UNIVERSITY OF BIRMINGHAM

Phenotypic and meiotic differences between diploid and polyploid plants

by

Anushree Choudhary

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

University of Birmingham

July 2019

ABSTRACT

Polyploidy is present in a large number of crop plants and is considered as one of the driving forces in the evolution of angiosperms. Unlocking genetic variation in various autopolyploid crop plants is highly relevant to crop breeders. Homologous recombination, a tightly controlled cell process during the production of gametes in meiosis, is responsible for creation of genetic variation. Owing to the presence of more than two homologous chromosomes, polyploid meiosis faces a variety of challenges, such as multivalent formation and mis-segregation.

Using a plant trial with more than 300 diploid and tetraploid *Arabidopsis thaliana* F2 individuals, significant differences were found in various traits between the two populations. Cytological analysis using FISH on diploid and tetraploid plants revealed an overall increase in meiotic recombination in tetraploids, although the per bivalent frequency was reduced. The process of meiotic recombination was further explored in potato (*Solanum tuberosum),* a globally important autotetraploid crop. Chiasma frequency and multivalent frequency for chromosomes 1 and 2 varied according to variety, where the diploid variety showed a reduced chiasma frequency compared with tetraploid varieties. Immunolocalisation of the axis and synaptonemal complex proteins, ASY1 and ZYP1, demonstrated the complexities that may arise during meiosis in an autotetraploid plant.

I dedicate this thesis to my family and the beech tree in my backyard.

Acknowledgements

I would like to thank my supervisor, Dr Lindsey Leach, for providing me an opportunity to work on this project. I am grateful for her support and guidance. I would at the same time like to thanks my funding body BBSRC, which enabled me to embark on this journey.

I thank Prof Zewei Luo for his help in getting my plant samples sequenced, and sharing interesting ideas. I would like to thank Dr Sue Armstrong for allowing me to use space and equipments in the cytology lab. I am grateful to Dr Eugenio Sanchez-Moran for guiding me with the cytological analysis and sharing his wisdom. I am also thankful to Dr Juliet Coates for various wise and useful advices.

I would like to thank Dr Daniel Gibbs, Dr Kim Osman, Dr Marina Martinez Garcia, Dr Chris Morgan, Dr Mark Bailey, Dr Harriet Davies, Dr Petra Stamn, Dr Jing Chen and Dr Garima Sharma for various help and moments. Special thanks to Steve Price for the technical support he provided during my PhD. Thanks to Karen Stapes and Adrian Breckles for taking care of my plants in the glasshouse. Thanks to all the fellow PhD students and people on the second and first floor who have enriched my journey one way or the other.

Finally, I want to thank my parents and my in-laws for always encouraging me during the project. I would like to thank my friends for their kind support during my crisis time, which enabled me to carry on. At last, though not the least, a big thanks to my husband and son for keeping patience and supporting me throughout.

Contributions - Publication and Public Engagement

Publication: "Varietal Variation in Meiotic Chromosome Behaviour in *Solanum tuberosum.*" Draft manuscript has been prepared and it is ready to be submitted to the journal 'Chromosoma'.

Public engagement: I undertook a module on public engagement in the second year of my PhD. I enjoyed the module, learning how to engage actively with our real science funders, the public at large, at the same time realising it to be a tool to inspire younger generation to develop curiosity in science. Since then, I have undertaken a few activities as follows:

1) Feb, 2015: I engaged with the children and adults alike, with an activity called 'Super-plants', in Thinktank Birmingham Science Museum. Talking about how plants can tolerate polyploidy, and how fascinating chromosomes, DNA and meiosis can be, children and adult were engaged with interactive games of identifying diploid and polyploid plants and simulating meiosis using plasticine.
2) April, 2015: After completing the public engagement module, I took up another activity in Thinktank. This time it involved meeting and inspiring Brownies of different age groups, involving activities and chats about importance of plants. One of the main highlight of the activity was to identify three famous women in science: Barbara McClintock, Lynn Margulis and Beatrix Potter through clues, such as stories illustrating their lives.
3) Sept, 2015: I participated in an event 'Girls in STEM', University of Birmingham to encourage year 9 girls from local Birmingham schools to consider their future career

in sciences. I gave a presentation about how they can aim for higher education and talked about my research work on polyploidy and meiosis. They participated in hands on session on strawberry DNA extraction and looking at Brassica pollen under the microscope.

4) Feb, 2017: I participated with my lab members in this day long activity in Thinktank museum, interacting with families, talking about potatoes and how genetics can help in breeding useful potato varieties. I illustrated strawberry DNA extraction, giving ideas to children how they can try this at home with daily household things, of course under parents supervision.
5) Feb, 2019: This event was undertaken in Thinktank again, with my lab members discussing how humble potatoes can be a great source of energy and various vitamins. There were hands on activities with plasticine, making potato clock and drawing various potato heads.

The engagement activities undertaken have been motivating and satisfying experience, developing my ability to think on the feet and try to answer the questions which are not in my comfortable zone.

Table of Contents

List of Appendices

List of Figures

List of tables

List of Abbreviations

Abscisic Acid	ABA
Amplified Fragment Length Polymorphism	AFLP
Asynaptic	ASY
Basal Branches	BB
Bloom Syndrome Helicase	BLM
Bovine Serum Albumin	BSA
Cauline Leaf Number	CLN
Cetyltrimethylammonium Bromide	CTAB
Columbia	Col
Crossover	CO
CtBP Interacting Protein	CTIP
Cyclin Dependent Kinase	CDK/cdc2
Cycling DOF Factor	CDF
Days to Flower	DTF
Days to Germination	DTG
4',6-diamidino-2-phenylindole	DAPI
Deoxyribonucleic acid	DNA
Diethyl Pyrocarbonate	DEPC
Displacement loop	D loop
Disrupted Meiotic cDNA 1	DMC1
Digoxigenin	DIG
Double Holliday junction	dHj
Double Stranded Breaks	DSB
DSB Forming	DFO
Epigenetic Recombinant Inbred Lines	EpiRIL
Ethylenediaminetetraacetic Acid	EDTA
Exonuclease 1	EXO1
Fanconi Anemia, Complementation Group M	FANCM
Fidgetin-Like 1	FIGL1
Fluorescence *in situ* Hybridisation	FISH
Fluorescein Isothiocyanate	FITC
E3 Ubiquitin Protein Ligase	HEI10
Homologous Pairing Protein	HOP
Hydrochloric Acid	HCl
Increased Recombination Centres 20	IRC20
Landsberg	Ler
Lateral Branches	LB
Life Cycle	LC
Meiotic Nuclear Division Protein	MND
MEIosis Specific	MEI
Methyl Methane Sulfonate Sensitivity 4	MMS4
Meiotic Recombination 11	MRE11

Mut L Homologue	MLH
Mut S Homologue	MSH
Sodium Chloride	NaCl
Next Generation Sequencing	NGS
Nijmegen Breakage Syndrome 1	NBS1
Non Crossover	NCO
Nucleolar Organizing Region	NOR
ParTing Dancer	PTD
Polymerase Chain Reaction	PCR
Phosphate Buffered Saline	PBS
Putative Recombination Initiation Defect	PRD
Pairing Homoeologue	Ph
Quantitative Trait Loci	QTL
Radiation Sensitive 51	RAD51
Random Amplified Polymorphic DNA	RAPD
Reactive Oxygen Species	ROS
REcombination	REC
Recombination Nodules	RNs
Reproductive Period	RP
Revolutions per Minute	RPM
Restriction Fragment Length Polymorphism	RFLP
Ribonucleic Acid	RNA
Ribosomal DNA	rDNA
Rosette Leaf Number	RLN
Saline Sodium Chloride	SSC
Shortage in Chiasma	SHOC
Single End Invasions	SEI
Simple Sequence Length Polymorphism	SSLP
Single Nucleotide Polymorphism	SNP
Slow Growth Suppressor	SGS
Super Killer	SKI
Suppressor of RAD six 2	SRS2
Sodium Chloride	NaCl
Sodium Dodecyl Sulphate	SDS
Sporulation in absence of Spo11 2	SAE2
Sporulation 11	Spo 11
Sterile Distilled Water	SDW
Strong Culm	SCM
Switch 1	SWI1
Synaptonemal Complex	SC
Synthesis Dependent Strand Annealing	SDSA
Synthetic Lethal of unknown(X) function	SLX
Archaeal Topoisomerase	TOPVI
Total Branches	TB

Total Leaf Number	TLN
Tris-boric Acid-EDTA	TBE
X-Ray Repair Cross Complementing 3	XRCC3
X-Ray Sensitive	XRS
Yeast Endodeoxyribonuclease 1	YEN1
Zip Mer Msh	ZMM
Zinc Transporter	ZIP

CHAPTER 1

GENERAL INTRODUCTION

1 General Introduction

1.1 Quantitative Genetics

Quantitative genetics refers to the study of the inheritance of traits that are affected by more than one gene and to a considerable extent by the environment. It is also referred to as biometrical genetics (Kearsey *et al.*, 1996). A few examples of quantitative traits are height, size, blood pressure and most diseases in humans, litter size, milk production in animals, yield and growth in plants. The variation between individuals for these traits is often high and a range of phenotypes showing a continuous distribution can be seen. This is in contrast to the simple Mendelian inheritance of qualitative traits where variation is due to the influence of usually one (or two) genes showing categorical distribution. Quantitative traits are also referred to as complex traits; complex because many genes as well as the environment influence them and they show continuous variation due to the segregation of the genes at many loci, effects of each of which may be small compared with the effects of the environment (Kearsey *et al.*, 1996). This makes analysis of an individual trait gene difficult or complex (Hill, 2010). Quantitative genetic analysis involves the interpretation of the phenotypic observations of the traits by the use of different statistical methods such as variance component analysis to determine the action of the quantitative genes (Hill, 2010).

The pioneers of quantitative genetics include Sir Ronald Fisher and Sewall Wright, who invented statistical methods such as analysis of variance and path coefficients to explain the

variance components of the traits linking genotype with the phenotype (Hill, 2010). Charles Darwin was aware about the accumulating variations responsible for evolution well before Fisher and Wright, but could not provide a model for inheritance of the accumulating trait variations (Darwin, 1859). Francis Galton and Karl Pearson also studied polygenic human traits in late 19th and early 20th centuries, but were not able to establish their inheritance patterns (Kearsey *et al.*, 1996). Around same time in the 19th century, Mendel also published his laws of inheritance, which went unnoticed at the time but were rejuvenated in 1900 by Hugo de Vries, Carl Correns and Tschermak and later by William Bateson. In the early 20th century, Morgan Hunt and Alfred Sturtevant provided necessary proofs to establish the chromosomal theory of inheritance and linkage along the chromosome, which explained the variability in the offspring population of fruit flies (Atherly, Girton and McDonald, 1999). There were two school of thoughts, one based on Mendel's work and the other on the biometrical methods of Galton and Pearson. Sir Ronald Fisher in 1918 unified both by studying the genetics of the quantitative traits and linking it with the phenotype variation as observed for various traits (Visscher and Walsh, 2017).

1.1.1 Quantitative Trait Loci (QTL)

Kenneth Mather (1949) coined the term polygenes for multiple genes affecting the quantitative/polygenic trait. These polygenes are now known as quantitative trait loci (QTL) (Kearsey *et al.*, 1996), and can be defined as regions of the genome which modulate the phenotypic variation of the quantitative trait (Abiola *et al.*, 2003). The knowledge of quantitative traits is important for plant and animal breeders in numerous ways. For example, they want to know the extent of genetic variation of a trait, which can be selected for

improvement by artificial selection. It is equally important for evolutionary biologists to understand how evolution occurs in natural populations, and for the study of human/animal behaviour or the susceptibility to a disease or pest resistance in plants and animals including humans (Kearsey *et al.*, 1996).

1.1.1.1 Major and minor genes

In classical genetics, most of the genes have been discovered by a chance mutation affecting the alleles of a phenotypic trait that could be clearly identified. A few examples of such alleles are scabrous, bobbed and scute bristle in *Drosophila* selected for bristle number, and the Booroola F gene selected in sheep for increased ovulation. These major genes were deleterious in nature and hence occurred in very low frequency in unselected populations. However, major alleles, which have positive effect on the quantitative traits can contribute to selection response, hence one of the major efforts in early QTL analysis was in finding a major gene affecting a quantitative trait (Falconer and Mackay, 1996).

QTL may include a mixture of major and minor genes, where each gene shows Mendelian inheritance. Major genes are identified as the alleles having a large effect on the phenotypic trait, while the minor genes are considered to contribute a small amount of variation to the phenotype. These genes/alleles can interact with each other. There are three types of gene/allele action based on their interactions (Falconer and Mackay, 1996):

1) Additive gene action, which is a result of additive effects of each allele, where the alleles contribute equally to a phenotype at a locus.

2) Dominance gene action, based on interaction between alleles at a locus resulting in complete, incomplete or over-dominance.

3) Epistasis, where there is interaction between alleles at different loci.

It is the additive gene action, which is of most interest to breeders, as it can lead to an increase in the value of that trait in the offspring through selection. The first step to quantify it, is by establishing the trait heritability.

1.1.1.2 *Trait Heritability*

It is important to know, especially for breeders, the relative contribution of the genetic or environmental factors to the phenotypic variation in a trait. Heritability helps in identifying this. It is a population concept, which holds true only under the environment for which it is calculated. It is defined as the proportion of genotypic variance to the total phenotypic variance of a trait in a specific population in a particular environment. The genotypic variance can be further partitioned into additive genetic, dominance and epistasis effects, giving estimates of narrow sense or broad sense heritability respectively (Visscher, Hill and Wray, 2008).

Thus, the observed phenotype (P) of a trait can be expressed as the sum total of unobserved genotype (G) and environmental factors (E):

Phenotype (P) = Genotype (G) + Environment (E).

This can be expressed in terms of variance as: $V_P = V_G + V_E$.

Broad sense heritability (H^2) is thus expressed as: $H^2 = V_G/V_P$.

Genetic variance can be further partitioned into additive, dominance or epistatic genetic effects as: $V_G = V_A + V_D + V_I$.

Narrow sense heritability (h^2) or the additive gene effects on which selection works can be expressed as: $h^2 = V_A / V_P$.

The values thus obtained are only an estimate of the heritability of the traits because they can change with a change in the population structure, environment and even by the differences in phenotype collection (Falconer and Mackay, 1996). The values range from 0 to 1 and the same character may show different values reflecting the conditions under which they were studied. Nevertheless, the values of the traits can be compared between different populations to draw practical insights into the biological processes responsible for the trait and its variation (Visscher, Hill and Wray, 2008). A comparison in different animal species and humans showed that morphological traits such as body size, weight and height had higher heritability estimates than fitness traits such as life history traits, fertility and calving success across different environments (Visscher, Hill and Wray, 2008).

1.1.1.3 Mapping Quantitative Trait Loci

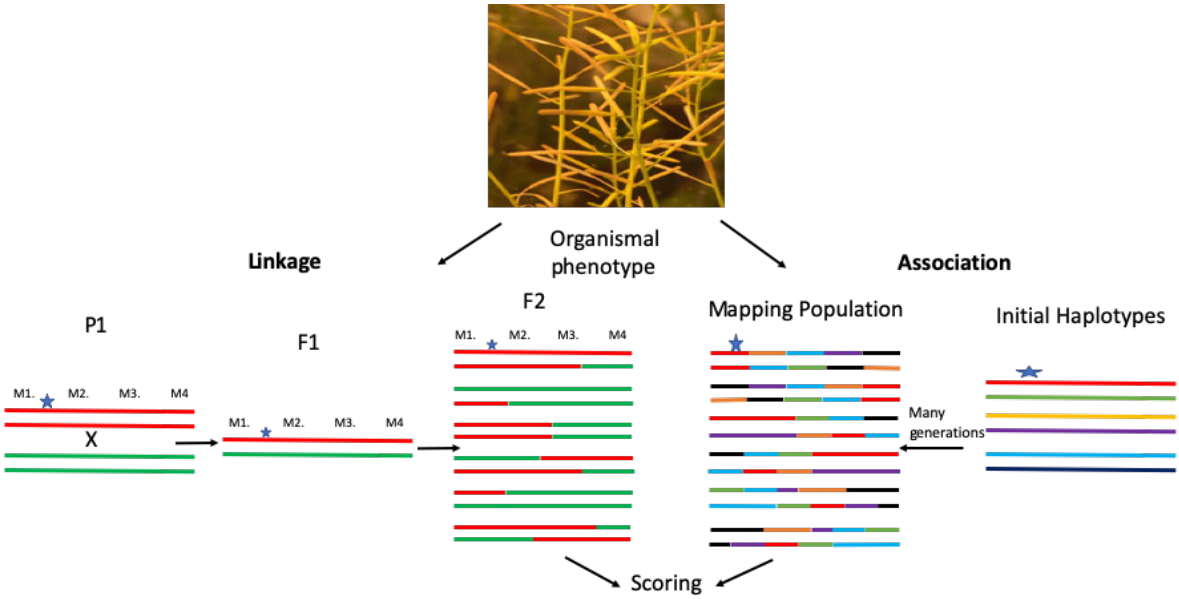
QTL mapping requires the integration of phenotype score of the trait with the marker genotype score in a mapping population using statistical methods. Mapping populations can involve segregating populations (plants and animals), pedigrees (domestic animals and humans) and natural populations (all species). Heritability serves as the first step towards QTL mapping analysis. The higher the heritability, the greater the chances of success in identifying the QTL. QTLs are identified by making use of the molecular markers, which are considered to be in linkage with the QTL region of the genome (Mackay, Stone and Ayroles, 2009). With the

advent of the genomics era and reduction in the cost of genotyping, hundreds of molecular markers can be made available for QTL detection.

Molecular markers refer to the variation in the DNA sequences, proteins or metabolites between two individuals, which can be used to identify a particular trait or disease due to the variation being inherited with the trait. DNA markers serve as an integral tool for QTL analysis. Various types of DNA molecular markers have been developed. They can be dominant markers, such as Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP) or codominant markers such as Restriction Fragment Length Polymorphisms (RFLPs), Simple Sequence Length Polymorphisms (SSLPs) and sequencing based Single Nucleotide Polymorphisms (SNP) (Gupta *et al.*, 1999). Dominant markers cannot distinguish between homozygotes and heterozygotes, while co-dominant markers can. Next generation sequencing technologies have changed the landscape of the availability of the molecular markers. They are high throughput technologies capable of generating thousands of markers without any previous knowledge of genome sequence. These technologies can be used to genotype whole genome, RNA, or only regions of interest, or reduce the genome complexity through use of restriction enzymes (Davey *et al.*, 2011).

Linkage analysis and Association Mapping/Genome Wide Association Study (GWAS) are the two most broadly used methods for QTL mapping. Linkage analysis involves segregating populations often created from two inbred parental lines, and is mostly used for QTL mapping in plants. The phenotype and the marker scores of the segregating populations are analysed

to find out which markers are associated with trait variation, as shown in **Figure 1-1** (Mackay, Stone and Ayroles, 2009).



Individual	M1	M2	M3	M4	Phenotype
1	2	2	2	1	20
2	1	0	0	0	8
3	2	2	0	0	24
4	1	1	2	1	4
5	1	0	0	0	14
6	1	1	2	1	12

Figure 1-1 Quantitative Trait Locus Mapping.

Figure adapted from Mackay *et al*. (2009). Markers 0 and 2 indicate homozygous and 1 is heterozygous.

Association studies tracks association between the genotype marker and the phenotypic trait in unrelated individuals sampled from large natural populations. Both linkage analysis as well as association mapping depends on the recombination between the locus of interest and the associated marker (Mackay, Stone and Ayroles, 2009).

1.1.1.4 QTL Mapping of Molecular Traits

A phenotype is a product of all the different processes involving gene expression, amount and quality of protein produced, and metabolites produced in the process. Hence, it is important

to understand the underlying mechanisms as well as inheritance patterns governing the three important shapers of phenotype. Gene expression, the process of reading the genetic code and its expression, involves the conversion of DNA into RNA and finally into protein and is similar in all organisms. However, the quantitative genetic expression which involves the amount of mRNA expressed can be different, not only for different organisms but for different individuals of the same species as well, depending on the variability of the underlying DNA (Jansen and Nap, 2001). The genomic regions governing the variation in expression of genes are known as expression QTL or eQTLs. They are divided into two classes: *cis* and *trans* eQTL, based on where they are present in the genome. *Cis*–eQTLs locate close to the target genomic region and control target gene expression, whereas *trans*-eQTLs are located away from the variable genomic region controlling the expression of the target either by itself or in coordination with other regulatory factors (Druka *et al.*, 2010). A gene regulatory network was constructed by conducting eQTL analysis and co-expression analysis of functionally related genes in *indica* rice identifying 5079 *cis*-eQTLs and 8568 *trans*-eQTLs. Out of these, 138 *trans*-eQTLs hotspots regulating expression variation of many genes were identified in the population, which included eight master regulators (Wang *et al.*, 2014). They identified various elements and pathways known to be involved in flowering regulation. This example suggests the utility of the approach to analyse and construct the gene regulatory networks and thus identifying the patterns of regulation of gene expression at the whole genome level.

Similar to the identification of eQTLs, there are several ways now to identify the differential protein and metabolic profile of individuals. As an important shaper of the phenotype, it is crucial to know and learn about the factors affecting the differential proteins and metabolic

profile. pQTL studies can help understand the genetic variations responsible for the proteomic variation in an organism. For example, Witzel *et al.* (2010) identified 51 pQTLs in barley and identified the underlying proteins involved in metabolism and defence processes, which can be directly used to improve barley grain quality. Genetic architecture of metabolic diversity was carried out in a RIL maize population to identify the variation in the primary metabolites in three tissues. 297 mQTLs with moderate to major effect were identified affecting carbohydrate metabolism, the tricarboxylic acid cycle and several amino acid synthesis and catabolic pathways (Wen *et al.*, 2015). Recently, four genes active in primary metabolism network in maize have been identified in a GWAS study, which are important in the essential amino acid production and signalling pathways in different tissues (Wen *et al.*, 2018). Identifying such metabolic variation in a population can inform breeders for improving nutritional quality of maize.

An integrated transcriptomic, proteomic and metabolomic genome wide dataset was generated using 162 RILs of *Arabidopsis thaliana*, which identified 6 different QTL hotspots acting together and applying to a wide range of phenotypes (Fu *et al.*, 2009). Though there were eQTLs for the 5000 transcripts analysed between the two parental lines, only a few actually produced a phenotype, indicating a system wide buffering of the effects of the genetic variation. This study shows that integrating the genetic variations with protein and metabolic variation and finally their effect on the phenotype can give us an idea about robust evolutionary mechanism developed by organisms.

1.1.1.5 QTL and Plant Breeding

QTL mapping has been used by plant breeders to select for the genes which affect the desirable trait positively. A study in potato describing QTL for late blight resistance on chromosomes 3 and 5, and QTL for foliage maturity on chromosome 5, found that the two QTLs on chromosome 5 were closely linked to the same genetic marker and could be one major gene with pleiotropic effect on both the traits (Visker *et al.*, 2003). Another study identified one major QTL for maturity on chromosome 5 in tetraploid potatoes and QTLs for yield and overall scores of yield on chromosomes 1 and 6. Six QTLs for after cooking blackening, four for regular tuber shape and four each for frying colour, quality and sprouting were identified, but they could only explain a minor proportion of the trait variance (Bradshaw *et al.*, 2008). In a separate association study in potato, plant maturity QTL was again identified on the north arm of chromosome 5, which was fine mapped to identify Cycling DOF Factor 1 (*CDF1*) gene, which is a circadian clock gene, with three different alleles affecting plant maturity and tuber development (Kloosterman *et al.*, 2013). Several alleles of various genes linked to tuber yield and starch content have been identified in a GWAS study on all the twelve potato chromosomes in a tetraploid potato population. Both the traits were found to be under linked genetic control, with a large number of differential SNPs identified having antagonistic effects, and some having synergistic effects (Schönhals *et al.*, 2017). The knowledge about QTLs of various traits can be utilized for improved and effective breeding. For example, identification of the *CDF1* alleles, will be important for developing cultivars with early or late maturity as required (Kloosterman *et al.*, 2013). Similarly, breeders can make use of the knowledge about underlying genes to enhance various traits such as protein and nutritional content.

Other studies have also identified loci that affect more than one trait. A QTL has been identified in rice for strong culm called *SCM2* QTL. It was found to have a positive effect on both panicle and spikelet number, thereby increasing the chances of improvement in rice by positively controlling yield as well as lodging resistance (Ookawa *et al.*, 2010). Yet another major QTL, *Ghd8* has been identified in rice, which plays several roles regulating grain productivity, height and heading date (Yan *et al.*, 2011). The pleiotropic effect of the semi dwarfing gene has been identified, having positive effect on the height, heading and flowering date in barley (Kuczyńska, Mikołajczak and Ćwiek, 2014). Identification of a quantitative pleiotropic locus can be effectively utilized in breeding programmes where manipulating a single locus can have various and wider positive effects on the desired phenotype.

These examples emphasize the importance of knowledge of the variation in the metabolic/transcriptomic/proteomic/genetic profile of the individuals in a population for molecular breeding. Genes underlying these trait variations can be targeted to improve the nutritional quality and/or yield of the desired crops.

1.2 Polyploidy

The history of quantitative genetics through the works of geneticists such as RA Fisher is rooted in the genetic analysis of diploid species, although many species of plants are polyploid. The word polyploidy in Greek means many fold. It refers to the presence of more than two sets of chromosomes in an organism. The moss, *Physcomitrella patens,* is haploid with only one set of chromosomes; the model angiosperm, *Arabidopsis thaliana* (hereafter referred to as *A. thaliana* or *Arabidopsis*), is diploid with two sets of chromosomes, while many crops such as the wheat we use in pasta or bread making is polyploid with more than two sets of chromosomes.

Polyploidy is known to occur in nature and has played an important role in the evolution of angiosperms (Soltis *et al*., 2009). The occurrence of polyploidy is more prevalent in plants than in animals. For example, oats, wheat, strawberries, blueberries, banana, cotton and coffee are all polyploids, suggesting the plasticity of plants. However, it is also present in lower vertebrates such as fishes and amphibians, and has in fact played an important role in vertebrate evolution (Comai, 2005).

1.2.1 Types and Formation

Polyploidisation can occur through genome doubling or by hybridisation between two genomes (Tayalé & Parisod, 2013). Somatic doubling in the meristem or the non-meristematic tissue can also lead to polyploid formation in plants (Ramsey and Schemske, 1998) as shown in **Figure 1-2**. However, the more important and sustainable model of polyploid formation is

the non-reduced gamete formation leading to the formation of 2n gametes. There are two types of polyploid plants formed through the route of unreduced gamete formation:

1) Autopolyploids, which develop by the doubling of the genome within the same species.

2) Allopolyploids, which develop by the hybridisation and subsequent doubling of the genomes between different species (Ramsey and Schemske, 1998).

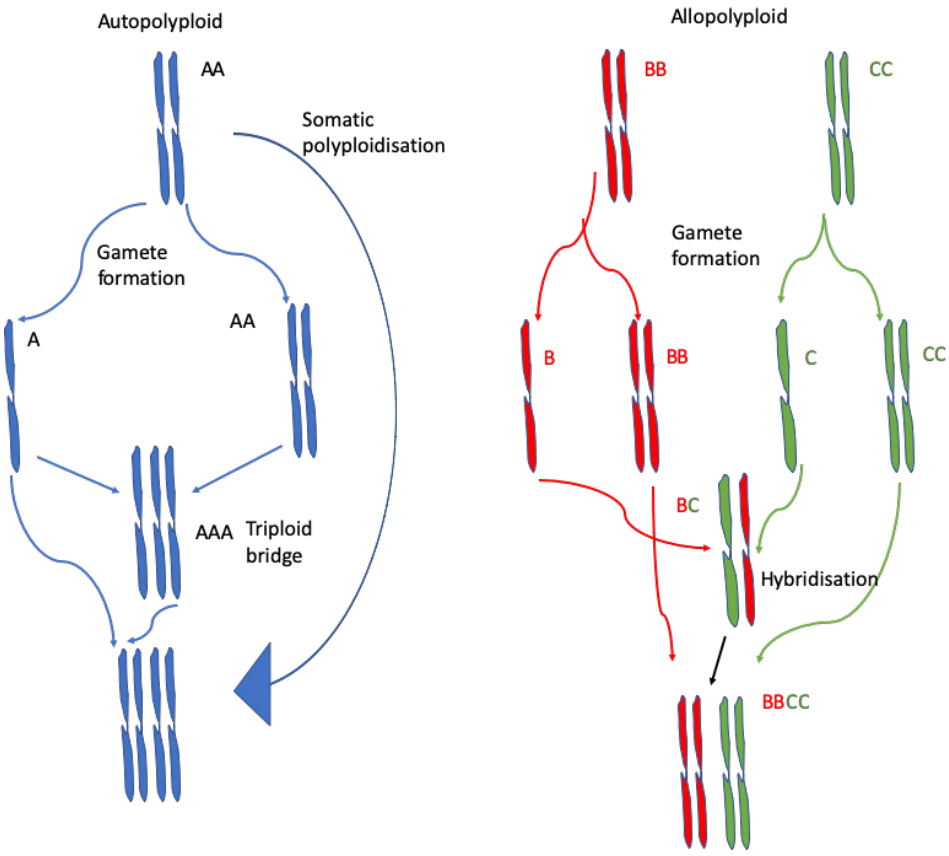


Figure 1-2 Pathways for polyploid formation.

Figure adapted from Bomblies and Madlung, (2014).

Autopolyploid formation may occur either through triploid bridge or in a one step process. Triploids may form spontaneously within the diploid population, which can form tetraploids by backcrossing with diploids, or tetraploids may form by the union of two unreduced 2n

gametes (Ramsey and Schemske, 1998). Allopolyploids may also form in a similar manner, but two different species hybridise with each other and the F1 so formed sometimes produce triploid progeny. The hybrid triploid can then backcross with a diploid or self fertilise to create a tetraploid. Secondly, an allopolyploid may directly form by the crossing of the hybrid F1 or the F2 progeny produced (Ramsey and Schemske, 1998). **Table 1-1** shows a few important polyploid plants along with their use and ploidy level (Sattler, Carvalho and Clarindo, 2016).

Name of plant	Use	Type and Ploidy level
Alfalfa	Forage crop	Autopolyploid, 4x = 42
Banana	Fruit crop	Autopolyploid, 3x = 33
Potato	Tuber crop	Autopolyploid, 4x = 48
Sweet Potato	Tuber crop	Autopolyploid, 6x = 90
Leek	Vegetable crop	Autopolyploid, 4x = 32
Yam	Vegetable crop	Autopolyploid, 3x = 60; 4x = 80
Kiwi Fruit	Fruit crop	Autopolyploid, 4x = 116, 6x = 174
Rapeseed	Oil crop	Allopolyploid, 4x = 38
Bread wheat	Grain crop	Allopolyploid, 6x = 42
Durum wheat	Grain crop	Allopolyploid, 4x = 28
Cotton	Industrial crop	Allopolyploid, 4x = 52
Coffee	Beverage crop	Allopolyploid, 4x = 44
Sugarcane	Industrial crop	Allopolyploid, 8x = 80
Peanut	Crop	Allopolyploid, 4x = 40
Oats	Grain crop	Allopolyploid, 6x = 42
Strawberry	Fruit crop	Allopolyploid, 8x = 56
Rhododendron	Ornamental	Autopolyploid, 4x = 52
Lilies	Ornamental	Autopolyploid, 3x = 36, 4x = 48

Table 1-1 Few examples of polyploid plants.

1.2.2 Advantages and Disadvantages

There are certain advantages and disadvantages of polyploidy to the plants as discussed by Comai (2005). Polyploid plants are often found to be more vigorous than their diploid

progenitors, a process called heterosis. This often results in larger plant parts such as fruits or grains, and is thus desirable in crop breeding. For example, hexaploid bread wheat *Triticum aestivum,* which formed by the hybridization of three diploid genomes, has more desirable bread making qualities and improved nutrition than the diploid relatives (Sattler, Carvalho and Clarindo, 2016). New allotetraploids, created by crossing *A. thaliana* and *Arabidopsis arenosa* autopolyploids showed higher vegetative growth, more rosette leaves, bigger seeds with much higher germination rates than in *A. arenosa* (Chen, 2010). Polyploidy also results in plants with higher fixed heterozygosity (Soltis and Soltis, 2000). This is especially true in allopolyploids, where prevention of intergenomic recombination helps in maintaining the same level of heterozygosity across generations (Comai, 2005). Another advantage is gene redundancy, which helps in masking deleterious alleles, and allows development of new functions for the duplicated gene. Polyploidy also increases the self-fertilisation ability and helps in gaining asexual reproduction, which can be useful in adverse conditions.

Polyploids have been found to be better suited to tolerate environmental stresses compared with diploids. *Arabidopsis* autotetraploids were found to be more resistant to drought and salt stress through abscisic acid (ABA) signalling and reactive oxygen species (ROS) dependent mechanisms (Del Pozo & Ramirez-Parra, 2014). Similarly, *Arabidopsis* tetraploids were found to be resistant to salinity by their ability to accumulate more potassium in their leaves. They were also found to produce more seeds than their diploid counterparts when under salt stress. Thus, it was concluded that the tetraploids have better ability to tolerate salt stress and may have reproductive advantages in saline conditions (Chao *et al*., 2013). Roots of tetraploid rice have also been found to be more resistant to salt stress than the diploid rice (Tu *et al*., 2014).

Despite several advantages, neopolyploids have to pass through a crucial stage of stabilizing the fertilisation process before they can be established. Fertility is often reduced in polyploid plants leading to a lower seed set, indicating difficulties in meiosis, a crucial process leading to the formation of gametes (Comai, 2005). Following polyploidisation, genomic rearrangements have been found to occur. While in autopolyploids they may occur over a long term, in allopolyploids they may happen in the early generations itself (Parisod *et al*., 2010). Epigenetic changes such as gene silencing mediated by DNA methylation also occurs in polyploids. Wang *et al*. (2004) found differential gene expression among the newly created synthetic allotetraploid *Arabidopsis suecica* plants and its progenitors, *A. thaliana* and *A. arenosa*. Gene silencing in different lines of newly formed tetraploids, which was maintained by DNA methylation was also found (Wang *et a*l., 2004). It was shown that ~1.3% of those 3% of genes that showed differential expression between parents and the allotetraploids, were silenced in more than one independent lines, indicating that the silencing of genes after polyploidisation was fast and largely random. Epigenetic regulation of protein coding genes has also been observed in *Arabidopsis suecica*, an allotetraploid (Lee and Chen, 2001). The silenced genes were found to be in hyper methylated regions, which could be reactivated by demethylation.

The genomic rearrangements might be necessary to maintain and establish the species of the newly created polyploid plant. For example, a rearrangement of 45S rDNA from two homologues of chromosome 4 to two homologues of chromosome 3 was found in autotetraploid Wilna ecotype of *A. thaliana* as compared to the diploid plant (Weiss and

Maluszynska, 2000). This chromosomal translocation helped in ensuring bivalent pairing amongst the chromosomes in meiosis to ensure fertility. Thus, the structural arrangements might serve as a tool for the establishment of the species and hence their evolutionary success.

