with BnaC5 rather than it's alternate homoeologue BnaC4, 65% and 35%, respectively. For others with more than one potential pairing partner, such as BnaA9, the exchanges were much more balanced between the two possible homoeologous partners BnaC8 and BnaC9, 45% and 55%, respectively. HeR events where the A genome replaced the C genome were 2.2 times more frequent than those where the C genome replaced the A genome.

Unpaired duplication and deletion events (with no apparent reciprocal exchange) were 50% more common than HeR events. Duplications were more common that deletions, three times more likely, though this ratio was exaggerated by several duplications that were inherited from the SGDH parents and were present in a large number of testcross individuals. Specifically, there is a duplication at the top of BnaC3 and another at the bottom of BnaA10 present in almost half of the testcross F_1 individuals. The unpaired deletions and duplications were more common in the A genome than the C genome, 65% and 56%, respectively. Chromosome BnaA1 had the highest number of deletions (12% of all deletions) and BnaC7 had the fewest (0.4%). Duplications were most common on chromosomes BnaA10 and BnaC3 due to the inherited duplications common to a large number of testcross individuals and were least common on BnaA8 (0.1% of all duplications). Aneuploids accounted for 10% of all unpaired deletions and 6% of all unpaired duplications in the SGDH testcross lines.

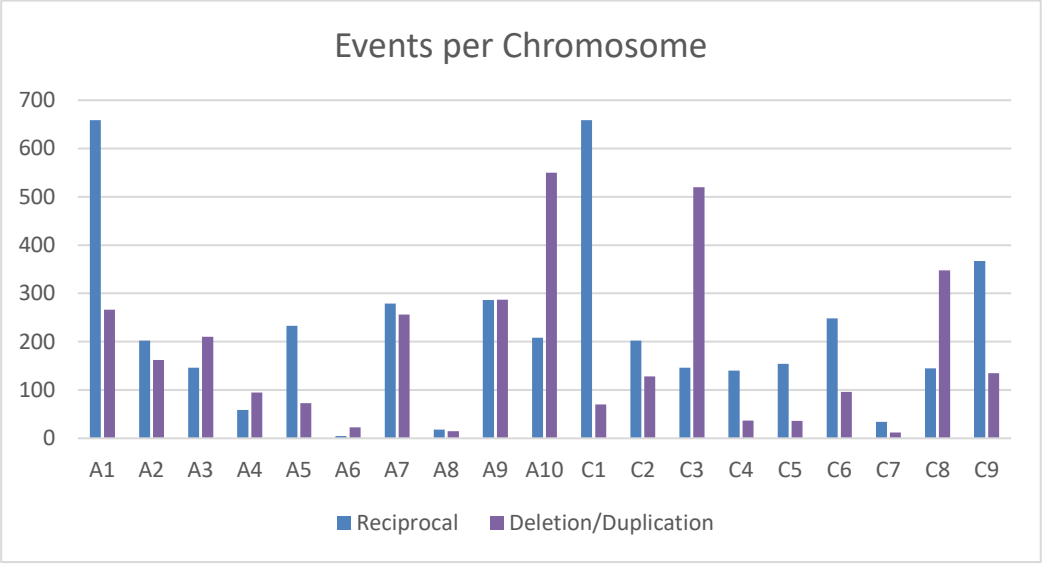


Figure 5-4: Distribution of events across the B. napus genome.
The number of reciprocal and deletion/duplication events in the testcross lines measured using the Brassica SNP array is shown. Blue bars represent the reciprocal events, purple bars show the events where only the deletion or duplication of the SNP allele could be identified.

5.4.3 Genetic mapping of QTL controlling level of homoeologous exchange

A genetic map of the SGDH population with 21118 SNP markers had previously been generated (CLARKE *et al.* 2016), genotype data for the SGDH lines used in the current analyses were extracted and QTL mapping was performed with phenotype data from both the cytogenetic and molecular estimates of homoeologous recombination. The cytogenetic analysis identified two significant QTL, one on BnaA3 (25.7-26.2 Mb) and the second on BnaA9 (11.1-23.9 Mb) (Table 5-**2**). Based on the SNP marker analysis, the average rate per individual of HeR and unpaired deletions and duplications for each SG testcross family were used independently to map QTL controlling homoeologous recombination. Significant QTL controlling the level of HeR exchange were identified on chromosomes BnaA3 (23.3-26.2 Mb), BnaA9 (11.1-23.9 Mb) and BnaC7 (42.2-43.4 Mb). In the analysis of unpaired duplications/deletions a QTL was identified on BnaA9 (10.3-23.9 Mb). The three independently mapped QTLs on chromosome BnaA9 were located in the same very large peri-centromeric region (Figure 5-**5**); and the two QTLs near the bottom of BnaA3 from the cytogenetic and HeR analysis, respectively also overlapped. The HeR BnaC7 locus was not verified potentially because fewer lines were used for the cytogenetic analysis.

Table 5-2: QTL loci controlling homoeologous pairing in Brassica napus.

Chromo-some	QTL name	Map position (Mbp)	LOD Score	R^2	Additive Effect	*B. napus* Candidates	*A. thaliana* Homologue	Common name(s)	RNASeq data	
									Meiocyte (cpm)	Leaf (cpm)
Reciprocal Exchange										
A3	*BnA3_HeR*	23.3-26.2	6.8	0.17	1.65	BnaN03g45420	AT4G20900	MS5, TDM1	0	0
						BnaN03g46530	AT4G22910	FZR2, CCS52A1	0	0.9
						BnaN03g46570	AT4G22970	RSW4	63.5	3.7
						BnaN03g48460	AT4G25540	MSH3	38.5	2
A9	*BnA9_HeR*	11.1-23.9	8	0.34	-2.63	BnaN09g19560	AT5G45400	RPA1C	26	1.7
						BnaN09g22360	AT4G05190	ATK5, KINESIN5	16.5	21.7
						BnaN09g23400	AT4G30870	MUS81	14	9.7
						BnaN09g25570	AT4G11920	CCS52A2, FZR1	3.7	1.1
						BnaN09g26800	AT1G34355	ATPS1	0	0
C7	*BnC7_HeR*	42.2-43.4	3.3	0.09	-1.76	BnaN17g45230	AT4G22910	FZR2, CCS52A1	0.1	0.5
						BnaN17g45270	AT4G22970	RSW4	51.5	5
						BnaN17g47090	AT4G25540	MSH3	49	5
Deletion/Duplication										
A9	*BnA9_DD*	10.3-23.9	4.8	0.32	-2.05	BnaN09g19560	AT5G45400	RPA1C	26	1.7
						BnaN09g22360	AT4G05190	ATK5, KINESIN 5	16.5	21.7
						BnaN09g23400	AT4G30870	MUS81	14	9.7

						BnaN09g25570	AT4G11920	CCS52A2, FZR1	3.7	1.1
						BnaN09g26800	AT1G34355	ATPS1	0	0
Synaptic Partner Switch										
A3	*BnA3_SPSC*	25.7-26.2	3.5	0.09	0.48	BnaN03g48460	AT4G25540	MSH3	38.5	2
A9	*BnA9_SPSC*	11.1-23.9	12.8	0.58	-1.51	BnaN09g19560	AT5G45400	RPA1C	26	1.7
						BnaN09g22360	AT4G05190	ATK5, KINESIN 5	16.5	21.7
						BnaN09g23400	AT4G30870	MUS81	14	9.7
						BnaN09g25570	AT4G11920	CCS52A2, FZR1	3.7	1.1
						BnaN09g26800	AT1G34355	ATPS1	0	0

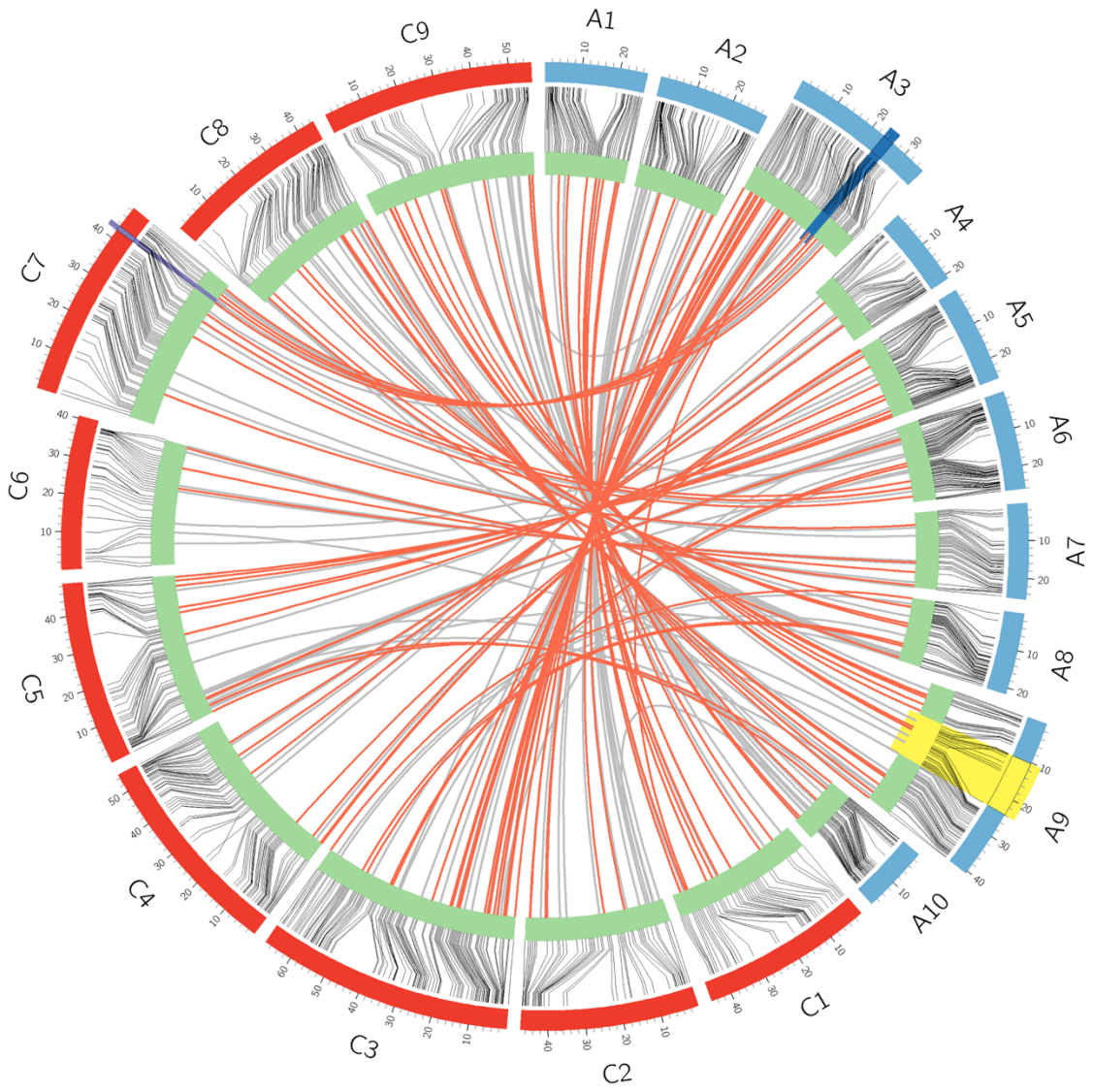


Figure 5-5: Map positions of QTL controlling homoeologous pairing events in Brassica napus.
The outer circle represents the physical length of the chromosomes (A genome in blue, C genome in red), the inner circle (green) the genetic linkage groups, the position of the markers on the physical chromosomes is shown by the linked grey lines. The positions of the QTL loci are shown by coloured blocks, with the colours representing the different phenotypes used to identify loci; purple – HeR only, blue – HeR and cytogenetics, and yellow – common to all phenotypes. The synteny between *B. napus* meiosis genes are shown as connecting lines across the centre of the circle, those genes with only two orthologues are shown in red.