1.3 Meiosis

Meiosis is a crucial two-step cell division process, occurring in the reproductive cells and leading to the formation of gametes. In the first division called Meiosis I, the homologous chromosomes segregate leading to a halving in the number of chromosomes in the cell. In the second division called Meiosis II, the sister chromatids segregate and finally four cells are formed, each containing half the original number of chromosomes (Cha and Hartsuiker, 2014). The chromosomes are replicated in the S phase and cohesion between the sister chromatids is maintained in G2 phase. This is followed by entry into meiosis, which includes a long prophase I, which can be divided into five stages-

1) Leptotene: chromosomes start condensing and double stranded breaks (DSB) initiated by endonuclease SPO11 along with the axis proteins are formed at different places on the chromosome (Kleckner, 2006).

2) Zygotene: further condensation of chromosomes occurs and formation of the synaptonemal complex (SC), a proteinaceous axis, starts between the homologues (Zickler and Kleckner, 1999). The synaptonemal complex (**Figure 1-3**) consists of two chromosome axes referred to as lateral elements connected by the central zipper proteins called transverse elements (Osman *et al*., 2011) .

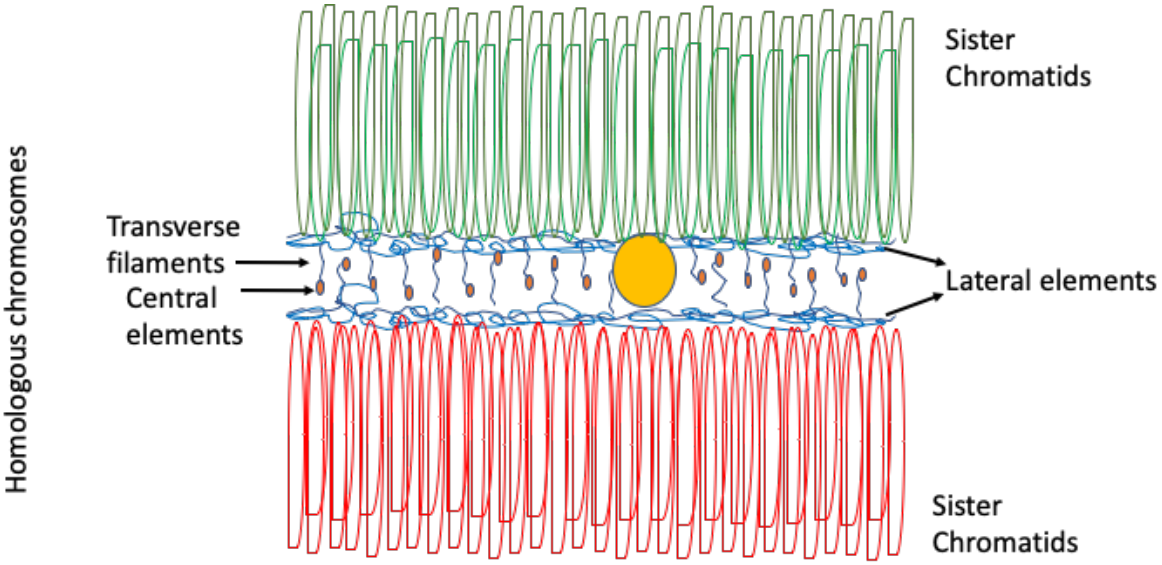


Figure 1-3 Synaptonemal complex diagrammatic representation.

Figure adapted from Osman *et al*. (2011).

3) Pachytene: the SC formation is complete between the homologues and crossing over occurs between the homologous chromosomes. This leads to the formation of chiasma where recombination occurs between the parental chromosomes.

4) Diplotene: homologues condense further and SC starts breaking down.

5) Diakinesis: homologues begin to separate except at chiasma and the nuclear envelope disintegrates.

The chromosomes still attached at chiasma align at the equatorial plate during metaphase I (M1) and appear as 5 bivalents in *A. thaliana*. Homologous chromosomes separate to opposite poles during anaphase I; followed by a brief interphase II and meiosis II that is similar to mitosis. Chromosomes again condense and align at metaphase II. During anaphase II centromeric cohesion is lost between the sister chromatids and they are separated to the opposite poles as seen in **Figure 1-4**, resulting in the formation of a tetrad.

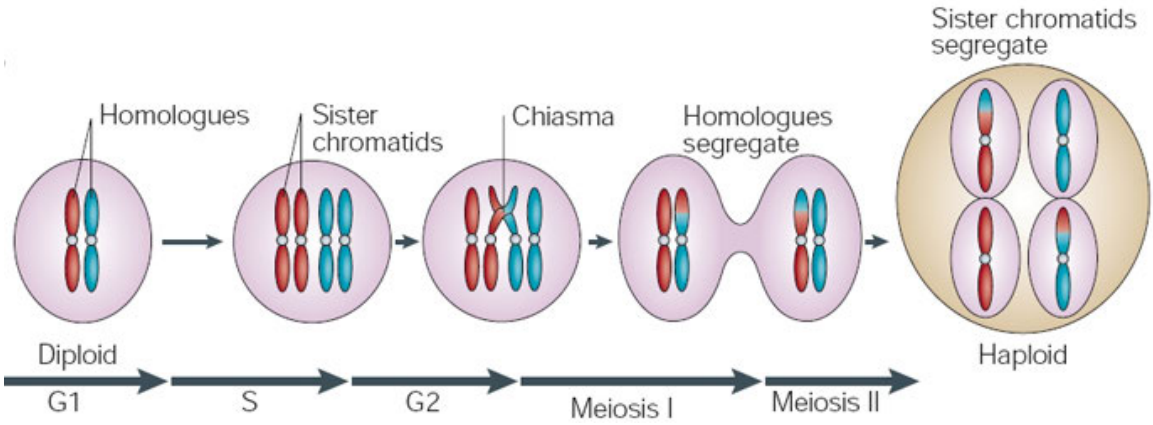


Figure 1-4 Meiosis in a diploid cell.

Figure reproduced from Marston and Amon, 2004. Obtained permission, License number 4622450558081.

1.3.1 Meiotic Recombination/Homologous Recombination

Chiasma formation or crossover resulting in the recombination of homologous chromosomes is the key process in meiosis. It not only creates new combination of alleles but is also responsible for correct chromosome segregation (Youds and Boulton, 2011). Numerous factors affect this process. It is started by the formation of double stranded breaks (DSBs) at various sites across the genome, by the conserved endonuclease SPO11 first characterised in yeast, *Saccharomyces cerevisiae* (Keeney, Giroux and Kleckner, 1997). Different proteins such as replication protein A, RAD51 and DMC1, bind 3' single strand DNA overhangs created at the DSB sites, which start to look for complementary strands on the homologous chromosome, thus creating the D loop. A double Holliday junction (dHj) can be formed after the second end capture, which can then be resolved either as a crossover or a non-crossover as seen in **Figure 1-5** (Youds and Boulton, 2011), or the D loop can be resolved by Synthesis Dependent Strand Annealing (SDSA).

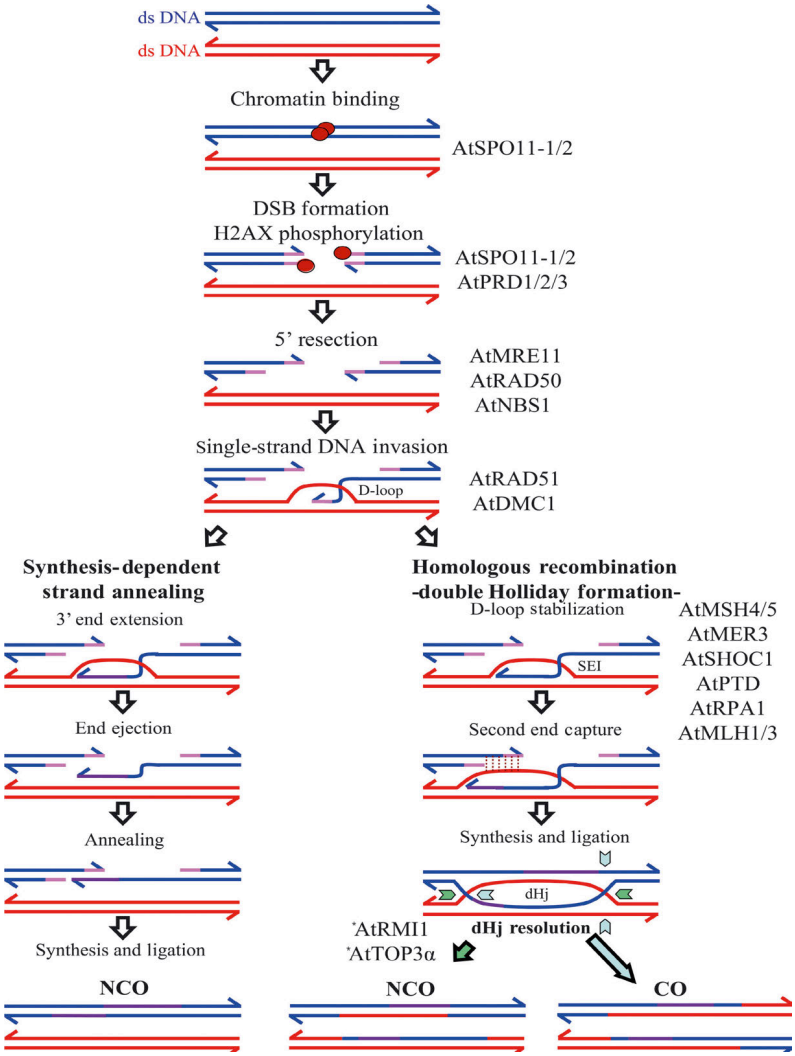


Figure 1-5 Homologous recombination during meiosis in Arabidopsis showing different outcomes.

Figure reproduced from Osman *et al.* (2011). Obtained permission, License number 4622451353169.

1.3.1.1 DSB formation

Homologous recombination between the sister chromatids start at the leptotene stage in prophase I by the formation of DSBs. SPO11, which is structurally similar to archaeal Topoisomerase VIA subunit (Keeney, 2008), catalyses the DSB formation through esterification by breaking the DNA backbone and forming a phosphodiester bond with the 5' terminal (Keeney, Giroux and Kleckner, 1997). SPO11 homologues have been identified in

various organisms. In *A. thaliana*, 3 homologues are known: AtSPO11-1, AtSPO11-2 and AtSPO11-3, but only 1 and 2 mediate DSB formation (Osman *et al.*, 2011).

SPO11 requires other accessory proteins to carry out its activities. In *S. cerevisiae*, a complement of 9 other proteins: SKI8, MER2, MEI4, REC102, REC104, REC114, MRE11, RAD50 and XRS2 are required for DSB formation (Keeney, 2008). In *A. thaliana* along with the two SPO11 proteins, other proteins including AtMTOPVIB, AtPRD1, AtPRD2, AtPRD3, AtDFO and AtSWI1 are required for meiotic DSB formation (Osman *et al.*, 2011; Zhang *et al.*, 2012; Vrielynck *et al.*, 2016). AtPRD2 is functionally similar to MEI4 in *S. cerevisiae*, AtPRD1 is similar to MEI1 in mouse, and AtPRD3 is a homologue of the rice PAIR1, and functions as an accessory protein required along with SPO11 (Osman *et al.*, 2011).

The location, formation and frequency of meiotic DSBs are tightly controlled. There are DSB hotspots in the genome regions where nucleosomes are dispersed to enable access for DSB machinery. In *S. cerevisiae*, they are formed adjacent to transcriptional start sites, while in humans and mice specific sequence motifs enable DSB formation (Gray and Cohen, 2016). Recently, cyclin dependent kinases (CDKs) have been found to regulate DSB formation in fission yeast (Bustamante-Jaramillo *et al.*, 2019).

After break formation, SPO11 remains covalently bonded to the 5’ ends of the DNA. Repair of these sites is important for chromosome integrity. A complex of MRE11-RAD50-XRS2/NBS1(MRX) along with COM1/SAE2/CTIP removes SPO11 from DNA strands in budding yeast and mammals, along with a specified length oligonucleotide from both the sides of

SPO11 (Neale, Pan and Keeney, 2005). In *A. thaliana*, AtMRE11*,* AtRAD50 and AtCOM1, the homologues of the respective proteins in *S. cerevisiae*, seem to play the same role (Osman *et al.*, 2011).

The 3' single stranded DNA (ssDNA) overhangs on both side of the breaks (see **Figure 1-5**) are then acted upon by EXO1 or SGS1-DNA2, which extend it by further resecting the 5' ends to play a role in strand exchange repair of the break sites (Mimitou and Symington, 2011). The 3' ssDNA ends attract RECA-related recombinases RAD51 and DMC1 to form nucleoprotein filaments, which invade duplex DNA to perform a homology search and strand exchange along with accessory proteins RAD54, RAD54/TID1, RAD52, RAD55 and RAD57 in yeast. In *A. thaliana* six RAD51 paralogues are present but only three, AtRAD51, AtRAD51D and AtXRCC3, along with AtDMC1, play a similar role in meiosis (Osman *et al.*, 2011).

Stable strand exchange is then promoted by the conserved MND1/HOP2 complex in *S. cerevisiae* and by AtMND1 and AtHOP2 in *A. thaliana*. The complex serves to stabilise the presynaptic filaments to promote duplex capture and synaptic complex formation (Osman *et al.*, 2011). MND1/HOP2 complex functions along with DMC1 and RAD51, and ssDNA forms a D loop by displacing the complementary DNA strand. This can then be resolved either by SDSA or by formation of double Holliday junction (dHJ), which can resolve either as a crossover or non-crossover (Gray and Cohen, 2016). SDSA involves the disruptions and reannealing of the invading strand with the other end of the DSB. DNA synthesis using the complementary strand and ligation repairs the DSB, resulting in the resolution of the D-loop as a non-crossover (**Figure 1-5**, left panel) (Gray and Cohen, 2016). IRC20 and SRS2 proteins have been found to

work in the D loop processes to enhance SDSA and prevent formation of the dHj in yeast (Miura *et al.*, 2012). Another way of DSB repair is the formation of a joint molecule by the capture of extending D-loop by the second end of the DSB resulting in a dHj, which can be resolved either as a crossover or a non-crossover (**Figure 1-5**, right panel) (Gray and Cohen, 2016).

1.3.1.2 CO formation and control

A proportion of the DSBs are repaired by crossover formation between the homologues. For example, roughly 200 - 250 DSBs are formed in *Arabidopsis,* but approximately 10 are repaired through crossover formation (Osman *et al.*, 2011; Serrentino and Borde, 2012). Crossover formation between the homologous chromosomes during meiotic prophase I not only creates variation but is also important for correct chromosome segregation. This is ultimately important for the fertility of an organism; it is therefore imperative that the process is tightly controlled ensuring that each chromosome gets at-least one crossover known as the obligate crossover (Jones and Franklin, 2006). The ability of chromosomes to get one obligate crossover is known as crossover assurance of the meiotic cells.

Crossover homeostasis and crossover interference are other important mechanisms governing CO distribution. The process of homeostasis maintains the formation of crossover at the expense of non-crossover when DSB formation is disturbed (Wang *et al.*, 2015). This process in turn functions to maintain crossover assurance. The interference mechanism is responsible for placing the crossovers further apart than they would appear by random distribution. Börner *et al.* (2004) found that crossover control does not depend on SC

formation and explained the stress model for interference and recombination. Different models elucidating interference have been described. The mechanical basis of the stress model has been explained in detail by Kleckner *et al*. (2004). According to the model, a stress is created by the contraction and expansion of the chromosomes during prophase I and hence the chromosome axis buckles under the pressure, which leads to the formation of crossover. This releases the stress along the axis and hence another crossover is less likely to form in the proximity.

Broadly two pathways of crossover formation have been described: Class I/Type I/Interference dependent pathway, which are affected by the process of interference and depend on ZIP-MSH-MER proteins and Class II/Type II/Interference independent pathway, which depend on MUS81-MMS4 proteins (Osman *et al.*, 2011).

1.3.1.2.1 Factors affecting CO frequency and distribution

Meiotic recombination, its frequency and distribution has been found to be affected by various internal and external factors. For example heterochiasmy, the sex related differences in recombination has been known in various species. Male crossovers were found to be more than the female in an *Arabidopsis* genome wide study. The difference was found to be statistically significant in sub-telomeric regions. A difference in genetic map lengths of male vs female was found to be correlated with the synaptonemal complex length in meiosis (Giraut *et al.*, 2011).

There are also a wide range of genetic factors involved. Recombination starts after the formation of DSBs, therefore distribution of DSBs have an effect on where a crossover occurs. DSB distribution, crossovers and chromatin architecture were mapped genome wide in *Arabidopsis* using SPO11-oligonucleotide. It was found that DSBs were enriched in nucleosome depleted regions in gene promoters and so were the crossovers (Choi *et al.*, 2018). These regions have previously been shown to be enriched with trimethylated lysine 4 on histone H3 (H3K4me3), which is a mark of open chromatin and is rich in crossovers (Choi *et al.*, 2013). Choi *et al.* (2018) also found AT rich regions associated with higher DSB and crossover formation in *Arabidopsis*. This is in contrast to maize, where DSB hotspots were found related with GC rich regions (He *et al.*, 2017).

Variation in meiotic recombination frequency among eight different accessions of *A. thaliana* has been observed (Sanchez-Moran *et al*., 2002). The study analysed chromosomes of all the different accessions with FISH using 45S and 5S chromosomal probes and identified the chromosome specific differences among all the accessions. Total chiasma frequency ranging from 7.90 for accessions Cape Verde (Cvi) to 9.36 for accession Feira (Fei-0) was found, with Cvi and Landsberg (Ler) having lower chiasma frequencies than the rest of the six accessions. At the chromosomal level, chromosome 4 was found to be least variable among all the accessions while chromosome 2 was the most variable. This indicated the presence of different genetic factors controlling the recombination in different varieties of the same species.

In maize, the presence of retrotransposon insertions and their organization into haplotype affects recombination frequency (He and Dooner, 2009). Related to this, heterozygosity was shown to have an effect on crossover distribution when juxtaposed to homozygous regions on the chromosome, increasing the crossover frequency in the heterozygous regions and decreasing it in the homozygous regions in *A. thaliana* F2 lines (Ziolkowski *et al.*, 2015). In another genome wide study in maize, highly variable recombination pattern was found across populations globally as well as at the chromosomal level and between different lines, with interference acting to reduce crossovers (Bauer *et al.*, 2013). Strength of interference is another important factor affecting crossover frequency and distribution. For example, it is variable across the chromosome 4 in male meiosis in *A. thaliana* F1 lines created by crossing Columbia and Landsberg. It correlated with the physical distance between the COs and was not affected by the centromere (Drouaud *et al.*, 2007). Polymorphisms in meiotic genes such as *HEI10*, may also be responsible for CO frequency variation as was found in the F2 population of *Arabidopsis* (Ziolkowski *et al.*, 2017).

Epigenetic control plays an important part in CO formation and distribution. Not only marking the open states of chromatin as described before (Choi *et al.*, 2018), methylation plays an important role in suppressing crossovers in the pericentromeric repeats which can lead to chromosomal segregation errors (Underwood *et al.*, 2018). Several transposable and repetitive elements exist in the pericentromeric regions in *A. thaliana*. They remain transcriptionally inactive by histone 3 lysine 9 dimethylation (H3K9me2) and DNA methylation in cytosine CG and non CG sequences, which also represses meiotic DSB and crossover in pericentromeric heterochromatin. It was shown that mutating H3K9me2 and non CG DNA

methylation led to an increase in crossovers in the pericentromeric regions (Underwood *et al.*, 2018). This was in contrast to the results obtained in epiRILs created by crossing mutant ddm1 lines (which have 70% reduced methylation overall) with WT *Arabidopsis*. An enhanced suppression of recombination in pericentromeric regions with increased recombination frequency in chromosome arms in the epiRILs was seen (Colome-Tatche *et al.*, 2012). In potato, open chromatin regions marked by H3K4me3 were found to be enriched with crossovers at all genic and intergenic regions. The crossovers were also found to be enriched with *Stowaway* transposons in the genic promoters (Marand *et al.*, 2017). This suggests how different epigenetic controls work in coherence with the genetic factors to maintain the recombination landscape.

Environmental factors also play an important role along with the genetic factors in shaping the recombination landscape of an organism, particularly temperature. Exposure of barley meiocytes to a temperature of 35 °C resulted in the failure of meiosis due to the failure of the formation of the synaptonemal complex. However, when the barley plants were exposed to a temperature of 30 °C, a reduction in chiasma frequency was observed along with a shift in the distribution of chiasma from distal to the proximal regions (Higgins *et al.*, 2012). An increase in recombination frequency as well as shift of Type I COs was seen in barley when grown at 25 °C in contrast to 15 °C. However, the increased recombination was found to be due to Type II COs, only in males and not in females, implying that temperature modulation affects chiasma frequency differentially between the sexes (Phillips *et al.*, 2015). This shift in chiasma formation can help recombine the genes on the chromosomal regions which are limited in recombination, thus creating new varieties of the crops. Similarly in *A. thaliana*, an elevated

temperature of 28 °C led to an increase in CO formation by increasing the number of class I crossovers (Modliszewski *et al.*, 2018). It was found that the increase in crossover numbers occurred by repairing more DSBs than usual as crossovers. A U shaped curve was found when *A. thaliana* plants were subjected to a range of temperatures ranging from 8 °C to 28 °C, with the lowest recombination frequency observed at 18 °C. The increase was caused by increasing Class I CO at both higher and lower temperatures. However, there was a negative relationship with the synaptonemal complex at high temperature, though a slight (14%) increase in SC at lower temperatures was observed which could not be explained (Lloyd *et al.*, 2018).

Several other biotic and abiotic factors such as pathogen infection, nutrition availability, UV exposure, can also affect the frequency and can shift the position of chiasmata (Lambing, Franklin and Wang, 2017) in several species. These factors can be utilized to modulate meiotic recombination accordingly to introgress traits such as disease resistance from wild crop relatives into the crop plants.

1.3.1.2.2 Types of crossovers

Class I/Interference dependent pathway

Interference has been described as a process that prevents the formation of two crossovers in an area close to each other on a chromosome. A group of proteins called ZMM (ZIP1, ZIP2, ZIP3, ZIP4, MSH4, MSH5 and MER3) were found to be required for the formation of interference sensitive crossovers in yeast (Börner, Kleckner and Hunter, 2004). They studied the recombination, SC formation and meiotic progression in wild type and mutant yeast lines and also compared them at the temperatures of 33 °C and 23 °C. It was concluded that the

crossover/non-crossover decision was made early on in the process of recombination by analyzing the ZMM mutants at high temperature. A defect in converting the DSBs to Single End Invasions (SEI) was implicated specifically in the formation of crossovers (CO). They found that in the ZMM mutants, crossovers were affected adversely while non-crossovers (NCO) formed normally at 33 °C. DSBs occurred normally but DSB repair via crossover and hence recombination was lower than in the wild type. At 23 °C the DSB formation was normal but crossovers formed at 40-60% of the wild type level, while non-crossovers formed at very high levels as compared to the WT. This implied that at the higher temperature, COs were not able to form due to defective SEI formation after DSBs have occurred, whereas at lower temperature the CO designated sites committed to form COs but matured inappropriately into non COs.

Homologues of the ZMM proteins have been identified in different eukaryotes including *A. thaliana*, where additional proteins are also required for class I CO formation (Osman *et al*., 2011). ZMM proteins MSH4 and MSH5 have been characterized in *Arabidopsis* (AtMSH4, AtMSH5) and found to be functional in early meiotic prophase I (Higgins *et al.*, 2004; Higgins *et al.*, 2008). They were found to be similar to the proteins found in yeast and humans, and co-localized on the meiotic axis from leptotene to pachytene. By counting the chiasma number during metaphase I in *Atmsh*4 and *Atmsh*5 separately, they found out that chiasma number per cell ranged from 0 to 7 with an average of 1.55 and 1.15 respectively, as compared to almost 10 per cell in the wild type. Double mutation did not yield significant differences from single mutation in terms of chiasmata formation, indicating that both proteins work in the same pathway. The reduced chiasmata were randomly distributed, pointing to the existence

of two crossover pathways in *Arabidopsis*. Thus, almost 80-90% of crossovers were found to be interference sensitive and only 10-15% of crossovers in *Arabidopsis* were interference insensitive (Higgins *et al.*, 2004; Higgins *et al.*, 2008).

MLH1, the homologue of bacterial MutL protein, is a DNA mismatch repair protein, which has been shown to be required for crossover formation, working in the Class I pathway in *S. cerevisiae* (Hunter and Borts, 1997). In *Arabidopsis*, three homologues of MutL, AtMLH1, AtMLH3 and AtPMS1 exist, where MLH1 and MLH3 function together as a heterodimer and are essential for crossing over (Jackson *et al.*, 2006). Another Class I CO protein identified in *A. thaliana* is AtMER3, very similar to the yeast MER3 protein. The number of crossovers was significantly reduced in *Atmer3* mutants as compared to the wild type and the remaining crossovers were found to be interference independent (Mercier *et al.*, 2005).

AtSHOC1, another protein in *A. thaliana* structurally similar to XPF endonucleases and orthologous with yeast ZIP2, has been identified to function in class I CO by stabilizing D-loop and single end invasion (SEI) leading to CO formation (Macaisne *et al.*, 2008). Mutation in *SHOC1* reduced the chiasma frequency approximately to 15% of the wild type (Macaisne, Vignard and Mercier, 2011). AtHEI10, an *A. thaliana* protein, structurally and functionally related to yeast ZIP3, has been characterized and shown to function indispensably in class I CO (Chelysheva *et al.*, 2012). With progression of meiosis, HEI10 sites were only retained at the sites designated to be Class I crossovers. Mutation and double mutation of any of the above mentioned proteins in *A. thaliana* reduces the crossover frequency by 80-90%,

suggesting that most of the CO are the result of Class I CO pathway (Lambing, Franklin and Wang, 2017).

Class II/Non-interfering COs pathway

It has been shown in budding yeast that a second pathway for CO formation exists, which is interference independent. This pathway depends on MUS81-MMS4 based complex proteins and accounts for 15% of all crossovers formed (de los Santos *et al.*, 2003). It was found that MMS4 acts more in the short chromosomes that have less interference. These occasionally occurring crossovers in *S. cerevisiae* serve to resolve the aberrant joint molecule structures using MUS81-MMS4, YEN1 and SLX1-SLX4, either producing crossovers or undergoing dissolution to form non crossovers (Zakharyevich *et al.*, 2012). Unusually, in the fission yeast, *S. pombe*, only Class II COs occur, which depend on MUS81-EME 1, and class I COs are unknown (Smith *et al.*, 2003).

Existence of the Class II crossover pathway has also been shown in *A. thaliana* (Higgins *et al.*, 2004). The study showed that roughly 15% of crossovers were not AtMSH4 (Class I ZMM protein) dependent. Of these 15%, at-least some of the crossovers were found to be dependent on AtMUS81, suggesting that it is not the only Class II pathway protein. However other proteins have not yet been identified (Higgins, Buckling, *et al.*, 2008).

Identification of the pathway through which each crossover has formed has been possible in wild type tomato. This task was done by superimposing the immunofluorescent light microscopy image of the synaptonemal complex spread showing MLH1 foci onto the electron

microscopy image of the same spread showing recombination nodule locations (Anderson *et al.*, 2014). Recombination nodules (RNs) are spherical protein structures associated with the synaptonemal complex during prophase I. They have been shown to be associated with the crossover sites (Stack and Anderson, 2002).

It has been suggested that the Class II pathways serve as a fall back mechanism in case CO formation fails through Class I pathway (Kohl & Sekelsky, 2013). This has happened over the course of meiotic evolution from mitosis. As shown in **Figure 1-5**, the recombination process can follow any of the three paths. The pathway to be chosen depends on yet another protein SGS1, the orthologue of Bloom syndrome helicase (BLM), in *S. cerevisiae,* which is an anti-crossover protein (De Muyt *et al*., 2012). SGS1 promoted the formation of early non-crossovers and crossovers through the Class I pathway. However, in the absence of SGS1 both CO and NCO were formed through Class II pathway (Kohl & Sekelsky, 2013), thus serving both anti-crossover as well as pro-crossover functions. A similar role has been identified for the FANCM helicase in *Arabidopsis* (Kohl & Sekelsky, 2013).

Propensity of the DSBs to be repaired as crossovers or non-crossovers also depend on the activities of the pro crossover as well as anti-crossover proteins. Three groups of genes promoting non crossover at the expense of crossover have been identified in *Arabidopsis*, which include FANCM, RTR complex and FIGL1, which function independently of each other (Lambing, Franklin and Wang, 2017). First is FANCM helicase and its two cofactors MHF1 and MHF2, which promote NCOs through the SDSA pathway (Crismani *et al.*, 2012). In *Arabidopsis fancm* mutants, an increased CO frequency comparable with the wild type was found, that

required MUS81 pathway, indicating FANCM helicase exercises a negative control over Class II crossovers (Girard *et al.*, 2014). The second pathway involves Topoisomerase 3α (AtTOP3α) and two BLM/SGS1 homologues RECQ4A and RECQ4B, which again affect the Class II crossover pathway to limit the crossover formation (Séguéla-Arnaud *et al.*, 2015). Third pathway is the AAA-Atpase FIDGETIN-LIKE-1 (FIGL1) anti CO protein, which along with its interacting partner FIDGETIN-LIKE-1 INTERACTING PROTEIN (FLIP) in a complex, works by counteracting DMC1/RAD51 mediated inter-homologue strand invasion and thus controlling CO formation (Girard *et al.*, 2015; Fernandes *et al.*, 2018). Disruption of all three pathways in *Arabidopsis* led to an increase in meiotic recombination, however it was similar to the effect observed in *recq4* and *figl1* double mutant in Col/Ler hybrid F1, which was 7.8 fold increase in crossovers as compared with the wild type (Fernandes *et al.*, 2017).

Knowledge about the type of crossovers and the different factors affecting the frequency and distribution of crossovers can help to manipulate genetic recombination as a useful tool in crop breeding ensuring food security. Managing recombination can help introduce elite traits into new varieties or prevent the elite traits to be lost from a variety and thus, informed breeding can be performed accordingly (Lambing, Franklin and Wang, 2017). For example, increasing *HEI10* dosage (affecting Class I crossovers) along with *RECQ4a* and *RECQ4b* mutations (affecting Class II crossovers) in *Arabidopsis* F2 lines led to a massive increase in the number of crossovers in mutant F2 lines compared with the wild type (Serra *et al.*, 2018). In crop plants rice, pea and tomato, RECQ4 was identified to be the main anti-crossover protein, mutation of which led to increased crossovers in the three crops by almost three fold (Mieulet *et al.*, 2018).

1.4 Meiosis in polyploids

A finer balance and control of the meiotic homologous chromosome pairing is required in polyploids compared with diploids. Since there are more than two sets of chromosomes, there is a higher chance of multivalent formation and mis-segregation of chromosomes. This is one of the major challenges, which polyploids face to ensure fertility and to establish and maintain their species.

In autopolyploids, where more than two homologous chromosomes are present in the cell, chiasma formation can occur between more than two pairing homologues at different positions. This can lead to the formation of various chromosomal conformations that include multivalent and univalent formation at Metaphase I along with the bivalent associations. For example, an autotetraploid (4x) having 4 sets of the same genome may resolve as 2 bivalents as shown in **Figure 1-6**, one quadrivalent, or one trivalent and one univalent, which may result in the formation of aneuploid gametes.

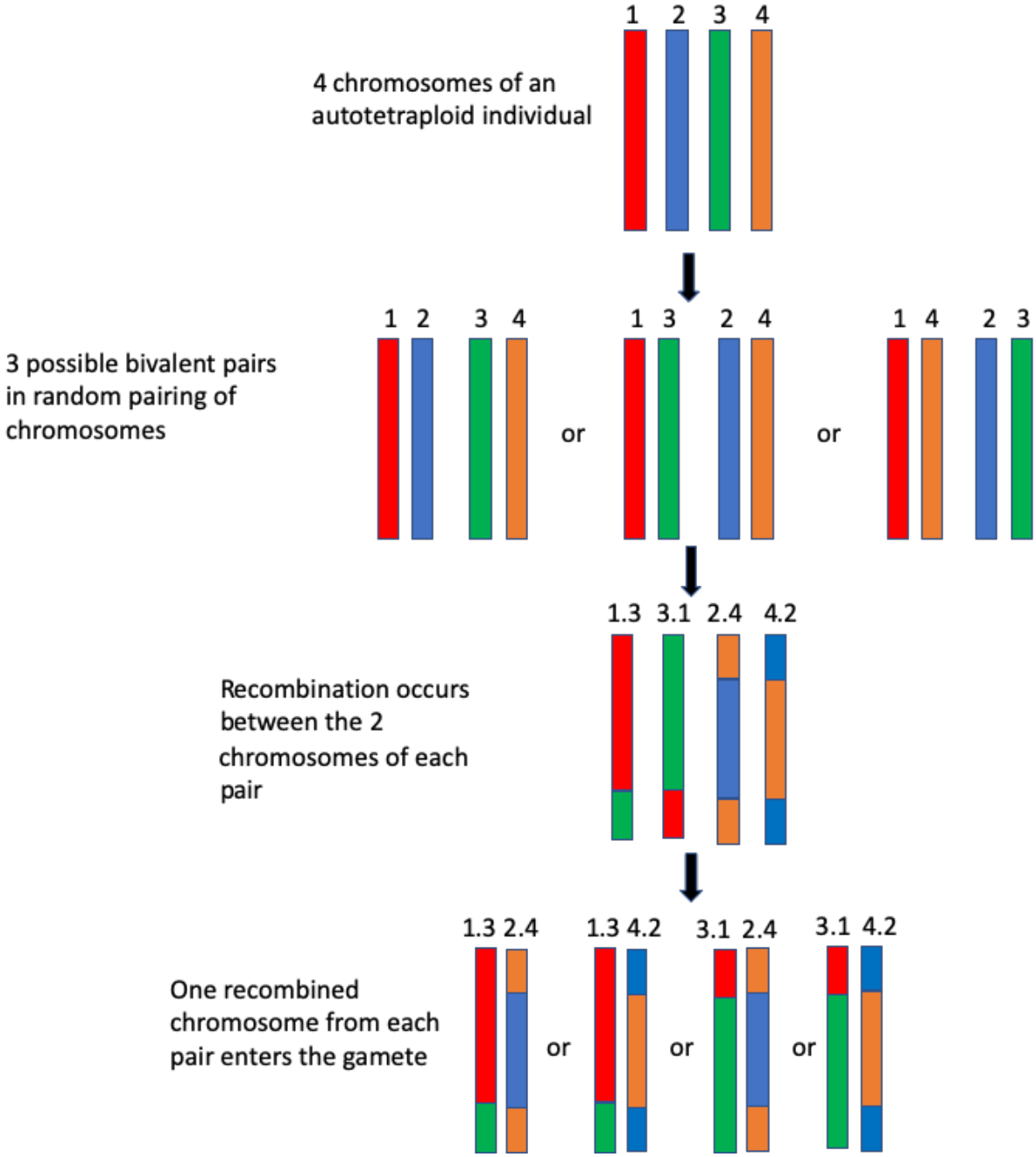


Figure 1-6 Chromosome segregation during bivalent meiosis of an autotetraploid species.

Figure adapted from Leach, 2008.

Santos *et al*. (2003) showed that multivalents were formed frequently in all 5 (4 established and 1 newly created) lines of the autotetraploid *A. thaliana* studied. The formation of multivalents suggested the presence of more than one Autonomous Pairing Site (APS) among

the homologues. It has been suggested that pairing between the homologues starts simultaneously from opposite ends, which can be a reason for multivalent formation in autopolyploids as one homologue can synapse with more than one homologous partner at different sites, called pairing partner switches (Zielinski and Mittelsten Scheid, 2012). However, in established tetraploids, metaphase I chromosomes occurred predominantly as bivalents, suggesting the partial diploidization of meiosis. Meiosis evolves in response to whole genome duplication (Bomblies *et al*., 2015). Naturally occurring and established autopolyploids such as *Arabidopsis arenosa* have evolved by reducing crossover rates genome wide to stabilize meiosis and maintain fertility. Though the telomeres and NORs in *A. arenosa* show quadrivalent association during early Prophase I, the centromeric regions show bivalent association throughout Prophase I. This could possibly be the reason for the bivalent pairing of metaphasic chromosomes (Carvalho et al., 2010).

Several interacting meiotic genes have been adapted to tetraploidy in *A. arenosa* (Yant *et al*., 2013). The group identified at least seven interacting meiotic genes showing ploidy specific selection. SNPs were identified between naturally occurring diploids and tetraploids of *A. arenosa,* suggesting their evolution in response to whole genome duplication. These included genes encoding chromosome axis components such as *ASY1, ASY3, ZIP1a* and others. For example, a SNP in *ASY1,* changing a single amino acid at a conserved site in the protein has been found to be highly prevalent in tetraploid *A. arenosa* compared with the diploid plants (Hollister *et al.*, 2012) indicating its role in successful polyploidisation. One possible reason attributed for reduction in crossing over is increased crossover interference, which resulted in only one CO per homologue, thus preventing multivalent formation. This increased

interference can be achieved by the decreased chromosome length, which was found to be the case in *A. arenosa*, alfalfa and male silk moths (Bomblies *et al.*, 2016). This, along with the *A. arenosa* genome wide scan indicating high degree of selection for axis and axis related proteins modulating interference, indicates the important role of the chromosome axis in interference modulation in autotetraploid *A. arenosa* (Yant and Bomblies, 2015; Bomblies *et al.*, 2016). It has been shown that diploids with lower crossover rates have more chances of forming a stable and fertile tetraploid (Lloyd and Bomblies, 2016) and can establish their species without facing the fertility bottleneck. This knowledge can be utilized for polyploid plant breeding.

Reduction in CO frequency in autotetraploids is one of the possible adaptations towards stability, however some natural autotetraploids, for example, *Parnassi palustris,* show multivalents with no significant reduction in fertility compared to diploids (Wentworth and Gornall, 1996). Similarly, the prevalence of quadrivalents was found to be associated with high fertility in tetraploid rye compared with bivalents (Hazarika and Rees, 1967). In autotetraploid *Lolium perenne*, high fertility was found to be associated with high quadrivalent frequency and distally placed chiasmata in quadrivalents, which resulted in regular segregation (Crowley and Rees, 1968). This indicates the genetic role, for example a change in chromatin state, which might have shifted the chiasma distally, leading to a regular chromosomal disjunction which helped in the maintenance of the fertility. Thus, it seems that the selection works to ensure the best survival of an organism whether by quadrivalent formation or by reducing the number of crossovers, to ensure fertility to maintain and propagate its species.

One of the important and unique features of an autopolyploid meiosis is the process of double reduction. Double reduction occurs when the identical alleles carried on the sister chromatids of a chromosome do not segregate and end up in the same gamete as shown in **Figure 1-7**. This happens when an autotetraploid forms a quadrivalent during meiosis. The four colours represent the four homologues in an autotetraploid individual. Two loci are shown, A and B, with A being close to centromere and B at a distance. Therefore, B can undergo a crossover with the centromere and can segregate via path Y with no double reduction or path Z with double reduction. Locus A will undergo path X with no double reduction (Wu *et al.*, 2001). The gametes resulting from double reduction are shown by small blue arrows.

Double reduction in tetraploids can lead to segregation distortion and needs to be taken into account in segregation analysis. The upper limit of the coefficient of double reduction has been identified as ¼ and the upper limit of recombination frequency among the homologues has been identified as ¾ in an autotetraploid compared to ½ in a diploid (Luo *et al.*, 2006). As can be seen in **Figure 1-7**, 12 out of 20 gametes are recombinants and 8 out of 20 gametes are non-recombinants which works out the probability to be ¾ (1/2 *1 + 1/2 * 8/16) for maximal recombination frequency and ¼ (1/2 * 8/16) for maximum double reduction in autotetraploids in an event of quadrivalent meiosis (Luo *et al.*, 2006).

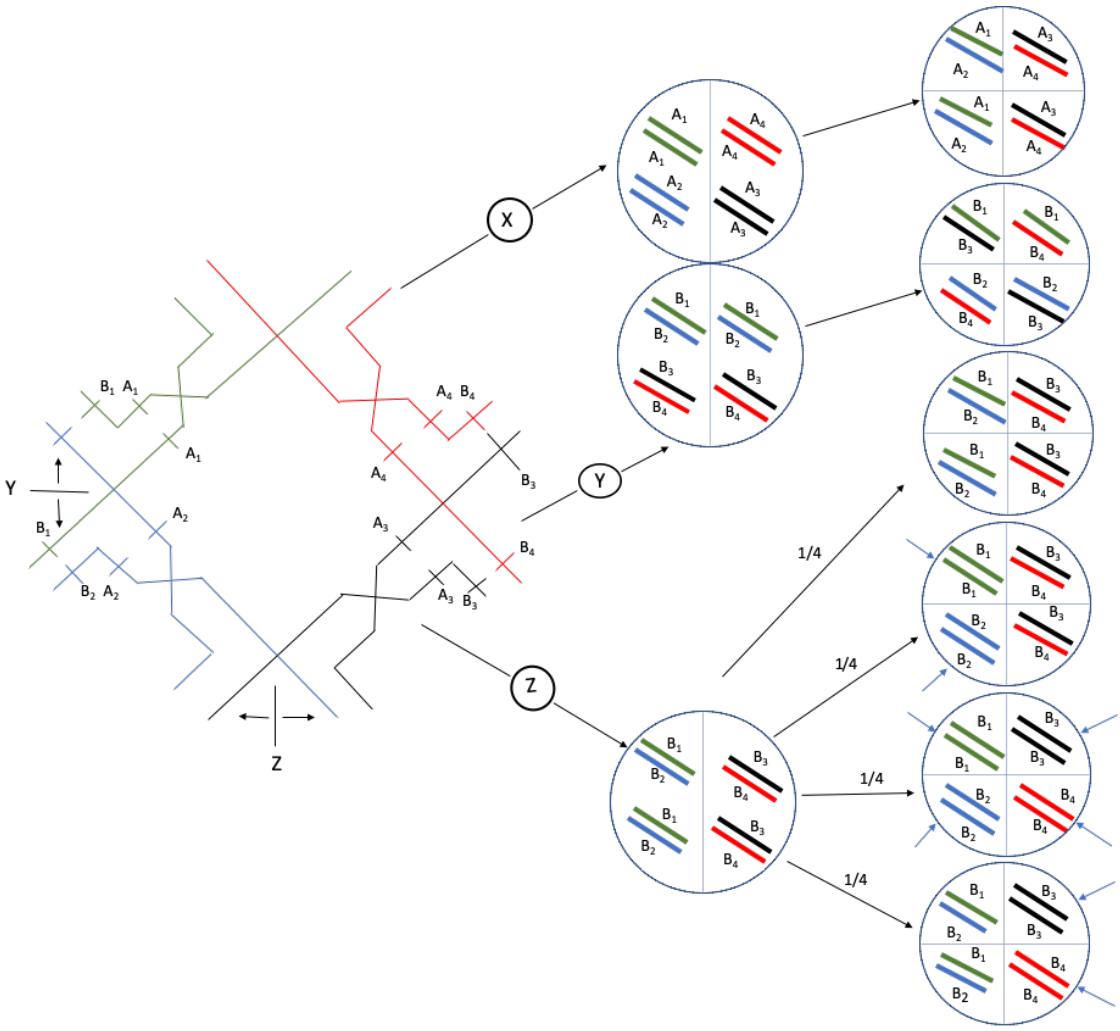


Figure 1-7 Segregation patterns of loci A and B in an autotetraploid meiosis.

Figure adapted from Wu *et al.* (2001). Red, blue, black and green colours represent the four homologues of the chromosome in an autotetraploid. X indicates the meiotic path with no crossover so that sister chromatids migrate to the same pole in anaphase I and get separated in anaphase II. Y indicates the meiotic path with crossover between locus and centromere with sister chromatids migrating to different poles and again separating into different gametes in anaphase II. Z indicates the meiotic path where sister chromatids migrate to same pole and end up in same gametes after anaphase II leading to double reduction. Blue arrows indicate gametes with sister chromatids.

A cytological comparison between the meiosis in a diploid and autotetraploid line of *Arabidopsis thaliana* can be seen in **Figure 1-8**.

Allopolyploids on the other hand, have to control the ectopic crossing over that can occur between the nearly homologous parts of the homoeologous (from two different parental

species) chromosomal sets. For example, homoeologous pairing is avoided in allohexaploid (6x) wheat by the Pairing homoeologous 1 (Ph1) region which ensures that pairing occurs only between the homologous sets (Riley, 1974). Ph1 was mapped to the chromosome 5B interstitial region consisting of *CDC2*-related genes with interspersed heterochromatin from 3B inserted into the cluster (Griffiths *et al.*, 2006). The locus Ph1 has evolved with polyploidisation and prevents recombination between homoeologues by preventing the MLH1 sites to mature into crossovers between the homoeologues (Martin *et al.*, 2014). Recently it was shown that Ph1 promoted the early homologous synapsis during and before the telomere bouquet stage, thus avoiding the homoeologous pairing. It was also shown that the level of MLH1 progression into COs between homoeologues could be manipulated using environmental factors such as nutrient composition and temperature (Martin *et al.*, 2017).

Brassica napus, also has a locus PrBn responsible for controlling the degree of homoeologous pairing (Jenczewski *et al.*, 2003). In *Arabidopsis suecica*, a similar locus called BYS controlling homoeologous recombination has been found (Hollister, 2015). It has been seen that alignment between homologous as well as homoeologous chromosomes occur in the early stages of prophase I for some chromosomes, but synapsis formation is not completed between the homoeologues and ectopic crossing over is thus prevented (Lloyd and Bomblies, 2016). In this way, allopolyploids have developed mechanisms to ensure their fertility and hence the establishment of their species.

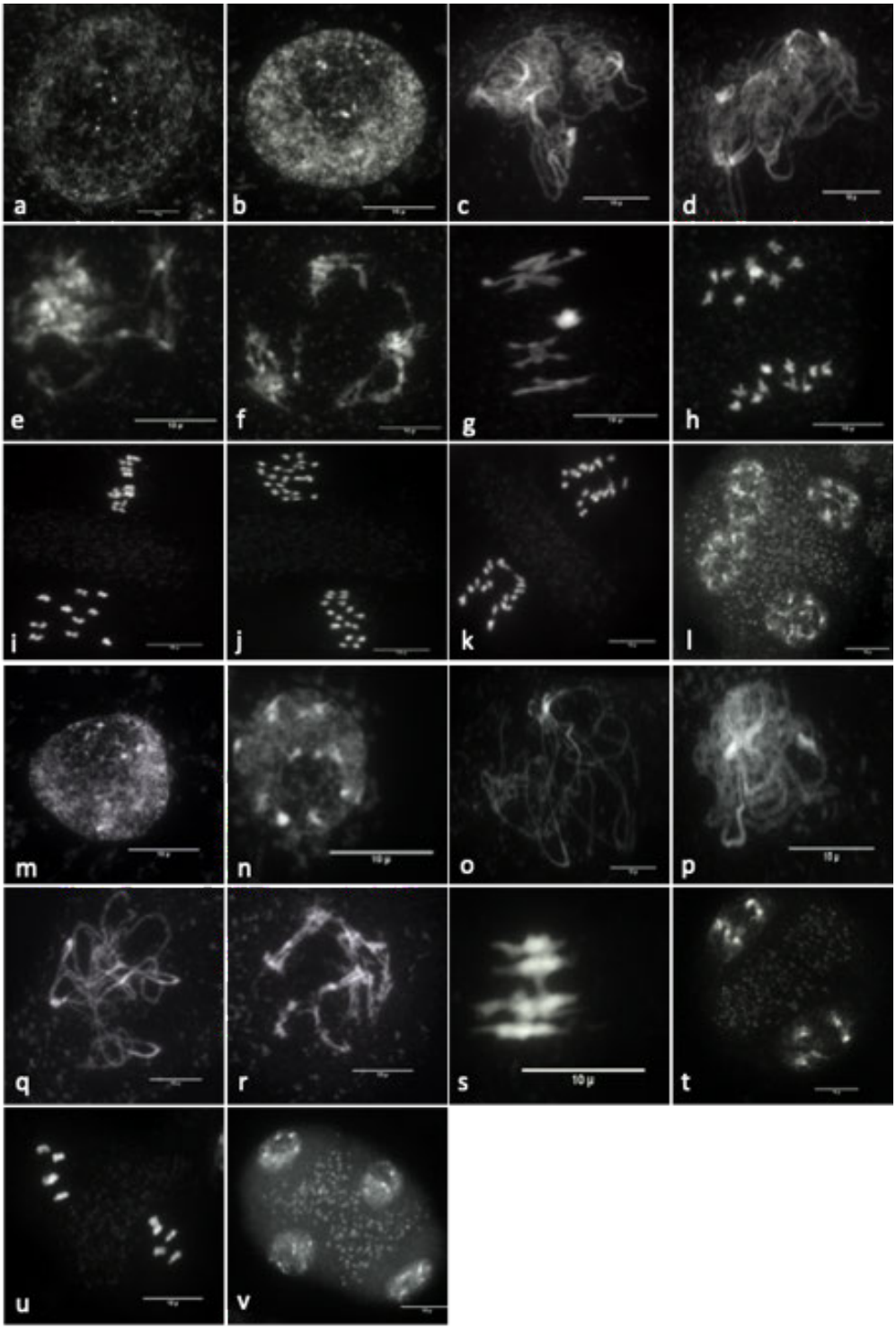


Figure 1-8 A Meiotic atlas of tetraploid and diploid Arabidopsis thaliana variety Landsberg.

a-l Tetraploid meiosis. m-v Diploid meiosis (a,m) Interphase, (b,n) Leptotene, (c,o) Zygotene, (d,p) Pachytene, (e,q) Diplotene, (f,r) Diakinesis, (g,s) Metaphase I (h) Anaphase I, (i,u) Metaphase II (t) Dyad stage (j) Anaphase II start, (k) Anaphase II, (l,v) Tetrad stage. Scale bar is 10 μm

1.5 Model Plants

1.5.1 Arabidopsis thaliana

Arabidopsis thaliana, common thale cress, was first discovered by Johannes Thal in the 16th century (Wixon, 2001). Since then it has become the plant of choice for genetic and molecular studies. It is an angiosperm and belongs to Brassicaceae family. It is one of the widely used model organisms and has several features which makes it easy to use:

- Easy to grow and maintain.
- Small genome size of 125 Mb.
- Genome completely sequenced.
- Excellent forward and reverse genetic resources.
- Self-fertilizing and good seed set.
- Established polyploid lines are available.

Columbia (Col) and Landsberg (Ler) are two of the most widely used accessions of *Arabidopsis thaliana.*

Brief History

Dr. Friedrich Laibach was the earliest proponent of *Arabidopsis* for plant genetics and established the number of chromosome pairs in meiotic cells as n = 5 during his PhD (Somerville and Koornneef, 2002). He liked collecting different strains of the plant. Dr. George Redei, obtained four different strains of these plants from Laibach's personal collection. He irradiated a few samples from one of the strains collected from Landsberg ander Warthe, Poland, which created a mutation in the erecta gene and is now known as Ler 0. He found that

the original non-irradiated Landsberg samples obtained were not isogenic and one of the plants showing a different phenotype was selected and named Columbia, Col 0 (lehleseeds.com).

Various morphological differences occur between Columbia (Col) and Landsberg (Ler) *Arabidopsis thaliana* (Passardi *et al.*, 2007) including:

1) Imbibed seeds of Col are bigger and rounder than Ler.

2) Main root is comparatively bigger in Col than Ler.

3) Col plants usually have a bigger rosette with more number of leaves at flowering. They usually flower later than Ler.

4) Ler leaves are rounded with small petioles, while Col has elongated leaves.

5) Ler plants are shorter than Col and have an erect morphology.

5) The siliques in Col are longer and tapered while in Ler they are short and blunt.

Various genome level differences and similarities between the two accessions have also been found. Differences in 5S rDNA between both lines have been found cytologically with the 5S rDNA cluster present in the small arm of chromosome 3 near the centromeric region in Col lines, while in Ler lines it occurs interstitially in the long arms (Sanchez-Moran *et al.*, 2002). The first large scale comparative analysis of Columbia and Landsberg identified 25,274 SNPs and 14,570 InDels both within the coding and non-coding regions of the genome (The Arabidopsis Genome Initiative, 2000). Another whole genome comparison between the two lines also identified 6636 insertions or deletions (Ziolkowski *et al.*, 2009). They estimated that the two lines diverged 200,000 years ago.

Zapata *et al*. (2016), carried out chromosome level assembly of Ler with respect to the Col reference genome. This study also found the Ler-specific rDNA locus in the arm of chromosome 3. They also found similar telomeric sequence repeats and pericentromeric regions in 8 arms of the chromosomes (except the NOR bearing arms of 2 and 4 chromosomes). Several inversions and chromosomal transpositions in the pericentromeric regions, and inversion in the chromosomal arms were found in the whole genome alignment of the Ler and Col assemblies (Zapata *et al.*, 2016). Inversion resulted in the deletion of several genes in Ler. The study found indels and highly divergent regions (HDRs) and suppression of COs in the rearranged regions. The inversion of 1.2 Mb on chromosome 4 in Ler resulted in reduced recombination in its short arm. Another 170 kb inversion on chromosome 3 was found to be specific to Ler lines. 40 and 63 unique genes specific for Ler and Col respectively were found to have evolved by gene deletions in one of the two genomes. Additional copies of 330 gene pairs were also found in either the Ler or Col genome, which evolved by duplication (Zapata *et al.*, 2016). Thus, a rich resource of natural genetic variation exists between the two lines, which can be utilized for various genetic studies.

1.5.2 Solanum tuberosum

The common potato that we often consume and relish, belongs to family Solanaceae, which includes many other plants such as tomato, pepper, tobacco, brinjal and petunia. The modified underground stem of the plant is called a tuber, which we consume as potato. The classification of species in genus *Solanum* has been a complex issue due to various genetic, morphological and geographical factors. Section petota of family Solanaceae includes the entire tuber bearing wild and cultivated potatoes. The most recent classification of species of cultivated potato includes four species: *S. tuberosum, S. ajanhuiri, S. juzepczukii and S. curtilobum*. *S. tuberosum* can be further divided into two subspecies: *Solanum tuberosum tuberosum*, which includes all the European and American cultivated varieties that originated from Chilean landraces, and *Solanum tuberosum andigenum*, which includes cultivated varieties in Central and South America (Ovchinnikova *et al.*, 2011; Spooner *et al.*, 2014). The modern cultivars of *S. tubero*sum, which are based on Chilean germplasm, have had crosses from wild species and are therefore hybrid in nature (Ovchinnikova *et al.*, 2011)

The basic chromosome number in the potato family is 12, though a ploidy series is shown by different species, both in wild as well as cultivated potatoes. Almost 70% of species exist as diploids with 2n = 24, but triploids (2n= 36), tetraploids (2n = 48), pentaploids (2n =60) and hexaploids (2n =72) also occur (Spooner *et al.*, 2014). The cultivated potato is mainly tetraploid with 2n = 48, though there are cultivated diploid species as well. For example, *S. ajanhuiri* is a cultivated diploid species formed by hybridisation between diploid cultivar *S. stenotomum* of the Andigenum group and wild diploid species *S. boliviense* (Rodríguez *et al.*, 2010).

Origin of potato

Domestication of the wild varieties of potatoes can be traced back to more than 8000 years ago, when the people in the Andes and Peruvian region used potato as food. There are two major and conflicting views about the origin of potatoes. A single origin of cultivated *S. tuberosum* from Peruvian *S. brevicaule* complex, which is a group of 20 morphologically similar wild species, has been suggested based on AFLP data analysis of various wild and landrace members of Solanum section petota (Spooner *et al.*, 2005). The proponents of this view proposed that Chilean potatoes originated from the hybridisation of wild species and tetraploid Andean varieties. These hybridised Andean varieties adapted into long day Chilean varieties, once they reached Chile from the Andes. Another view proposes independent origin of the Andean and Chilean landraces from different diploid and tetraploid wild species in South American regions of Peru and Chile (reviewed in Spooner *et al.*, 2014). Hybrid origin between closely related species has also been suggested for *S. tuberosum* using DNA sequencing data of waxy gene (Rodríguez *et al.*, 2010). The question is still not resolved completely.

Potato in Europe

It is believed that Spanish settlers from Latin America first introduced potatoes in Europe in the 16th century to the Canary Islands. Again, there are two competing views about the origins of cultivated potato in Europe. One view considered the origination of European potato from lowland Chile. A second view considered them to be of Andean origin, and that Chilean varieties were introduced only when late blight epidemic affected the Andean varieties in 1845-1892. It was a largely accepted view, until it was shown that the Chilean variety was

introduced in 1811 in Europe long before the late blight epidemic of 1845. This was confirmed by sequencing the plastid DNA from 64 herbarium species collected between 1600 and 1910. A 241 bp deletion has been known to be present in plastid DNA in Chilean varieties, which corresponded with the herbarium species. This showed that all the modern cultivars originated from Chilean landraces (Ríos *et al.*, 2007; Ames and Spooner, 2008), though both Andean and Chilean landraces were present in the Canary Islands.

Genetic variation

Cultivated potatoes are considered to be highly diverse species as characterised using different molecular markers. Introgression of various traits from wild relatives has helped in increasing the diversity of the different varieties. 107 wild species of diploid, tetraploid and hexaploid potatoes are known to be distributed from Mexico to Equador (Spooner *et al.*, 2014). 67 samples representing wild species, landraces and different cultivated varieties were sequenced, where the sequence diversity between the diploids and tetraploids was found to be highest among any resequenced crops. Candidate genes responsible for domestication of potatoes were found to be under selection with various wild type alleles introgressed in the cultivated species (Hardigan *et al.*, 2017). Copy number variations (CNV) affecting growth and development of the plants have been found in different cultivated varieties of potatoes (Iovene *et al.*, 2013). Carputo *et al.* (2013), found out differences in response to Potato Virus Y between different species as well as between the varieties in the same species indicating a genetic variation responsible for the resistance (Carputo *et al.*, 2013). In a separate study, the modern cultivated potato cultivars were found to be genetically related to the Chiloe Island landraces, preserving the genetic diversity present in those landraces. Thus, the Chiloean

landraces germplasm can be utilized for genetic improvement and breeding in the modern cultivars (Esnault *et al.*, 2014). The wild species have been found to be present in a wide geographical range. For example, polyploid wild species were found in very wet areas from Colombia to Costa Rica, where diploids were absent, in addition to other areas. Diploids were present in a wide variety of other areas ranging from South western USA to Argentina, Chile and Uruguay (Hijmans *et al.*, 2007). These species are therefore adapted to a wide variety of environments and are therefore highly diverse. This diversity can be utilized for breeding new and resistant varieties of cultivated potatoes.

1.6 Aims

A crucial goal of plant breeding is to unlock or release the existing genetic variation in plant genomes through meiotic recombination, thus facilitating the creation of new crop varieties with enhanced agricultural traits such as yield, or disease resistance. It is known that chromosome pairing and recombination are subject to highly stringent control, which results in at least one crossover per chromosome pair and a non-uniform distribution of crossover events along chromosome arms. However, the direct causal factors affecting the crossover rate variation remain poorly understood. This research project aims to analyse the effects of polyploidisation on the structure and function of the model plant *Arabidopsis thaliana*, with main focus on meiosis and meiotic recombination.

This project will test the effects of polyploidy on meiotic recombination and understand its potential to manipulate and redistribute meiotic recombination in our major diploid and polyploid crops, using *Arabidopsis* as an exemplar. We aim to provide a genome-wide characterization of meiotic recombination frequency in both diploid and autotetraploid genomes of *Arabidopsis thaliana*. This will clarify whether polyploidisation can change the frequency and/ or distribution of meiotic recombination on either a genome-wide or a local scale, and will therefore advance our understanding of the factors controlling meiotic recombination in plants.

A further aim is to transfer the skills gained in a model plant to study meiotic recombination directly in a real-world crop plant. Potato is an important autopolyploid food-crop, however

detailed work on cytological analysis of its meiotic prophase is lacking in the literature. Methodologies and knowledge base are mostly developed using the model plant organisms, with the intention to transfer the knowledge to complex crop species and learn more about the ways of enhancing the crop species. A better understanding about the meiotic processes of the cultivated potato can help in breeding programmes. The aims of this part of the project is to develop cytological methods that can be applied reliably and to analyse if the methodologies used for preparation of chromosomal spreads in *Arabidopsis* can be transferred to potato for characterising meiotic recombination frequency for individual chromosomes in cultivated diploid and tetraploid potato varieties. This will give an idea about the frequency and/or distribution of the meiotic recombination that can be utilised for introgression of desirable traits in the cultivated tetraploid potato plants.

1.7 References

Abiola, O. *et al.* (2003) 'The nature and identification of quantitative trait loci: A community's view', *Nature Reviews Genetics*, 4(11), pp. 911-916. doi: 10.1038/nrg1206.

Ames, M. and Spooner, D. M. (2008) 'DNA from herbarium specimens settles a controversy about origins of the European potato', *American Journal of Botany*, 95(2), pp. 252-257. doi: 10.3732/ajb.95.2.252.

Anderson, L. K. *et al.* (2014) 'Combined fluorescent and electron microscopic imaging unveils the specific properties of two classes of meiotic crossovers', *Proceedings of the National Academy of Sciences*, 111(37), pp. 13415–13420. doi: 10.1073/pnas.1406846111.

Atherly, A. G., Girton, J. R. and McDonald, J. F. (1999) *The Science of Genetics*. Saunders College Publishing. Available at: https://books.google.co.uk/books?id=2skmVdmu2RMC.

Bauer, E. *et al.* (2013) 'Intraspecific variation of recombination rate in maize', *Genome Biology*, 14(9), p. R103. doi: 10.1186/gb-2013-14-9-r103.

Bomblies, K. *et al.* (2016) 'The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role?', *Chromosoma*, 125(2), pp. 287–300. doi: 10.1007/s00412-015-0571-4.

Bomblies, K., Higgins, J. D. and Yant, L. (2015) 'Meiosis evolves: adaptation to external and internal environments', *New Phytologist*, 208(2), pp. 306–323. doi: 10.1111/nph.13499.

Bomblies, K. and Madlung, A. (2014) 'Polyploidy in the Arabidopsis genus', *Chromosome Research*, 22(2), pp. 117-134. doi: 10.1007/s10577-014-9416-x.

Börner, G. V., Kleckner, N. and Hunter, N. (2004) 'Crossover/Noncrossover Differentiation, Synaptonemal Complex Formation, and Regulatory Surveillance at the Leptotene/Zygotene Transition of Meiosis', *Cell*, 117(1), pp. 29–45. doi: 10.1016/S0092-8674(04)00292-2.

Bradshaw, J. E. *et al.* (2008) 'QTL mapping of yield, agronomic and quality traits in tetraploid potato (Solanum tuberosum subsp. tuberosum)', *Theoretical and Applied Genetics*, 116(2), pp. 193-211. doi: 10.1007/s00122-007-0659-1.

Bustamante-Jaramillo, L. F. *et al.* (2019) 'CDK contribution to DSB formation and recombination in fission yeast meiosis', *PLoS Genetics*, 15(1) p. e1007876. doi: 10.1371/journal.pgen.1007876.

Carputo, D. *et al.* (2013) 'Genetic diversity among potato species as revealed by phenotypic resistances and SSR markers', *Plant Genetic Resources: Characterisation and Utilisation*, 11(2), pp. 131-139. doi: 10.1017/S1479262112000500.

Carvalho, A. *et al.* (2010) 'Chromosome and DNA methylation dynamics during meiosis in the autotetraploid Arabidopsis arenosa', *Sexual Plant Reproduction*, 23(1), pp. 29–37. doi: 10.1007/s00497-009-0115-2.

Cha, R. S. and Hartsuiker, E. (2014) 'Meiosis', in *eLS*. American Cancer Society. doi: 10.1002/9780470015902.a0001359.pub3.

Chao, D.-Y. *et al.* (2013) 'Polyploids Exhibit Higher Potassium Uptake and Salinity Tolerance in Arabidopsis', *Science*, 341(6146), pp. 658–659. doi: 10.1126/science.1240561.

Chelysheva, L. *et al.* (2012) 'The Arabidopsis HEI10 is a new ZMM protein related to Zip3', *PLoS Genetics*, 8(7). doi: 10.1371/journal.pgen.1002799.

Chen, Z. J. (2010) 'Molecular mechanisms of polyploidy and hybrid vigor', *Trends in Plant Science*, pp. 57–71. doi: 10.1016/j.tplants.2009.12.003.

Choi, K. *et al.* (2013) 'Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters', *Nature Publishing Group*, 45(11). doi: 10.1038/ng.2766.

Choi, K. *et al.* (2018) 'Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis thaliana transposons and gene regulatory regions', *Genome Research*, 28(4), pp. 532-546. doi: 10.1101/gr.225599.117.

Colome-Tatche, M. *et al.* (2012) 'Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation', *Proceedings of the National Academy of Sciences*, 109(40), pp. 16240-16245. doi: 10.1073/pnas.1212955109.

Comai, L. (2005) 'The advantages and disadvantages of being polyploid.', *Nature reviews. Genetics*, 6(11), pp. 836–46. doi: 10.1038/nrg1711.

Crismani, W. *et al.* (2012) 'FANCM Limits Meiotic Crossovers', *Science*, 336(6088), pp. 1588–1590. doi: 10.1126/science.1220381.

Crowley, J. G. and Rees, H. (1968) 'Fertility and selection in tetraploid Lolium', *Chromosoma*, 24(3), pp. 300–308. doi: 10.1007/BF00336197.

Darwin, C. (1859) *On the Origin of the Species*, *Darwin*. doi: 10.1016/S0262-4079(09)60380-8.

Davey, J. *et al.* (2011) 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', *Nature Reviews Genetics*, 12(7), pp. 499–510. doi: 10.1038/nrg3012.

De Muyt, A. *et al.* (2012) 'BLM Helicase Ortholog Sgs1 Is a Central Regulator of Meiotic Recombination Intermediate Metabolism', *Molecular Cell*. Elsevier Inc., 46(1), pp. 43–53. doi: 10.1016/j.molcel.2012.02.020.

Drouaud, J. *et al.* (2007) 'Sex-specific crossover distributions and variations in interference level along Arabidopsis thaliana chromosome 4', *PLoS Genetics*, 3(6), p. e106. doi: 10.1371/journal.pgen.0030106.

Druka, A. *et al.* (2010) 'Expression quantitative trait loci analysis in plants', *Plant Biotechnology Journal*, 8(1), pp. 10–27. doi: 10.1111/j.1467-7652.2009.00460.x.

Esnault, F. *et al.* (2014) 'Genetic diversity analysis of a potato (Solanum tuberosum L.) collection including Chiloé Island landraces and a large panel of worldwide cultivars', *Plant Genetic Resources: Characterisation and Utilisation*, 12(1), pp. 74-82. doi: 10.1017/S1479262113000300.

Falconer, D. S. and Mackay, T. F. C. (1996) *Introduction to Quantitative Genetics (4th Edition)*, *Trends in Genetics*. Available at: http://www.amazon.com/Introduction-Quantitative-Genetics-Douglas-Falconer/dp/0582243025.

Fernandes, J. B. *et al.* (2017) 'Unleashing meiotic crossovers in hybrid plants', *Proceedings of the National Academy of Sciences* 115(10), pp. 2431-2436. doi: 10.1073/pnas.1713078114.

Fernandes, J. B. *et al.* (2018) 'FIGL1 and its novel partner FLIP form a conserved complex that regulates homologous recombination', *PLoS genetics*, 14(4). doi: 10.1371/journal.pgen.1007317.

Fu, J. *et al.* (2009) 'System-wide molecular evidence for phenotypic buffering in Arabidopsis', *Nature Genetics*, 41(2), pp. 166–167. doi: 10.1038/ng.308.

Girard, C. *et al.* (2014) 'FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers', *Nucleic Acids Research*, 42(14), pp. 9087–9095. doi: 10.1093/nar/gku614.

Girard, C. *et al.* (2015) 'AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms', *PLOS Genetics*. Edited by M. Lichten, 11(7), p. e1005369. doi: 10.1371/journal.pgen.1005369.

Giraut, L. *et al.* (2011) 'Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes', *PLoS Genetics*, 7(11). doi: 10.1371/journal.pgen.1002354.

Gray, S. and Cohen, P. E. (2016) 'Control of Meiotic Crossovers: From Double-Stand Break Formation to Designation', *Annual Review of Genetics*, 50(1), p. annurev-genet-120215-035111. doi: 10.1146/annurev-genet-120215-035111.

Griffiths, S. *et al.* (2006) 'Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat', *Nature*, 439(7077), pp. 749–752. Available at: http://dx.doi.org/10.1038/nature04434.

Gupta, P. K. *et al.* (1999) 'Molecular markers and their applications in wheat breeding', *Plant Breeding*, pp. 369–390. doi: 10.1046/j.1439-0523.1999.00401.x.

Hardigan, M. A. *et al.* (2017) 'Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato', *Proceedings of the National Academy of Sciences*, 114(46), pp. E9999-E10008. doi: 10.1073/pnas.1714380114.

Hazarika, M. H. and Rees, H. (1967) 'Genotypic control of chromosome behaviour in rye X. Chromosome pairing and fertility in autotetraploids', *Heredity*. The Genetical Society of Great Britain, 22(3), pp. 317–332. Available at: http://dx.doi.org/10.1038/hdy.1967.44.

He, L. and Dooner, H. K. (2009) 'Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for Helitron and retrotransposon insertions', *Proceedings of the National Academy of Sciences*, 106(21), pp. 8410-8416. doi: 10.1073/pnas.0902972106.

He, Y. *et al.* (2017) 'Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize', *Proceedings of the National Academy of Sciences*, 120(1-3), pp. 206-209. doi: 10.1016/0304-8853(93)91323-Y.

Higgins, J. D. *et al.* (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(20), pp. 2557–2570. doi: 10.1101/gad.317504.

Higgins, J. D., Vignard, J., *et al.* (2008) 'AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in Arabidopsis thaliana, but is not required for synapsis', *Plant Journal*, 55(1), pp. 28–39. doi: 10.1111/j.1365-313X.2008.03470.x.

Higgins, J. D., Buckling, E. F., *et al.* (2008) 'Expression and functional analysis of AtMUS81 in Arabidopsis meiosis reveals a role in the second pathway of crossing-over', *Plant Journal*, 54(1), pp. 152–162. doi: 10.1111/j.1365-313X.2008.03403.x.

Higgins, J. D. *et al.* (2012) 'Spatiotemporal Asymmetry of the Meiotic Program Underlies the Predominantly Distal Distribution of Meiotic Crossovers in Barley', *The Plant Cell*, 24(10), pp. 4096–4109. doi: 10.1105/tpc.112.102483.

Hijmans, R. J. *et al.* (2007) 'Geographical and environmental range expansion through polyploidy in wild potatoes (Solanum section Petota)', *Global Ecology and Biogeography*, 16(4), pp. 485-495 doi: 10.1111/j.1466-8238.2007.00308.x.