Of the identified QTLs, at the BnaA9 and BnaC7 loci alleles from the resynthesized line PSA12 contribute to high levels of HeR, but at the BnaA3 locus DH12075 alleles are associated with elevated HeR. A previous study showed that even elite *B. napus* lines exhibit generally low but varying levels of HeR (HIGGINS *et al.* 2018) so it is perhaps not surprising to detect alleles for increased rates of HeR coming from the natural *B. napus* line. It was noted in the cytogenetic analysis when lines were ranked according to their mean scores, the nine highest had inherited the faint 45S rDNA signal attributed to the PSA12 parent that is close to the pericentromeric region on BnaA9. The only other lines with this signal, namely, SG-155, SG-309, and SG-318, were ranked among the next ten. In all cases the BnaA9 locus was the strongest QTL representing 34, 32 and 58% of the variation in the HeR, deletion/duplication and cytogenetic analyses, respectively. The BnaA3 locus in the HeR events accounted for 17%, the BnaC7 locus for 9%, and the BnaA3 locus from the cytogenetic analysis was responsible for 9% of the observed variation (Table 5-**2**).

5.4.4 Distribution of meiotic genes across the *Brassica napus* genome

In order to determine if any of the QTL loci could be associated with annotated meiosis-related genes, homologues of documented meiosis genes were identified in the *B. napus* genome. Meiosis has been studied extensively in *Arabidopsis thaliana*, such that 141 meiosis-related genes could be collated from various sources and are listed in Table **S6** (KAUR *et al.* 2006; MA 2006; GEUTING *et al.* 2009; BAUKNECHT AND KOBBE 2014; DE *et al.* 2014; OH *et al.* 2014; LARIO *et al.* 2015; MERCIER *et al.* 2015; WRIGHT *et al.* 2015; VRIELYNCK *et al.* 2016; BOLANOS-VILLEGAS *et al.* 2018; CHAMBON *et al.* 2018; FERNANDES *et al.* 2018; ROHRIG *et al.* 2018). Sequence homology searches against available gene annotations identified 416 and 410 homologous gene copies in two *B. napus* reference genomes, that of DH12075 and Darmor, respectively (Table **S6**). There was no obvious change in gene copy number since the formation of *B. napus*, since the progenitor genomes of *B. oleracea* and *B. rapa* contained 204 and 207 annotated genes, respectively (Table **S6**). Over half (53.9%) of the *A. thaliana* genes were found in only two homoeologous copies in the allopolyploid *B. napus*. This

is significantly higher than expected, since only 19.6% of all *A. thaliana* genes have only two potential orthologues in *B. napus*. The result mirrors that found in a range of angiosperms which showed a biased loss of meiotic gene duplicates subsequent to WGD (LLOYD *et al.* 2014). The progenitor A and C genomes of *B. napus* are suggested to have evolved through a two-step process with the third ancestral genome, which maintained a higher number of gene copies over time (least fractionated), presumed to be hybridised last (CHENG *et al.* 2012). Interestingly this evolutionary path is not obvious from the meiosis genes since of those genes maintained in two copies, only 30.3% are found in the least fractionated genome, suggesting much more balanced fractionation or preferential maintenance of gene copies from the first two genomes.

For each of the QTL regions the sequence underlying the QTL confidence interval was searched for potential meiosis related genes and all possible candidates are listed in Table 5-**2**. In each instance between one (BnaA3_SPSC) and five genes were identified (BnaA9 loci), which did not necessarily reflect the physical size of the QTL region (Figure 5-**5**). Since a number of genes with common function were identified for QTL BnaA3_HeR and BnaC7_HeR it suggests they are derived from orthologous regions. RNASeq libraries were made from isolated meiocytes and leaf tissue from DH12075. Read counts for all libraries are given in Table **S4** and expression of the meiosis genes underlying the QTL regions for leaf and meiocyte tissue in DH12075 is summarized in Table 5-**2**. Since it is not clear that all homologous duplicated gene copies found in *B. napus* would maintain a common function as that predicted from *A. thaliana* gene annotation, the *B. napus* RNASeq data was used to indicate whether the particular gene copy was being expressed in meiotically active tissue. Of the 12 candidates, three showed no observable expression in meiocyte tissue of the stable parental line. For BnaA3_HeR and its homoeologous region on BnaC7, two genes were relatively highly expressed in meiocyte tissue, orthologues of At4g22970, a separase, and At4g22540, MSH3. For the strongest QTL region on BnaA9, mapped by all phenotypes, the most obvious difference in expression levels between leaf and meiocyte tissue was detected for an orthologue of At5g45400 (RPA1C).

5.5 Discussion

Structural variation in the form of copy number or presence/absence variants and even major chromosomal exchanges are being increasingly identified as natural sources of important trait variation (GABUR *et al.* 2019). The recent advances in whole genome sequencing have begun to expose the extent of such variation within many crop species and the importance of homoeologous recombination (HeR) as a mechanism causing such changes in allopolyploid crops is now evident (CHALHOUB *et al.* 2014; STEIN *et al.* 2017). Genetic control of HeR has either been clearly mapped or suggested for several important crop species including wheat (MARTIN *et al.* 2018), canola (JENCZEWSKI *et al.* 2003), coffee (LASHERMES *et al.* 2016) and cotton (FLAGEL *et al.* 2012), and the ability to control and manipulate chromosome rearrangements could be an effective tool for crop improvement. In this study both cytogenetic analysis and molecular quantification of homoeologous recombination were used to identify loci important for controlling mispairing and subsequent homoeologous recombination in *B. napus*. One major QTL on chromosome BnaA9 (*BnaPh1*, *B. napus Pairing homoeologous* 1) was independently verified using each method and was by far the largest contributor to variation in each QTL analysis, explaining 32-58% of the total variation. These QTLs are the first to be mapped in *B. napus* using homoeologous recombination rate in a segregating allotetraploid population.

Prior to this study, the only major locus controlling chromosome pairing in *B. napus, PrBn,* was identified by measuring different levels of chromosome pairing between the Brassica A and C genomes in allohaploid *B. napus* (JENCZEWSKI *et al.* 2003). The *PrBn* QTL was mapped to chromosome BnaC9 and an additional six minor QTL were further identified on chromosomes BnaA1, BnaC1, BnaC3 and BnaC6 in segregating allohaploid lines (LIU *et al.* 2006). None of these previously mapped minor QTL correspond to any of the QTL mapped in the current study, though interestingly the BnaC9 locus appears to locate to a homoeologous region to that of the BnaA9 locus identified in this study. More recently, GAEBELEIN *et al.* (2019) studied synthetic

Brassica allohexaploid plants (A, B and C genomes) and used GWAS to identify SNP markers on A3, A4, A10, C3 as well as three B genome loci that correlated with seed yield as a proxy for meiotic stability, but similarly none of these loci overlap with the QTLs identified in this study. However, the authors note that in one of their allohexaploid populations there was evidence that the loss of both copies of chromosome A9 could not be tolerated, though other aneuploids persisted through multiple generations. In this study aneuploids were reasonably common in the testcross F_1 lines, accounting for 10% of all deletions and 6% of all duplications though there was no bias toward or against BnaA9 for such events. It should be noted however the testcross plants used in this study were not grown to full maturity so it is unknown if they would have survived and produced seed making the complete effect of aneuploidy unknown.

The large QTL on BnaA9 was identified using cytogenetic and molecular identification of homoeologous exchange but it was also identified when unpaired deletions and duplications were analysed. These events could be caused by the dissolution of complex multivalents which form as a result of crossovers between homoeologues and therefore it might be expected that they would identify a common QTL. However, it also raises the question of whether the QTL could play a role beyond just restricting homoeologous recombination such that it may also be important for chromosome stability. Despite selection of additional lines with crossovers near the BnaA9 QTL, the region is large, 12.8 Mb, and includes the presumed centromeric region that will have low homologous crossover rates and high linkage disequilibrium. This would make the genes within the pericentromeric region less likely to undergo rearrangement, which could be essential for an important meiotic gene that prevents homoeologous chromosome pairing and recombination. A summary of known meiotic genes underlying each of the QTLs identified in this study are presented in Table 5-**2** and all annotated genes within the BnaA9 QTL region are listed in Table **S7**. Because homoeologous chromosome sorting occurs early in Prophase I (GRANDONT *et al.* 2014), any candidate gene(s) would presumably have an

important influence early in the meiotic pathway. Of the five genes underlying the BnaA9 QTL, two are of interest: RPA1C (Replication Protein A 1C) and MUS81 (MMS and UV Sensitive 81). RPA1C plays a role in double-stranded break repair in early meiosis in *A. thaliana* (AKLILU *et al.* 2014) and might therefore be a candidate gene. The well-characterized endonuclease MUS81 is another DNA repair protein involved in the interference free crossover pathway in meiosis (BERCHOWITZ *et al.* 2007; HIGGINS *et al.* 2008). The BnaC9 copies of MUS81 and RPA1C were investigated previously in relation to the *PrBn* locus but researchers did not find significant differential expression of either gene between their high and low homoeologous pairing lines and thus concluded that neither gene was responsible for the *PrBn* phenotype (BLARY 2016). However, the *BnaPh1* locus was mapped in an allotetraploid population rather than allohaploids and moreover, in wheat it has been shown that levels of meiotic transcription are stable in the presence and absence of the *Ph1* gene (MARTIN *et al.* 2018), so either MUS81 or RPA1C or very possibly an as yet uncharacterised gene could be responsible for the QTL identified in this study. For the *BnaPh1* locus, the corresponding region in diploid *B. rapa* was assessed to determine if genes present in the presumed progenitor had been deleted from *B. napus*; however, there was no evidence of missing meiosis related genes for this region.

The reciprocal exchange (HeR) analysis identified two smaller QTL, one on BnaA3 and the other in the homologous region on BnaC7 that contain four and three meiosis genes, respectively (Table 5-**2**). The most interesting of these genes is MSH3, a homologue of the MutS gene that is the major controller of mismatch repair in *E. coli* (KUNKEL AND ERIE 2005). There are seven known MutS homologues in *A. thaliana,* of which MSH4 and MSH5 have clearly identified roles in meiotic recombination (HIGGINS *et al.* 2004; HIGGINS *et al.* 2005) while MSH2, MSH3, MSH6 and MSH7 are important for DNA repair in somatic cells of *A. thaliana* (CULLIGAN AND HAYS 2000). More recent studies demonstrate a role for MSH2 in suppression of homoeologous recombination during meiosis in tomato (TAM *et al.* 2011), similarly MSH6 in rice

(JIANG *et al.* 2020), and in Arabidopsis disruption of MSH7 led to an increase in meiotic recombination (LARIO *et al.* 2015). Additionally, in wheat MSH7 has been identified as a candidate *Ph2* gene (DONG *et al.* 2002; LLOYD *et al.* 2007). In *B. napus*, one of the candidate meiotic instability loci identified by GAEBELEIN *et al.* (2019) was MSH2 on chromosome C3. GONZALO *et al.* (2019) demonstrated that when MSH4 was reduced to only one functional copy homoeologous recombination was decreased in *B. napus* allohaploids, but homologous recombination in allotetraploids was unaffected. The effect of MSH4 reduction on the rate of homoeologous recombination in allotetraploids could not be determined due to the low rate of naturally occurring homoeologous recombination. This work suggested that reducing the number of functional gene copies for meiotic genes may be an important evolutionary adaptation for meiotic stability in polyploids.