Hill, W. G. (2010) 'Understanding and using quantitative genetic variation.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1537), pp. 73–85. doi: 10.1098/rstb.2009.0203.

Hollister, J. D. *et al.* (2012) 'Genetic Adaptation Associated with Genome-Doubling in

Autotetraploid Arabidopsis arenosa', *PLoS Genetics*, 8(12). doi: 10.1371/journal.pgen.1003093.

Hollister, J. D. (2015) 'Polyploidy: Adaptation to the genomic environment', *New Phytologist*, 205(3), pp. 1034-1039. doi: 10.1111/nph.12939.

Hunter, N. and Borts, R. H. (1997) 'Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis', *Genes and Development*, 11(12), pp. 1573-1582. doi: 10.1101/gad.11.12.1573.

Iovene, M. *et al.* (2013) 'Copy number variation in potato - An asexually propagated autotetraploid species', *Plant Journal*, 75(1), pp. 80-89. doi: 10.1111/tpj.12200.

Jackson, N. *et al.* (2006) 'Reduced meiotic crossovers and delayed prophase I progression in AtMLH3-deficient Arabidopsis', *EMBO Journal*, 25(6), pp. 1315–1323. doi: 10.1038/sj.emboj.7600992.

Jansen, R. C. and Nap, J. P. (2001) 'Genetical genomics: The added value from segregation', *Trends in Genetics*, 17(7), pp. 388–391. doi: 10.1016/S0168-9525(01)02310-1.

Jenczewski, E. *et al.* (2003) 'PrBn, a major gene controlling homeologous pairing in oilseed rape (Brassica napus) haploids', *Genetics*, 164(2), pp. 645–653.

Jones, G. H. and Franklin, F. C. H. (2006) 'Meiotic Crossing-over: Obligation and Interference', *Cell*, 126(2), pp. 246–248. doi: 10.1016/j.cell.2006.07.010.

Kearsey, M. J. *et al.* (1996) 'Introduction', in *The Genetical Analysis of Quantitative Traits*. doi: 10.1007/978-1-4899-4441-2_1.

Keeney, S. (2008) 'Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis', in *Recombination and Meiosis*. Springer Berlin Heidelberg, pp. 81–123. doi: 10.1007/7050_2007_026.

Keeney, S., Giroux, C. N. and Kleckner, N. (1997) 'Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family', *Cell*, 88(3), pp. 375–384. doi: 10.1016/S0092-8674(00)81876-0.

Kleckner, N. *et al.* (2004) 'A mechanical basis for chromosome function.', *Proceedings of the National Academy of Sciences of the United States of America*, 101(34), pp. 12592–12597. doi: 10.1073/pnas.0402724101.

Kleckner, N. (2006) 'Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex', *Chromosoma*, 115(3), pp. 175–194. doi: 10.1007/s00412-006-0055-7.

Kloosterman, B. *et al.* (2013) 'Naturally occurring allele diversity allows potato cultivation in northern latitudes', *Nature*, 495(7440), pp. 246-250. doi: 10.1038/nature11912.

Kohl, K. P. and Sekelsky, J. (2013) 'Meiotic and Mitotic Recombination in Meiosis', *Genetics*, 194(2), pp. 327–334. doi: 10.1534/genetics.113.150581.

Kuczyńska, A., Mikołajczak, K. and Ćwiek, H. (2014) 'Pleiotropic effects of the sdw1 locus in barley populations representing different rounds of recombination', *Electronic Journal of Biotechnology*, 17(5), pp. 217–223. doi: 10.1016/j.ejbt.2014.07.005.

Lambing, C., Franklin, F. C. H. and Wang, C.-J. R. (2017) 'Understanding and manipulating meiotic recombination in plants', *Plant Physiology*, 173(3) pp. 1530-1542. doi: 10.1104/pp.16.01530.

Leach, L. J. (2008) 'Genetical genomics of yeast duplicate genes and autotetrasomic linkage analysis / by Lindsey Jane Leach.' Birmingham: University of Birmingham.

Lee, H. S. and Chen, Z. J. (2001) 'Protein-coding genes are epigenetically regulated in Arabidopsis polyploids.', *Proceedings of the National Academy of Sciences of the United States of America*, 98(12), pp. 6753–6758. doi: 10.1073/pnas.121064698.

Lloyd, A. *et al.* (2018) 'Plasticity of meiotic recombination rates in response to temperature in arabidopsis', *Genetics*, 208(4), pp. 1409-1420. doi: 10.1534/genetics.117.300588.

Lloyd, A. and Bomblies, K. (2016) 'Meiosis in autopolyploid and allopolyploid Arabidopsis', *Current Opinion in Plant Biology*. doi: 10.1016/j.pbi.2016.02.004.

de los Santos, T. *et al.* (2003) 'The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast.', *Genetics*, 164(1), pp. 81–94. Available at: http://www.genetics.org/content/164/1/81.abstract.

Luo, Z. W. *et al.* (2006) 'Constructing genetic linkage maps under a tetrasomic model', *Genetics*, 172(4), pp. 2635-2645. doi: 10.1534/genetics.105.052449.

Macaisne, N. *et al.* (2008) 'SHOC1, an XPF Endonuclease-Related Protein, Is Essential for the Formation of Class I Meiotic Crossovers', *Current Biology*, 18(18), pp. 1432–1437. doi: 10.1016/j.cub.2008.08.041.

Macaisne, N., Vignard, J. and Mercier, R. (2011) 'SHOC1 and PTD form an XPF-ERCC1-like complex that is required for formation of class I crossovers', *Journal of Cell Science*, 124(16), pp. 2687–2691. doi: 10.1242/jcs.088229.

Mackay, T. F. C., Stone, E. a and Ayroles, J. F. (2009) 'The genetics of quantitative traits: challenges and prospects.', *Nature reviews. Genetics*, 10(8), pp. 565–577. doi:

10.1038/nrg2612.

Marand, A. P. *et al.* (2017) 'Meiotic crossovers are associated with open chromatin and enriched with Stowaway transposons in potato', *Genome Biology*, 18(1), p. 203. doi: 10.1186/s13059-017-1326-8.

Marston, A. L. and Amon, A. (2004) 'Meiosis: cell-cycle controls shuffle and deal.', *Nature reviews. Molecular cell biology*, 5(12), pp. 983–97. doi: 10.1038/nrm1526.

Martín, A. C. *et al.* (2014) 'Licensing MLH1 sites for crossover during meiosis.', *Nature communications*, 5, p. 4580. doi: 10.1038/ncomms5580.

Martín, A. C. *et al.* (2017) 'Dual effect of the wheat Ph1 locus on chromosome synapsis and crossover.', *Chromosoma*, 126(6), pp. 669–680. doi: 10.1007/s00412-017-0630-0.

Mercier, R. *et al.* (2005) 'Two meiotic crossover classes cohabit in Arabidopsis: One is dependent on MER3, whereas the other one is not', *Current Biology*, 15(8), pp. 692–701. doi: 10.1016/j.cub.2005.02.056.

Mieulet, D. *et al.* (2018) 'Unleashing meiotic crossovers in crops', *Nature Plants*, 4(12), pp. 1010-1016. doi: 10.1038/s41477-018-0311-x.

Mimitou, E. P. and Symington, L. S. (2011) 'DNA end resection-Unraveling the tail', *DNA Repair*, 10(3), pp. 344–348. doi: 10.1016/j.dnarep.2010.12.004.

Miura, T. *et al.* (2012) 'Homologous recombination via synthesis-dependent strand annealing in yeast requires the IRC20 and SRS2 DNA helicases', *Genetics*, 191(1), pp. 65–78. doi: 10.1534/genetics.112.139105.

Modliszewski, J. L. *et al.* (2018) 'Elevated temperature increases meiotic crossover frequency via the interfering (Type I) pathway in Arabidopsis thaliana', *PLoS Genetics*, 14(5). doi: 10.1371/journal.pgen.1007384.

Neale, M. J., Pan, J. and Keeney, S. (2005) 'Endonucleolytic processing of covalent protein-linked DNA double-strand breaks.', *Nature*, 436(7053), pp. 1053–7. doi: 10.1038/nature03872.

Ookawa, T. *et al.* (2010) 'New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield', *Nature Communications*. Nature Publishing Group, 1(8), p. 132. doi: 10.1038/ncomms1132.

Osman, K. *et al.* (2011) 'Pathways to meiotic recombination in Arabidopsis thaliana', *New Phytologist*, 190(3), pp. 523–544. doi: 10.1111/j.1469-8137.2011.03665.x.

Ovchinnikova, A. *et al.* (2011) 'Taxonomy of cultivated potatoes (Solanum section Petota:

Solanaceae)', *Botanical Journal of the Linnean Society*, 165(2), pp. 107-155. doi: 10.1111/j.1095-8339.2010.01107.x.

Parisod, C., Holderegger, R. and Brochmann, C. (2010) 'Evolutionary consequences of autopolyploidy', *New Phytologist*, 186(1), pp. 5–17. doi: 10.1111/j.1469-8137.2009.03142.x.

Passardi, F. *et al.* (2007) 'Morphological and physiological traits of three major Arabidopsis thaliana accessions', *Journal of Plant Physiology*, 164(8), pp. 980–992. doi: 10.1016/j.jplph.2006.06.008.

Phillips, D. *et al.* (2015) 'The effect of temperature on the male and female recombination landscape of barley', *New Phytologist*, 208(2), pp. 421-429. doi: 10.1111/nph.13548.

del Pozo, J. C. and Ramirez-Parra, E. (2014) 'Deciphering the molecular bases for drought tolerance in A rabidopsis autotetraploids', *Plant, Cell & Environment*, 37(12), pp. 2722–2737. doi: 10.1111/pce.12344.

Ramsey, J. and Schemske, D. W. (1998) 'Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants', *Annual Review of Ecology and Systematics*, 29(1), pp. 467–501. doi: 10.1146/annurev.ecolsys.29.1.467.

Riley, R. (1974) 'Cytogenetics of chromosome pairing in wheat.', *Genetics*, 78(1), pp. 193–203. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17248662.

Ríos, D. *et al.* (2007) 'What is the origin of the European potato? Evidence from Canary Island landraces', *Crop Science*, 47(3), p 1271. doi: 10.2135/cropsci2006.05.0336.

Rodríguez, F. *et al.* (2010) 'Hybrid origins of cultivated potatoes', *Theoretical and Applied Genetics*, 121(6), pp 1187-1198. doi: 10.1007/s00122-010-1422-6.

Sanchez-Moran, E. *et al.* (2002) 'Variation in chiasma frequency among eight accessions of Arabidopsis thaliana', *Genetics*, 162(3), pp. 1415–1422.

Santos, J. L. *et al.* (2003) 'Partial Diploidization of Meiosis in Autotetraploid', *Genetics*, 165(3), pp. 1533–1540.

Sattler, M. C., Carvalho, C. R. and Clarindo, W. R. (2016) 'The polyploidy and its key role in plant breeding', *Planta*, pp. 281–296. doi: 10.1007/s00425-015-2450-x.

Schönhals, E. M. *et al.* (2017) 'Physical mapping of QTL for tuber yield, starch content and starch yield in tetraploid potato (Solanum tuberosum L.) by means of genome wide genotyping by sequencing and the 8.3 K SolCAP SNP array', *BMC Genomics*, 18(1), p 642. doi: 10.1186/s12864-017-3979-9.

Séguéla-Arnaud, M. *et al.* (2015) 'Multiple mechanisms limit meiotic crossovers: TOP3α and

two BLM homologs antagonize crossovers in parallel to FANCM', *Proceedings of the National Academy of Sciences*, 112(15), pp. 4713–4718. doi: 10.1073/pnas.1423107112.

Serra, H. *et al.* (2018) 'Massive crossover elevation via combination of HEI10 and recq4a recq4b during Arabidopsis meiosis.', *Proceedings of the National Academy of Sciences of the United States of America*, 115(10), pp. 2437-2442. doi: 10.1073/pnas.1713071115.

Serrentino, M. E. and Borde, V. (2012) 'The spatial regulation of meiotic recombination hotspots: Are all DSB hotspots crossover hotspots?', *Experimental Cell Research*. doi: 10.1016/j.yexcr.2012.03.025.

Smith, G. R. *et al.* (2003) 'Fission Yeast Mus81·Eme1 Holliday Junction Resolvase Is Required for Meiotic Crossing over but Not for Gene Conversion', *Genetics*, 165(4), pp. 2289–2293.

Soltis, D. E. *et al.* (2009) 'Polyploidy and angiosperm diversification.', *American journal of botany*, 96(1), pp. 336–48. doi: 10.3732/ajb.0800079.

Soltis, P. S. and Soltis, D. E. (2000) 'The role of genetic and genomic attributes in the success of polyploids.', *Proceedings of the National Academy of Sciences of the United States of America*, 97(13), pp. 7051–7057.

Somerville, C. and Koornneef, M. (2002) 'A fortunate choice: the history of Arabidopsis as a model plant.', *Nature reviews. Genetics*, 3(11), pp. 883–889. doi: 10.1038/nrg927.

Spooner, D. M. *et al.* (2005) 'A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping', *Proceedings of the National Academy of Sciences*, 102(41), pp 14694-14699. doi: 10.1073/pnas.0507400102.

Spooner, D. M. *et al.* (2014) 'Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes', *Botanical Review*, 80(4), pp. 283-383. doi: 10.1007/s12229-014-9146-y.

Stack, S. M. and Anderson, L. K. (2002) 'Crossing over as assessed by late recombination nodules is related to the pattern of synapsis and the distribution of early recombination nodules in maize', *Chromosome Research*, 10(4), pp. 329–345. doi: 10.1023/A:1016575925934.

Tayalé, a and Parisod, C. (2013) 'Natural pathways to polyploidy in plants and consequences for genome reorganization.', *Cytogenetic and genome research*, 140(2–4), pp. 79–96. doi: 10.1159/000351318.

The Arabidopsis Genome Initiative (2000) 'Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.', *Nature*, 408(6814), pp. 796–815. doi: 10.1038/35048692.

Tu, Y. *et al.* (2014) 'Genome duplication improves rice root resistance to salt stress.', *Rice (New York, N.Y.)*, 7(1), p. 15. doi: 10.1186/s12284-014-0015-4.

Underwood, C. J. *et al.* (2018) 'Epigenetic activation of meiotic recombination near Arabidopsis thaliana centromeres via loss of H3K9me2 and non-CG DNA methylation', *Genome Research*, 28(4), pp. 519-531. doi: 10.1101/gr.227116.117.

Visker, M. H. P. W. *et al.* (2003) 'Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type?', *Theoretical and Applied Genetics*, 106(2), pp. 317-325. doi: 10.1007/s00122-002-1021-2.

Visscher, P. M. and Bruce Walsh, J. (2017) 'Commentary: Fisher 1918: the foundation of the genetics and analysis of complex traits', *International Journal of Epidemiology*, 48(1), pp. 10-12. doi: 10.1093/ije/dyx129.

Visscher, P. M., Hill, W. G. and Wray, N. R. (2008) 'Heritability in the genomics era--concepts and misconceptions.', *Nature reviews. Genetics*, 9(4), pp. 255–266. doi: 10.1038/nrg2322.

Vrielynck, N. *et al.* (2016) 'A DNA topoisomerase VI-like complex initiates meiotic recombination.', *Science (New York, N.Y.)*, 351(6276), pp. 939–43. doi: 10.1126/science.aad5196.

Wang, J. *et al.* (2004) 'Stochastic and Epigenetic Changes of Gene Expression in Arabidopsis Polyploids', *Genetics*, 167(4), pp. 1961–1973. doi: 10.1534/genetics.104.027896.

Wang, J. *et al.* (2014) 'An expression quantitative trait loci-guided co-expression analysis for constructing regulatory network using a rice recombinant inbred line population', *Journal of Experimental Botany*, 65(4), pp. 1069–1079. doi: 10.1093/jxb/ert464.

Wang, S. *et al.* (2015) 'Meiotic crossover patterns: Obligatory crossover, interference and homeostasis in a single process', *Cell Cycle*, 14(3), pp. 305–314. doi: 10.4161/15384101.2014.991185.

Weiss, H. and Maluszynska, J. (2000) 'Chromosomal rearrangement in autotetraploid plants of Arabidopsis thaliana.', *Hereditas*, 133(3), pp. 255–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11433970.

Wen, W. *et al.* (2015) 'Genetic Determinants of the Network of Primary Metabolism and Their Relationships to Plant Performance in a Maize Recombinant Inbred Line Population', *The Plant Cell*, 27(7), pp. 1839–1856. doi: 10.1105/tpc.15.00208.

Wen, W. *et al.* (2018) 'An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize', *Plant Journal*, 93(6), pp. 1116-1128. doi: 10.1111/tpj.13835.

Wentworth, J. and Gornall, R. (1996) 'Cytogenetic evidence for autopolyploidy in *Parnassia palustris*', *New Phytologist*, 134, pp. 641–648. doi: 10.1111/j.1469-8137.1996.tb04929.x.

Witzel, K. *et al.* (2010) 'Mapping of quantitative trait loci associated with protein expression variation in barley grains', *Molecular Breeding*, 27(3), pp. 301–314. doi: 10.1007/s11032-010-9432-2.

Wixon, J. (2001) 'Featured organism: Arabidopsis thaliana.', *Comparative and functional genomics*, 2(2), pp. 91–98. doi: 10.1002/cfg.75.

Wu, S. S. *et al.* (2001) 'A multivalent pairing model of linkage analysis in autotetraploids', *Genetics*, 159(3), pp. 1339–1350.

Yan, W. H. *et al.* (2011) 'A major QTL, Ghd8, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice', *Molecular Plant*, 4(2), pp. 319–330. doi: 10.1093/mp/ssq070.

Yant, L. *et al.* (2013) 'Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa', *Current Biology*. Elsevier Ltd, 23(21), pp. 2151–2156. doi: 10.1016/j.cub.2013.08.059.

Yant, L. and Bomblies, K. (2015) 'Genome management and mismanagement - cell-level opportunities and challenges of whole-genome duplication', *Genes and Development*, 29(23), pp. 2405-2419. doi: 10.1101/gad.271072.115.

Youds, J. L. and Boulton, S. J. (2011) 'The choice in meiosis - defining the factors that influence crossover or non-crossover formation.', *Journal of cell science*, 124(Pt 4), pp. 501–13. doi: 10.1242/jcs.074427.

Zakharyevich, K. *et al.* (2012) 'Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase', *Cell*, 149(2), pp. 334–347. doi: 10.1016/j.cell.2012.03.023.

Zapata, L. *et al.* (2016) 'Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms', *Proceedings of the National Academy of Sciences*, 113(28), pp. E4052–E4060. doi: 10.1073/pnas.1607532113.

Zhang, C. *et al.* (2012) 'The Arabidopsis thaliana DSB formation (AtDFO) gene is required for meiotic double-strand break formation', *Plant Journal*, 72(2), pp. 271–281. doi: 10.1111/j.1365-313X.2012.05075.x.

Zickler, D. and Kleckner, N. (1999) 'Meiotic chromosomes: integrating structure and function.', *Annual review of genetics*, 33, pp. 603–754. doi: 10.1146/annurev.genet.33.1.603.

Zielinski, M.-L. and Mittelsten Scheid, O. (2012) 'Meiosis in Polyploid Plants', in *Polyploidy and Genome Evolution*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 33–55. doi: 10.1007/978-3-642-31442-1_3.

Ziolkowski, P. A. *et al.* (2009) 'Genome sequence comparison of Col and Ler lines reveals the dynamic nature of Arabidopsis chromosomes', *Nucleic Acids Research*, 37(10), pp. 3189–3201. doi: 10.1093/nar/gkp183.

Ziolkowski, P. A. *et al.* (2015) 'Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis', *eLife*. doi: 10.7554/eLife.03708.

Ziolkowski, P. A. *et al.* (2017) 'Natural variation and dosage of the HEI10 meiotic E3 ligase control Arabidopsis crossover recombination', *Genes and Development*, 31(3), pp. 306-317. doi: 10.1101/gad.295501.116.

CHAPTER 2

MATERIALS AND METHODS

2 Materials and Methods

2.1 Plant material

2.1.1 Arabidopsis

The seeds for two genotypes of *Arabidopsis thaliana,* Columbia and Landsberg, were sourced from NASC. They include diploid and tetraploid Landsberg stock ABRC; CS20, and stock ABRC; CS3900 respectively, diploid Columbia stock ABRC; CS3176 and tetraploid Columbia stock ABRC; CS3151. The seeds were surface sterilized using 50% ethanol and 0.5% Triton 100 followed by washes in sterile distilled water and stratified at 4 ˚C for two nights before being sown. The seeds were then grown in the glass house facility on 3rd floor in the School of Biosciences, in small pots in soil composed of 4 parts Compost 2 parts Vermiculite and 1 part Silvaperl or in the growth room. The temperature was maintained at 20-22 °C with 16 hours of light. For the second plant trial described in **2.3**, the seeds were sown in 10 cm pots in the growth room with 12 hours of day and night maintained. The temperature was maintained between 20-24 ˚C.

2.1.2 Potato

Four autotetraploid and two diploid varieties of *Solanum tuberosum*, have been used. Sante, Sarpo Mira, Cara and Maris Peer are the autotetraploid varieties from the association panel of 350 varieties as described in Sharma *et al.* (2018). The diploid varieties used were Mayan-gold and Scapa (http://varieties.ahdb.org.uk/varieties). The potatoes were allowed to sprout by placing them in egg boxes in a light place with the rose end upwards – the process of

chitting. After the shoots were 3 cm long, the sprouted potatoes were planted in an 18 litre quadgrow planter, with one to two tubers in each pot, using humax compost. The potatoes were covered with the compost and earthed up regularly once they started to produce shoots leaving only 2 inches of stem visible above the ground. The plants were grown in the glass-house and growth room in the School of Biosciences. The temperature was maintained between 20–24˚C.

2.2 *Arabidopsis* and potato crosses

The diploid and tetraploid Columbia and Landsberg lines of *Arabidopsis thaliana* were crossed with each other. Both the times, Columbia line was used as the maternal parent and Landsberg line as the male parent. The F1 seeds were collected from the crossed flowers and were grown to collect F2 seeds.

Potato crossing was tried using varieties Sante and Sarpo Mira. The pollen was collected from the donor plant by gentle vibration behind the flowers onto a black paper. It was then used to pollinate the stigma of the receiving plant gently using a brush.

2.3 *Arabidopsis* Growth Trial

For the first trial in 2015, 920 soil pots were prepared in 23 trays in a glasshouse compartment. The sterilized and stratified seeds of diploid and tetraploid Columbia and Landsberg parents, F1 and F2 were sown into pots in a randomised block design. The 23 trays served as the blocks and the plants were randomly allocated to slots within each tray. The trays were moved around the different places in the glasshouse on a weekly basis to enable a homogenous

environment for all the plants. In total 391 F2 diploid, 391 F2 tetraploid, 23 each of diploid and tetraploid parent and 23 each of diploid and tetraploid F1 seeds were sown. Each tray contained 40 pots, which had 17 F2 each of diploid and tetraploid and 1 each of parents and F1 of each diploid and tetraploid variety. Numbered sticks labeled the plants so that data collection was free of any bias.

From the first trial, enough leaf material to carry out sequencing analysis could not be collected because the plants did not grow well and were infested by flies, which did not let them produce enough leaves. Buds for chromosomal analysis also could not be collected for the same reason. The variance, a measure of dispersion around mean, in the F1 was found to be higher than the F2 variance for a few traits in both diploid and tetraploid plants. This high variance in F1 affect the heritability calculations as can be seen in **2.10.** All these factors necessitated conducting second plant trial which was conducted in 2016.

For the second trial in 2016 (**Figure 2-1**), 980, 10 cm soil pots were prepared in growth room, where 12 hours of day and light was maintained. Bigger pots were used to enable the plants to root well and grow healthily. The sterilized and stratified seeds of diploid and tetraploid Columbia and Landsberg parents, F1 and F2 were sown into pots in a randomised manner, but the pots could not be moved around as for the first trial. In total 401 F2 diploid, 401 F2 tetraploid and 28 each of diploid and tetraploid parent and 33 each of diploid and tetraploid F1 seeds were sown.

Figure 2-1 Arabidopsis thaliana plant growth trial 2016.

2.3.1 Phenotype scoring

Different phenotype traits were collected in the plant trial and are shown in **Table 2-1**.

Phenotype traits

Trait	Definition
Days to Germinate (DTG)	Time taken from seed sown to the first visible cotyledons
Days to Flowering 1 (DTF1)	Days from germination till the first visible flower buds in the apical meristem
Days to Flowering 2 (DTF2)	Days from germination till the main stalk is 1 cm long
Days to Flowering 3 (DTF3)	Days from germination till the first open flower is visible
Rosette Leaves Number (RLN)	Number of leaves counted at DTF1
Number of Cauline Leaves (CLN)	Leaves on the main stem
Total Leaves Number (TLN)	Sum of RLN and CLN
Number of Lateral Branches (LB)	Branches count on main stem
Number of Basal Branches (BB)	Apart from the main branch
Total Branches (TB)	Sum of LB and BB
Reproductive Period (RP)	Days between DTF1 and complete senescence
Life Cycle (LC)	Days a plant take for its life cycle from germination to complete senescence
Silique Length (FERT 1)	Average of 10 random siliques
Seed Number (FERT 2)	Average number of seeds from 10 siliques

Table 2-1 Different phenotype traits collected in Arabidopsis thaliana 2016 trial.

2.3.2 Sample collection

Leaf Collection: the rosette leaves of the plants were collected after the DTF2. The leaves were carefully cut using scissors and were placed in labeled microfuge tubes in liquid nitrogen. The leaves were collected in triplicate and then stored in cryoracks at -80 °C for future molecular biology experiments (for both trials).

Bud Collection: the unopened buds for the plants were collected and fixed immediately in the fixative as mentioned in **2.4.1.1**. The buds were kept at 4 °C for future cytology experiments (for second trial only).

Silique collection: The plants were allowed to grow till maturity. 10 siliques from individual plants were randomly selected, carefully picked from the plant using forceps and kept separately in labeled microfuge tubes for silique length measurement and seed counts.

Plant height: pictures were taken for all the plants from a particular fixed height. The length/height for a few plants were calculated using the freehand line feature of software Image J on the longest branch of the plant (for first trial only).

Seed count and silique length: The length of each individual silique was measured using a ruler/vernier caliper to the nearest mm. It was then carefully opened using fine forceps and seeds were kept on a clean paper and were counted. This was done for both diploid and tetraploid varieties of *Arabidopsis thaliana*.

2.4 Cytological Methods

The flowers/inflorescence of *A. thaliana* and the buds of *S. tuberosum* were fixed and chromosome spreads were prepared according to Armstrong *et al.* (2009).

2.4.1 Chromosomal spreads for Arabidopsis *thaliana*

2.4.1.1 Fixing buds

A. thaliana inflorescence was fixed in a 3:1 ratio of ethanol: glacial acetic acid fixative at 4 °C Inflorescences from different plants were fixed separately in small universals. The fixed inflorescence was kept at 4 °C for 3 hours and then it was changed with new cold fixative. The fixed buds were then stored at 4 °C.

2.4.1.2 Slide for chromosomal spreads

Inflorescence was placed in the fresh and cold fixative in a watch glass. All the opened and big buds with yellow anthers were removed, the fixative was replaced and the buds were washed by citrate buffer (pH 4.5) thrice. The buds were then incubated in an enzymatic mixture containing 0.3% w/v pectolyase and 0.3% w/v cellulase in citrate buffer, in a humid environment at 37 °C for at least 90 minutes. Tetraploids were incubated for at least 120 minutes. After incubation the enzyme was replaced by cold citrate buffer and single bud was placed on a clean slide and quickly macerated using the brass rod/mounted needle. 7-10 µL of 60% acetic acid for diploids and 80% for tetraploids was added on the macerated bud on the slide, which was then kept on a hot plate at 45 °C for a minute while stirring with a needle. Another drop of acetic acid was added in between and mixed to prevent drying out of material. For tetraploids, the process was done for 2 minutes. The slide was then removed from the hot plate and cells were fixed with 200 µL of cold 3:1 fixative and dried. The spread was stained with DAPI (4',6-diamidino-2-phenylindole, 10 µL/mL DAPI at 1 mg/mL in an antifade mounting medium VectaShield) and the chromosomes were then visualised using the Fluorescent microscope, Olympus BX61 fluorescent microscope using Smart Capture3 software.

2.4.2 Chromosomal spreads / protocol modification for *Solanum tuberosum*

The solutions and enzymatic mixture used were exactly same as for *A. thaliana* except DAPI concentration, which was used at a concentration of 5 µL/mL DAPI at 1 mg/mL in an antifade mounting medium (VectaShield). Fixation of either the buds or the anthers was done in the same way. When the anthers were fixed, the fixative was changed after 24 hours in addition to the 2nd change on the same day after 3 hours. When potato buds were fixed, the fixative was changed after 3 hours, 24 hours and after a week from first day. This was done because the buds kept losing chlorophyll even after the 2nd day change.

For making spreads, individual anthers were used after measuring them under the dissecting microscope. The anthers were digested in the enzyme for 105 minutes in the moist chamber at 37 °C. Any less, gave poor cell wall digestion and any more, cells were digested, which gave only fewer cells for analysis. After digestion, enzyme was replaced by citrate buffer and an individual anther was placed on the slide with 2 µL of 60% acetic acid. It was macerated using a brass rod and placed on the hot plate at 45 °C with 10 µL of 80% acetic acid for 3 minutes, with 2 more drops of 80% acetic acid added in between to prevent drying of the cells. Longer time helped remove more cytoplasm and spread the sticky metaphase chromosomes better. The rest of the protocol was as described in **2.4.1.2**.

2.4.3 Fluorescence *in situ* hybridisation

The chromosomal spread slides having metaphase I (M1) stages were used to label the specific sites with probes. It is a two-day process based on the protocol from Professor Chris Franklin's lab, which has been adapted from Armstrong, Sanchez-moran and Franklin (2009). It involved

washing slides in 100% ethanol for 10 minutes to remove coverslip followed by washing slides in 4T (4x SSC buffer and 0.05% Tween 20) for more than one hour on a shaker. 100 mL 0.01M HCl (200 µL 5M HCl in 100 mL SDW) was incubated at 37 °C while the slides were being washed. Next the slides were washed in 2x SSC for 10 minutes at room temperature when pepsin solution (0.01% pepsin in 0.01M HCl) was prepared. Slides were then washed in pepsin for 90 seconds only at 37 °C. This was followed by sequential washing in 2x SSC for 10 minutes, paraformaldehyde (4% pH 8) for 10 mins and two times in sterile distilled water for 5 minutes each. The material on the slides was then dehydrated by washing for 2 minutes each in an alcohol series of 70%, 90% and 100% ethanol followed by air drying for at least 15 minutes. In the meantime, probes were prepared by mixing 14 µL master mix and 6 µL probe (3 µL 5S BIO, 3 µL 45S DIG labelled by nick translation with digoxygenin or biotin attached to dUTP) for each slide, and heating it at 94 °C for 10 minutes in a PCR machine. The probes were kept on ice straight out of PCR machine to keep them denatured. The probe was then applied on the material on the slides, which was then covered by coverslips and kept on the hotplate at 75 °C for four minutes during which the rubber seal was applied around the coverslip. The slides were place in a humid chamber at 37 °C overnight.

The following day, the slides were taken out and washed thrice in 50% Formamide 2x SSC at 45 °C for 5 minutes each. It was followed by a 4 minute wash in 2x SSC at 45 °C and 5 minutes each in 4T at 45 °C followed by 4T at room temperature. 80 µL of secondary antibodies, anti-biotin Cy3, diluted 1 in 200 in milk block, and anti-digoxygenin FITC, diluted 1 in 50 in DIG block, were then applied one at a time on the slides with incubation in moist chamber at 37 °C in dark for 30 minutes each time. After the incubation with each secondary, slides were

washed thrice in 4T for 5 minutes in dark after which they were mounted in 8 μL of DAPI and checked under the fluorescent microscope.

2.4.4 Immunolocalisation in potato

2.4.4.1 On fixed material on DAPI slides

Immunolocalisation of AtASY1 and AtZYP1 on prepared slides was carried out as per the protocol from Prof Chris Franklin's lab and described in Armstrong, Sanchez-moran and Franklin (2009). The coverslips were removed by dipping the slides in 10 mM citrate buffer after cleaning the excess oil using 70% ethanol from the coverslip. Next the slides were kept in boiled citrate buffer for 45 seconds followed by a wash in 1% phosphate buffered saline (PBS), 0.1% triton X-100 solution for 10 minutes at room temperature. The blocking buffer (1% PBS. 0.1% triton X-100, 3% bovine serum albumin) was then applied on the slides and they were kept in a moist box for 10 minutes at room temperature. 100 μL of primary antibodies, rat anti-ASY1 and rabbit anti-ZYP 1c diluted 1 in 400 of the blocking buffer were then applied on the slide on a parafilm coverslip and incubated at 4 °C overnight in a moist box. Next day, the slides were washed in PBS triton buffer 3 times for 5 minutes each. The 100 μL secondary antibodies, anti-rat ASY1-FITC diluted 1 in 50 and anti-rabbit CY3 diluted 1 in 200 blocking buffer, were applied on the slides on a parafilm coverslip. The slides were incubated at 37 °C for 30 minutes in the dark and then washed again thrice in PBS triton buffer in the dark and stained with DAPI for visualisation.

2.4.4.2 Immunolocalisation on fresh material

Using fresh buds from potato plant, suitable size anthers were dissected on moist filter paper. Few anthers were kept in 20 µL of digestion mix (0.4% cytohelicase, 1.5% sucrose and 1% polyvinylpyrollidone) in a cavity slide and incubated in a moist box on a hot plate at 37 °C for 4 minutes. Using a brass rod, gentle pressure was applied on the anthers to exude the meiocytes into the digestion mix. The slide was again incubated in a moist box for 3 more minutes. Next, the cavity slide was removed from the box and 2 µL of meiocyte suspension was added to 10 µL of 1% lipsol on a slide and spread with a fine needle. The material was then fixed with 12 µL of cold 4% paraformaldehyde (pH 8) and air dried for 2 hours in a fume hood. After drying, the slides were washed briefly twice in 1X PBS, 0.1% triton X-100. 100 µL of blocking solution (3% BSA in 1X PBS, 0.1% triton X-100) was applied on slides using parafilm strips and the slides were incubated in a moist box at room temperature for 15-30 minutes. The blocking strips were then removed and primary antibodies were applied with the following dilutions: rat anti-ASY1 and rabbit anti-ZYP 1, 1 in 400 blocking buffer. The slides were then incubated in a moist box at 4 °C overnight. Next day, the slides were washed thrice for 5 minutes each in 1X PBS, 0.1% triton X-100 and secondary antibodies anti-rat FITC and anti-rabbit Cy3 diluted 1 in 50 and 1 in 200 blocking buffer respectively, were applied. The slides were incubated in a moist box at 37 °C for 30 minutes in the dark. They were washed again as above and stained with DAPI for visualisation.

2.4.5 Alexander pollen staining

Pollen viability for potato variety Sante was checked using Alexander staining (Alexander, 1969). Pollen collected was placed on a slide in a drop of the stain and pressed down using a

coverslip sealed with the rubber solution. Slide was then placed at 50 °C for 1 hour and was then checked using fluorescent microscope. Viable pollen should appear dark red under UV light.

2.5 DNA Extraction

DNA extraction of *A. thaliana* leaves was carried out using different methods for different purposes as explained below.