The need for neo-allopolyploids to quickly minimize mispairing between chromosomes and restrict recombination to homologues is paramount for ensuring the survival and adaptation of the new species. *Brassica napus* is a relatively recent allopolyploid (<10,000 years old) and yet it already shows control of pairing and recombination, specifically when compared to a newly resynthesized line. The rapid adaption of this phenotype would perhaps be suggestive of gene deletion, yet all the identified meiosis related genes were present in the respective diploids, although it is possible that either the substantive progenitor for *B. napus* is no longer available or there is untapped natural variation for this trait within the diploid germplasm. Because diploidization is essential for allopolyploid survival it is not surprising that it has evolved in other major crop species, but based on current knowledge the mechanism for achieving this may vary. For example, in wheat recent evidence points to ZIP4, a homologue of yeast Spo22, as the major gene controlling homoeologous chromosome pairing (MARTIN *et al.* 2018), but this is unlikely to be the case in *Brassica*. The phenotype for *Ph1* is extreme and is only seen in deletion lines, no natural variation has been found, unlike *B. napus* which shows a low but measurable amount of homoeologous recombination even in elite lines (HIGGINS *et al.* 2018). The

current study exploited the difference in recombination rates in a segregating population to identify loci important for this trait and evaluated the recognized meiotic genes underlying those as possible candidates. Further analysis of these loci in other populations may provide a more precise position for the major QTL on BnaA9 and new long read sequencing technologies can further elucidate the genes in the pericentromeric region that will be underrepresented in current short read genome assemblies. Combined, these two approaches will help determine if plant species have found a common solution to one of the challenges brought about by allopolyploidy or if species have developed unique ways to deal with this problem.

6. Conclusions and Discussion

6.1 Summary and Limitations

6.1.1 Homoeologous recombination can be measured using a genome-wide SNP array

High density SNP arrays have become standard for routine genetic mapping in bi-parental populations and GWAS studies, and I have developed a method to assess homoeologous recombination events across the genome of the allotetraploid *Brassica napus* using the Brassica 60K Illumina Infinium SNP array. Brassica is a particularly good system for measuring homoeologous recombination because the diploid species, *B. rapa* and *B. oleracea,* whose genomes combined to form the allotetraploid are still extant and can be crossed to "resynthesize" *B. napus*. These resynthesized lines are known to have higher rates of HeR (PARKIN *et al.* 1995; SHARPE *et al.* 1995; UDALL *et al.* 2005; GAETA *et al.* 2007; SAMANS *et al.* 2017), but a comprehensive assessment of the rate and distribution of such recombination was previously limited by low throughput genetic markers such as RFLPs and SSRs. The high-density array format allows a quick and (relatively) inexpensive way to assess a large number of lines simultaneously. One of the issues with SNP markers in a polyploid such as *B. napus* is the amplification of more than one locus, often two homoeologous loci. After the Brassica 60K array was developed I analysed a diverse set of *B. napus* lines to identify those markers which amplified a single locus and created a custom cluster file that can be applied to any dataset to filter out poor SNPs prior to analysis. I used this cluster file to create a genetic map from ~20,000 SNP markers in a doubled haploid *B. napus* population that was created by crossing a natural *B. napus* line, DH12075 with a resynthesized *B. napus* line, PSA12.

While analyzing the diverse set of Brassica lines to validate the 60K array, it was clear that for assessing homoeologous recombination although intuitively markers which amplify two homoeologous loci would seem desirable, the array technology makes it impossible to separate the loci because they use the same two fluorophores (Figure 3-1). Therefore, it was necessary to identify a large number of single locus SNPs from the array to comprehensively cover the *B. napus* genome. Using the most significant alignment position in both the A and C genomes for each SNP array probe, the chromosomes were aligned into homoeologous pairs (Figure 4-2). I created testcross F_1 plants and hybridized the DNA to the array to demonstrate it was possible to visualize reciprocal allele gain and loss at homoeologous loci.

Detection of HeR through alignment of whole genome shotgun sequencing data has been proven, most notably in the *B. napus* genome sequencing paper (CHALHOUB *et al.* 2014). The method developed using a SNP array is less sensitive than alignment of WGS data as I showed when the ZhongYou821 parental line was resequenced where only five HeR events were detected with SNPs arrays but eight were identified by sequence alignment. One of the limitations of the SNP array method is the lines used for the testcross must be heterozygous at both homoeologues to visualize the event and the markers on the array are only a snapshot of the total genome. But I needed to measure *de novo* events in at least 16 lines from each testcross to get a value of the rate of recombination in each parental line. Currently, the cost of genome sequencing is still too expensive to make WGS a practical option for assessing the number of lines used in this study. Since I needed to measure a large number of lines to get an average rate of recombination, the SNP array was the most practical method to use but if you wanted to identify every HeR in a plant line alignment of WGS data would be required.

6.1.2 Homoeologous recombination occurs in natural *B. napus*

The *B. napus* genomic sequencing (Chaloub 2014) revealed the high level of homoeologous recombination that has occurred in *B. napus* and is now fixed in modern canola lines. Alignment of the *B. napus* sequence to the diploid progenitors *B. rapa* and *B. oleracea* revealed that the low glucosinolate (GSL) profile in *B. napus* is due to a homoeologous recombination event which resulted in the replacement of the C genome copy of the allele with the A genome copy which is carrying a non-functional GSL gene. This exchange is the basis for all modern canola varieties which must have low glucosinolates. The method I developed using the SNP array measures not only these fixed events but *de novo* HeR that has occurred in the most recent meiosis. Initially ten *B. napus* lines from a worldwide collection were chosen and I crossed each with the Australian cultivar Rainbow and analysed the F_1 lines using the SNP array method. This demonstrated that HeR occurs at a low but measurable rate even in established *B. napus* elite lines.

These data also revealed that the location of exchanges is biased toward certain homoeologous chromosome pairs, namely A1/C1 and A2/C2. These two pairs of chromosomes are syntenic along their entire length so this is not surprising but does indicate there are physical factors which influence where HeR occurs. The position on the chromosome also affected the number of recombinations and unsurprisingly exchange events were more likely to occur at the distal regions of the chromosomes. Homoeologous exchanges where both the allele gain and loss could be identified were actually less common than events where only the duplication or deletion of an allele could be confirmed. This could be due to a lack of heterozygosity or poor marker coverage in one of the two homoeologues but several of the unpaired events were very large (Fig 4-2) and 5% were aneuploids. For some of the large events the breakpoint coincides with the predicted centromere position of the chromosome. One hypothesis for this is that the homoeologous chromosomes formed a multivalent when pairing (examples of which can be seen in the chromosome spreads in Fig 5-1) and the complicated dissolution of these multivalents resulted in the two ends of the

chromosome being pulled to opposite poles of the dividing cell and breakage of the chromosome at the centromere.

In this study I only analyzed 11 *B. napus* lines and all had at least one duplication, deletion or HeR event though the number of testcross F_1 lines analyzed is insufficient to comprehensively measure the variation in rate of HeR. However, this work shows that HeR occurs much more frequently in modern, elite *B. napus* cultivars than was previously thought. A more thorough examination of a larger number of lines would give an idea of the level of natural variation for homoeologous recombination rate in the *B. napus* genepool.

6.1.3 QTLs were mapped in a resynthesized population and there are candidate meiotic genes underlying the QTL

One of the parents of the DH population I used to validate the Brassica 60K array was a resynthesized line so the DH individuals segregate for rate of homoeologous recombination. Using the same testcross F_1 method used for the natural *B. napus* cultivars, 48 DH lines from this population were used for QTL mapping and I identified one major and two minor QTL. These are the first QTL controlling homoeologous recombination mapped in allotetraploid *B. napus*. The only other major locus, *PrBn,* was mapped in allohaploid *B. napus* but no phenotype was observed in allotetraploid lines (JENCZEWSKI *et al.* 2003).

The major QTL located on chromosome A9 had an extremely high LOD score and explained 38-52% of the variation. However, it is located in the very large pericentromeric region of the chromosome and spans 12 Mb. The centromeric regions are well known to have low rates of homologous recombination making it difficult to shrink the size of the QTL. Initially 30 lines were selected for mapping and once the A9 QTL was identified 18 additional lines from the population were chosen to try and reduce the size of the QTL but this had only a marginal effect. Because only 48 lines were used in this study, it is possible that QTL mapping in another resynthesized

population, particularly focusing on lines with crossovers near the centromere, could reduce the size of the QTL. Other populations may also have a higher density of informative SNP markers in the A9 region which could also improve the QTL mapping. Regardless, it is likely that this will only improve the accuracy to a minimal extent because of the low recombination rate within the centromeric region.

Genomic sequence analysis of the A9 QTL region and gene expression data from isolated meiocytes identified two candidate meiosis genes, *RPA1C* and *MUS81* based on their expected function early in the meiotic pathway and expression in isolated meiocytes. RPA1 is one of three proteins that make up the ssDNA binding RPA complex. Arabidopsis has five RPA1 proteins (RPA1A-E), and RPA1C is suggested to play an early role in meiosis particularly with crossing over and repair (AKLILU *et al.* 2014). MUS81 is an endonuclease that processes recombination intermediates as part of the interference-free crossover pathway and is strongly expressed in early meiosis (HIGGINS *et al.* 2018). The two minor QTL that were identified are on chromosome A3 and the homologous region of chromosome C7. These QTL are significantly smaller 2.9 Mb and 1.2 Mb, respectively, and have one interesting candidate meiosis gene, MSH3, a homologue of the MutS gene from *E. coli* which is a major controller of mismatch repair (KUNKEL AND ERIE 2005). MSH3 partners with another MutS homologue, MSH2 which has been identified to play a role in suppressing homoeologous recombination (TAM *et al.* 2011).

6.2 Discussion and Future Work

6.2.1 Investigation of candidate genes underlying QTL

This work represents the first major QTL for controlling chromosome pairing and recombination in allotetraploid *B. napus* and lays the foundation for detailed analysis of the genes underlying the QTL, initially with the known meiotic genes but this could be expanded to other genes with unknown or otherwise defined function(s). Preliminary RNASeq analysis of meiocytes from DH12075 to confirm expression was done as part of this study but analysis of meiocytes from the resynthesized line and

differential expression analysis could provide additional information about the candidate genes or identify other genes of interest. A more detailed RNASeq analysis of meiocytes at different stages of meiosis, particularly Prophase I when chromosome pairing and recombination is known to occur would provide confirmation that the timing of expression for the candidate genes fits within the expected stages. Similar work to this is ongoing in maize and wheat (SHUNMUGAM *et al.* 2018; DUKOWIC-SCHULZE *et al.* 2020).

In addition to gene expression studies, genome editing of candidate genes using the CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats, CRISPR associated protein9) system could be done and transformed plants analyzed for changes in recombination rate. Inactivation of genes by CRISPR use is now relatively routine and has been used in Brassica to knock out genes involved in pod shatter, seed production, flower colour (ZHAI *et al.* 2019; KHAN *et al.* 2020; LIU *et al.* 2020) and many others. Standard protocols and methods for using CRISPR are well established, including in *B. napus* (PFALZ *et al.* 2020) so modification of the candidate genes and subsequent analysis should be relatively straightforward. However, it should be noted that although *Ph1* in wheat was identified in 1958 (RILEY AND CHAPMAN 1958; SEARS AND OKAMOTO 1958) and much work has been done analyzing the genes in the region on chromosome 5B, at least three different genes have been postulated as the cause of the homoeologous pairing phenotype (GRIFFITHS *et al.* 2006; BHULLAR *et al.* 2014; REY *et al.* 2017). Therefore, despite the relative ease of generating CRISPR mutants in Brassica, a definitive answer as to what causes the *BnaPh1* phenotype could still be a long way off.

6.2.2 Further analysis of QTL region on A9

While the candidate meiotic genes on A9 are worth exploring in greater detail, the QTL region itself also warrants closer scrutiny. As with many plant genomes, the current *B. napus* assembly was created from short-read sequencing data. While highly accurate, short-read sequencing is often inadequate for complete genome

assembly particularly in repeat-rich regions including centromeres. Often short-read assemblies do not accurately capture structural variants (SV) including duplications, deletions, insertions, inversions and translocations and there is increasing evidence that many important traits in crops are due to SV (GABUR *et al.* 2019).

New long-read sequencing technologies such as Oxford Nanopore and PacBio single molecule real time (SMRT), are capable of routinely producing reads >100kb from a single strand of DNA and are therefore able to sequence through some of the long repetitive regions which are difficult to assemble with short-read data. Assemblies can also be improved with new physical mapping technologies including optical mapping and chromosome conformation capture through proper ordering of scaffolds (see MICHAEL AND VANBUREN (2020)) for a recent review of these technologies and their current use in plant genome assembly).

Recent studies have highlighted the importance of long-read assemblies in tying SV with previously mapped QTLs that could not be resolved with short-read data or mutant analysis. A major study in tomato used Oxford Nanopore sequencing of 100 tomato varieties and identified >200,000 SVs and the researchers were able to directly link SVs with QTLs important for tomato domestication and improvement traits that could not be resolved with short-read data (ALONGE *et al.* 2020). One of the traits, smoky flavour, was originally mapped using GWAS and a candidate gene was identified but could not be confirmed with additional mutant analysis (TIEMAN *et al.* 2017), but the new assemblies revealed a complex locus containing copy number variants and coding sequence mutations that was responsible for the flavour profile. The authors were also able to identify causal variants for previously identified but unexplained variation in fruit size (CHAKRABARTI *et al.* 2013) and harvesting traits (SOYK *et al.* 2019). In *B. napus* CHAWLA *et al.* (2020) investigated SV in flowering time pathway genes which are known to be affected by copy number variation, but they found ~25% of genes had small to medium insertions or deletions. Among others, they identified a 90bp insertion in a vernalization gene which was undetectable with short-

read data. Other recent work has identified SV being responsible for disease resistance in rice (READ *et al.* 2020), aphid resistance in wheat (TULPOVÁ *et al.* 2019), and quality traits in maize (LI *et al.* 2020).

Several reference genomes have been completed using these long-read and physical mapping technologies including the *Brassica* species *B. rapa* and *B. oleracea* (BELSER *et al.* 2018), *B. nigra* (PERUMAL *et al.* 2020) and winter-type *B. napus* (LEE *et al.* 2020). Comparative analysis of the A9 regions from new long-read assemblies of *B. napus* and *B. rapa* could help to identify SV between the two and potentially tie them to the QTL identified in this study.

6.2.3 Refinement of QTL regions and confirmation of minor loci on A3/C7

Chromosome pairing control in wheat has identified one major QTL, *Ph1,* but also several other genes with minor effects (SEARS 1976). Similarly, the *PrBn* locus was identified as the major control of pairing in *B. napus* allohaploids but several smaller QTL were identified in the same lines (LIU *et al.* 2006) so we could expect there are also other minor QTL in allotetraploid *B. napus.*

The two minor QTL mapped in this study are on chromosome A3 and the homoeologue C7, and MSH3 was identified as a potential candidate gene. The LOD scores for these two QTL are low but significant, though the cytological assessment only identified the A3 QTL, perhaps because fewer lines were phenotyped in that study compared to assessment with the SNP array. Though mapping in additional populations is unlikely to alter the QTL on A9, mapping in at least one more resynthesized population could confirm the A3 and C7 QTL, which would make further investigation of the *MSH3* genes more justifiable. Mapping of HeR in additional population(s) could also identify other minor QTL, particularly if the population size(s) are larger. Similar to the A9 QTL, these regions could be analysed with new long-read assemblies and CRISPR/Cas9 editing of the candidate genes could be done.

6.2.4 Analysis of HeR in natural *B. napus* lines

An important finding in this work is that homoeologous recombination is ongoing in *B. napus,* which was thought to be a stable allotetraploid and thus opens up the possibility of generating novel genetic variants through HeR that may have desirable traits for this important oilseed crop. To that end, further investigation of *B. napus* cultivars to measure HeR rates could identify variation within natural *B. napus* populations.

A current limitation to measuring recombination rate is having to assess the progeny of a line which means crossing to another line and analyzing the F_1 progeny or analyzing microspores from the parental line. But a recent study in Arabidopsis used linked-read sequencing to identify homologous recombination events in pollen grains from F_1 plants (SUN *et al.* 2019) and a similar method using linked-read sequencing of pollen from an F_1 tomato hybrid also showed the ability to detect recombinant molecules (ROMMEL FUENTES *et al.* 2020). This approach could make it more efficient to screen lines for diversity of homoeologous recombination rate within natural *B. napus* collections rather than having to make testcross populations as was done in this study. This method would also be useful in screening transformants from gene editing experiments for changes in recombination rate.

One caveat is that both Arabidopsis and tomato are diploid species so it may be more complicated to detect recombinants in a highly duplicated polyploid like *B. napus* because it will depend on the ability to uniquely map the short-read data to the genome. Additionally, the required sequencing depth would be significantly higher than was required in the Arabidopsis and tomato studies due to the much lower rate of homoeologous recombination as compared to homologous recombination, particularly in natural *B. napus* lines. So while the pollen sequencing method would undoubtedly save time because crossing and growing a new generation of plants

would not have to be done, the financial cost of achieving sufficient sequencing depth to properly assess recombination rates would have to be considered.

6.2.5 Genome scanning of *B. napus* vs *B. rapa* and *B. oleracea*

In addition to extensive studies on homoeologous recombination in wheat and brassica, research in autopolyploids, particularly model species, is relevant to the work presented here. *Arabidopsis arenosa* is a close relative of *A. thaliana* and exists naturally in both diploid and autotetraploid forms (KOCH AND MATSCHINGER 2007), and similar to *B. napus* natural tetraploids are meiotically stable but neopolyploids created from the diploid are unstable (YANT *et al.* 2013). Yant *et al* (2013) reasoned that alleles in the tetraploid which contribute to proper chromosome segregation will have low allelic diversity and show higher differentiation between the diploid and tetraploid genomes. They used whole genome shotgun sequencing of eight diploid individuals and 16 tetraploid individuals to scan the genome and identify regions which met those criteria. Meiotic genes were significantly over-represented in their results: eight of the 44 genes identified in the analysis have early meiotic functions of chromosome structure, alignment, synapsis and crossover formation. Investigation of two of these genes, ASY1 and ASY3, showed they are associated with decreased multivalent formation and changes in bivalent shape and length of chromosome axes (MORGAN *et al.* 2020). However, the effect for both genes was subtle leading to the conclusion that there are likely many meiotic genes with small effects that have been adapted to create stable autotetraploid *A. arenosa* (MORGAN *et al.* 2020). It is possible that the minor *Ph2* locus in wheat and the A3 and C7 QTLs identified in *B. napus* are also part of a larger set of genes which have been modified for genome stability in allotetraploids.

A similar type of genome scanning analysis could be carried out in Brassica by looking for regions of low allelic diversity and high differentiation between *B. napus* and the diploids, *B. rapa* and *B. oleracea.* However, the study in Arabidopsis had the advantage of diploid and tetraploid individuals of the same species, so the genome

scanning in Brassica would undoubtedly be more complex and identified regions would need to be more rigorously analysed before making strong conclusions.

References

Aklilu, B. B., R. S. Soderquist and K. M. Culligan, 2014 Genetic analysis of the Replication Protein A large subunit family in Arabidopsis reveals unique and overlapping roles in DNA repair, meiosis and DNA replication. Nucleic Acids Research 42**:** 3104-3118.

Alix, K., P. R. Gerard, T. Schwarzacher and J. S. P. Heslop-Harrison, 2017 Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Annals of Botany 120**:** 183-194.

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

Allers, T., and M. Lichten, 2001 Differential timing and control of noncrossover and crossover recombination during meiosis. Cell 106**:** 47-57.

Alonge, M., X. Wang, M. Benoit, S. Soyk, L. Pereira *et al.*, 2020 Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell.

Armstrong, S. J., F. C. Franklin and G. H. Jones, 2001 Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*. Journal of Cell Science 114**:** 4207-4217.

Attia, T., and G. Röbbelen, 1986a Cytogenetic relationship within cultivated Brassica analyzed in amphihaploids from the three diploid ancestors. Canadian Journal of Genetics and Cytology 28**:** 323-329.

Attia, T., and G. Röbbelen, 1986b Meiotic pairing in haploids and amphidiploids of spontaneous versus synthetic origin in rape, *Brassica napus* L. Canadian Journal of Genetics and Cytology 28**:** 330-334.

Azumi, Y., D. Liu, D. Zhao, W. Li, G. Wang *et al.*, 2002 Homolog interaction during meiotic prophase I in Arabidopsis requires the SOLO DANCERS gene encoding a novel cyclin-like protein. EMBO Journal 21**:** 3081-3095.

Basten, C. J., B. S. Weir and S.-B. Zeng, 1999 *QTL Cartographer. A reference manual and tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC.

Bauknecht, M., and D. Kobbe, 2014 AtGEN1 and AtSEND1, two paralogs in Arabidopsis, possess holliday junction resolvase activity. Plant Physiology 166**:** 202-216.

Belser, C., B. Istace, E. Denis, M. Dubarry, F.-C. Baurens *et al.*, 2018 Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nature Plants 4**:** 879-887.

Berchowitz, L. E., and G. P. Copenhaver, 2010 Genetic interference: don't stand so close to me. Current Genomics 11**:** 91-102.

Berchowitz, L. E., K. E. Francis, A. L. Bey and G. P. Copenhaver, 2007 The Role of *AtMUS81* in Interference-Insensitive Crossovers in *A. thaliana*. PLOS Genetics 3**:** e132.

Bhullar, R., R. Nagarajan, H. Bennypaul, G. K. Sidhu, G. Sidhu *et al.*, 2014 Silencing of a metaphase I-specific gene results in a phenotype similar to that of the Pairing homeologous 1 (Ph1) gene mutations. Proceedings of the National Academy of Sciences U S A 111**:** 14187-14192.

Bishop, D. K., and D. Zickler, 2004 Early decision; meiotic crossover interference prior to stable strand exchange and synapsis. Cell 117**:** 9-15.

Blanc, G., and K. H. Wolfe, 2004 Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16**:** 1679-1691.

Blary, A., 2016 Towards a functional characterization of meiotic recombination in rapeseed : analysis of the meiotic transcriptome and hyper-recombinant mutants, pp. Université Paris-Saclay.

Blary, A., A. Gonzalo, F. Eber, A. Bérard, H. Bergès *et al.*, 2018 FANCM Limits Meiotic Crossovers in Brassica Crops. Frontiers in Plant Science 9**:** 368.

Bleuyard, J. Y., M. E. Gallego and C. I. White, 2006 Recent advances in understanding of the DNA double-strand break repair machinery of plants. DNA Repair (Amst) 5**:** 1-12.

Bleuyard, J. Y., and C. I. White, 2004 The *Arabidopsis* homologue of Xrcc3 plays an essential role in meiosis. EMBO Journal 23**:** 439-449.

Bolanos-Villegas, P., W. Xu, M. Martinez-Garcia, M. Pradillo and Y. Wang, 2018 Insights Into the Role of Ubiquitination in Meiosis: Fertility, Adaptation and Plant Breeding. Arabidopsis Book 16**:** e0187.