2.5.1 DNA extraction for PCR Genotyping

This process was based on a homemade protocol provided from the lab of Dr Eugenio Sanchez Moran. The leaf was grounded well with 40 µL of extraction buffer (200 mM Tris HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) using autoclaved plastic pestles in an Eppendorf tube on ice. 400 µL of extraction buffer was added to the ground material in the tube and vortexed. The supernatant was collected in a fresh and clean tube after the grounded lysate was centrifuged at 13000 rpm for 5 minutes. 400 µL of isopropanol was gently mixed by inverting the tubes and incubated at room temperature for 2 minutes. It was centrifuged again at 13000 rpm for 10 minutes and the supernatant was discarded. The DNA pellet was washed twice with 70% ethanol and centrifuged again at 13000 rpm for 5 minutes. Finally, the DNA pellet was air dried and dissolved in 30-50 µL of RNase and DNase free water and kept at 95 °C in a heat block for 3 minutes and spinned down briefly. The DNA was stored at -20 °C.

2.5.2 DNA extraction for NGS

For Next Generation Sequencing (NGS), DNA was extracted using the Sigma plant DNA extraction kit G2N350 as per the manufacturer's protocol.

2.5.3 DNA quantification and quality assessment

DNA was quantified on the Nanodrop in ng/µL by measuring the peak absorbance at 260 nm. A260/280 and A260/230 readings were recorded to also check the quality of the DNA. DNA was also quantified using Qubit 2 fluorimeter as per the manufacturer's protocol.

2.6 Genotyping by PCR using SSLP markers

Genotyping to distinguish and to confirm the heterozygosity of the diploid and tetraploid F1 generation created by crossing Columbia and Landsberg parental diploid and tetraploid *A. thaliana* lines was carried out using the SSLP markers as described in Pacurar *et al.* (2012) and Hou *et al.* (2010). PCR with the following conditions was carried out to confirm the marker heterozygosity in the F1s created: For chromosomes 1-4, annealing takes place at 60 °C for 30 minutes followed by an extension of 1 minute at 72 °C. For chromosome 5, annealing occurs at 51 °C followed by extension of 1 minute at 72 °C. Primers and Polymorphism Lengths for these markers are given in **Table 2-2**.

Primer pair (5’ – 3’)	Chromosome	Length Col fragment	Length Ler fragment
UPSC_1-1220-f TTTAGGGATGGGTCATGGTC UPSC_1-1220-r TTGGTTCTTCTTTCGGATTTTC	1	211	232
UPSC_2-19330-f ACGTATGCACCGCAACAAT UPSC_2-19330-r GGCGAGGGATACGAAAATGT	2	155	135
UPSC_3-7140-f CTCCAGCTCCACCACCAG UPSC_3-7140-r CCAAAAGACATTCTTCCACCA	3	170	190
UPSC_4-17544-f CACCATTGACATTTGATGCAC UPSC_4-17544-r CCGTAGCTCCATTGGCTTAT	4	214	234
AB010070-0918f CTCTGTTGGGGCAAAACC AB010070-0918r GATGCTGGAGAGTAGCTTAG	5	220	146

Table 2-2 Primers and polymorphism length for SSLP chromosome markers in Arabidopsis thaliana.

2.7 Agarose gel electrophoresis

Bioline agarose powder was used to prepare 2% (w/v) and 1% (w/v) gel in 1x TBE buffer to check the polymorphic markers after PCR, or genomic DNA after extraction. Gel Red was used as the nucleic acid stain. The gels were allowed to run at 100 V for more than 5 hours for SSLP markers or 1 hour for genomic DNA and were then checked using the Bio Rad ChemiDoc trans-illuminator under UV light and visualized with Imagescan software.

2.8 Sanger sequencing

The PCR products obtained were given a clean up using Axygen PCR clean up kit as per their protocol or the markers were gel extracted using Qiagen gel extraction kit. It was then Sanger sequenced using the same PCR primers to confirm the marker sequences, which were assembled using CAP 3, and aligned with *A. thaliana* sequence using Multialin (http://multalin.toulouse.inra.fr/multalin/).

2.9 Statistical analysis

Different kind of statistical tests were performed depending on data distribution. They include Ryan joiner, to test for the normality of the data distribution, Bartlett, to test for the homogeneity of variance between the data for different lines, Mann Whitney, two sample proportion test, to find out if the difference between two groups was significant at p = 0.05, Kruskal-Wallis, to compare the difference between more than two groups, Post hoc Dunn, to find out what groups analysed in Kruskal-Wallis are actually significantly different. χ^2 test of association and goodness of fit tests to explore associations and deviations. Data exploratory analysis including mean, variance and range was performed on the phenotype data collected. All the tests were carried out using R, Minitab and Excel. Graphs were made either in Excel, R or in Minitab. Outliers were recognized in the dataset of various traits by using the Boxplot.stats () command in R. R considers any value which is less than or greater than 1.5 times the interquartile range as an outlier. The recognized outliers were manually removed and the trait distribution was compared with and without the outliers.

2.10 Heritability analysis

The genetic variability in a heterozygous F1 is fixed, hence any variation in phenotype shown by F1 can be considered to equate to the environmental variance. Alternatively, an average of variances of parents and F1 (or the weighted average) can be taken as equal to environmental variance (provided the variances between the parents and F1 do not differ significantly). Thus, the genetic variance in F2 and hence the heritability estimates of F2 can be calculated as follows:

$V_P = V_G + V_E$ (1)

Where V_P, V_G and V_E are the phenotypic, genotypic and environmental variance of the trait respectively.

$V_E = V_{F1}$ (2)

or $V_E = (V_{P1}+V_{P2}+V_{F1})/3$ (3)

where V_E is the weighted combined variance calculated using the combined variance formula:

$$s_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{\sum_{i=1}^{k}(n_i - 1)}$$

When using combined variance of the three, it was ensured that the three variances did not differ significantly from each other. This was checked using the Bartlett test for homogeneity of variances.

In equations 2 and 3, V_{F1} (V_{P1} or V_{P2}) is the phenotypic variance observed in F1 (P1 or P2) generation plants.

$V_{F2} = V_G + V_{F1}$ (4)

Where V_{F2} is the phenotypic variance observed for F2 plants

$H^2 = V_G/V_P$ or $H^2 = (V_{F2} – V_{F1})/V_{F2}$ (5)

H^2 is the Broad sense heritability

This method gives us a broad estimate, which includes both the additive as well as the dominance variance.

$H^2 = V_G/V_P = (V_A + V_D)/(V_A + V_D + V_E)$ (6)

Where V_A is the additive variance and V_D is the dominance variance.

Selection can only be applied for the additive genetic component of the total phenotypic variance; hence it is important for breeders to know about the additive genetic component V_A of the total variance V_P. This estimation is known as narrow sense heritability and is calculated as follows:

$h^2 = V_A / (V_A + V_D + V_E)$ (7)

h^2 is the narrow sense heritability.

2.10.1 Trait Segregation Analysis

Another statistical method developed for the F2 trait segregation analysis for estimation of major gene effects, was utilized for estimating different components (additive, dominance and environmental variance) of the total phenotypic variance for each trait, in both diploids and tetraploids (Chen *et al.*, 2018). These were then used to calculate both, the broad sense and the narrow sense heritability. The genetic model assumes the phenotype follows a mixed normal distribution in the sample population, with *m* component distributions, each corresponding to a different genotype at the QTL. For an F2 diploid population there are three possible genotypes (AA, Aa, aa) while in an F2 tetraploid population, there are five possible genotypes (AAAA, AAAa, AAaa, Aaaa, aaaa).

This method utilizes the Likelihood Ratio test for the maximum likelihood estimate of the major genetic effects (additive, dominance or interactive in the case of diploids; monogenic, digenic, trigenic or quadrigenic in tetraploids) of the quantitative traits estimated using EM algorithm. The likelihood function for the trait can be given as:

$$L(G, \sigma^2 | X, G_{P1}, G_{P2}, \alpha) = \prod_{j=1}^{n} \sum_{i=0}^{m-1} f_i(G_{P1}, G_{P2}, \alpha)\, g_i(x; G_i, \sigma^2)$$

where G is genotypic value vector, $G = (G_0 \ldots G_{m-1})$, σ^2 is the residual variance, G_{P1} and G_{P2} are the parental QTL genotypes. $X = (x_1, x_2, \ldots x_n)$ represents offspring trait phenotype data, α being the coefficient of double reduction, $f_i(G_{P1}, G_{P2}, \alpha)$ $(i = 0, \ldots m-1)$ indicates the genotypic frequency of genotype Qq in a diploid and $Q_i q_{4-i}$ in a tetraploid and $g_i(x; G_i, \sigma^2)$ represents the probability density function of a normal distribution with mean G_i and variance σ^2.

The EM algorithm involves iterating the E step to estimate the conditional probability of the i^{th} individual with j^{th} genotype, and using this estimated probability in the M step to calculate the maximum likelihood estimates (MLEs) of parameters G and σ^2, until the values obtained from E and M step converge, which indicates the maximization of the parameter values. The LOD score is then used to compare the estimates of genotypes and variances when QTL exists, to the means of genotypes and residual variance when there is no QTL. For a tetraploid, multiple LOD scores are obtained depending on the parental genotypes and the coefficient of double reduction. The LOD score with the highest value is considered to show the highest probability for the existence of a QTL. Based on the MLEs obtained, the estimated genetic variance components (V_A, V_D and V_E) are calculated using an orthogonal contrast scales model and used here to obtain both the broad sense and the narrow sense heritability for the trait, using equations (6) and (7) respectively.

2.11 RAD Sequencing

Extracted DNA samples from the leaves of the diploid and tetraploid *A. thaliana* collected during the second plant trial were sent to Fudan University in China, where library preparation for RAD sequencing was carried out according to the protocol described in Jiang *et al.* (2016) and shown in **Figure 2-2**. The library was further sent for Illumina sequencing to the Gene Energy company (www.genergy.cn). 24 samples were paired end sequenced on HiSeq 3000, which included 8 F2s, 2 F1s and the two parents each for diploids and tetraploids as a pilot to develop methodologies for sequence analysis. The samples were demultiplexed and sequences were made available for individual samples in Fasta file format, which were bioinformatically analysed further as shown later in Chapter 6.

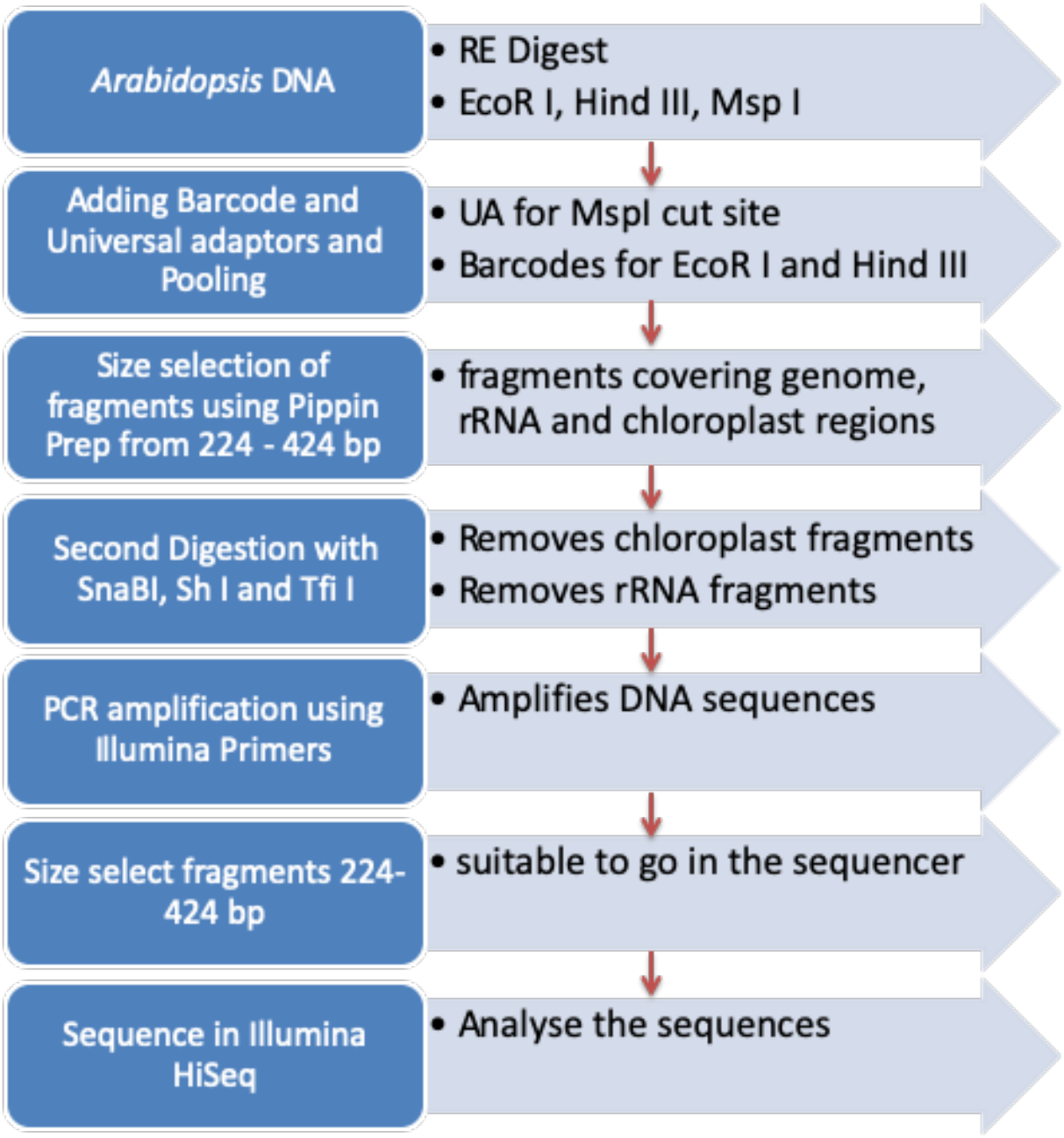


Figure 2-2 RAD sequencing protocol.

Figure adapted from Jiang *et al*. (2016).

2.12 References

Alexander, M. P. (1969) 'Differential staining of aborted and nonaborted pollen', *Biotechnic and Histochemistry*, 44(3), pp. 117-122. doi: 10.3109/10520296909063335.

Armstrong, S. J., Sanchez-moran, E. and Franklin, F. C. H. (2009) 'Cytological Analysis of Arabidopsis thaliana Meiotic Chromosomes', *Methods Molecular Biology*. Edited by S. Keeney. Totowa, NJ: Humana Press (Methods in Molecular Biology), 558, pp. 131–145. doi: 10.1007/978-1-60761-103-5_9.

Chen, J. *et al.* (2018) 'Orthogonal contrast based models for quantitative genetic analysis in autotetraploid species', *New Phytologist*, 220(1), pp. 332-346. doi: 10.1111/nph.15284.

Hou, X. *et al.* (2010) 'A platform of high-density INDEL/CAPS markers for map-based cloning in Arabidopsis', *The Plant Journal*, 63(5), pp. 880–888. doi: 10.1111/j.1365-313X.2010.04277.x.

Jiang, N. *et al.* (2016) 'A highly robust and optimized sequence-based approach for genetic polymorphism discovery and genotyping in large plant populations', *Theoretical and Applied Genetics*, 129(9), pp. 1739-1757. doi: 10.1007/s00122-016-2736-9.

Pacurar, D. I. *et al.* (2012) 'A collection of INDEL markers for map-based cloning in seven Arabidopsis accessions', *Journal of Experimental Botany*, 63(7), pp. 2491–2501. doi: 10.1093/jxb/err422.

Sharma, S. K. *et al.* (2018) 'Linkage Disequilibrium and Evaluation of Genome-Wide Association Mapping Models in Tetraploid Potato', *G3: Genes|Genomes|Genetics*, 8(10), pp. 3185-3202. doi: 10.1534/g3.118.200377.

CHAPTER 3

GENOTYPING AND CYTOLOGICAL ANALYSIS OF PARENTS FOR THE CREATION OF DIPLOID AND TETRAPLOID *ARABIDOPSIS THALIANA* F2 POPULATIONS

3 Genotyping and cytological analysis of parents for the creation of diploid and tetraploid *Arabidopsis thaliana* F2 populations

Landsberg and Columbia diploid as well as tetraploid genotypes of *Arabidopsis thaliana* as described in materials and methods (**2.1.1**) were crossed to create F1 progeny, which produced an F2 population. A series of genotyping and cytological tests were performed on the parental and F1 lines as described below.

3.1 Genotyping

3.1.1 Introduction

Natural phenotypic and genotypic variations between different individuals within species is the basis of genetic studies that help in understanding the variation in development and physiology of an organism. These differences can be identified as polymorphisms at DNA level for a certain trait. These nucleotide level polymorphisms have been utilized in the development of molecular markers, which are the biggest pillar of the molecular genetic studies in different organisms.

Accessions of *Arabidopsis thaliana* (hereafter referred to as *A. thaliana* or *Arabidopsis*) differ in a number of traits and an array of molecular markers have been developed to study the differences/polymorphisms. These molecular markers help to identify the gene or the genic regions responsible for the trait variation. The first stage in mapping the genic region is to create the mapping population by crossing two parental accessions and developing an F1 generation, which can be selfed to produce a segregating F2 progeny. Molecular markers at

the first stage can help to identify the heterozygosity of the F1 plants. Different kind of markers can be used. We have used the Simple Sequence Length Polymorphisms (SSLP)/InDeLs to differentiate between Columbia and Landsberg alleles in the F1 population and check their heterozygosity. SSLPs are repeated sequences of different lengths in different accessions and InDeLs refer to insertion deletion polymorphisms between two genotypes. These molecular markers have been well identified and they occur in non-genic region. They are easy for PCR analysis and give a straightforward analysis for the F1 generation (Pacurar *et al.*, 2012).

3.1.1.1 SSLP markers

Simple sequence length polymorphisms (SSLP) are the PCR based microsatellite markers. The markers utilize the polymorphism in tandem repeats contained in eukaryotic genomes. They are codominant genetic markers and the differences can be visualized using gel electrophoresis after amplifying the polymorphic fragments by PCR. Their discovery accelerated the process of genetic mapping in mammals (Tautz, 1989). In *A. thaliana*, 30 new mono and di nucleotide SSLP markers were identified and assigned to the linkage map for first time by Bell and Ecker (1994). Since then, a large number of SSLP markers have been developed in *Arabidopsis.*

3.1.2 Results

3.1.2.1 SSLP marker genotyping of diploids

SSLP markers were selected as described in methods and materials (**2.6**). The difference in parents and the polymorphism in the F1 created were checked on a 2% agarose gel and the results are shown in **Figure 3-1**.

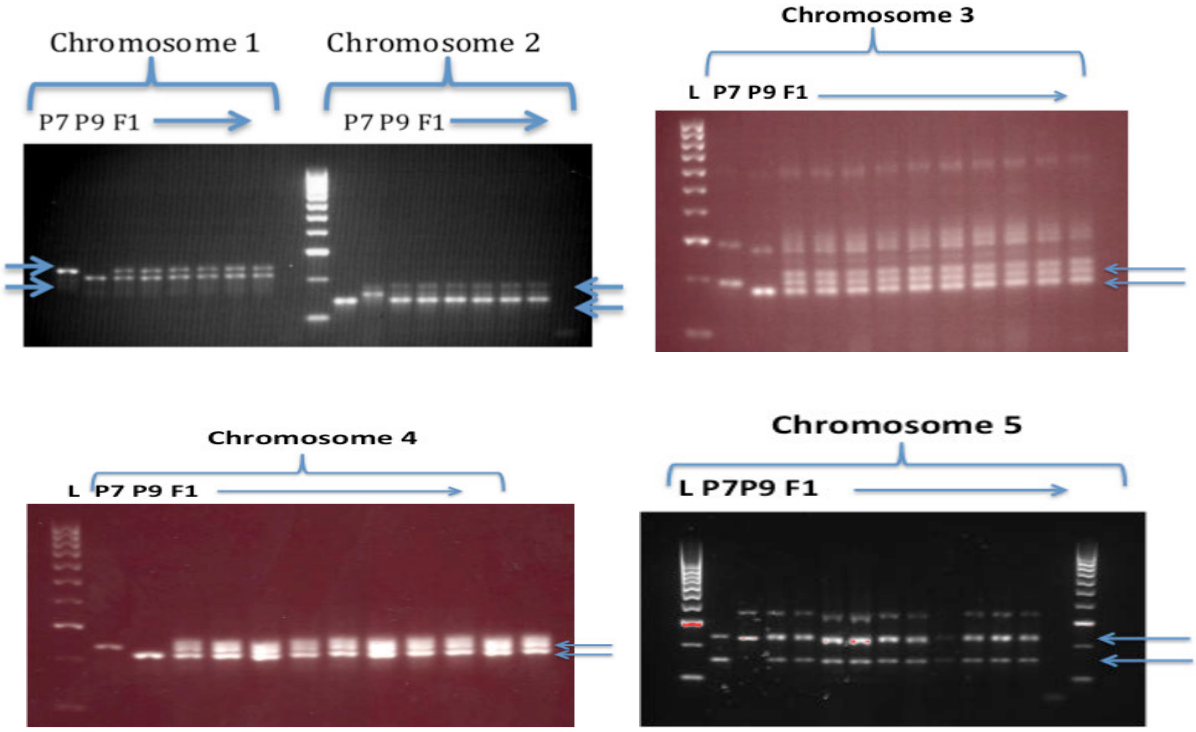


Figure 3-1 2% Agarose gel for resolving marker fragments in different chromosomes of diploid parental and F1 diploid plants.

Agarose gel for resolving fragments of 232 bp and 211 bp on chromosome 1, 135 and 155 bp on chromosome 2, 190 bp and 170 bp on chromosome 3, 234 bp and 214 bp on chromosome 4 and 146 bp and 220 bp on chromosome 5 for diploid parental lines of *Arabidopsis thaliana* and F1s resulting from their cross. P7 is Landsberg *Arabidopsis thaliana*, P9 is Columbia *Arabidopsis thaliana*, F1 is Progeny from the cross of Landsberg and Columbia parents. The blue arrows point out the two fragments.

The desired marker bands are visible for the parental lines and the same two different Columbia and Landsberg bands are visible in the F1 lines in **Figure 3-1**. For chromosomes 3 and 5 there are more than the desired bands, but all the bands visible in the parental lines are

also visible in the F1, again confirming the polymorphism in the F1s. For other chromosomes such as 2 and 4, stuttering is visible. These are the shadow bands, which are one size above the desired band. They are normally formed during PCR of short repeats; however, we are able to differentiate between the two different bands of the parents in the F1. Thus, the polymorphism is confirmed in the F1s.

3.1.2.2 SSLP marker genotyping of tetraploids

F1 plants were created by crossing the tetraploid *Arabidopsis* Columbia and Landsberg lines. They were genotyped using all the five chromosome markers as was done for diploids. While different distinct bands were expected in the parental lines, F1 line was expected to show heterozygosity. In **Figure 3-2**, distinct bands in Landsberg and Columbia tetraploid parents can be seen, and the same two distinct bands are visible in the F1 created by the cross.

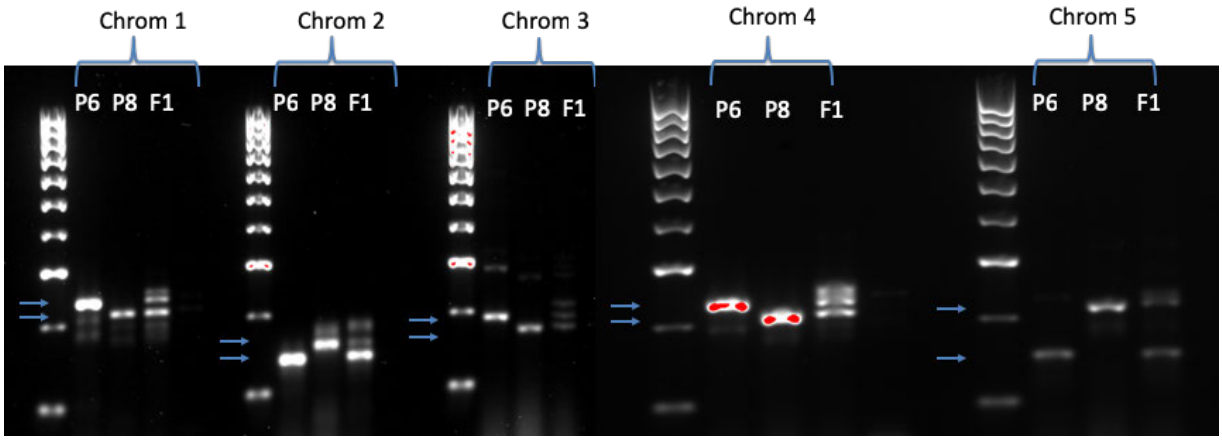


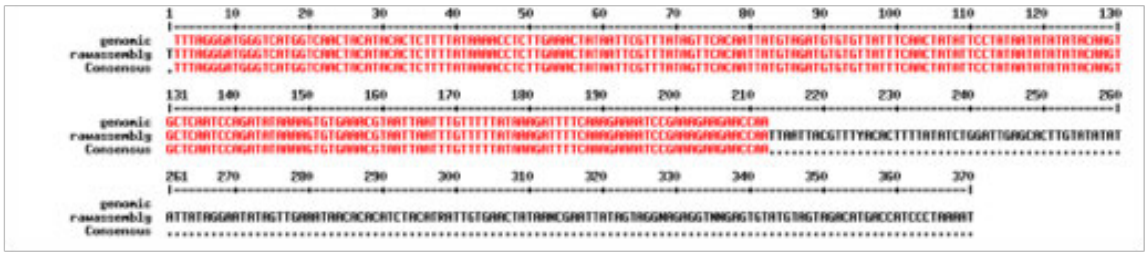
Figure 3-2 2% Agarose gel marker identification in different chromosomes of tetraploid parental and F1 lines.

2% Agarose gel for resolving fragments of 232 bp and 211 bp on chromosome 1, 135 and 155 bp on chromosome 2, 190 bp and 170 bp on chromosome 3, 234 bp and 214 bp on chromosome 4 and fragments of 146 bp and 220 bp on chromosome 5 for tetraploid parental lines of *Arabidopsis thaliana* and F1 resulting from their cross. P6 is tetraploid Landsberg parent. P8 is tetraploid Columbia parent. The blue arrows point out the two fragments.

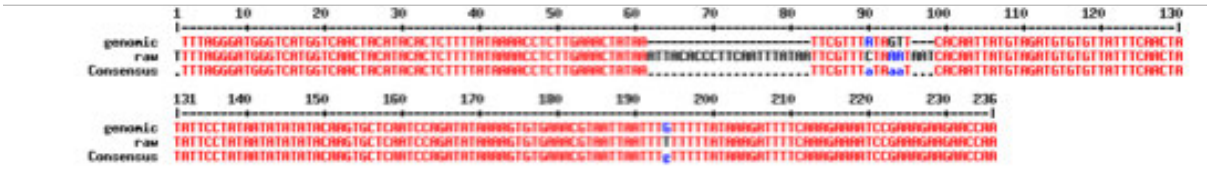
3.1.2.3 Sequencing of the SSLP fragments

The PCR products of F1 were sequenced to confirm the parental sequences and aligned with the reference Col *Arabidopsis* sequence using Multialin as described in materials and methods (**2.8**).

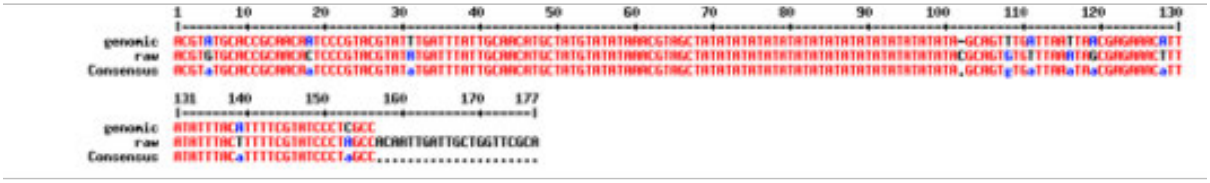
Chromosome 1 F1 Fragment from Col (211 bp)



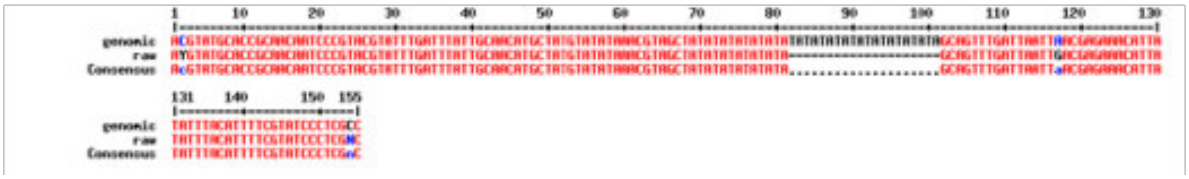
Chromosome 1 F1 Fragment from Ler (232 bp)



Chromosome 2 F1 fragment from Col (155 bp)



Chromosome 2 F1 fragment from Ler (135 bp)



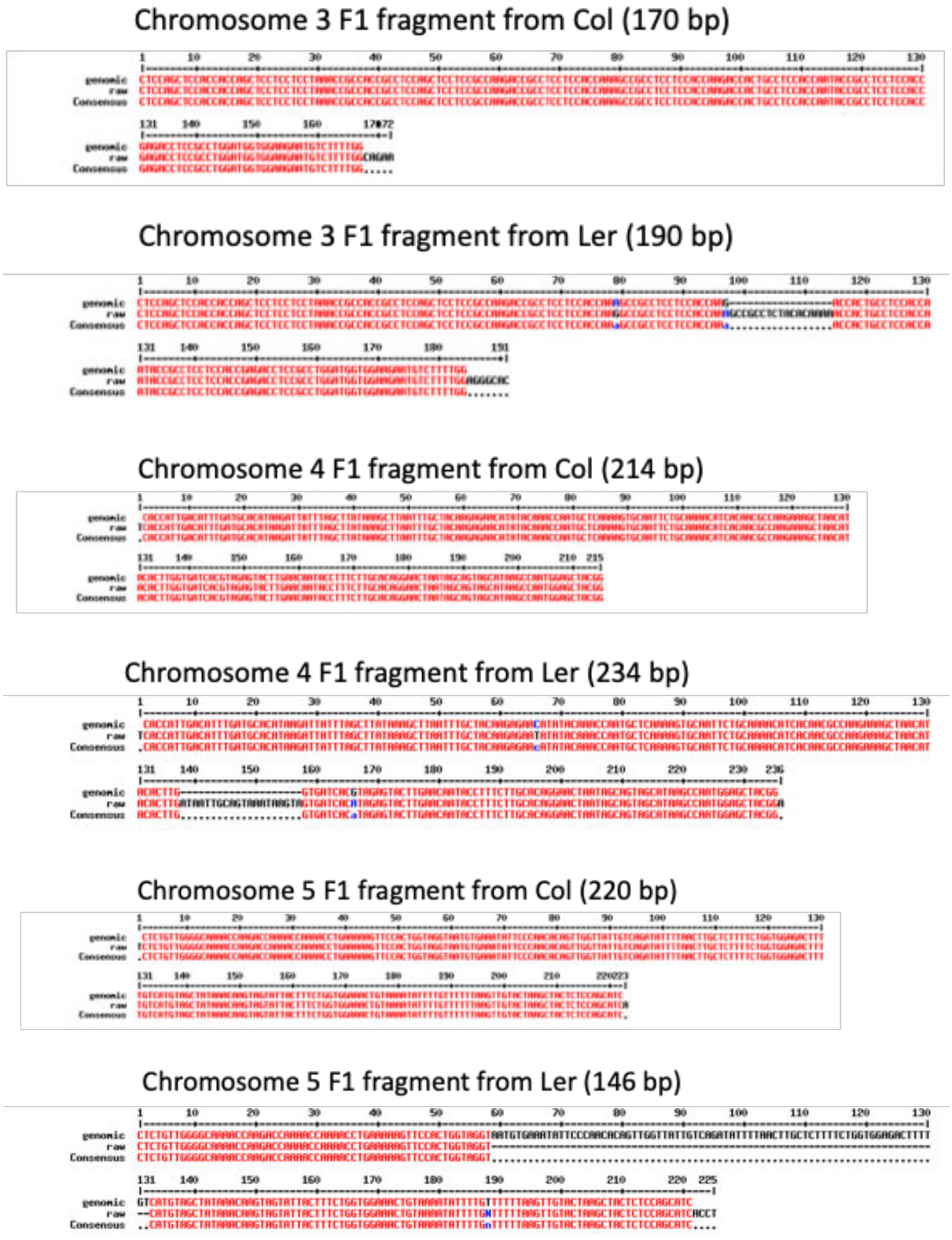


Figure 3-3 Heterozygous F1 sequencing showing SSLP fragments of respective Columbia and Landsberg parent.

The above sequences in **Figure 3-3** confirm the polymorphisms between the two lines.

3.2 Cytological analysis of parental lines and F1s

3.2.1 Introduction

Arabidopsis thaliana has been the model of choice for investigating the molecular mechanisms governing meiosis in plants. Diploid *Arabidopsis* consists of 5 pairs or 10 chromosomes. There are rare naturally occurring established autotetraploid lines of *A. thaliana* (Weiss and Maluszynska, 2000). However, the tetraploid lines which have been artificially synthesized and advanced through several generations are available to buy from the stock centre (NASC). These plants were sourced as tetraploids from the stock centre, but it is important to establish their tetraploidy before carrying on any further analysis. To do this, chromosomal spread of different plants of Columbia and Landsberg tetraploid variety were prepared and were stained using DAPI. This enables counting of chromosomes in mitotic as well as meiotic pollen mother cells (PMC). If 20 chromosomes could be counted in at-least ten cells for each tetraploid plant, their tetraploidy was confirmed.

FISH analysis using 5S and 45S rDNA probes on the metaphase I (M1) chromosomes is used as one of the cytological methods of quantifying recombination (Moran *et al.*, 2001). It can also help in identifying the chromosomes and hence establishing, that the correct number of chromosomes have doubled up in a tetraploid. FISH using 45S and 5S rDNA can be used to identify the individual chromosomes of *A. thaliana*. Chromosomes 2 and 4 carry 45S sequences in their short arms, while chromosomes 3, 4 and 5 carry 5S sequences as seen in **Figure 3-4**. Chromosome 1 does not have these signals and thus the process enables identification of all the chromosomes. In Columbia ecotype, the 5S signal on chromosome 3 is

found in the short arms, while it is found in the long arm in Landsberg (Sanchez-Moran *et al*., 2002).

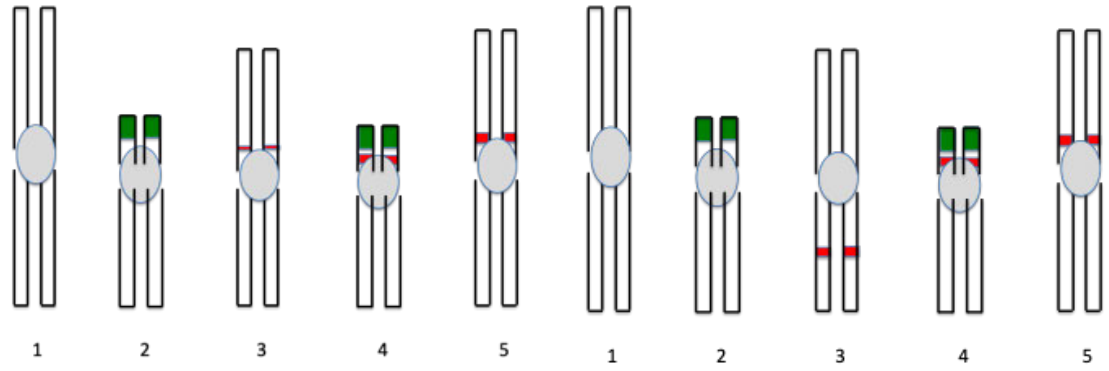


Figure 3-4 FISH 45S and 5S signals in Arabidopsis thaliana.