Bolger, A. M., M. Lohse and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30**:** 2114-2120.

Bonnet, S., A. Knoll, F. Hartung and H. Puchta, 2013 Different functions for the domains of the *Arabidopsis thaliana* RMI1 protein in DNA cross-link repair, somatic and meiotic recombination. Nucleic Acids Research 41**:** 9349-9360.

Borde, V., 2007 The multiple roles of the Mre11 complex for meiotic recombination. Chromosome Research 15**:** 551-563.

Borner, G. V., N. Kleckner and N. Hunter, 2004 Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. Cell 117**:** 29-45.

Boulton, S. J., 2006 Cellular functions of the BRCA tumour-suppressor proteins. Biochem Soc Trans 34**:** 633-645.

Bowers, J. E., B. A. Chapman, J. Rong and A. H. Paterson, 2003a Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422**:** 433-438.

Bowers, J. E., B. A. Chapman, J. Rong and A. H. Paterson, 2003b Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422**:** 433.

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. Theor Appl Genet 127**:** 2051-2064.

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

Bus, A., N. Korber, R. J. Snowdon and B. Stich, 2011 Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theoretical and Applied Genetics 123**:** 1413-1423.

Chakrabarti, M., N. Zhang, C. Sauvage, S. Muños, J. Blanca *et al.*, 2013 A cytochrome P450 regulates a domestication trait in cultivated tomato. Proceedings of the National Academy of Sciences 110**:** 17125.

Chalhoub, B., F. Denoeud, S. Liu, I. A. Parkin, H. Tang *et al.*, 2014 Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345**:** 950-953.

Chambon, A., A. West, D. Vezon, C. Horlow, A. De Muyt *et al.*, 2018 Identification of ASYNAPTIC4, a Component of the Meiotic Chromosome Axis. Plant Physiology 178**:** 233-246.

Chawla, H. S., H. Lee, I. Gabur, P. Vollrath, S. Tamilselvan-Nattar-Amutha *et al.*, 2020 Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. Plant Biotechnology Journal n/a.

Chelysheva, L., D. Vezon, K. Belcram, G. Gendrot and M. Grelon, 2008 The Arabidopsis BLAP75/Rmi1 homologue plays crucial roles in meiotic double-strand break repair. PLoS Genetics 4**:** e1000309.

Chelysheva, L., D. Vezon, A. Chambon, G. Gendrot, L. Pereira *et al.*, 2012 The Arabidopsis HEI10 is a new ZMM protein related to Zip3. PLoS Genetics 8**:** e1002799.

Chen, C., W. Zhang, L. Timofejeva, Y. Gerardin and H. Ma, 2005 The Arabidopsis ROCK-N-ROLLERS gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. Plant Journal 43**:** 321-334.

Cheng, F., J. Wu, L. Fang, S. Sun, B. Liu *et al.*, 2012 Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. PLoS One 7**:** e36442.

Ciccia, A., N. McDonald and S. C. West, 2008 Structural and functional relationships of the XPF/MUS81 family of proteins. Annual Review of Biochemistry 77**:** 259-287.

Cifuentes, M., L. Grandont, G. Moore, A. M. Chèvre and E. Jenczewski, 2010 Genetic regulation of meiosis in polyploid species: New insights into an old question. New Phytologist 186**:** 29-36.

Clarke, W. E., E. E. Higgins, J. Plieske, R. Wieseke, C. Sidebottom *et al.*, 2016 A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. Theoretical and Applied Genetics 129**:** 1887-1899.

Clarke, W. E., I. A. Parkin, H. A. Gajardo, D. J. Gerhardt, E. Higgins *et al.*, 2013 Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus L*. PLoS One 8**:** e81992.

Cole, F., S. Keeney and M. Jasin, 2010 Evolutionary conservation of meiotic DSB proteins: more than just Spo11. Genes and Development 24**:** 1201-1207.

Comai, L., 2005 The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6**:** 836.

Cook, J. P., M. D. McMullen, J. B. Holland, F. Tian, P. Bradbury *et al.*, 2012 Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158**:** 824-834.

Couteau, F., F. Belzile, C. Horlow, O. Grandjean, D. Vezon *et al.*, 1999 Random chromosome segregation without meiotic arrest in both male and female meiocytes of a dmc1 mutant of *Arabidopsis*. Plant Cell 11**:** 1623-1634.

Crismani, W., C. Girard, N. Froger, M. Pradillo, J. L. Santos *et al.*, 2012 FANCM Limits Meiotic Crossovers. Science 336**:** 1588.

Crismani, W., V. Portemer, N. Froger, L. Chelysheva, C. Horlow *et al.*, 2013 MCM8 is required for a pathway of meiotic double-strand break repair independent of DMC1 in Arabidopsis thaliana. PLoS Genetics 9**:** e1003165.

Culligan, K. M., and J. B. Hays, 2000 Arabidopsis MutS homologs-AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7-form three distinct protein heterodimers with different specificities for mismatched DNA. Plant Cell 12**:** 991-1002.

Dalton-Morgan, J., A. Hayward, S. Alamery, R. Tollenaere, A. S. Mason *et al.*, 2014 A high-throughput SNP array in the amphidiploid species *Brassica napus* shows diversity in resistance genes. Funct Integr Genomics 14**:** 643-655.

Dangel, N. J., A. Knoll and H. Puchta, 2014 MHF1 plays Fanconi anaemia complementation group M protein (FANCM)-dependent and FANCM-independent roles in DNA repair and homologous recombination in plants. Plant Journal 78**:** 822-833.

De, K., L. Sterle, L. Krueger, X. Yang and C. A. Makaroff, 2014 *Arabidopsis thaliana* WAPL is essential for the prophase removal of cohesin during meiosis. PLoS Genetics 10**:** e1004497.

De Muyt, A., L. Pereira, D. Vezon, L. Chelysheva, G. Gendrot *et al.*, 2009 A high throughput genetic screen identifies new early meiotic recombination functions in *Arabidopsis thaliana*. PLoS Genetics 5**:** e1000654.

De Muyt, A., D. Vezon, G. Gendrot, J. L. Gallois, R. Stevens *et al.*, 2007 AtPRD1 is required for meiotic double strand break formation in *Arabidopsis thaliana*. EMBO Journal 26**:** 4126-4137.

Delourme, R., C. Falentin, B. F. Fomeju, M. Boillot, G. Lassalle *et al.*, 2013 High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus L*. BMC Genomics 14**:** 120.

Deschamps, S., V. Llaca and G. D. May, 2012 Genotyping-by-Sequencing in Plants. Biology (Basel) 1**:** 460-483.

Dewitte, W., and J. A. Murray, 2003 The plant cell cycle. Annual Review of Plant Biology 54**:** 235-264.

Dixon, G., 2006 Origins and diversity of Brassica and its relatives. Vegetable brassicas and related crucifers**:** 1-33.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29**:** 15-21.

Dong, C., R. Whitford and P. Langridge, 2002 A DNA mismatch repair gene links to the *Ph2* locus in wheat. Genome 45**:** 116-124.

Doutriaux, M. P., F. Couteau, C. Bergounioux and C. White, 1998 Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana*. Molecular and General Genetics 257**:** 283-291.

Doyle, J. J., L. E. Flagel, A. H. Paterson, R. A. Rapp, D. E. Soltis *et al.*, 2008 Evolutionary genetics of genome merger and doubling in plants. Annual Review of Genetics 42**:** 443-461.

Dukowic-Schulze, S., N. Garcia, A. S. K. Shunmugam, S. Kagale and C. Chen, 2020 Isolating Male Meiocytes from Maize and Wheat for "-Omics" Analyses. Methods Mol Biol 2061**:** 237-258.

Duroc, Y., A. Lemhemdi, C. Larcheveque, A. Hurel, M. Cuacos *et al.*, 2014 The kinesin AtPSS1 promotes synapsis and is required for proper crossover distribution in meiosis. PLoS Genetics 10**:** e1004674.

Endow, S. A., F. J. Kull and H. Liu, 2010 Kinesins at a glance. Journal of Cell Science 123**:** 3420-3424.

Ferdous, M., J. D. Higgins, K. Osman, C. Lambing, E. Roitinger *et al.*, 2012 Inter-homolog crossing-over and synapsis in Arabidopsis meiosis are dependent on the chromosome axis protein AtASY3. PLoS Genetics 8**:** e1002507.

Fernandes, J. B., M. Duhamel, M. Seguela-Arnaud, N. Froger, C. Girard *et al.*, 2018 FIGL1 and its novel partner FLIP form a conserved complex that regulates homologous recombination. PLoS Genetics 14**:** e1007317.

Flagel, L., J. Udall, D. Nettleton and J. Wendel, 2008 Duplicate gene expression in allopolyploid Gossypium reveals two temporally distinct phases of expression evolution. BMC Biology 6**:** 16.

Flagel, L. E., J. F. Wendel and J. A. Udall, 2012 Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. BMC Genomics 13**:** 302.

Franklin, F. C., J. D. Higgins, E. Sanchez-Moran, S. J. Armstrong, K. E. Osman *et al.*, 2006 Control of meiotic recombination in Arabidopsis: role of the MutL and MutS homologues. Biochem Soc Trans 34**:** 542-544.

Friebe, B., P. Zhang, G. Linc and B. S. Gill, 2005 Robertsonian translocations in wheat arise by centric misdivision of univalents at anaphase I and rejoining of broken centromeres during interkinesis of meiosis II. Cytogenetic and Genome Research 109**:** 293-297.

Gabur, I., H. S. Chawla, R. J. Snowdon and I. A. P. Parkin, 2019 Connecting genome structural variation with complex traits in crop plants. Theoretical and Applied Genetics 132**:** 733-750.

Gaebelein, R., S. V. Schiessl, B. Samans, J. Batley and A. S. Mason, 2019 Inherited allelic variants and novel karyotype changes influence fertility and genome stability in Brassica allohexaploids. New Phytologist 223**:** 965-978.

Gaeta, R. T., and J. C. Pires, 2010 Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytologist 186**:** 18-28.

Gaeta, R. T., J. C. Pires, F. Iniguez-Luy, E. Leon and T. C. Osborn, 2007 Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 19**:** 3403-3417.

Ganal, M. W., T. Altmann and M. S. Röder, 2009 SNP identification in crop plants. Curr Opin Plant Biol 12**:** 211-217.

Ganal, M. W., G. Durstewitz, A. Polley, A. Bérard, E. S. Buckler *et al.*, 2011 A large maize (*Zea mays L.*) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6.

Ganal, M. W., A. Polley, E. M. Graner, J. Plieske, R. Wieseke *et al.*, 2012 Large SNP arrays for genotyping in crop plants. J Biosci 37**:** 821-828.

Geuting, V., D. Kobbe, F. Hartung, J. Durr, M. Focke *et al.*, 2009 Two distinct MUS81-EME1 complexes from Arabidopsis process Holliday junctions. Plant Physiology 150**:** 1062-1071.

Gill, K. S., B. S. Gill, T. R. Endo and Y. Mukai, 1993 Fine physical mapping of Ph1, a chromosome pairing regulator gene in polyploid wheat. Genetics 134**:** 1231-1236.

Golubovskaya, I. N., L. C. Harper, W. P. Pawlowski, D. Schichnes and W. Z. Cande, 2002 The *pam1* gene is required for meiotic bouquet formation and efficient homologous synapsis in maize (*Zea mays L*.). Genetics 162**:** 1979-1993.

Gonzalo, A., M. O. Lucas, C. Charpentier, G. Sandmann, A. Lloyd *et al.*, 2019 Reducing *MSH4* copy number prevents meiotic crossovers between non-homologous chromosomes in *Brassica napus*. Nature Communications 10**:** 2354.