Figure adapted from Sanchez-Moran *et al.* (2002). Green 45S and red 5S signals shown in respective chromosomes in Columbia line (1st five) and Landsberg line (last five). The number of 5S repeats on chromosome 3 in Columbia is less than the number of 5S repeats on chromosome 5.

Crossing over between the homologous chromosome pair is identified as chiasma (the point of crossing over). The process initiates by the chromosome pairing when two homologues come together and exchange the genetic material. The chiasma then serves as the tethering point between the two homologues, which can be identified cytologically during M1 of meiosis. This helps in the proper orientation of the bivalents during M1 and hence in the proper segregation of the chromosomes. Chiasma, the points of recombination, can be counted for each bivalent. Fluorescence *in situ* hybridisation (FISH) helps to correctly identify the configurations of the chromosomes in metaphase, which in turn can help in counting the number of chiasmata in a particular bivalent.

The bivalent configurations in *Arabidopsis* have been assigned into two categories: rods and rings (Moran *et al.*, 2001). Rods are considered bound by chiasma in one arm only, while rings

are considered to be bound by both their arms. **Figure 3-5** helps in identifying the configurations.

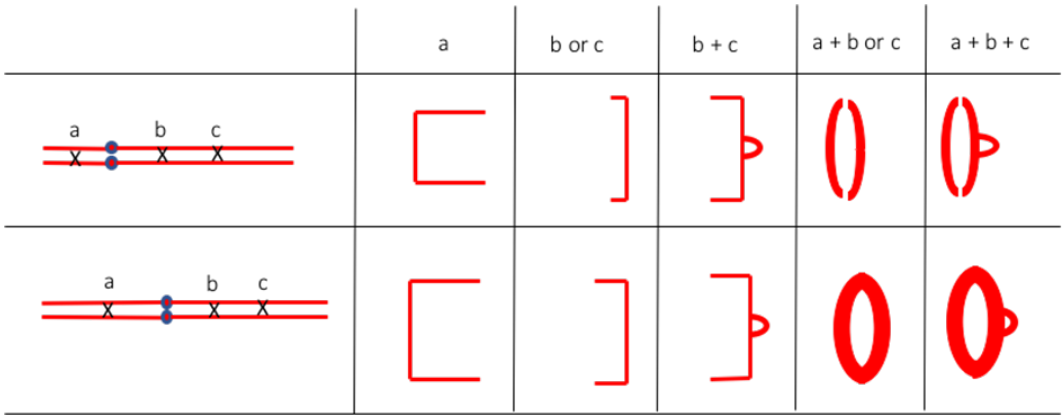


Figure 3-5 Possible bivalent configurations in diploid Arabidopsis thaliana in different chromosomes.

Figure adapted from Moran *et al.* (2001). 'a' represents a crossover between the homologues in the short arm of the chromosomes, 'b' and 'c' represent a crossover in the longer arm of the chromosome.

In chromosomes 2 and 4 (first row), which are acrocentric chromosomes, 'a' represents a chiasma in the shorter arm and therefore a rod with free longer, arms can be seen. However, when it occurs in the longer arm represented by 'b' or 'c', the rod is slightly longer and the free arms are comparatively smaller than in 'a'. If the chiasma forms at 'b' as well as 'c' in the longer arm, there appears a knob in the rod as can be seen in **Figure 3-5**. When the chiasma forms in both the arms it appears like a ring as seen in 'a+b or c', or a ring with a knob in 'a+b+c'. Similar kinds of chiasma formation can occur in the longer sub-metacentric chromosomes 1, 3 and 5 (second row) (**Figure 3-5**). Thus, if a rod occurs, we can assume that there is at-least one chiasma and if there is a ring, at-least two chiasmata have formed, one in each arm of the chromosome.

In an autotetraploid, four homologues of the same chromosome are present and crossover formation can occur between two, three or even all four homologues, resulting in a

quadrivalent formation. A few of the chiasma configurations that can occur in an autotetraploid are shown in **Figure 3-6**.

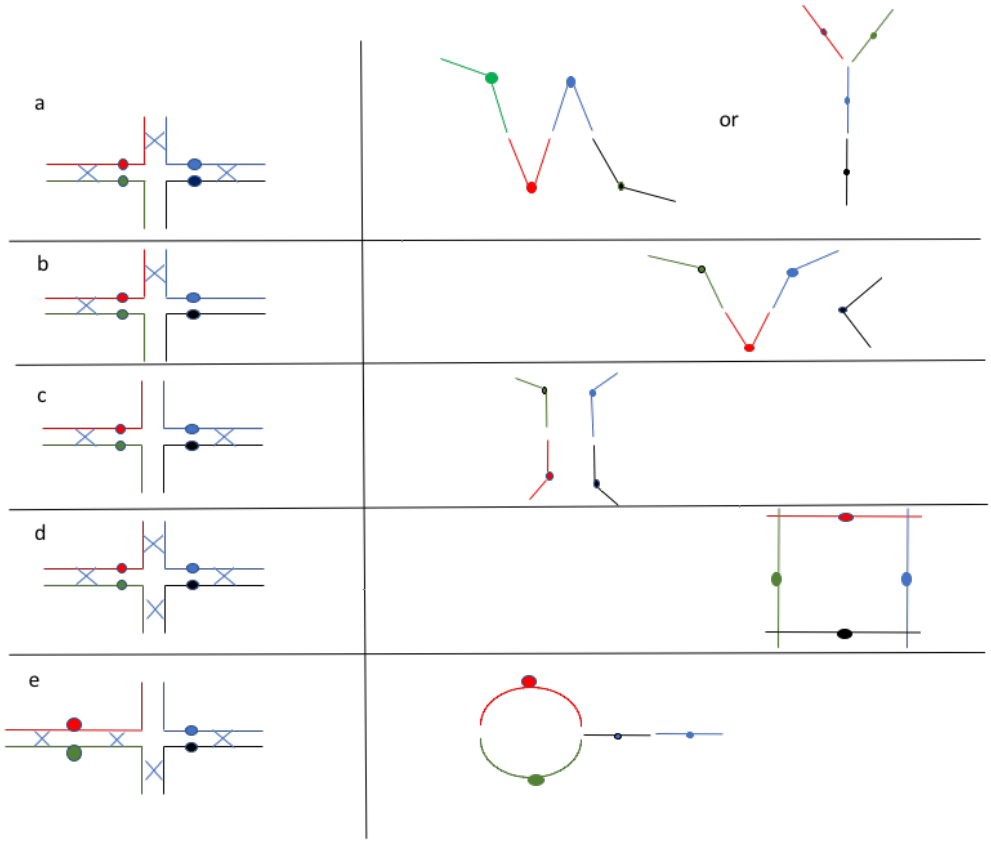


Figure 3-6 Few of the possible chiasma configurations in an autotetraploid meiosis.

Figure adapted from Sybenga (1975).

The four homologues of the same chromosome are represented by four different colours. The possible configurations depend on where crossovers occur. Quadrivalent configurations can be described as a chain or a Y in 'a', a ring in 'd', a trivalent with a univalent formation in 'b' or a spoon formation in 'e'. The segregation of the chromosomes can occur as bivalents, as shown in 'c'. Here, the results of the chromosome count for the parental as well as the F1 lines and the chiasma analysis of the diploid and tetraploid Columbia and Landsberg *A. thaliana* is presented.

Note on methodology of chiasma analysis: Cells in M1 were used to score chiasma and a conservative scoring was undertaken based on the configurations presented in the cell for a chromosome. A ring was considered to have 2 chiasmata and a rod to have 1 chiasma in diploids, unless it was visually very clear otherwise. Similarly for tetraploids, a ring quadrivalent was considered to have 4 chiasmata and chain quadrivalent to have 3. This kind of scoring can underestimate the crossover frequency. However, the consistency was maintained across all scorings, which can allow of useful comparisons. The number of cells with quadrivalents and trivalents were scored separately, but were grouped together as multivalents when doing any comparisons between multivalents and bivalents.

Hypothesis: According to the null hypothesis, there should be no significant differences in number of crossovers between diploids and tetraploid lines. However, the alternative hypothesis will be to have a significant difference in the number of crossovers between diploids and tetraploids. Here, we hypothesize an increase in the number of crossovers beyond the simple doubling in tetraploids which could be expected since the number of copies of each chromosome has been doubled.

3.2.2 Results

3.2.2.1 Chromosome counting for tetraploid plants

Chromosome spreading was carried out on the inflorescence of the plants and stained with DAPI to count the number of chromosomes under the fluorescent microscope and thus to confirm the tetraploidy of the plants. A minimum of 10 cells were counted, which included both the mitotic as well as the meiotic cells to establish the chromosome count of the plant.

A chromosome count was carried out on 10 separate plants of each of the tetraploid Columbia and Landsberg parental lines. Only those lines for which 20 chromosomes could be counted were used to create F1 lines. For the 10 plants counted, 6 out of 10 (60%) were found to have 20 chromosomes for Columbia and 8 out of 10 (80%) for Landsberg lines as can be seen in **Figure 3-7**. Three plants gave a chromosome count of 21 in Columbia lines and 1 plant had a count of 19 chromosomes. In Landsberg plants, one plant counted 21 and one 18. Again, the F2 seeds from only those F1 lines with a full set of 20 chromosomes were used for the plant trial.

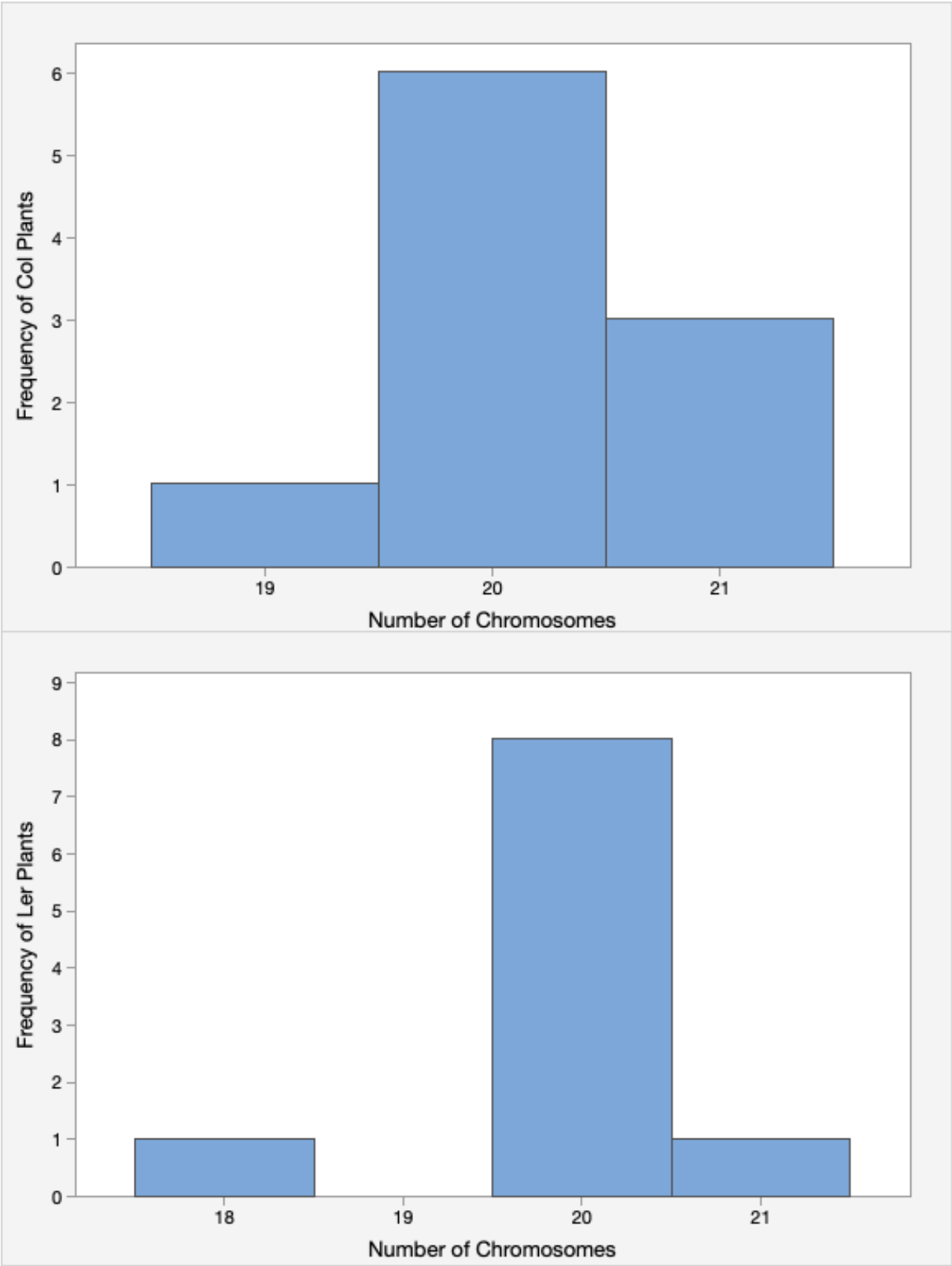


Figure 3-7 Chromosomal count in Parents, Columbia (Col) and Landsberg plants (Ler).

Arabidopsis thaliana tetraploid Columbia and Landsberg parental lines

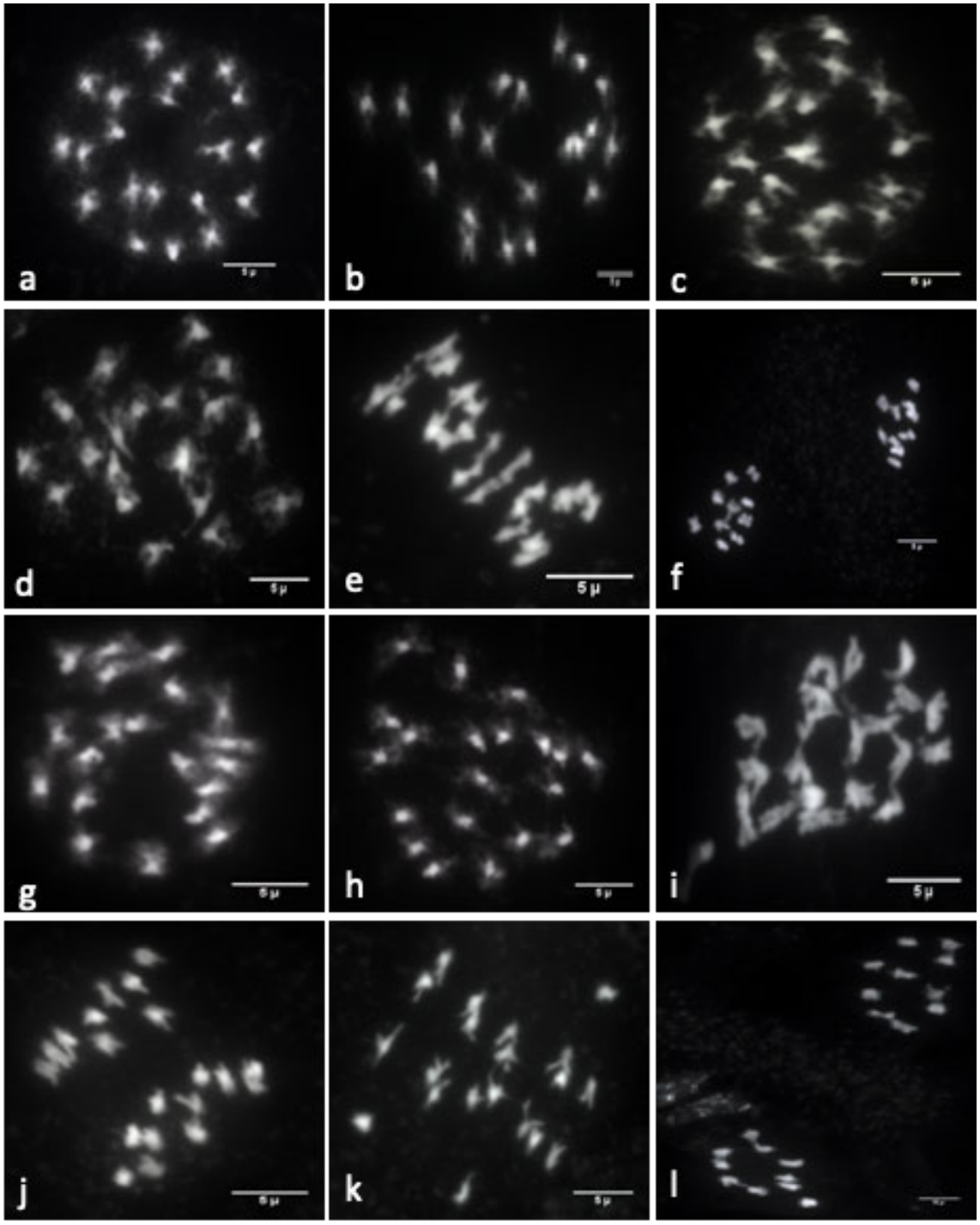


Figure 3-8 Tetraploid chromosome spreads of parental lines.

a-f Columbia chromosome spreads. a-d Mitotic Prophase, e Meiotic Anaphase I, f Meiotic late Metaphase II. g-l Landsberg chromosome spreads. g-h Mitotic Prophase, i Mitotic Metaphase, j-k Meiotic Anaphase I, l Late Meiotic Metaphase II. Scale bar is 5 μm.

Panels e and f in **Figure 3-8** show meiotic cells where separation of homologues is taking place in 'e' and metaphase II has been completed in 'f'. 20 chromosomes can be counted in both the cells. The other four panels (a,b, c, and d) show mitotic prophases. The highly condensed chromosomes can be seen as bright spots after they have been stained with DAPI. On counting these chromosomes, they are 20 in number and since the haploid number of chromosomes in *A. thaliana* is 5, it confirms that these Columbia cells are tetraploid (4n).

The three panels g, h, i, in **Figure 3-8** are the cells in different mitotic stages, where 20 distinct chromosomes can be counted. The panels j and k are meiotic anaphase I, where 10 bivalents are separating in individual homologues and thus 20 chromosomes can be counted. The cell in 'l' is in late metaphase II, where again 20 chromosomes can be counted. This confirms the tetraploid status of the Landsberg plants.

F1 tetraploid created by crossing the Landsberg and Columbia tetraploid parents

Different phases of mitosis and meiosis can be seen in **Figure 3-9**. 20 chromosomes can be counted in all the cells. In the first three panels (a, b and c), the mitotic prophase and metaphase can be seen, where 20 distinct chromosomes can be counted. In panels d and e, 10 distinct chromosomes can be seen in meiotic metaphase II on either side of the cytoplasm. Similarly, 10 bivalents are visible in the last meiotic M1 picture in panel f. At least 10 different cells were counted to confirm the tetraploidy in each F1 line.

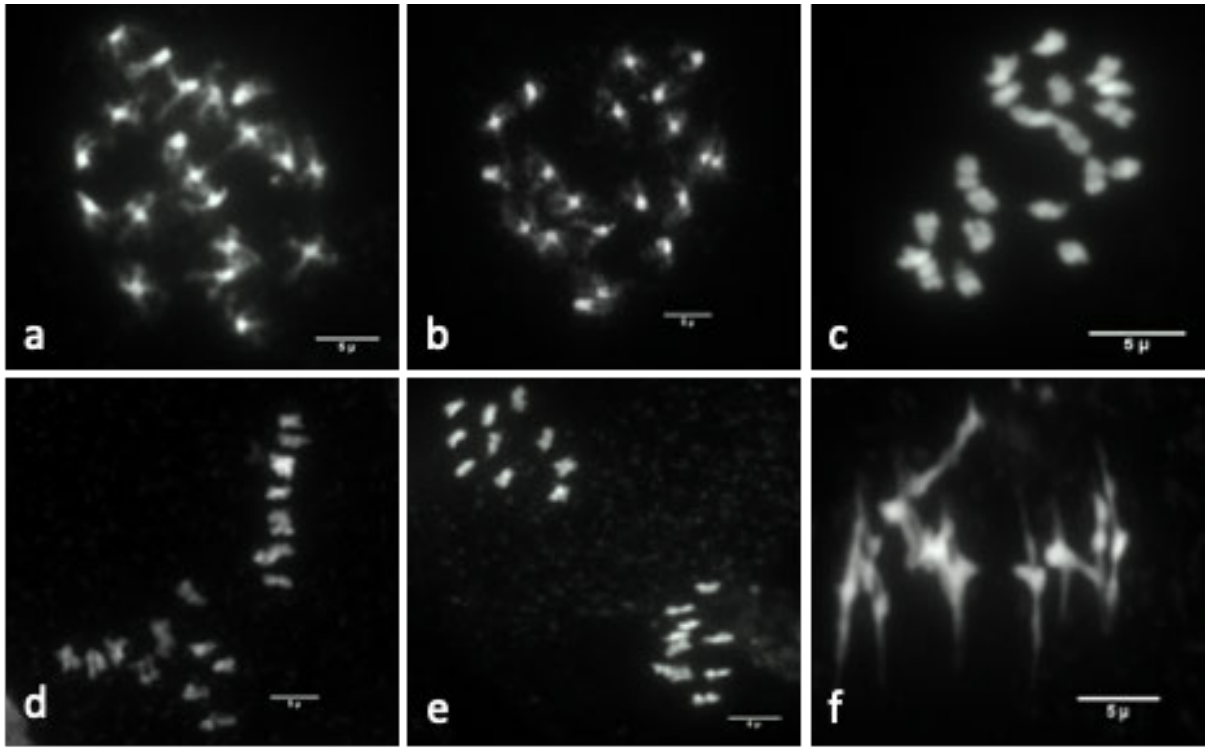


Figure 3-9 Chromosomal spreads of one of the F1 lines created by crossing Columbia and Landsberg parental lines.

Panels a, b and c depict mitotic cells, while d, e and f depict meiotic cells**.** Scale bar is 5 µm

3.2.2.2 Chromosome identification and counting

FISH using 5S and 45S rDNA probes along with chromosome morphology can help identify each chromosome in *Arabidopsis* accessions, both in diploid as well as tetraploid plants (**3.2.1**). It thus helps in confirming the correct number of chromosomes in a polyploid. In tetraploid *A. thaliana,* there are 20 chromosomes. This can be identified in a meiotic or in a mitotic cell. In a mitotic prophase cell, four of each chromosome should be counted, whereas in a metaphase II (M2) cell, 2 sets of each homologue of a chromosome should be seen on either side of the cytoplasm arranged on a spindle.

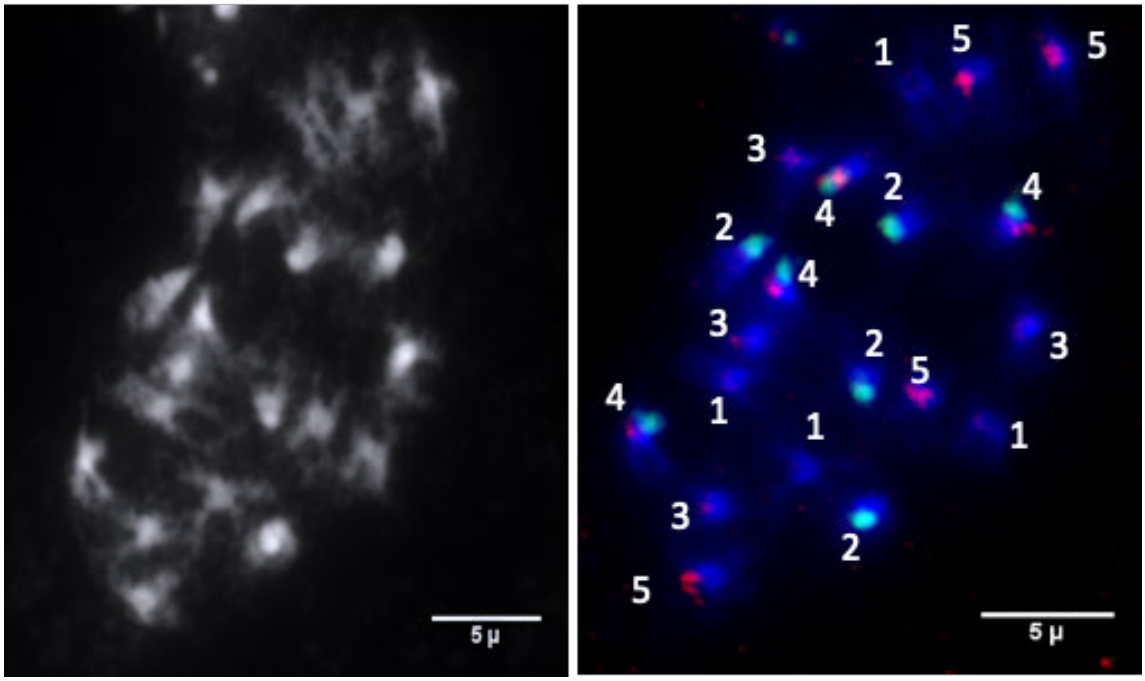


Figure 3-10 Mitotic prophase cell of Columbia tetraploid plant.

Left panel shows a cell stained with DAPI and the same cell with FISH probes on right panel. All the chromosomes can be counted based on the presence and absence of the FISH probes. The numbers in the right panel indicate the chromosome number. Green and red spots are 45S and 5S FISH probes respectively. Scale bar is 5 μm.

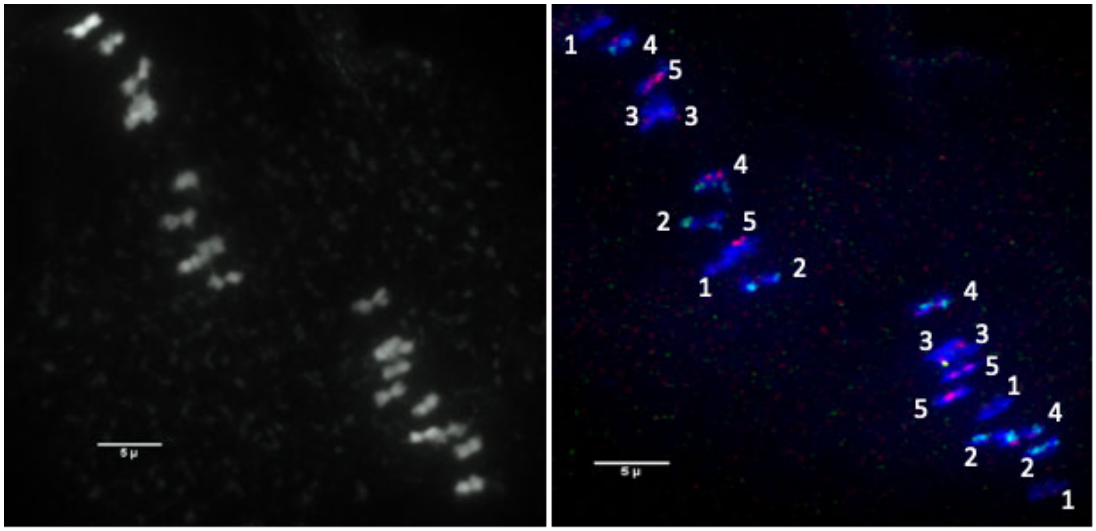


Figure 3-11 Metaphase II cell in a F1 plant.

Figure shows a M2 cell stained with DAPI (left) and the same cell with FISH probes (right) Green is 45S and red is 5S. Numbers indicate the chromosome number. Scale bar is 5 μm.

It can be seen in **Figure 3-10** that all four homologues of each chromosome can be identified and counted in a mitotic cell of the Columbia tetraploid parental plant. Similarly, **Figure 3-11** represents a meiotic cell in metaphase II in one of the F1 lines. 2 sets of each chromosome separated on two sides of the cytoplasm in the same cell can be clearly identified, confirming the chromosome count as well as correct number of each chromosome in the plant.

3.2.2.3 Chiasma analysis in parental lines

In *A. thaliana*, FISH using 45S and 5S rDNA probes can help in understanding the configurations of the bivalent chromosomes. The presence or absence of the signal firstly helps in identifying the chromosome number; secondly the presentation of the signal helps in identifying the bivalent configuration, and hence counting the chiasmata. Chiasma analysis was carried out in both parental lines but the same could not be performed in F1 lines because enough M1s were not found in the spreads prepared. A line drawing of different bivalents in a Landsberg tetraploid line in **Figure 3-12**, shows how the analysis is done by drawing out the chromosomes and inspecting the signals along with the arms and positions of centromeric attachment to the spindle (inferred). The points of CO are shown as crosses within the bivalents. All the chromosomes can be identified as bivalents in this tetraploid cell.

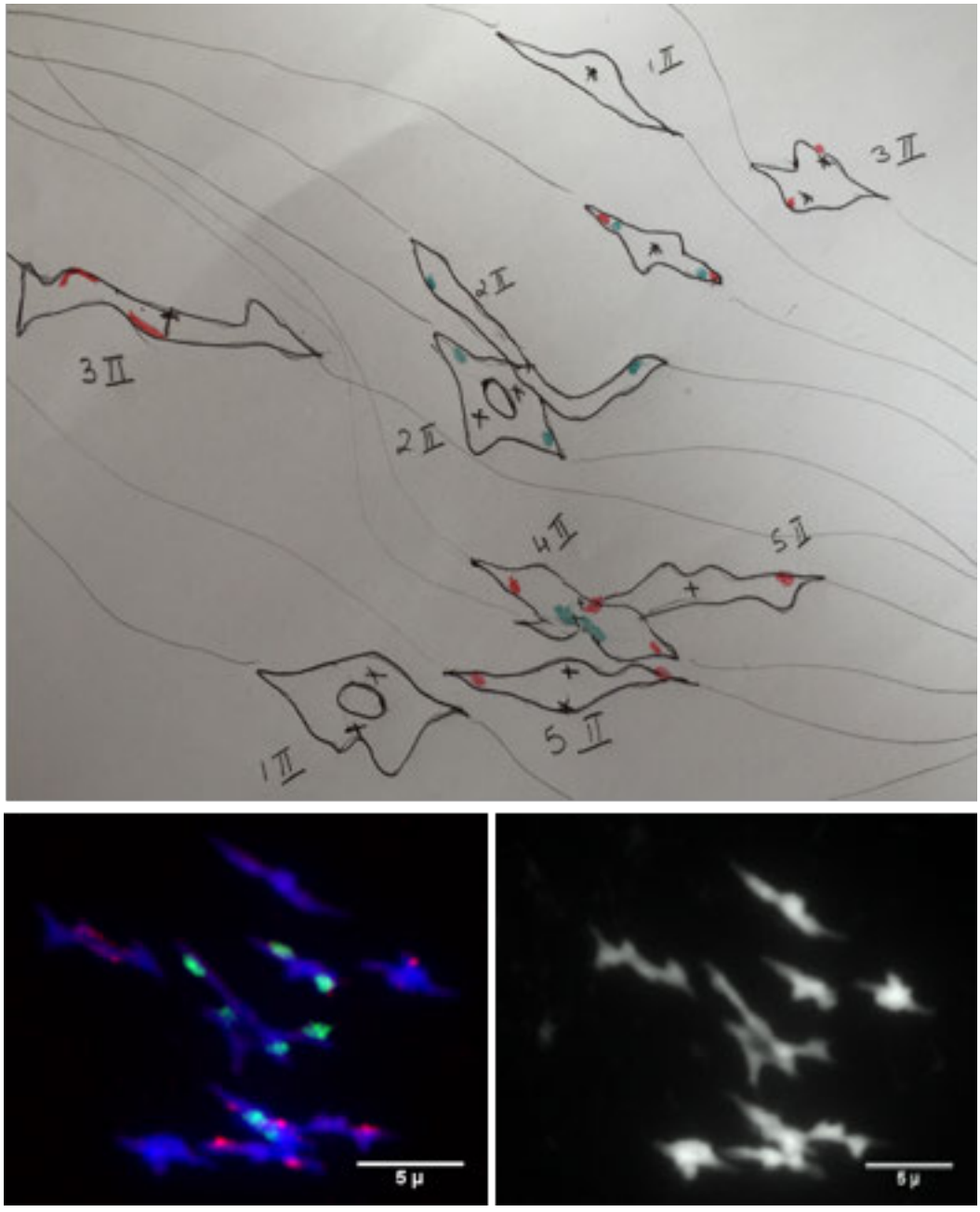


Figure 3-12 Representing chiasma count methodology.

Hand drawn bivalents of a tetraploid Landsberg plant showing 5S and 45S FISH signals and the corresponding DAPI and FISH pictures in meiotic M1. Scale bar is 5 µm.

3.2.2.3.1 Chiasma analysis in Columbia line

Chromosomal spreads of diploid as well as tetraploid *A. thaliana* ecotype Columbia, with DAPI stain and with 45S and 5S FISH probes are shown in **Figure 3-13**. It represents a meiotic cell in

M1 stage. In a diploid cell, five individual bivalents represent 10 homologues bound by chiasma in the top two panels, hence visible as five structures oriented on the poles.

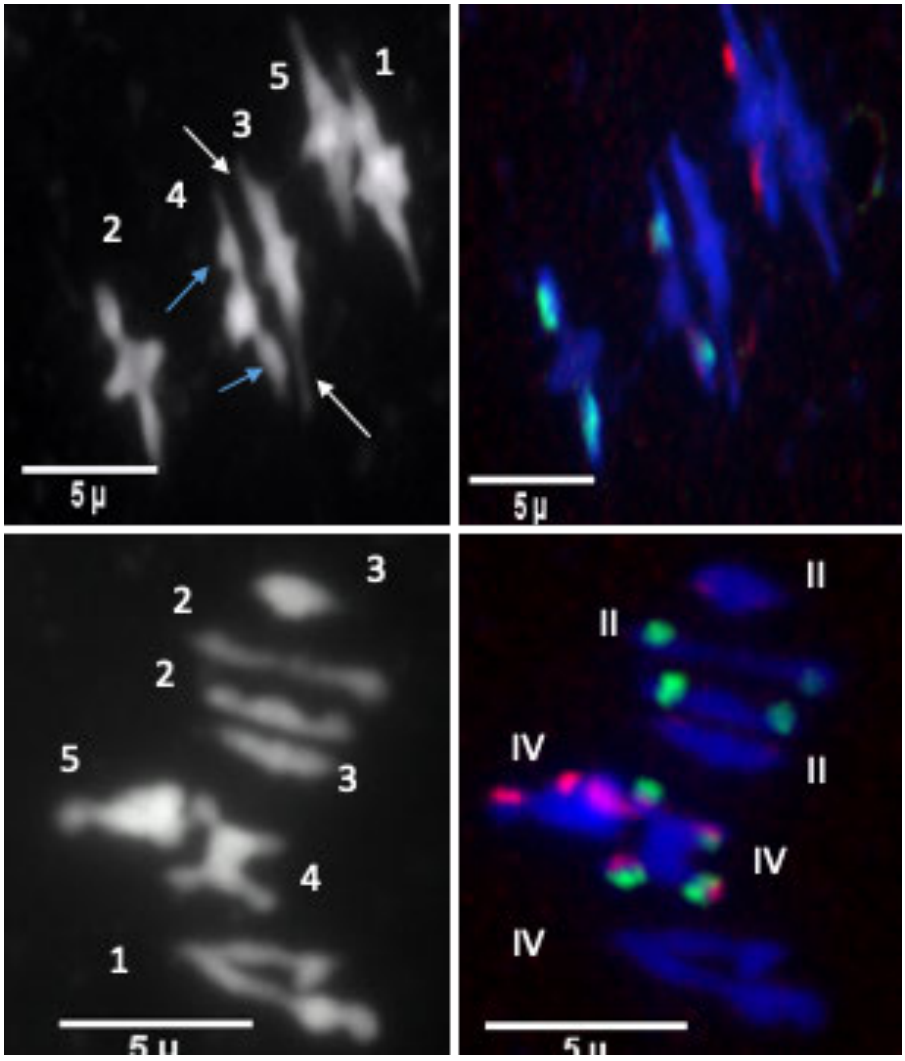


Figure 3-13 Comparison between a DAPI stained M1 cell (left) and the same cell (right) also showing 5S and 45S FISH probes in Columbia plants.

Figure represents cells in diploid (top) and tetraploid (bottom). Blue arrows indicate chromosome arms and white arrows indicate the centromere position. The numbers in the left panel are the chromosome numbers while the Roman numerals in the right panel indicate whether the chromosome is a bivalent II or quadrivalent IV. Scale bar is 5 μm.

The FISH signals help to identify the configurations. Chromosome 1 shows a ring bivalent thus having at-least two chiasmata, one in each arm. Chromosome 2 is a rod and shows two 45S signals, thus showing a chiasma in its long arm. Chromosome 3 is again a ring, though 5S

signals are not very clear, therefore we can estimate that there are two chiasmata. Chromosome 4 shows two 45S and two 5S signals, thus indicating chiasma formation in the long arm only. Chromosome 5 again shows a ring bivalent with 2 5S signals, thus it is bound by two chiasmata, one in each arm. Here, we can count 8 chiasmata in this metaphasic cell in diploid line. In this way, the chiasma analysis was performed for at-least 20-40 cells for each variety.

Chiasma configurations are complex in tetraploids. There are more than two homologous partners to pair with and exchange the genetic material. This can lead to the multivalent formation and some complex configurations. The bottom panel in **Figure 3-13** shows a DAPI stained spread from tetraploid Columbia along with FISH on the same spread using 5S and 45S probes. Chromosome 1, which has no signals shows a chain quadrivalent with 3 chiasmata. Chromosome 2 has two bivalents lying side by side each bound by one chiasma. Chromosome 3 also shows two bivalents, both being ring bivalents bound by two chiasmata each. Chromosomes 4 and 5 are quadrivalents bound by four and three chiasmata respectively. A total of 16 chiasmata are present in this metaphasic cell.

3.2.2.3.2 Chiasma analysis in Landsberg line

Chromosomal spreads of Landsberg ecotype of *A. thaliana* were prepared to analyse their chiasma configurations. FISH using 45S and 5S probes was performed with the only difference in the 5S signal in chromosome 3. While in Columbia, 5S signal is visible in the short arm close to the centromere, in Landsberg the 5S site is situated interstitially in the long arm (Sanchez-Moran *et al.*, 2002).

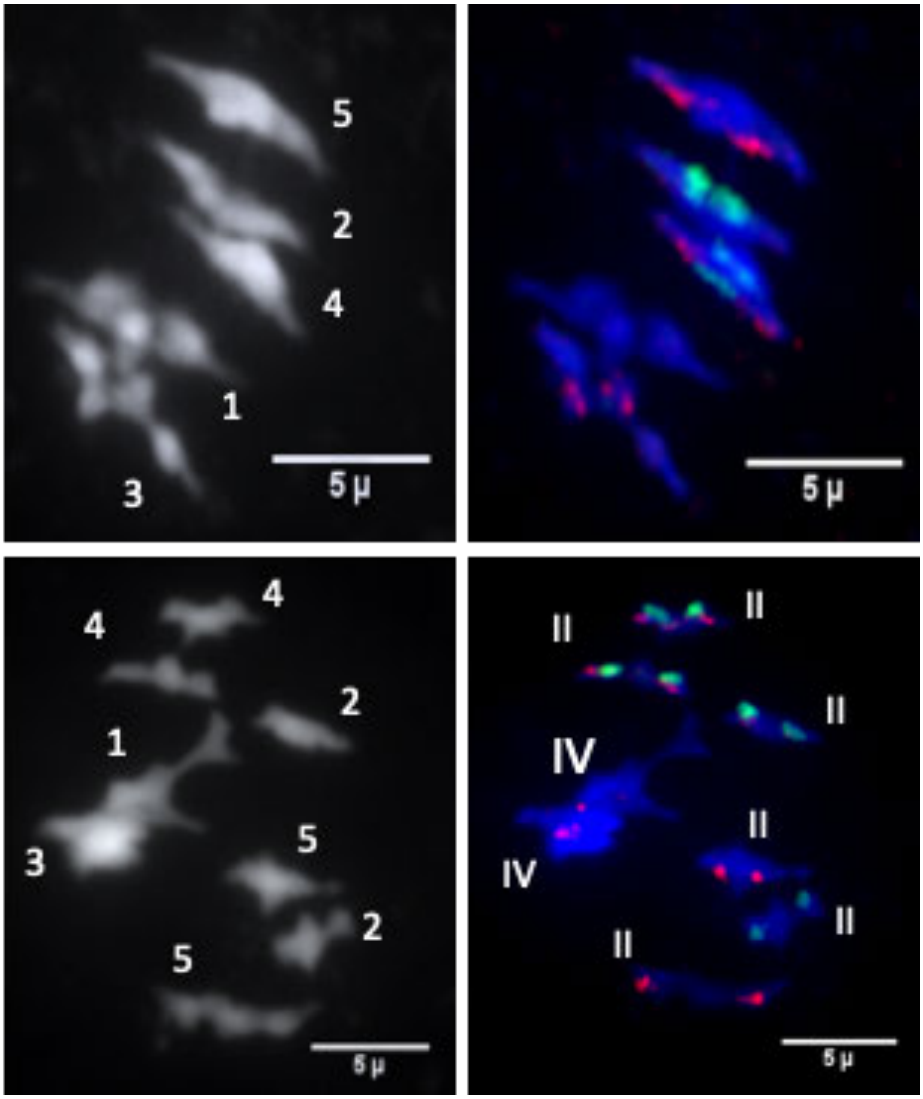


Figure 3-14 Comparison between a DAPI stained cell (left) and the same cell (right) also showing 5S and 45S FISH probes in Landsberg plants.

Figure shows cells in M1 in diploid (top) and tetraploid (bottom). The numbers in the left panel are the chromosome numbers while the Roman numerals in the right panel indicate whether the chromosome is a bivalent II or quadrivalent IV. Scale bar is 5 µm.

Figure 3-14 (top panel), shows the chromosomal spreads of diploid Landsberg, stained with DAPI on the left and with FISH using 45S and 5S probes on the right. Chromosome 1 shows a bivalent, with two chiasmata in the short arms, with a knob and free long arms clearly visible. Chromosome 2 shows a ring bivalent with two chiasmata, one in each arm. Chromosome 3 shows two interstitial 5S signals and a bivalent with two chiasmata in the longer arms.

Chromosome 4 is a ring bivalent hence bound by a chiasma each in both the arms. Chromosome 5 again is a ring bivalent bound by both the arms, hence shows two chiasmata. Thus, we can count 10 chiasmata in this metaphasic cell.

Figure 3-14 (bottom panel) shows a DAPI stained spread along with FISH on the same spread using 5S and 45S probes on tetraploid Landsberg *A. thaliana*. Chromosome 1 presents a chain quadrivalent with 3 chiasmata. Chromosome 2 shows two rod bivalents, with one chiasma each. Chromosome 3 again shows a chain quadrivalent with 3 chiasmata. There are two rod bivalents for chromosome 4, and one rod and another ring bivalent for chromosome 5. In this way, there is a total of 13 chiasmata present in this M1 cell.

In **Figure 3-13** (bottom panel), chromosome 4 is present as a ring quadrivalent while chromosomes 1 and 3 occur in a chain or linear formation. In **Figure 3-14** (bottom panel), chromosomes 1 and 3 occur in a chain formation with three chiasmata, where 2 bivalents seem to be interconnected by the unbound arm in their respective bivalents.

Chiasma scoring was carried out for both diploid and tetraploid Landsberg and Columbia parental lines for each chromosome and is presented in **Table 3-1**. It can be seen that the chiasma count for the longer arm is greater than the shorter arm, for all the chromosomes in all the four varieties. Chromosome 1 shows the highest chiasma frequency for all the varieties, followed by chromosome 5. This is followed by chromosome 3 except in 4n Columbia where chromosome 4 has a higher chiasma frequency than chromosome 3.

	Chr I (mean)	Chr II (mean)	Chr III (mean)	Chr IV (mean)	Chr V (mean)	Total
4n Ler	3.39	2.96	3.09	2.74	3.17	15.35
(n=23)	(1.39 2)	(1.09 1.87)	(1.09 2)	(0.87 1.87)	(1.13 2.04)	
4n Ler	3.44 (n_1=16)	3.23 (n_2=13)	3.36 (n_3=11)	3 (n_4=5)	3.33 (n_5=15)	-
(mult)	(1.44 2)	(1.46 1.77)	(1.36 2)	(1.4 1.6)	(1.33. 2)	
4n Ler	3.29 (n_1=7)	2.6 (n_2=10)	2.83 (n_3=12)	2.71 (n_4=17)	2.71 (n_5=7)	-
(biv)	(1.29 2)	(0.6 2)	(0.83 2)	(0.71 2)	(0.71 2)	
2n Ler	1.91	1.09	1.65	1.39	1.91	7.96
(n=23)	(0.87 1.04)	(0.09 1)	(0.65 1)	(0.48 0.91)	(0.87 1.04)	
4n Col	3.74	2.95	3.18	3.38	3.53	16.74
(n=39)	(1.74 2)	(1.32 1.92)	(1.26 1.92)	(1.4 2.03)	(1.55 1.97)	
4n Col	3.91 (n_1=23)	3.63 (n_2=16)	3.13 (n_3=15)	3.78 (n_4=18)	3.53 (n_5=17)	-
(mult)	(1.91 2)	(1.75 1.88)	(1.33 1.8)	(1.72 2.06)	(1.53 2)	
4n Col	3.5 (n_1=16)	2.55 (n_2=22)	3.25 (n_3=24)	3.05 (n_4=21)	3.52 (n_5=21)	-
(biv)	(1.5 2)	(0.55 2)	(1.25 2)	(1.05 2)	(1.52 2)	
2n Col	1.96	1.62	1.58	1.42	1.71	8.28
(n=52)	(0.90 1.06)	(0.60 1.02)	(0.58 1)	(0.42 1.0)	(0.73 0.98)	

Table 3-1 Mean chiasma frequency for each chromosome in Metaphase I stage.

Ler represents Landsberg and Col represents Columbia. 2n and 4n represents diploid and tetraploid respectively. The numbers in the parenthesis indicate the mean chiasma count for the short and long arm of the chromosome respectively. The number of cells analysed is given by n below the variety. The three rows for each tetraploid line shows mean chiasma count for total cells, extracted bivalents only and extracted multivalents only cells. n_1, n_2, n_3, n_4 and n_5 shows the number of cells for chromosomes 1, 2, 3, 4 and 5 respectively when bivalents only and multivalents only cells are separated out.

The chiasma count data was tested for normality and all the four sample counts violated the normality assumption as seen in **Figure 3-15**. Parametric tests were therefore considered unsuitable for the analysis of differences between varieties. Non parametric tests were then performed for subsequent analysis.

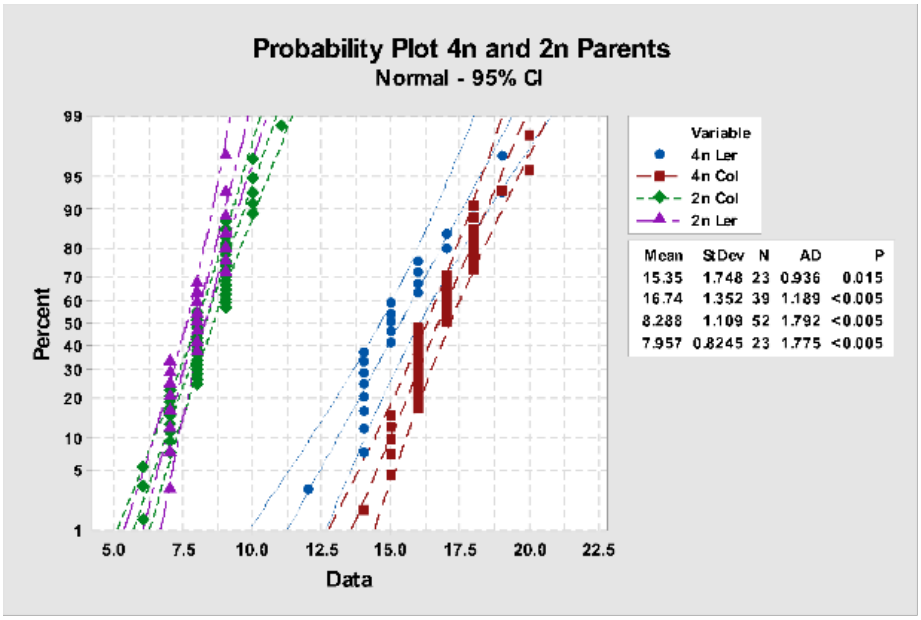


Figure 3-15 Normality test for chiasma count data in four parental lines.

The Kruskal-Wallis test showed a significant difference in the observed chiasma count between different lines ($\chi^2_{df=3}$, p-value < 2.2e-16). Post hoc Dunn test with Bonferroni corrections showed a significant difference between diploid and tetraploid Columbia (p-value < 2e-16) as well as diploid and tetraploid Landsberg lines (p-value = 2.6e-07). Columbia and Landsberg tetraploids were also significantly different from Landsberg and Columbia diploids (p-value = 1.2e-13, p-value = 5.7e-08). No significant difference occurred between the diploid lines as well as the tetraploid lines. The distribution of the crossovers across the short and long arms was found to be significantly associated with the genotype only for chromosome 2 ($\chi^2_{df=3}$ = 34.91, p-value < 0.0001). For all other chromosomes, it was independent of the genotype. Though, the difference between the diploids and tetraploids was found to be significant, it can be seen from **Table 3-1** that the overall crossover frequency is just about doubled in tetraploids as compared with diploids, except for chromosome 2 in Landsberg where it has reached 2.7 times that of diploid in tetraploid.

A per bivalent chiasma comparison between different parental lines was carried out by extracting cells in tetraploid parents, which had two bivalents (**Table 3-1**). Now a direct comparison of per bivalent chiasma frequency in tetraploids (obtained by dividing the chiasma count for each cell with 2 bivalents by 2) could be made with the diploids for each chromosome, as the cell count for each chromosomes with two bivalents in tetraploids was different. Kruskal-Wallis test produced a significant difference in the per bivalent chiasma count frequency between different lines for chromosomes 1 ($\chi^2_{df=3}$, p-value = 0.0001), 2 ($\chi^2_{df=3,}$ p-value = 5.039e-05) and 5 ($\chi^2_{df=3,}$ p-value = 0.0041). Post hoc Dunn test showed the significant difference between the different lines as shown in **Table 3-2**. The per bivalent chiasma frequency is significantly lower in tetraploid Landsberg than the diploid Landsberg as well as diploid Columbia in chromosomes 1 and 5. It is also significantly lower in tetraploid Columbia line than the diploid Columbia line for chromosome 1 and borderline significantly lower for chromosome 2. Though no significant difference was seen for total chiasma count between the two diploid lines, the chiasma count frequency for chromosome 2 in the Landsberg diploid was significantly lower than the Columbia diploid line.

		Ler 4n	**Ler 2n**	**Col 4n**
Ler 2n	Chr 1	**0.0100**	-	-
	Chr 2	0.956	-	-
	Chr 5	**0.0017**	-	-
Col 4n	Chr 1	1.0000	0.1086	-
	Chr 2	1.0000	0.701	-
	Chr 5	0.0594	0.8575	-
Col 2n	Chr 1	**0.0017**	1.000	**0.0135**
	Chr 2	0.490	**3.6e-05**	**0.053**
	Chr 5	**0.0265**	0.5538	1.0000

Table 3-2 Post hoc Dunn test after Kruskal-Wallis in parents.

The test is for difference in per bivalent chiasma frequency. Ler 2n and Ler 4n are diploid and tetraploid Landsberg, Col 2n and Col 4n are diploid and tetraploid Columbia.

The number and percentage of bivalents, quadrivalents, trivalents and univalents observed in tetraploid parental lines can be seen in **Table 3-3**. For both Columbia and Landsberg plants, chromosome 1 shows the highest number of multivalents followed by chromosome 5 in Landsberg but chromosome 4 in Columbia. Chromosomes 2 and 4 show higher number of multivalents than chromosome 3 in Columbia. In Landsberg, chromosome 2 shows higher number of multivalents than chromosome 3 and chromosome 4 shows by far the lowest. All the chromosomes except chromosome 3 in Columbia showed more closed quadrivalents than the open quadrivalents, while in Landsberg, chromosomes 2, 3 and 4 showed more closed quadrivalents. Both Columbia and Landsberg showed higher overall percentages of closed quadrivalents than open quadrivalents (62.6% vs 35% in Col, 52% vs 42% in Ler). Overall, Landsberg showed a higher percentage of quadrivalents than Columbia (**Table 3-3)**. Occasional univalents were also found in diploid Landsberg. Out of the 23 cells analysed for diploid Landsberg parental line, one cell had chromosome 4 as two univalents.

The two sample proportion test comparing the proportion of multivalent pairing between the two tetraploid parents for each chromosome was carried out. It did not differ significantly except for chromosome 4 between the Columbia and Landsberg (chr 1 (Z value= -0.83, p-value = 0.40), chr 2 (Z value = -1.18, p-value = 0.24), chr 3 (Z value = -1.34, p-value = 0.18), **chr 4 (Z value = 2.29, p-value = 0.02**), and chr 5(Z value = -1.65, p-value = 0.1)). This indicates except for chromosome 4, where Landsberg formed significantly lower multivalents than Columbia, the parents do not differ in pairing pattern for other chromosomes.

Genotype	Chr	2II n (%)	1IV n (%)	1III + 1I n (%)	1II + 2I n (%)	IV Ring n (%)	IV Chain n (%)	IV Other n (%)	n
Col	1	16 (41)	23 (59)	0 (0)	0 (0)	19 (82.6)	4 (17.4)	0 (0)	39
Col	2	22 (56.4)	15 (38.5)	1 (2.6)	1 (2.6)	9 (60)	5 (33.33)	1 (6.66)	39
Col	3	24 (61.5)	12 (30.8)	3 (7.7)	0 (0)	5 (41.7)	7 (58.3)	0 (0)	39
Col	4	21 (54)	18 (46.2)	0 (0)	0 (0)	13 (72.2)	4 (22.2)	1 (5.6)	39
Col	5	21 (54)	16 (41)	1 (2.6)	1 (2.6)	9 (56.25)	7 (43.75)	0 (0)	39
Col Avg n (%)	-	21 (53.38)	16.8 (43.1)	5 (2.58)	0.4 (1.03)	- (62.6)	- (35.0)	- (2.58)	39
Ler	1	7 (30.4)	16 (69.6)	0 (0)	0 (0)	7 (43.75)	9 (56.25)	0 (0)	23
Ler	2	10 (43.5)	10 (43.5)	3 (13)	0 (0)	6 (60)	3 (30)	1 (10)	23
Ler	3	12 (52.2)	11 (47.8)	0 (0)	0 (0)	6 (54.5)	5 (45.5)	2 (18.2)	23
Ler	4	17 (74)	3 (13)	2 (8.7)	1 (4.3)	2 (66.66)	1 (33.33)	0 (0)	23
Ler	5	7 (30.4)	15 (65.2)	0 (0)	1 (4.4)	5 (33.33)	10 (66.66)	0 (0)	23
Ler Avg n (%)	-	10.6 (46.1)	11 (47.8)	1 (4.3)	0.4 (1.7)	- (51.65)	- (46.35)	- (3.54)	23

Table 3-3 Chromosome configurations in Metaphase I in Columbia and Landsberg tetraploid parental lines.

Table shows the number of cells showing chromosome configurations in M1. II indicates bivalents, IV indicates quadrivalents, III indicates trivalents and I indicate univalents. The percentages are shown in parentheses.

A two sample proportion test was also carried out for comparing multivalent proportions between each chromosomes within each parental tetraploid variety (**Table 3-4**). No significant difference can be seen between chromosomes in Columbia (p-value > 0.05). However, chromosome 4 multivalent formation is significantly lower than chromosome 1 as well as chromosome 5 (p-value < 0.05, after accounting for Bonferroni corrections) in Landsberg.

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5
Chr 1 Col	-				
Ler					
Chr 2 Col	1.59, 0.113	-			
Ler	0.92, 0.359				
Chr 3 Col	1.81, 0.070	0.23, 0.817	-		
Ler	1.50, 0.134	0.59, 0.555			
Chr 4 Col	1.13, 0.257	-0.46, 0.648	-0.69, 0.492	-	
Ler	**3.26, 0.001**	2.42, 0.016	1.86, 0.063		
Chr 5 Col	1.36, 0.174	-0.23, 0.819	-0.46, 0.645	0.23, 0.819	-
Ler	0.31, 0.753	-0.60, 0.546	-1.19, 0.234	**-2.97, 0.003**	

Table 3-4 Results of a 2 sample proportion test for multivalents in parents.

The test is for within chromosomal differences of multivalent proportions in Columbia and Landsberg lines. Z value followed by p value is reported.

3.3 Discussion

To create an F2 mapping population, two different tetraploid and diploid *Arabidopsis* lines were crossed. Columbia and Landsberg were chosen as they differed in several phenotypic traits as described in the introduction (**1.5.1**). A comprehensive marker information data for diploids is available from several studies for the two lines. This made them a suitable choice for our study. Genotyping using SSLP markers along with the Sanger sequencing of the marker sequences confirmed the heterozygosity of the F1 lines created, which were used for creation of the F2 population.

Crossovers between homologues during meiosis helps in shuffling the genes, creating new combinations of alleles and new genotypes. Several genes involved in the process have been identified and various studies have enhanced the understanding of the process. However, a similar understanding of the crossovers and meiotic processes is required for the polyploid species, considering various important food crops are polyploid. Methods to understand the crossing over and processes related with it have been established for the diploid model plant *A. thaliana*. However, the effects of polyploidy on the process have not been studied thoroughly, with only few studies exploring the behavior of meiosis in polyploids. Pecinka *et al.* (2011) analysed the meiotic recombination frequency in *Arabidopsis* diploids vs auto and allo tetraploids using fluorescent seed markers in a specific chromosomal segment. They found an increase in the meiotic recombination frequency in both auto and allo tetraploids as compared with diploids. Another autotetraploid where a few studies about polyploid behaviour has been made is *Arabidopsis arenosa.* Chiasma frequency has been found to be

higher in diploids than tetraploid *A. arenosa* (Yant *et al.*, 2013) and it has been attributed to increased interference across chromosomes in tetraploids (Bomblies *et al.*, 2016). Here, we have used chiasma analysis using FISH probes to analyse the differences between the crossover frequencies in diploids vs tetraploids *A. thaliana.*

FISH analysis for chiasma counting in parental lines was carried out on the plants in which an initial chromosome count of 20 was established using DAPI staining. For Columbia parents, 4 out of 10 plants had counts of either 19 or 21. A count of 19 may indicate the loss of one chromosome, aneuploidy, or it could also be two chromosomes lying on top of each other in the mitotic or meiotic cell, thus making it difficult to count 20. However, where a count of 21 is found, it indicates the gain of a chromosome or trisomy. For the Landsberg parent, 1 out of 10 plants gave a count of 18 chromosomes, which can indicate either one of the chromosomes has not doubled up or there is a loss of two different chromosomes. It can be interesting to analyse which chromosomes have been lost or gained in the tetraploids. This can give us an insight about the differences that exist between the chromosomes which can lead to the loss or gain of a particular chromosome. However, due to time constraints, only those plants where the full complement of 20 chromosome could be counted were analysed using FISH for chiasma scoring.

Crossover frequency using chiasma analysis has been compared between eight diploid *Arabidopsis* accessions, where a mean chiasma count of 9.10 for Columbia and 8.70 for Landsberg line and a significant difference between them was established based on analysis of 50 cells (Sanchez-Moran *et al.*, 2002). In our study, mean chiasma count for diploid

Columbia was found out to be 8.28 and for diploid Landsberg was found out to be 7.96. There was no significant difference between the Landsberg and Columbia lines except in chromosome 2. The difference between the mean chiasma count in diploids observed here compared with the previous literature could be due to the conservative scoring method or due to fewer cells being analysed here.

For tetraploids, the mean chiasma count was found to be 16.74 for Columbia and 15.35 for Landsberg in our study. This is again different from the chiasma frequency observed previously in the established lines of autotetraploid Columbia, which was found to be in a range from 17.94 to 18.80 for four different established lines (Santos *et al*., 2003). The reasons for this difference could be similar to those explained for diploids above. Many cells were lost in carrying out FISH experiments where 45S and 5S probes did not work. Due to time constraints it was not possible to grow more plants and collect more buds and therefore, the number of cells analysed was less than 50. Environmental differences may have also played a role. Though the temperature is maintained in the glasshouse, fluctuations do occur in peak winter or summer. These differences in the environment can affect the crossover frequency (Börner, Kleckner and Hunter, 2004). An increase in crossover frequency was seen in *A. thaliana* when the plants normally grown at 20 °C were shifted to 28 °C, through modulation of Class 1 crossovers (Modliszewski *et al.*, 2018). However, this is unlikely to explain the differences we observed as temperature usually exceeded 20 °C in the glasshouse, therefore our estimates might have been expected to be higher.

A significant difference in chiasma frequency was found between the diploid and tetraploid parental lines, however the magnitude of the difference did not exceed two fold (**Table 3-1**). Doubling of crossovers in a tetraploid may be expected based on the doubling of the chromosomes, considering that an obligate crossover is formed between the homologues. However, this is a "naïve" expectation which would assume the homologues to pair only as bivalents in the tetraploids, which is not the case. Any increment beyond doubling may be considered as a substantial increase due to tetraploidy. This shows that there was no substantial increase in the number of crossovers in tetraploids beyond a simple doubling, except for chromosome 2 in Landsberg tetraploids where a fold change of 2.7x was observed. This increase in crossover frequency cannot only be accounted for by increased quadrivalent formation. This might indicate that polyploidisation may play a role in increasing the crossovers in the small chromosomes, may be by modulating the associated chromatin structure giving access to the recombination machinery to operate. Thus, it may serve as a means for local meiotic recombination manipulation rather than a global phenomenon. The distribution of the crossovers across the arms was found to be significantly associated with the genotype for chromosome 2. Previously, Sanchez Moran *et al*. (2002) have found chromosome 2 chiasma frequency to be more variable between the wild type and mutant plants than chromosome 4 in *A. thaliana,* ascribing it to chromatin states associated with rDNA transcriptional levels. Here as well, chromosome 2 was found to be more variable between the varieties. In both diploids and tetraploid lines, higher chiasma frequency was found for longer chromosomes 1, 5 and 3; and lower for the smaller chromosomes 2 and 4, with an exception of 4n Columbia line.

A per bivalent crossover frequency comparison between diploids and tetraploids illustrated a significant reduction in crossovers in tetraploid Landsberg line for chromosomes 1 and 5 and chromosomes 1 and 2 in Columbia. Rest of the other chromosomes in both tetraploid lines showed a slight decrease or a slight increase which was not significant. This indicates a trend towards reduced crossover formation in tetraploids, may be to stabilise segregation pattern. Noteworthy is the reduction in longer chromosomes 1 and 5 indicating genetic factors playing to stabilise segregation patterns in autotetraploids.

In tetraploids, complex multivalent configurations can be observed as rings, chains, linear, or saucepan, where all four or three chromosomes are bound, or as trivalent where three chromosomes are chiasma bound and one chromosome is present as a univalent. In the case of a ring tetravalent, at-least 4 chiasmata are present, leading to a closed conformation. It can be considered as two rod bivalents which are connected to each other through their unbound arms. In a chain and linear configurations, the four homologues are connected in a chain formation either through the short or the long arms and are therefore open conformations. A saucepan looks like a ring bivalent connected with a rod bivalent, which is a complex pattern. The kind of quadrivalents present in both the parental lines seem to follow same pattern of more closed than open quadrivalents except for chromosome 1, where Landsberg parent had more open quadrivalents than the Columbia parent. More open quadrivalents might indicate diploidisation of chromosomes. Though multivalent formation leads to an increase in the crossover frequency in a plant, but they can also lead to mis-segregation and thus disturb the chromosomal balance. However, despite a substantial proportion of closed quadrivalents observed here, it can be noted that mis-segregation of the chromosomes was observed only

rarely overall (5%). Cytological diploidisation of chromosomes can lead to proper segregation, ensuring fertility of the plant. A decrease in chiasma frequency through increased crossover interference leading to complete diploidisation, has been suggested as a mechanism of successful establishment of autotetraploid *Arabidopsis arenosa* (Bomblies *et al.,* 2016).

Both the tetraploid lines look like they have already started to diploidise. This is evident with high number of bivalent formations (~ 54% and 46% averaged across all chromosomes across for Columbia and Landsberg respectively). The bivalent formation varied from 41 to 61.5% across the five chromosomes in Columbia, with 41% for chromosome 1 and 61.5% for chromosome 3. In Landsberg, it varied from 30 to 74%, with 30.4% for chromosome 1 and 74% for chromosome 4. Landsberg had more multivalents than Columbia, though the genotype effect was only significant for chromosome 4 and the chromosome effect was also only significant for chromosome 4 in Landsberg. The data observed here, show the differences between genotypes and chromosomes, inferring that bivalent and multivalent formation is genotype and chromosome dependent, though statistical significant testing proves otherwise, may be because of low number of cells considered. Line dependent multivalent formation and partial diploidisation has been shown before (Santos *et al.*, 2003). They also showed a reduction in multivalent formation across the chromosomes based on their size, with shortest chromosomes 2 and 4 showing lowest multivalent formation, bigger chromosomes 3 and 5 varying across lines and the largest chromosome 1 having the highest proportion of multivalents. In our experiments, chromosome 3 in Columbia shows the lowest percentage of multivalents (38.5%), though it is bigger in size than chromosome 2 and 4. Chromosome 2 and 4 are nearly the same size, but chromosome 2 showed lower multivalent frequency than

chromosome 4. Compared to Columbia, in Landsberg, chromosome 4 showed the lowest percentage of multivalent formation (21.7%), which is significantly different from chromosome 1 and 5 both (**Table 3-4**). Chromosome 1 shows the highest number of multivalents, while chromosome 2 which is similar in size to chromosome 4 shows higher multivalent formation than chromosome 3. This result is similar as well as different to those of Santos *et al*. (2003), where marked reductions in multivalent formation were seen in established autotetraploid lines for chromosomes 2 and 4, however in our study these smaller chromosomes show lower multivalent formation than chromosomes 1 and 5 but not chromosome 3 in Columbia.

After the heterozygosity confirmation of the F1 plants for both diploid and tetraploid variety and the chromosome counting of the parental and F1 plants was established, F2 seeds from confirmed tetraploid F1s were used further to make the F2 population required for the plant trial collecting various phenotype traits.

3.4 References

Bell, C. J. and Ecker, J. R. (1994) 'Assignment of 30 microsatellite loci to the linkage map of Arabidopsis', *Genomics*, 19(2), pp. 137–144. doi: 10.1006/geno.1994.1023.

Bomblies, K. *et al.* (2016) 'The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role?', *Chromosoma*, 125(2), pp. 287–300. doi: 10.1007/s00412-015-0571-4.

Börner, G. V., Kleckner, N. and Hunter, N. (2004) 'Crossover/Noncrossover Differentiation, Synaptonemal Complex Formation, and Regulatory Surveillance at the Leptotene/Zygotene Transition of Meiosis', *Cell*, 117(1), pp. 29–45. doi: 10.1016/S0092-8674(04)00292-2.

Modliszewski, J. L. *et al.* (2018) 'Elevated temperature increases meiotic crossover frequency via the interfering (Type I) pathway in Arabidopsis thaliana', *PLoS Genetics*. doi: 10.1371/journal.pgen.1007384.

Moran, E. S. *et al.* (2001) 'Chiasma formation in Arabidopsis thaliana accession Wassileskija and in two meiotic mutants', *Chromosome Research*, 9(2), pp. 121–128. doi: 10.1023/A:1009278902994.

Pacurar, D. I. *et al.* (2012) 'A collection of INDEL markers for map-based cloning in seven Arabidopsis accessions', *Journal of Experimental Botany*, 63(7), pp. 2491–2501. doi: 10.1093/jxb/err422.

Pecinka, A. *et al.* (2011) 'Polyploidization increases meiotic recombination frequency in Arabidopsis', *BMC Biology*, 9(1), pp. 24. doi: 10.1186/1741-7007-9-24.

Sanchez-Moran, E. *et al.* (2002) 'Variation in chiasma frequency among eight accessions of Arabidopsis thaliana', *Genetics*, 162(3), pp. 1415–1422.

Santos, J. L. *et al.* (2003) 'Partial Diploidization of Meiosis in Autotetraploid', *Genetics*, 165(3), pp. 1533–1540.

Sybenga, J. (1975) 'The Analysis of Crossing-over', in *Meiotic Configurations: A Source of Information for Estimating Genetic Parameters*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 13–133. doi: 10.1007/978-3-642-80960-6_2.

Tautz, D. (1989) 'Hypervariabflity of simple sequences as a general source for polymorphic DNA markers', *Nucleic Acids Research*, 17(16), pp. 6463–6471. doi: 10.1093/nar/17.16.6463.

Weiss, H. and Maluszynska, J. (2000) 'Chromosomal rearrangement in autotetraploid plants of Arabidopsis thaliana.', *Hereditas*, 133(3), pp. 255–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11433970.

Yant, L. *et al.* (2013) 'Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa', *Current Biology*. Elsevier Ltd, 23(21), pp. 2151–2156. doi: 10.1016/j.cub.2013.08.059.

CHAPTER 4

COMPARATIVE PHENOTYPIC ANALYSIS OF DIPLOID VS TETRAPLOID *ARABIDOPSIS THALIANA*

4 Comparative phenotypic analysis of diploid vs tetraploid *Arabidopsis thaliana*

4.1 Introduction

Most of the natural variations that exist between individuals of the same species are considered to be of polygenic origin. Not only the environmental effects govern these variations but there is an interaction between the genes/alleles affecting the trait as well. Most of the crop production and breeding and fitness traits are complex or continuous in nature (Holland, 2007). QTL mapping using marker linkage analysis is one way to identify the region of the genome responsible for the quantitative trait variation.

QTLs identifying various morphological and physiological traits have been identified in the model plant *Arabidopsis thaliana* (hereafter referred to as *A. thaliana* or *Arabidopsis*), as well as in crop plants such as tomato, wheat and rice. QTL affecting eight floral traits were mapped in an *A. thaliana* RIL population derived from Columbia and Landsberg parents (Juenger, Purugganan and Mackay, 2000). Another example is the fine mapping of FRI and FLC loci controlling flowering time, which was mapped using a linkage disequilibrium based association analysis in *Arabidopsis* using 196 accessions (Bevan and Walsh, 2005). Gene *DOG1*, involved in natural variation in seed dormancy was identified by QTL studies in *A. thaliana* (Alonso-Blanco *et al.*, 2009). Similarly, different QTLs for seed dormancy control and germination, flowering time, plant architecture, vegetative growth and physiology have been identified in *Arabidopsis* and also in cereals, rice, pea and lettuce (Alonso-Blanco *et al.*, 2009).