Grandont, L., N. Cuñado, O. Coriton, V. Huteau, F. Eber *et al.*, 2014 Homoeologous Chromosome Sorting and Progression of Meiotic Recombination in *Brassica napus*: Ploidy Does Matter! The Plant Cell 26**:** 1448-1463.

Greer, E., A. C. Martin, A. Pendle, I. Colas, A. M. Jones *et al.*, 2012 The Ph1 locus suppresses Cdk2-type activity during premeiosis and meiosis in wheat. Plant Cell 24**:** 152-162.

Griffiths, S., R. Sharp, T. N. Foote, I. Bertin, M. Wanous *et al.*, 2006 Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature 439**:** 749-752.

Harper, A. L., M. Trick, J. Higgins, F. Fraser, L. Clissold *et al.*, 2012 Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nat Biotechnol 30**:** 798-802.

Harper, L., I. Golubovskaya and W. Z. Cande, 2004 A bouquet of chromosomes. Journal of Cell Science 117**:** 4025-4032.

Hartung, F., S. Suer, A. Knoll, R. Wurz-Wildersinn and H. Puchta, 2008 Topoisomerase 3α and RMI1 suppress somatic crossovers and are essential for resolution of meiotic recombination intermediates in *Arabidopsis thaliana*. PLoS Genetics 4**:** e1000285.

Hartung, F., R. Wurz-Wildersinn, J. Fuchs, I. Schubert, S. Suer *et al.*, 2007 The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in *Arabidopsis*. Plant Cell 19**:** 3090-3099.

Hatzig, S. V., M. Frisch, F. Breuer, N. Nesi, S. Ducournau *et al.*, 2015 Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. Frontiers in Plant Science 6**:** 221.

He, Z., L. Wang, L. Harper Andrea, L. Havlickova, K. Pradhan Akshay *et al.*, 2017 Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotechnology Journal 15**:** 594-604.

Higgins, E. E., W. E. Clarke, E. C. Howell, S. J. Armstrong and I. A. P. Parkin, 2018 Detecting *de novo* Homoeologous Recombination Events in Cultivated *Brassica napus* Using a Genome-Wide SNP Array. G3: Genes|Genomes|Genetics 8**:** 2673.

Higgins, J. D., S. J. Armstrong, F. C. Franklin and G. H. Jones, 2004 The *Arabidopsis MutS* homolog *AtMSH4* functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. Genes and Development 18**:** 2557-2570.

Higgins, J. D., E. F. Buckling, F. C. H. Franklin and G. H. Jones, 2008 Expression and functional analysis of *AtMUS81* in Arabidopsis meiosis reveals a role in the second pathway of crossing-over. The Plant Journal 54**:** 152-162.

Higgins, J. D., E. Sanchez-Moran, S. J. Armstrong, G. H. Jones and F. C. Franklin, 2005 The *Arabidopsis* synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. Genes and Development 19**:** 2488-2500.

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.

Howell, E. C., M. J. Kearsey, G. H. Jones, G. J. King and S. J. Armstrong, 2008 A and C genome distinction and chromosome identification in *Brassica napus* by sequential fluorescence in situ hybridization and genomic in situ hybridization. Genetics 180**:** 1849-1857.

Iniguez-Luy, F. L., and M. L. Federico, 2011 The genetics of *Brassica napus*, pp. 291-322 in *Genetics and Genomics of the Brassicaceae*. Springer.

Ip, S. C., U. Rass, M. G. Blanco, H. R. Flynn, J. M. Skehel *et al.*, 2008 Identification of Holliday junction resolvases from humans and yeast. Nature 456**:** 357-361.

Jannink, J. L., A. J. Lorenz and H. Iwata, 2010 Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9**:** 166-177.

Jenczewski, E., and K. Alix, 2004 From Diploids to Allopolyploids: The Emergence of Efficient Pairing Control Genes in Plants. Critical Reviews in Plant Sciences 23**:** 21-45.

Jenczewski, E., F. Eber, A. Grimaud, S. Huet, M. O. Lucas *et al.*, 2003 *PrBn*, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. Genetics 164**:** 645-653.

Jeon, J.-S., Y.-Y. Chung, S. Lee, G.-H. Yi, B.-G. Oh *et al.*, 1999 Isolation and characterization of an anther-specific gene, RA8, from rice (Oryza sativa L.). Plant Molecular Biology 39**:** 35-44.

Jiang, M., X. Wu, Y. Song, H. Shen and H. Cui, 2020 Effects of *OsMSH6* Mutations on Microsatellite Stability and Homeologous Recombination in Rice. Frontiers in Plant Science 11**:** 220.

Jolivet, S., D. Vezon, N. Froger and R. Mercier, 2006 Non conservation of the meiotic function of the Ski8/Rec103 homolog in *Arabidopsis*. Genes Cells 11**:** 615-622.

Jones, G. H., and F. C. Franklin, 2006 Meiotic crossing-over: obligation and interference. Cell 126**:** 246-248.

Kaur, J., J. Sebastian and I. Siddiqi, 2006 The *Arabidopsis-mei2-like* genes play a role in meiosis and vegetative growth in *Arabidopsis*. Plant Cell 18**:** 545-559.

Keeney, S., C. N. Giroux and N. Kleckner, 1997 Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 88**:** 375-384.

Kent, W. J., 2002 BLAT—The BLAST-Like Alignment Tool. Genome Research 12**:** 656-664.

Khan, M. H. U., L. Hu, M. Zhu, Y. Zhai, S. U. Khan *et al.*, 2020 Targeted mutagenesis of *EOD3* gene in *Brassica napus* L. regulates seed production. Journal of Cellular Physiology n/a.

Kim, K. P., B. M. Weiner, L. Zhang, A. Jordan, J. Dekker *et al.*, 2010 Sister cohesion and structural axis components mediate homolog bias of meiotic recombination. Cell 143**:** 924-937.

Kleckner, N., 2006 Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. Chromosoma 115**:** 175-194.

Knight, E., E. Greer, T. Draeger, V. Thole, S. Reader *et al.*, 2010 Inducing chromosome pairing through premature condensation: Analysis of wheat interspecific hybrids. Functional and Integrative Genomics 10**:** 603-608.

Knoll, A., J. D. Higgins, K. Seeliger, S. J. Reha, N. J. Dangel *et al.*, 2012 The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in Arabidopsis. Plant Cell 24**:** 1448-1464.

Koch, M. A., and M. Matschinger, 2007 Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences 104**:** 6272.

Körber, N., A. Bus, J. Li, J. Higgins, I. Bancroft *et al.*, 2015 Seedling development traits in *Brassica napus* examined by gene expression analysis and association mapping. BMC Plant Biology 15**:** 136.

Kumar, R., H. M. Bourbon and B. de Massy, 2010 Functional conservation of Mei4 for meiotic DNA double-strand break formation from yeasts to mice. Genes and Development 24**:** 1266-1280.

Kunkel, T. A., and D. A. Erie, 2005 DNA Mismatch Repair. Annual Review of Biochemistry 74**:** 681-710.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1**:** 174-181.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nature methods 9**:** 357-359.

Lario, L. D., P. Botta, P. Casati and C. P. Spampinato, 2015 Role of AtMSH7 in UV-B-induced DNA damage recognition and recombination. Journal of Experimental Botany 66**:** 3019-3026.

Lashermes, P., Y. Hueber, M.-C. Combes, D. Severac and A. Dereeper, 2016 Inter-genomic DNA Exchanges and Homeologous Gene Silencing Shaped the Nascent Allopolyploid Coffee Genome (*Coffea arabica* L.). G3: Genes|Genomes|Genetics 6**:** 2937.

Lee, H., H. S. Chawla, C. Obermeier, F. Dreyer, A. Abbadi *et al.*, 2020 Chromosome-Scale Assembly of Winter Oilseed Rape Brassica napus. Frontiers in Plant Science 11**:** 496.

Leitch, A. R., and I. J. Leitch, 2008 Genomic Plasticity and the Diversity of Polyploid Plants. Science 320**:** 481.

Li, C., X. Xiang, Y. Huang, Y. Zhou, D. An *et al.*, 2020 Long-read sequencing reveals genomic structural variations that underlie creation of quality protein maize. Nature Communications 11**:** 17.

Li, W., C. Chen, U. Markmann-Mulisch, L. Timofejeva, E. Schmelzer *et al.*, 2004 The *Arabidopsis* AtRAD51 gene is dispensable for vegetative development but required for meiosis. Proceedings of the National Academy of Science U S A 101**:** 10596-10601.

Li, W., X. Yang, Z. Lin, L. Timofejeva, R. Xiao *et al.*, 2005 The AtRAD51C gene is required for normal meiotic chromosome synapsis and double-stranded break repair in Arabidopsis. Plant Physiology 138**:** 965-976.

Libby, B. J., L. G. Reinholdt and J. C. Schimenti, 2003 Positional cloning and characterization of Mei1, a vertebrate-specific gene required for normal meiotic chromosome synapsis in mice. Proceedings of the National Academy of Sciences U S A 100**:** 15706-15711.

Lim, K. Y., A. Kovarik, R. Matyasek, W. Chase Mark, J. Clarkson James *et al.*, 2007 Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. New Phytologist 175**:** 756-763.

Liu, L., C. Qu, B. Wittkop, B. Yi, Y. Xiao *et al.*, 2013 A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L. PLoS One 8**:** e83052.

Liu, Y., S. Ye, G. Yuan, X. Ma, S. Heng *et al.*, 2020 Gene silencing of BnaA09.ZEP and BnaC09.ZEP confers orange color in *Brassica napus* flowers. The Plant Journal n/a.

Liu, Z., K. Adamczyk, M. Manzanares-Dauleux, F. Eber, M. O. Lucas *et al.*, 2006 Mapping *PrBn* and other quantitative trait loci responsible for the control of homeologous chromosome pairing in oilseed rape (*Brassica napus L.*) haploids. Genetics 174**:** 1583-1596.

Lloyd, A., A. Blary, D. Charif, C. Charpentier, J. Tran *et al.*, 2017 Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. New Phytologist 217**:** 367-377.

Lloyd, A. H., A. S. Milligan, P. Langridge and J. A. Able, 2007 *TaMSH7*: a cereal mismatch repair gene that affects fertility in transgenic barley (*Hordeum vulgare* L.). BMC Plant Biology 7**:** 67.

Lloyd, A. H., M. Ranoux, S. Vautrin, N. Glover, J. Fourment *et al.*, 2014 Meiotic Gene Evolution: Can You Teach a New Dog New Tricks? Molecular Biology and Evolution 31**:** 1724-1727.

Lorenz, A., F. Osman, W. Sun, S. Nandi, R. Steinacher *et al.*, 2012 The fission yeast FANCM ortholog directs non-crossover recombination during meiosis. Science 336**:** 1585-1588.

Lysak, M. A., M. A. Koch, A. Pecinka and I. Schubert, 2005 Chromosome triplication found across the tribe Brassiceae. Genome Research 15**:** 516-525.

Ma, H., 2006 A molecular portrait of Arabidopsis meiosis. Arabidopsis Book 4**:** e0095.

Macaisne, N., M. Novatchkova, L. Peirera, D. Vezon, S. Jolivet *et al.*, 2008 SHOC1, an XPF endonuclease-related protein, is essential for the formation of class I meiotic crossovers. Current Biology 18**:** 1432-1437.

Macaisne, N., J. Vignard and R. Mercier, 2011 SHOC1 and PTD form an XPF-ERCC1-like complex that is required for formation of class I crossovers. Journal of Cell Science 124**:** 2687-2691.

Macqueen, A. J., and G. S. Roeder, 2009 Fpr3 and Zip3 ensure that initiation of meiotic recombination precedes chromosome synapsis in budding yeast. Current Biology 19**:** 1519-1526.

Maere, S., S. De Bodt, J. Raes, T. Casneuf, M. Van Montagu *et al.*, 2005 Modeling gene and genome duplications in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America 102**:** 5454-5459.

Mannuss, A., S. Dukowic-Schulze, S. Suer, F. Hartung, M. Pacher *et al.*, 2010 RAD5A, RECQ4A, and MUS81 have specific functions in homologous recombination and define different pathways of DNA repair in Arabidopsis thaliana. Plant Cell 22**:** 3318-3330.

Martin, A. C., P. Borrill, J. Higgins, A. Alabdullah, R. H. Ramirez-Gonzalez *et al.*, 2018 Genome-Wide Transcription During Early Wheat Meiosis Is Independent of Synapsis, Ploidy Level, and the *Ph1* Locus. Frontiers in Plant Science 9**:** 1791.

Martín, A. C., M.-D. Rey, P. Shaw and G. Moore, 2017 Dual effect of the wheat *Ph1* locus on chromosome synapsis and crossover. Chromosoma 126**:** 669-680.

Martini, E., R. L. Diaz, N. Hunter and S. Keeney, 2006 Crossover homeostasis in yeast meiosis. Cell 126**:** 285-295.

Mason, A. S., J. Batley, P. E. Bayer, A. Hayward, W. A. Cowling *et al.*, 2014 High-resolution molecular karyotyping uncovers pairing between ancestrally related Brassica chromosomes. New Phytologist 202**:** 964-974.

Mason, A. S., E. E. Higgins, R. J. Snowdon, J. Batley, A. Stein *et al.*, 2017 A user guide to the Brassica 60K Illumina Infinium™ SNP genotyping array. Theoretical and Applied Genetics 130**:** 621-633.

Mason, A. S., J. Zhang, R. Tollenaere, P. Vasquez Teuber, J. Dalton-Morgan *et al.*, 2015 High-throughput genotyping for species identification and diversity assessment in germplasm collections. Molecular Ecology Resources 15**:** 1091-1101.

Masterson, J., 1994 Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264**:** 421-424.

Mello-Sampayo, T., 1971 Genetic regulation of meiotic chromosome pairing by chromosome 3D of *Triticum aestivum*. Nature: New Biology 230**:** 22-23.

Mercier, R., C. Mezard, E. Jenczewski, N. Macaisne and M. Grelon, 2015 The molecular biology of meiosis in plants. Annual Review of Plant Biology 66**:** 297-327.

Mercier, R., D. Vezon, E. Bullier, J. C. Motamayor, A. Sellier *et al.*, 2001 SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. Genes and Development 15**:** 1859-1871.

Michael, T. P., and R. VanBuren, 2020 Building near-complete plant genomes. Current Opinion in Plant Biology 54**:** 26-33.

Moore, G., 2000 Cereal Chromosome Structure, Evolution, and Pairing. Annual Review of Plant Physiology and Plant Molecular Biology 51**:** 195-222.

Morgan, C., H. Zhang, C. E. Henry, F. C. H. Franklin and K. Bomblies, 2020 Derived alleles of two axis proteins affect meiotic traits in autotetraploid *Arabidopsis arenosa*. Proceedings of the National Academy of Sciences 117**:** 8980.

Murray, M. G., and W. F. Thompson, 1980 Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8**:** 4321-4326.

Nakagawa, T., H. Flores-Rozas and R. D. Kolodner, 2001 The MER3 helicase involved in meiotic crossing over is stimulated by single-stranded DNA-binding proteins and unwinds DNA in the 3' to 5' direction. Journal of Biological Chemistry 276**:** 31487-31493.

Nasmyth, K., 2002 Segregating sister genomes: the molecular biology of chromosome separation. Science 297**:** 559-565.

Neale, M. J., and S. Keeney, 2006 Clarifying the mechanics of DNA strand exchange in meiotic recombination. Nature 442**:** 153-158.

Nguepjop, J. R., H.-A. Tossim, J. M. Bell, J.-F. Rami, S. Sharma *et al.*, 2016 Evidence of Genomic Exchanges between Homeologous Chromosomes in a Cross of Peanut with Newly Synthetized Allotetraploid Hybrids. Frontiers in Plant Science 7**:** 1635.

Nicolas, S. D., M. Leflon, H. Monod, F. Eber, O. Coriton *et al.*, 2009 Genetic Regulation of Meiotic Cross-Overs between Related Genomes in *Brassica napus* Haploids and Hybrids. The Plant Cell 21**:** 373.

Nicolas, S. D., H. Monod, F. Eber, A. M. Chèvre and E. Jenczewski, 2012 Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization. The Plant Journal 70**:** 691-703.

Nonomura, K., A. Morohoshi, M. Nakano, M. Eiguchi, A. Miyao *et al.*, 2007 A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell 19**:** 2583-2594.

Obeso, D., R. J. Pezza and D. Dawson, 2014 Couples, pairs, and clusters: mechanisms and implications of centromere associations in meiosis. Chromosoma 123**:** 43-55.

Oh, S. A., V. Bourdon, H. G. Dickinson, D. Twell and S. K. Park, 2014 Arabidopsis Fused kinase TWO-IN-ONE dominantly inhibits male meiotic cytokinesis. Plant Reproduction 27**:** 7-17.

Osman, F., J. Dixon, C. L. Doe and M. C. Whitby, 2003 Generating crossovers by resolution of nicked Holliday junctions: a role for Mus81-Eme1 in meiosis. Molecular Cell 12**:** 761-774.

Osman, K., J. D. Higgins, E. Sanchez-Moran, S. J. Armstrong and F. C. Franklin, 2011 Pathways to meiotic recombination in *Arabidopsis thaliana*. New Phytologist 190**:** 523-544.

Osman, K., E. Sanchez-Moran, S. C. Mann, G. H. Jones and F. C. Franklin, 2009 Replication protein A (AtRPA1a) is required for class I crossover formation but is dispensable for meiotic DNA break repair. EMBO Journal 28**:** 394-404.

Page, S. L., and R. S. Hawley, 2004 The genetics and molecular biology of the synaptonemal complex. Annual Review of Cell and Developmental Biology 20**:** 525-558.

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.

Parkin, I. A., A. G. Sharpe, D. J. Keith and D. J. Lydiate, 1995 Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome 38**:** 1122-1131.

Parkin, I. A. P., W. E. Clarke, C. Sidebottom, W. Zhang, S. J. Robinson *et al.*, 2010 Towards unambiguous transcript mapping in the allotetraploid *Brassica napus*. Genome 53**:** 929-938.

Parkin, I. A. P., S. M. Gulden, A. G. Sharpe, L. Lukens, M. Trick *et al.*, 2005 Segmental structure of the Brassica napus genome based on comparative analysis with Arabidopsis thaliana. Genetics 171**:** 765-781.

Perumal, S., C. S. Koh, L. Jin, M. Buchwaldt, E. E. Higgins *et al.*, 2020 A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral Brassica genome. Nature plants 6**:** 929-941.

Pezza, R. J., R. D. Camerini-Otero and P. R. Bianco, 2010 Hop2-Mnd1 condenses DNA to stimulate the synapsis phase of DNA strand exchange. Biophysical Journal 99**:** 3763-3772.

Pfalz, M., A. Gonzalo, N. Christophorou, A. Blary, A. Berard *et al.*, 2020 Identifying and Isolating Meiotic Mutants in a Polyploid Brassica Crop, pp. 303-318 in *Plant Meiosis: Methods and Protocols*, edited by M. Pradillo and S. Heckmann. Springer New York, New York, NY.

Puizina, J., J. Siroky, P. Mokros, D. Schweizer and K. Riha, 2004 Mre11 deficiency in *Arabidopsis* is associated with chromosomal instability in somatic cells and Spo11-dependent genome fragmentation during meiosis. Plant Cell 16**:** 1968-1978.

Qian, L., W. Qian and R. J. Snowdon, 2014 Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. BMC Genomics 15**:** 1170.

Ramsey, J., and D. W. Schemske, 1998 PATHWAYS, MECHANISMS, AND RATES OF POLYPLOID FORMATION IN FLOWERING PLANTS. Annual Review of Ecology and Systematics 29**:** 467-501.

Read, A. C., M. J. Moscou, A. V. Zimin, G. Pertea, R. S. Meyer *et al.*, 2020 Genome assembly and characterization of a complex zfBED-NLR gene-containing disease resistance locus in Carolina Gold Select rice with Nanopore sequencing. PLoS Genetics 16**:** e1008571.

Rey, M.-D., A. C. Martín, J. Higgins, D. Swarbreck, C. Uauy *et al.*, 2017 Exploiting the *ZIP4* homologue within the wheat *Ph1* locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. Molecular Breeding 37**:** 95.

Riley, R., and V. Chapman, 1958 Genetic control of the cytologically diploid behaviour of hexaploid wheat. Nature 182**:** 713-715.

Roberts, N. Y., K. Osman and S. J. Armstrong, 2009 Telomere distribution and dynamics in somatic and meiotic nuclei of *Arabidopsis thaliana*. Cytogenetics and Genome Research 124**:** 193-201.

Rohrig, S., A. Dorn, J. Enderle, A. Schindele, N. J. Herrmann *et al.*, 2018 The RecQ-like helicase HRQ1 is involved in DNA crosslink repair in Arabidopsis in a common pathway with the Fanconi anemia-associated nuclease FAN1 and the postreplicative repair ATPase RAD5A. New Phytologist 218**:** 1478-1490.

Rommel Fuentes, R., T. Hesselink, R. Nieuwenhuis, L. Bakker, E. Schijlen *et al.*, 2020 Meiotic recombination profiling of interspecific hybrid F1 tomato pollen by linked read sequencing. Plant J 102**:** 480-492.

Rousseau-Gueutin, M., J. Morice, O. Coriton, V. Huteau, G. Trotoux *et al.*, 2017 The Impact of Open Pollination on the Structural Evolutionary Dynamics, Meiotic Behavior, and Fertility of Resynthesized Allotetraploid *Brassica napus* L. G3: Genes|Genomes|Genetics 7**:** 705.

Samans, B., B. Chalhoub and R. J. Snowdon, 2017 Surviving a Genome Collision: Genomic Signatures of Allopolyploidization in the Recent Crop Species *Brassica napus*. The Plant Genome 10**:** 1-15.

Sanchez-Moran, E., S. J. Armstrong, J. L. Santos, F. C. Franklin and G. H. Jones, 2002 Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana*. Genetics 162**:** 1415-1422.

Sanchez-Moran, E., J. L. Santos, G. H. Jones and F. C. Franklin, 2007 ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis. Genes and Development 21**:** 2220-2233.

Schranz, M. E., M. A. Lysak and T. Mitchell-Olds, 2006 The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends in Plant Science 11**:** 535-542.

Sears, E. R., 1976 Genetic control of chromosome pairing in wheat. Annual Review of Genetics 10**:** 31-51.

Sears, E. R., 1977 GENETICS SOCIETY OF CANADA AWARD OF EXCELLENCE LECTURE AN INDUCED MUTANT WITH HOMOEOLOGOUS PAIRING IN COMMON WHEAT. Canadian Journal of Genetics and Cytology 19**:** 585-593.

Sears, E. R., and M. Okamoto, 1958 Intergenomic chromosome relationship in hexaploid wheat, pp. 258-259 in *Proceedings of the X International Congress of Genetics*

Sebastian, J., M. Ravi, S. Andreuzza, A. P. Panoli, M. P. Marimuthu *et al.*, 2009 The plant adherin AtSCC2 is required for embryogenesis and sister-chromatid cohesion during meiosis in *Arabidopsis*. Plant Journal 59**:** 1-13.

Seeliger, K., S. Dukowic-Schulze, R. Wurz-Wildersinn, M. Pacher and H. Puchta, 2012 BRCA2 is a mediator of RAD51- and DMC1-facilitated homologous recombination in *Arabidopsis thaliana*. New Phytologist 193**:** 364-375.

Serra, H., R. Svačina, U. Baumann, R. Whitford, T. Sutton *et al.*, 2021 *Ph2* encodes the mismatch repair protein MSH7-3D that inhibits wheat homoeologous recombination. Nature Communications 12**:** 803.

Shah, R., R. J. Bennett and S. C. West, 1994 Genetic recombination in E. coli: RuvC protein cleaves Holliday junctions at resolution hotspots in vitro. Cell 79**:** 853-864.

Shalev, G., Y. Sitrit, N. Avivi-Ragolski, C. Lichtenstein and A. A. Levy, 1999 Stimulation of homologous recombination in plants by expression of the bacterial resolvase ruvC. Proceedings of the National Academy of Sciences U S A 96**:** 7398-7402.

Sharpe, A. G., I. A. Parkin, D. J. Keith and D. J. Lydiate, 1995 Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38**:** 1112-1121.

Sheehan, M. J., and W. P. Pawlowski, 2009 Live imaging of rapid chromosome movements in meiotic prophase I in maize. Proceedings of the National Academy of Sciences U S A 106**:** 20989-20994.

Shunmugam, A. S. K., V. Bollina, S. Dukowic-Schulze, P. K. Bhowmik, C. Ambrose *et al.*, 2018 MeioCapture: an efficient method for staging and isolation of meiocytes in the prophase I sub-stages of meiosis in wheat. BMC Plant Biology 18**:** 293.

Sidhu, G. K., S. Rustgi, M. N. Shafqat, D. von Wettstein and K. S. Gill, 2008 Fine structure mapping of a gene-rich region of wheat carrying Ph1, a suppressor of crossing over between homoeologous chromosomes. Proceedings of the National Academy of Sciences U S A 105**:** 5815-5820.

Singh, D. K., S. Andreuzza, A. P. Panoli and I. Siddiqi, 2013 AtCTF7 is required for establishment of sister chromatid cohesion and association of cohesin with chromatin during meiosis in *Arabidopsis*. BMC Plant Biology 13**:** 117.

Smith, G. R., M. N. Boddy, P. Shanahan and P. Russell, 2003 Fission yeast Mus81.Eme1 Holliday junction resolvase is required for meiotic crossing over but not for gene conversion. Genetics 165**:** 2289-2293.

Snowden, T., S. Acharya, C. Butz, M. Berardini and R. Fishel, 2004 hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. Molecular Cell 15**:** 437-451.

Snowdon, R. J., 2007 Cytogenetics and genome analysis in Brassica crops. Chromosome Research 15**:** 85-95.

Snowdon, R. J., and F. L. Iniguez Luy, 2012 Potential to improve oilseed rape and canola breeding in the genomics era. Plant breeding 131**:** 351-360.

Soyk, S., Z. H. Lemmon, F. J. Sedlazeck, J. M. Jimenez-Gomez, M. Alonge *et al.*, 2019 Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. Nature Plants 5**:** 471-479.

Sprink, T., and F. Hartung, 2014 The splicing fate of plant SPO11 genes. Frontiers in Plant Science 5**:** 214.

Starr, D. A., and H. N. Fridolfsson, 2010 Interactions between nuclei and the cytoskeleton are mediated by SUN-KASH nuclear-envelope bridges. Annual Review of Cell and Developmental Biology 26**:** 421-444.

Stein, A., O. Coriton, M. Rousseau-Gueutin, B. Samans, S. V. Schiessl *et al.*, 2017 Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. Plant Biotechnology Journal 15**:** 1478-1489.

Storlazzi, A., S. Gargano, G. Ruprich-Robert, M. Falque, M. David *et al.*, 2010 Recombination proteins mediate meiotic spatial chromosome organization and pairing. Cell 141**:** 94-106.

Sun, H., B. A. Rowan, P. J. Flood, R. Brandt, J. Fuss *et al.*, 2019 Linked-read sequencing of gametes allows efficient genome-wide analysis of meiotic recombination. Nature Communications 10**:** 4310.

Sym, M., J. A. Engebrecht and G. S. Roeder, 1993 ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. Cell 72**:** 365-378.

Symonds, V. V., P. S. Soltis and D. E. Soltis, 2010 Dynamics of polyploid formation in Tragopogon (Asteraceae): recurrent formation, gene flow, and population structure. Evolution 64**:** 1984-2003.

Szadkowski, E., F. Eber, V. Huteau, M. Lode, C. Huneau *et al.*, 2010 The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytologist 186**:** 102-112.

Tam, S. M., J. B. Hays and R. T. Chetelat, 2011 Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato. Theoretical and Applied Genetics 123**:** 1445-1458.

Tieman, D., G. Zhu, M. F. R. Resende, T. Lin, C. Nguyen *et al.*, 2017 A chemical genetic roadmap to improved tomato flavor. Science 355**:** 391.

Tulpová, Z., H. Toegelová, N. L. V. Lapitan, F. B. Peairs, J. Macas *et al.*, 2019 Accessing a Russian Wheat Aphid Resistance Gene in Bread Wheat by Long-Read Technologies. The Plant Genome 12**:** 180065.

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

Uanschou, C., A. Ronceret, M. Von Harder, A. De Muyt, D. Vezon *et al.*, 2013 Sufficient amounts of functional HOP2/MND1 complex promote interhomolog DNA repair but are dispensable for intersister DNA repair during meiosis in Arabidopsis. Plant Cell 25**:** 4924-4940.

Uanschou, C., T. Siwiec, A. Pedrosa-Harand, C. Kerzendorfer, E. Sanchez-Moran *et al.*, 2007 A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. EMBO Journal 26**:** 5061-5070.

Udall, J. A., P. A. Quijada and T. C. Osborn, 2005 Detection of chromosomal rearrangements derived from homologous recombination in four mapping populations of *Brassica napus L*. Genetics 169**:** 967-979.

Vrielynck, N., A. Chambon, D. Vezon, L. Pereira, L. Chelysheva *et al.*, 2016 A DNA topoisomerase VI-like complex initiates meiotic recombination. Science 351**:** 939-943.

Wang, M., K. Wang, D. Tang, C. Wei, M. Li *et al.*, 2010 The central element protein ZEP1 of the synaptonemal complex regulates the number of crossovers during meiosis in rice. Plant Cell 22**:** 417-430.

Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao *et al.*, 2014 Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnol J 12**:** 787-796.

Wang, X., H. Wang, J. Wang, R. Sun, J. Wu *et al.*, 2011 The genome of the mesopolyploid crop species *Brassica rapa*. Nature genetics 43**:** 1035-1040.

Wang, X., K. Yu, H. Li, Q. Peng, F. Chen *et al.*, 2015 High-Density SNP Map Construction and QTL Identification for the Apetalous Character in *Brassica napus L*. Frontiers in Plant Science 6**:** 1164.

Wang, Y., Z. Cheng, J. Huang, Q. Shi, Y. Hong *et al.*, 2012 The DNA replication factor RFC1 is required for interference-sensitive meiotic crossovers in *Arabidopsis thaliana*. PLoS Genetics 8**:** e1003039.

Wendel, J. F., 2000 Genome evolution in polyploids. Plant Molecular Biology 42**:** 225-249.

Wijeratne, A. J., C. Chen, W. Zhang, L. Timofejeva and H. Ma, 2006 The *Arabidopsis thaliana* PARTING DANCERS gene encoding a novel protein is required for normal meiotic homologous recombination. Molecular Biology of the Cell 17**:** 1331-1343.

Wright, K. M., B. Arnold, K. Xue, M. Surinova, J. O'Connell *et al.*, 2015 Selection on meiosis genes in diploid and tetraploid *Arabidopsis arenosa*. Molecular Biology and Evolution 32**:** 944-955.

Wu, Y., P. R. Bhat, T. J. Close and S. Lonardi, 2008 Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet 4**:** e1000212.

Xiong, Z., R. T. Gaeta and J. C. Pires, 2011 Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proceedings of the National Academy of Sciences 108**:** 7908.

Xiong, Z., and J. C. Pires, 2011 Karyotype and identification of all homoeologous chromosomes of allopolyploid *Brassica napus* and its diploid progenitors. Genetics 187**:** 37-49.

Yang, X., L. Timofejeva, H. Ma and C. A. Makaroff, 2006 The *Arabidopsis* SKP1 homolog ASK1 controls meiotic chromosome remodeling and release of chromatin from the nuclear membrane and nucleolus. Journal of Cell Science 119**:** 3754-3763.

Yang, Y., Y. Shen, S. Li, X. Ge and Z. Li, 2017 High Density Linkage Map Construction and QTL Detection for Three Silique-Related Traits in *Orychophragmus violaceus* Derived *Brassica napus* Population. Frontiers in Plant Science 8**:** 1512.

Yant, L., Jesse D. Hollister, Kevin M. Wright, Brian J. Arnold, James D. Higgins *et al.*, 2013 Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa. Current Biology 23**:** 2151-2156.

Yao, X., K. L. Anderson and D. W. Cleveland, 1997 The microtubule-dependent motor centromere-associated protein E (CENP-E) is an integral component of kinetochore corona fibers that link centromeres to spindle microtubules. Journal of Cell Biology 139**:** 435-447.

Zakharyevich, K., S. Tang, Y. Ma and N. Hunter, 2012 Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. Cell 149**:** 334-347.

Zamariola, L., N. De Storme, K. Vannerum, K. Vandepoele, S. J. Armstrong *et al.*, 2014a SHUGOSHINs and PATRONUS protect meiotic centromere cohesion in *Arabidopsis thaliana*. Plant Journal 77**:** 782-794.

Zamariola, L., C. L. Tiang, N. De Storme, W. Pawlowski and D. Geelen, 2014b Chromosome segregation in plant meiosis. Frontiers in Plant Science 5**:** 279.

Zhai, Y., S. Cai, L. Hu, Y. Yang, O. Amoo *et al.*, 2019 CRISPR/Cas9-mediated genome editing reveals differences in the contribution of *INDEHISCENT* homologues to pod shatter resistance in *Brassica napus* L. Theoretical and Applied Genetics 132**:** 2111-2123.

Zhang, C., Y. Song, Z. H. Cheng, Y. X. Wang, J. Zhu *et al.*, 2012 The *Arabidopsis thaliana* DSB formation (AtDFO) gene is required for meiotic double-strand break formation. Plant Journal 72**:** 271-281.

Zhang, L., X. Cai, J. Wu, M. Liu, S. Grob *et al.*, 2018 Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. Horticulture Research 5**:** 50.

Zhang, L., S. Wang, S. Yin, S. Hong, K. P. Kim *et al.*, 2014 Topoisomerase II mediates meiotic crossover interference. Nature 511**:** 551-556.

Zhao, D., X. Yang, L. Quan, L. Timofejeva, N. W. Rigel *et al.*, 2006 ASK1, a SKP1 homolog, is required for nuclear reorganization, presynaptic homolog juxtaposition and the proper distribution of cohesin during meiosis in *Arabidopsis*. Plant Molecular Biology 62**:** 99-110.

Zhao, K., C. W. Tung, G. C. Eizenga, M. H. Wright, M. L. Ali *et al.*, 2011 Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2**:** 467.

Zhou, S., Y. Wang, W. Li, Z. Zhao, Y. Ren *et al.*, 2011 Pollen semi-sterility1 encodes a kinesin-1-like protein important for male meiosis, anther dehiscence, and fertility in rice. Plant Cell 23**:** 111-129.