A number of major effect QTLs affecting the flowering time variation were identified in another study employing QTL detection in F2 population generated from 18 distinct accessions in *A. thaliana* (Salomé *et al.*, 2011).

The first step in a QTL mapping is the development of a QTL mapping population. An F2 population, which can be easily developed in plants by selfing the F1 hybrid produced by crossing two parental plants, is the quickest and cheapest to develop (Falconer and Mackay, 1996). Once the mapping population is ready, the plants are normally grown in a randomised fashion under specific environmental conditions and different phenotypic traits are accurately recorded. The phenotypes so recorded are then analysed using various statistical methods before the plant samples can be used to generate the molecular markers. Then the phenotype and the genotype data are combined to detect QTL (Falconer and Mackay, 1996).

A large plant growth trial with 920 plants was conducted in Spring-Summer 2015 and various phenotype traits were collected. The plants did not grow well and there was a fly infestation in the glasshouse because of which an adequate amount of leaf samples could not be collected. To enable the plant growth cycle completion, buds were not collected for cytology as well. A few traits analysed had F1 variance greater than F2 (**Appendix A**). All this necessitated undertaking a second trial. A second growth trial with 980 plants consisting of 401 F2, 33 F1s, and 28 parental lines of diploid and tetraploid *A. thaliana* (Columbia and Landsberg ecotypes) was carried out in Spring-Summer 2016. Various qualitative and quantitative phenotypic traits were scored (**Table 2-1**), and leaf and bud samples were collected for further analysis.

4.2 Results

4.2.1 Trait phenotype distribution in diploids and tetraploids

The distribution of different traits in diploids and tetraploids is shown below in the histograms in **Figure 4-1**, **Figure 4-2** and **Figure 4-3**. The data distribution looks symmetrical for most of the traits except days to germination, rosette leaf number, total leaf number and silique length. Days to germination, rosette and total leaf number distribution is right tailed while silique length distribution is left tailed. There is a clear shift in the distribution in the diploids vs tetraploids for seed number, with diploids having a greater number of seeds than tetraploids.

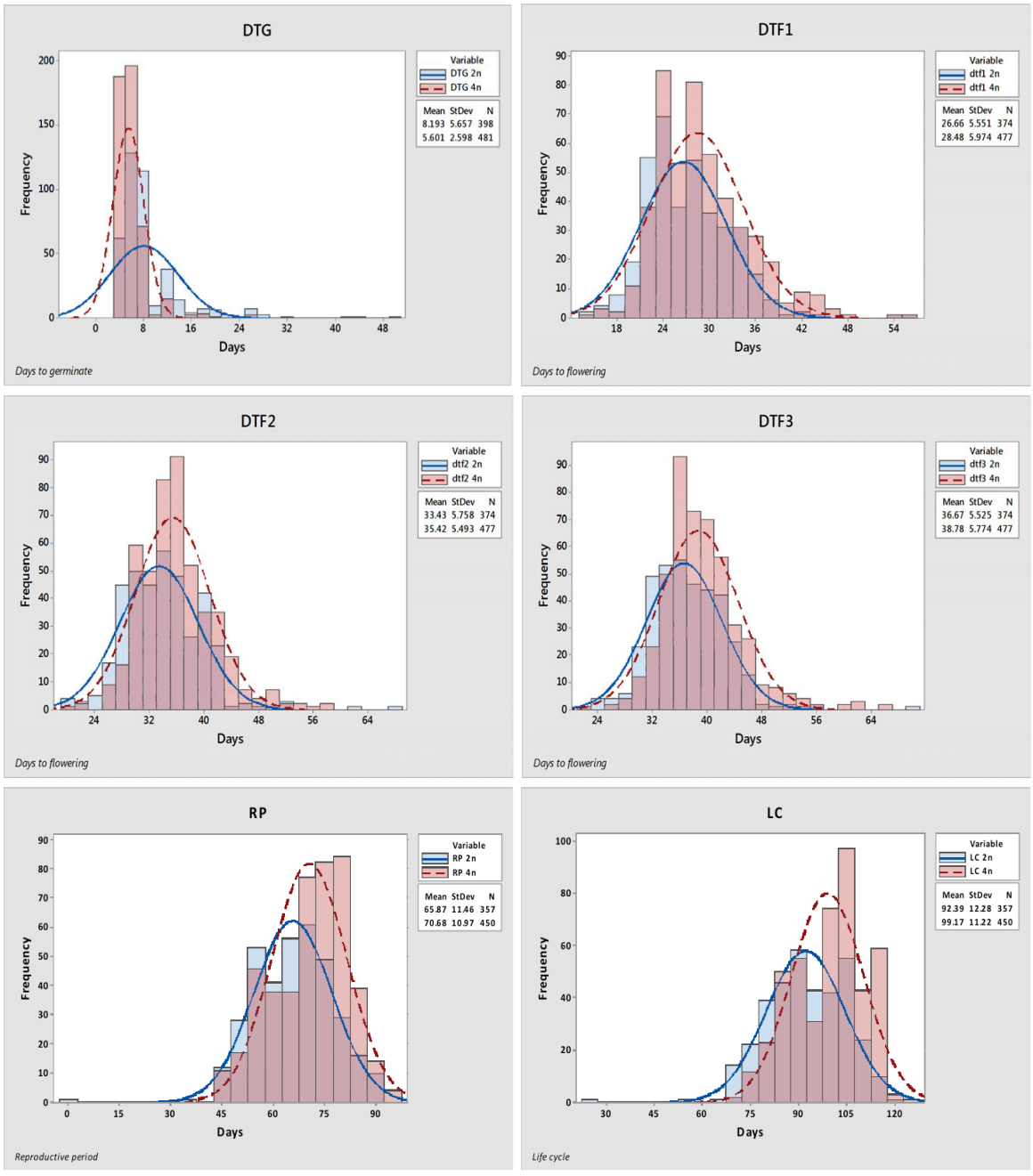


Figure 4-1 Histogram of diploids vs tetraploids for germination, flowering and fitness traits.

Histogram of diploids vs tetraploids for days to germinate (DTG), days to flower (meristem becomes ready (DTF1), main stalk is 1 cm long (DTF2) and first flower opens (DTF3)), reproductive period (days from DTF1 till senescence) and life cycle (days from germination till senescence).

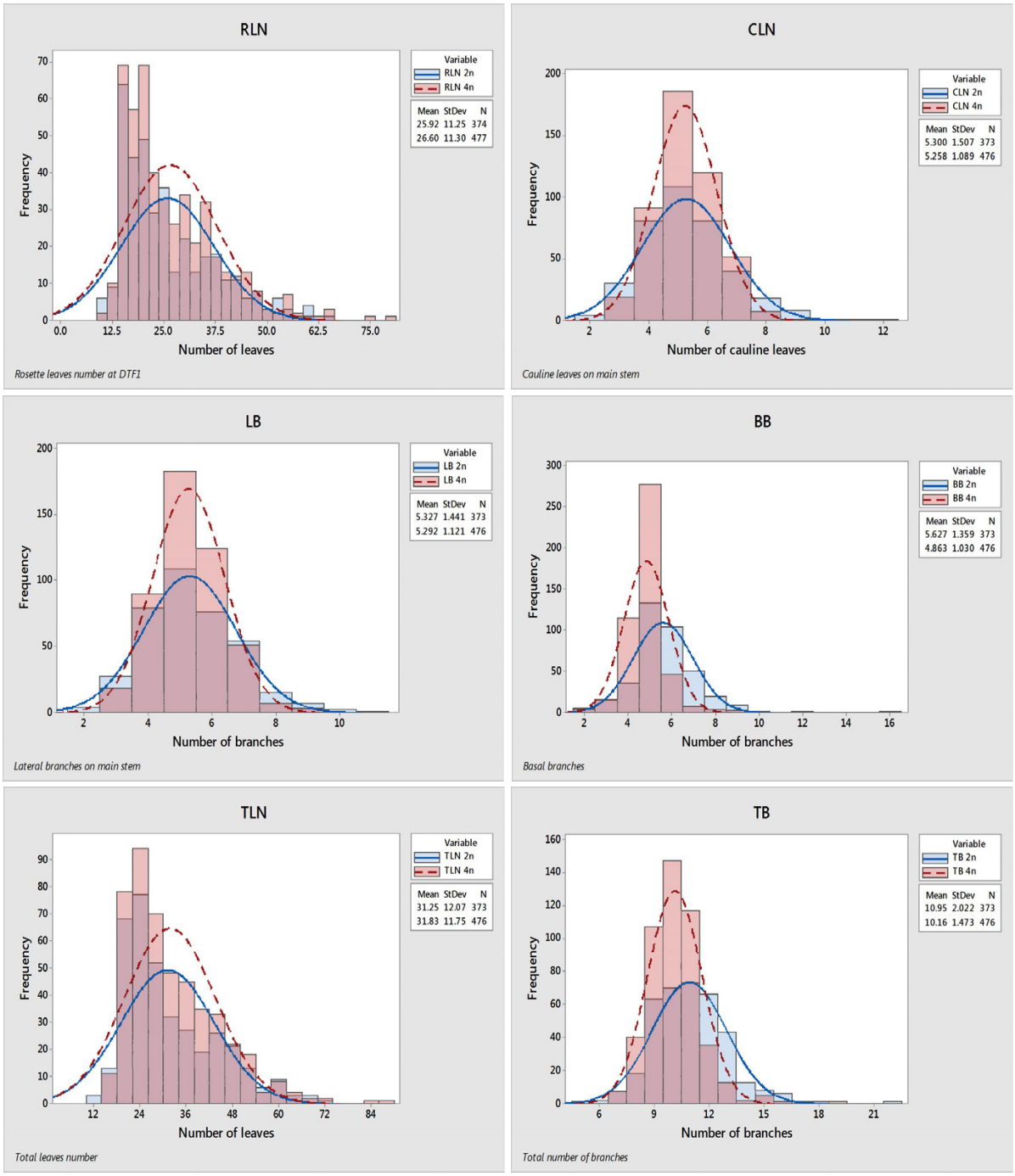


Figure 4-2 Histogram of diploid vs tetraploids for leaf and branch related traits.

Histogram of diploids vs tetraploids for rosette leaves (counted at DTF1) cauline leaves (on main stalk), lateral branches (on main stalk) and basal branches (apart from main stalk), total leaves (cauline + rosette) and total branches (lateral + basal).

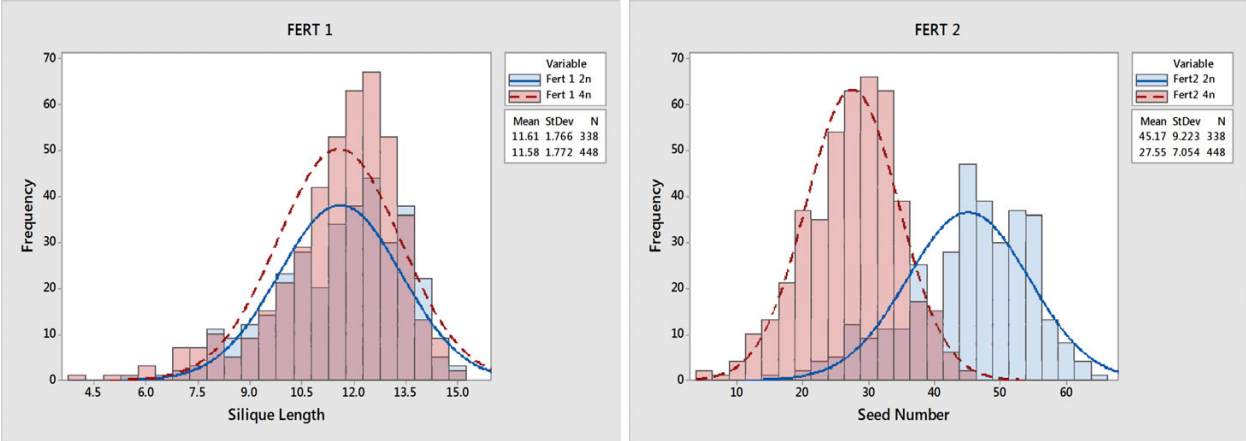


Figure 4-3 Histogram of diploids vs tetraploids for fertility traits.

FERT1 is silique length and FERT2 is seed numbers.

Histograms comparing diploid and tetraploid F2 with F1 and parental lines for different traits are shown in **Figure 4-4**, to compare how the trait variation is distributed in the segregating (F2) versus the non-segregating (P1, P2, F1) generations. The distribution for each trait for all the genotypes was checked after removing the outliers from the data (**Appendix B**). The distribution looked almost the same as that without the outliers removed, hence all of the following analysis has been based on the complete data set. The genotypes used are Diploid F2 (F2D), Tetraploid F2 (F2T), Diploid Columbia parent (ColD), Tetraploid Columbia parent (ColT), Landsberg Diploid parent (LerD), Landsberg Tetraploid parent (LerT), Diploid F1 (F1D) and tetraploid F1 (F1T).

For days to germination, it can be clearly seen that the F2 are more variable than the F1, whereas the germination in Landsberg is more variable than Columbia for both diploids as well as tetraploids. The F2 flowering distribution for DTF1 shows more than one peak both in diploids and tetraploids, indicating a mixed distribution, more variable than the F1, whereas for the parental lines the distribution shows a clear shift between Columbia and Landsberg

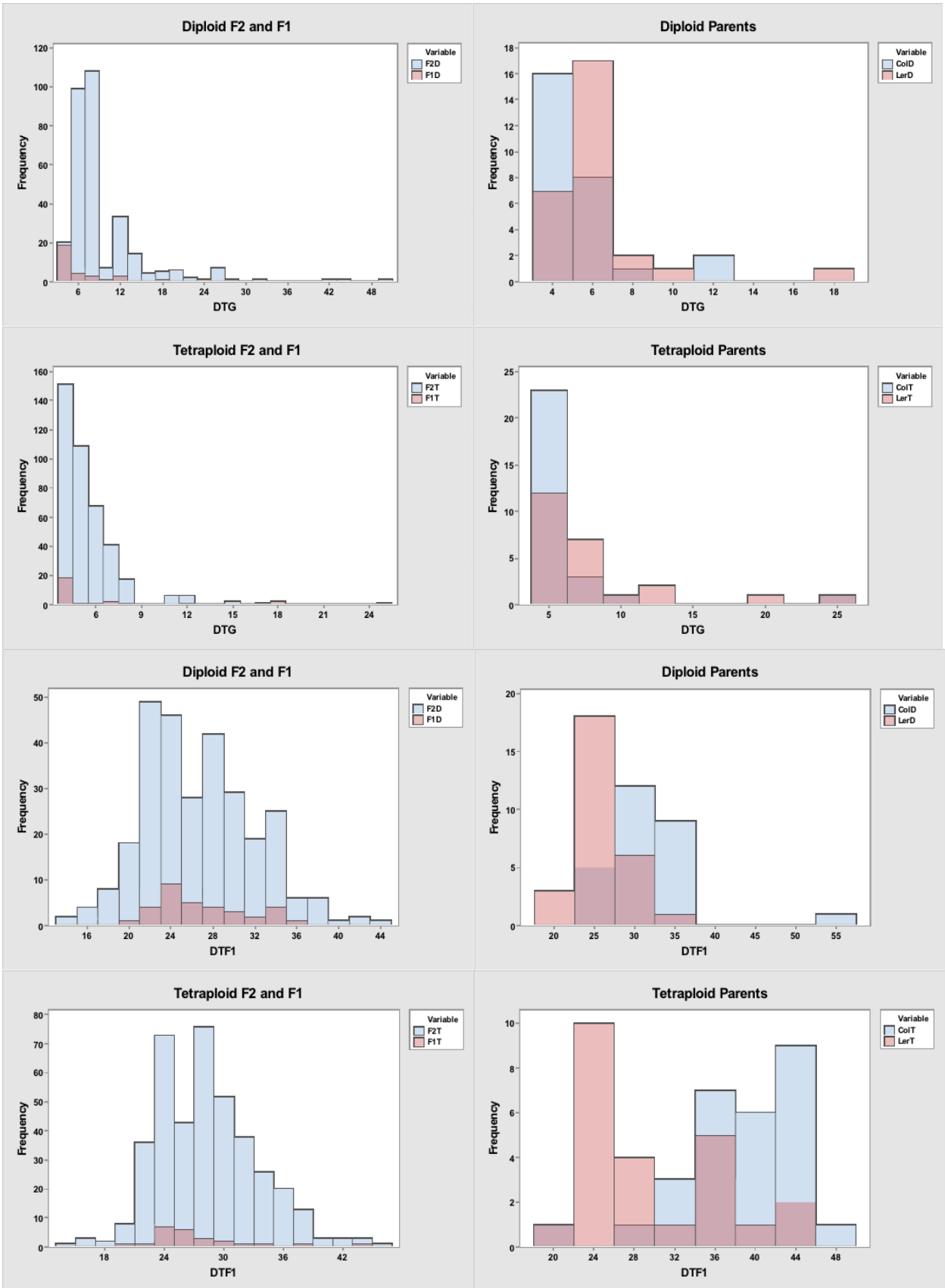
lines for both diploids and tetraploids, with Landsberg flowering earlier than the Columbia lines. The F2 flowering distribution for DTF2 shows more than one peak in diploids and tetraploids, indicating a mixed distribution (though it is less pronounced in tetraploids here) and is more variable than the F1. In the parental lines the distribution shows a clear shift between Columbia and Landsberg for diploids, with Landsberg flowering earlier than Columbia and for tetraploid parents the distribution shows more Landsberg parents flowering earlier. The F2 flowering distribution for DTF3 is symmetrical for both diploids and tetraploids and more variable than the F1. The distribution for parental lines is similar to the DTF2 distribution.

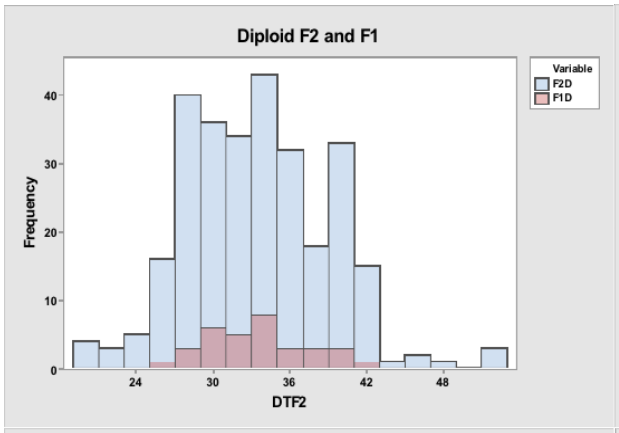
The F2 rosette leaf distribution is right tailed, with many small peaks on the right. The F1 is equally variable with the F2 in diploids, though less variable in tetraploids but following the same distribution as the F2s. In diploid parents, Landsberg shows a single peak indicating that most of the plants ranged in that particular leaf number (12-18 leaves), whereas Columbia diploids show a uniform distribution with Columbia showing more leaves than Landsberg. In tetraploid parents as well, Columbia parents tend to have a higher number of rosette leaves. The F2 distribution for cauline leaves look symmetrical and slightly more variable than the F1 for both diploids as well as tetraploids. There is however a clear difference between the diploid and tetraploid parents, with tetraploid Landsberg having more cauline leaves and showing more variation than Columbia. The distribution of total leaf number in the F2 is right tailed, but less pronounced than RLN and more variable than the F1 for both diploids as well as tetraploids. The Columbia parent shows more leaves and a more uniform distribution than the Landsberg parent, particularly in diploids

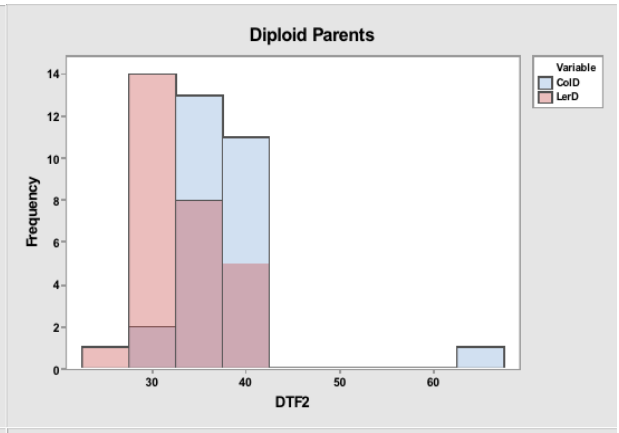
For lateral branches, the F2 data looks symmetrically distributed and more variable than the F1 in diploids and tetraploids. Landsberg tetraploid shows a higher number of lateral branches than Columbia, though the diploid parents are similar. The F2 distribution looks symmetrical and more variable than the F1 in both diploids and tetraploids for basal branches. Landsberg tetraploid parents have a broader distribution than the Columbia parents, while for diploids they show similar distributions. For the total number of branches, the F2 distribution looks symmetrical and more variable than F1, with diploid F2 showing more variation than the tetraploid F2. Landsberg parents show more variation than the Columbia parents.

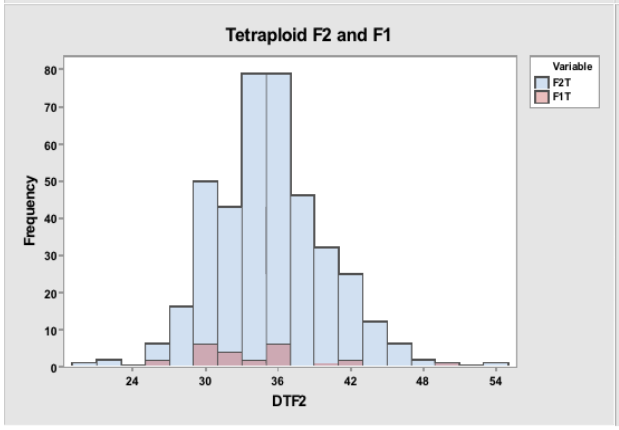
For reproductive period, F2 diploids show more than one peak and more variation than the F1, with a few F2s having a particularly short reproductive period. The tetraploid F2 clearly show a multimodal distribution with three clear peaks. The F1 shows a similar but narrower distribution. For life cycle, the F2 is more variable than the F1 and show more than one peak for both diploids and tetraploids, indicating a mixed distribution. The diploid parents show a similar distribution, while in tetraploid parents Columbia clearly show a longer life cycle.

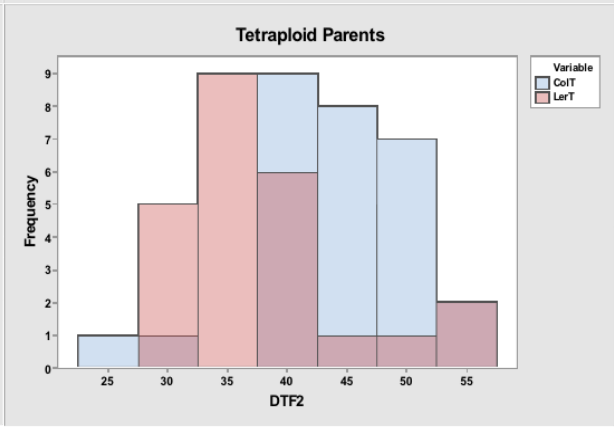
The F2 is more variable than the F1 for silique length in both diploid and tetraploids. The F2 distribution is left skewed in tetraploids and diploids. For parents, the difference between the two lines is clearly visible, with Columbia being at the higher end of distribution for both diploids and tetraploids. The diploid F2 shows more than one peak and more variability than the F1 for seed number. The tetraploid F2 shows a symmetrical distribution, slightly more variable than the F1. For the parents, Landsberg show more variation than Columbia in diploids.

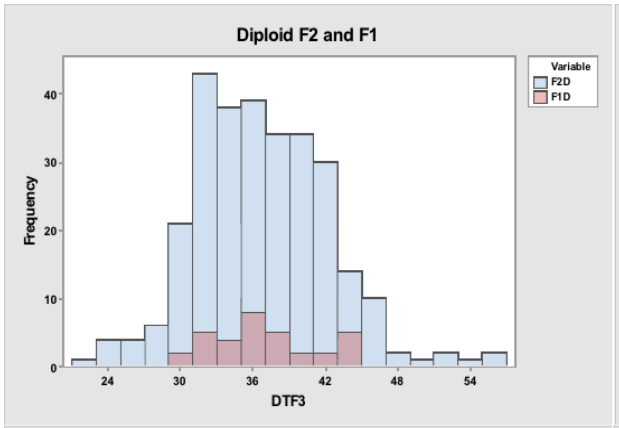

Diploid F2 and F1
Frequency
DTG
Variable
F2D
F1D
Diploid Parents
ColD
LerD
Tetraploid F2 and F1
F2T
F1T
Tetraploid Parents
ColT
LerT
DTF1

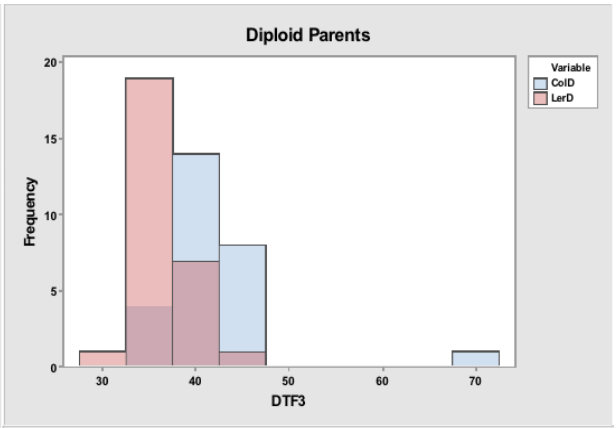

Diploid F2 and F1
Variable
F2D
F1D
Frequency
DTF2

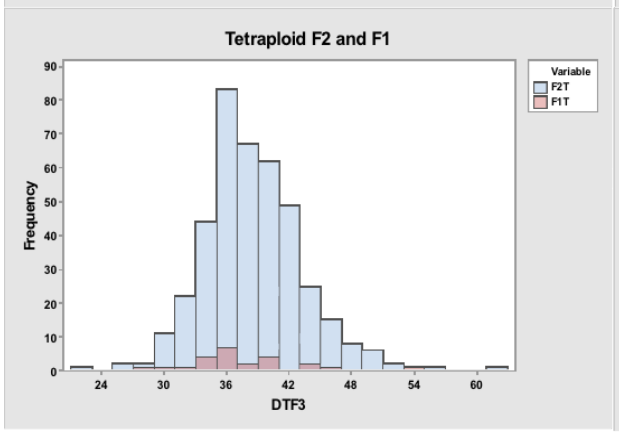

Diploid Parents
Variable
ColD
LerD
Frequency
DTF2

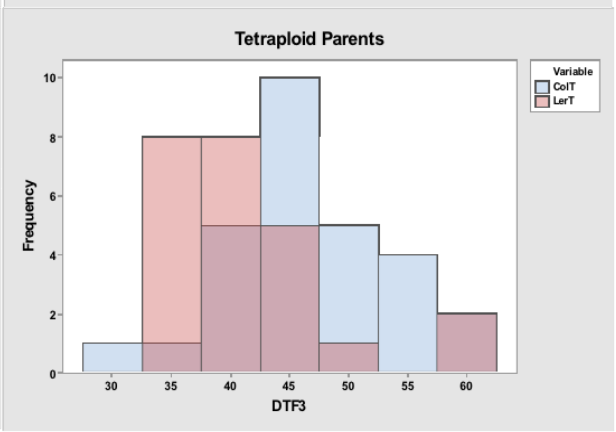

Tetraploid F2 and F1
Variable
F2T
F1T
Frequency
DTF2

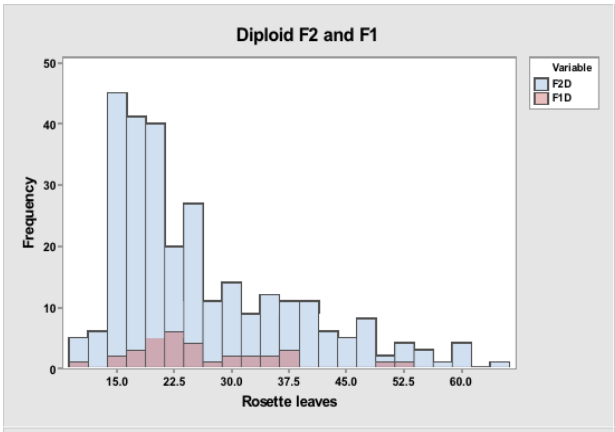

Tetraploid Parents
Variable
ColT
LerT
Frequency
DTF2

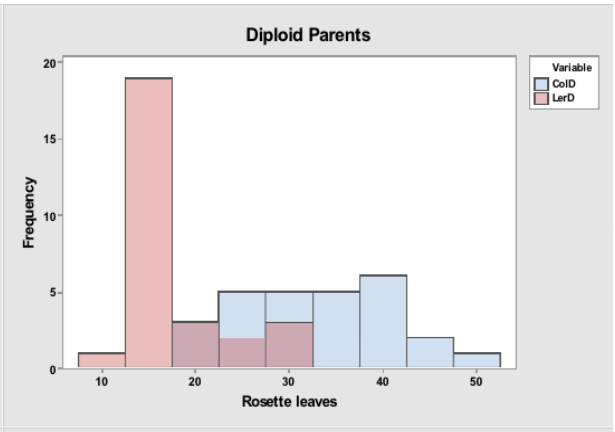

Diploid F2 and F1
Variable
F2D
F1D
Frequency
DTF3

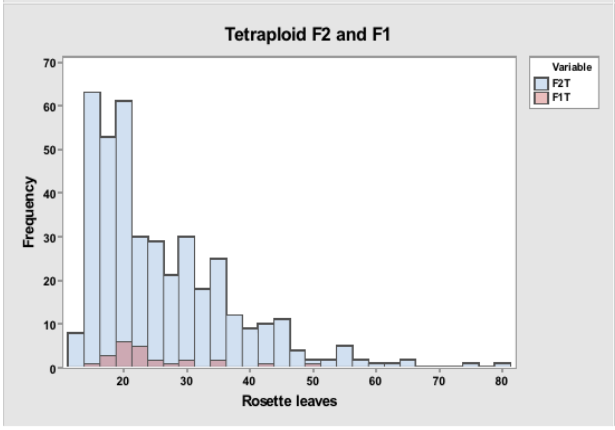

Diploid Parents
Variable
ColD
LerD
Frequency
DTF3

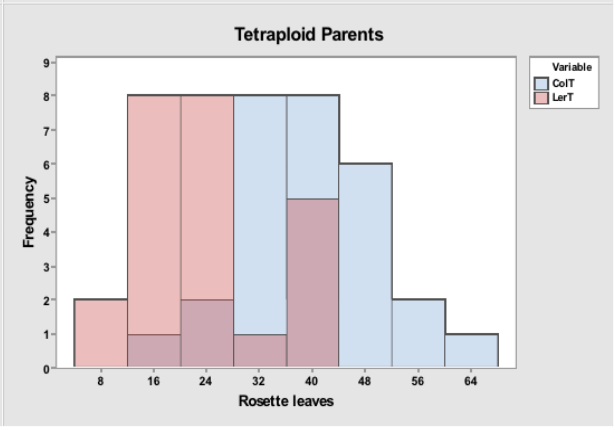

Tetraploid F2 and F1
Variable
F2T
F1T
Frequency
DTF3

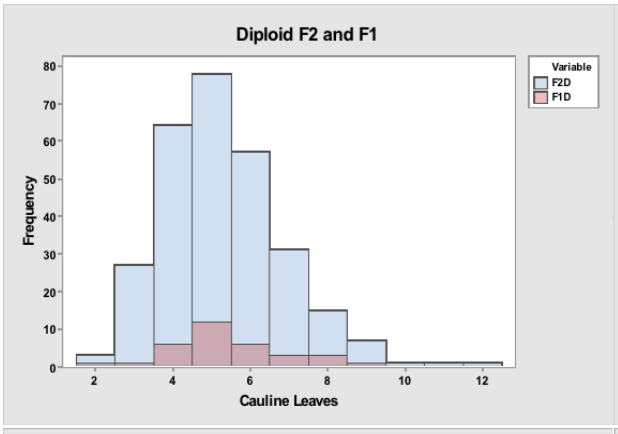

Tetraploid Parents
Variable
ColT
LerT
Frequency
DTF3

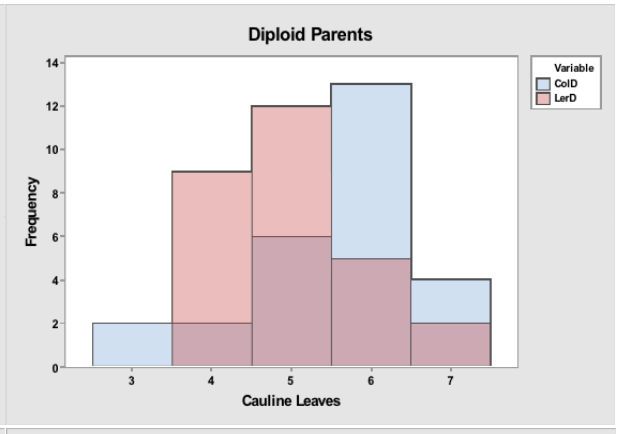

Diploid F2 and F1
Variable
F2D
F1D
Frequency
Rosette leaves

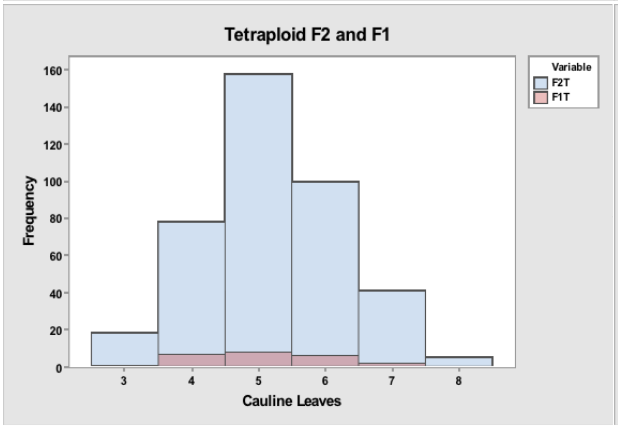

Diploid Parents
Variable
ColD
LerD
Frequency
Rosette leaves

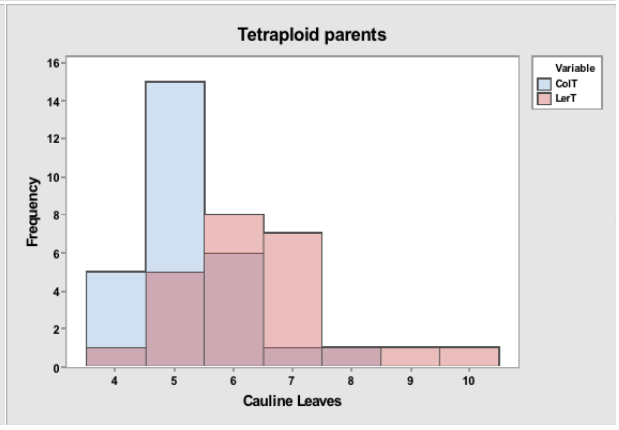

Tetraploid F2 and F1
Variable
F2T
F1T
Frequency
Rosette leaves

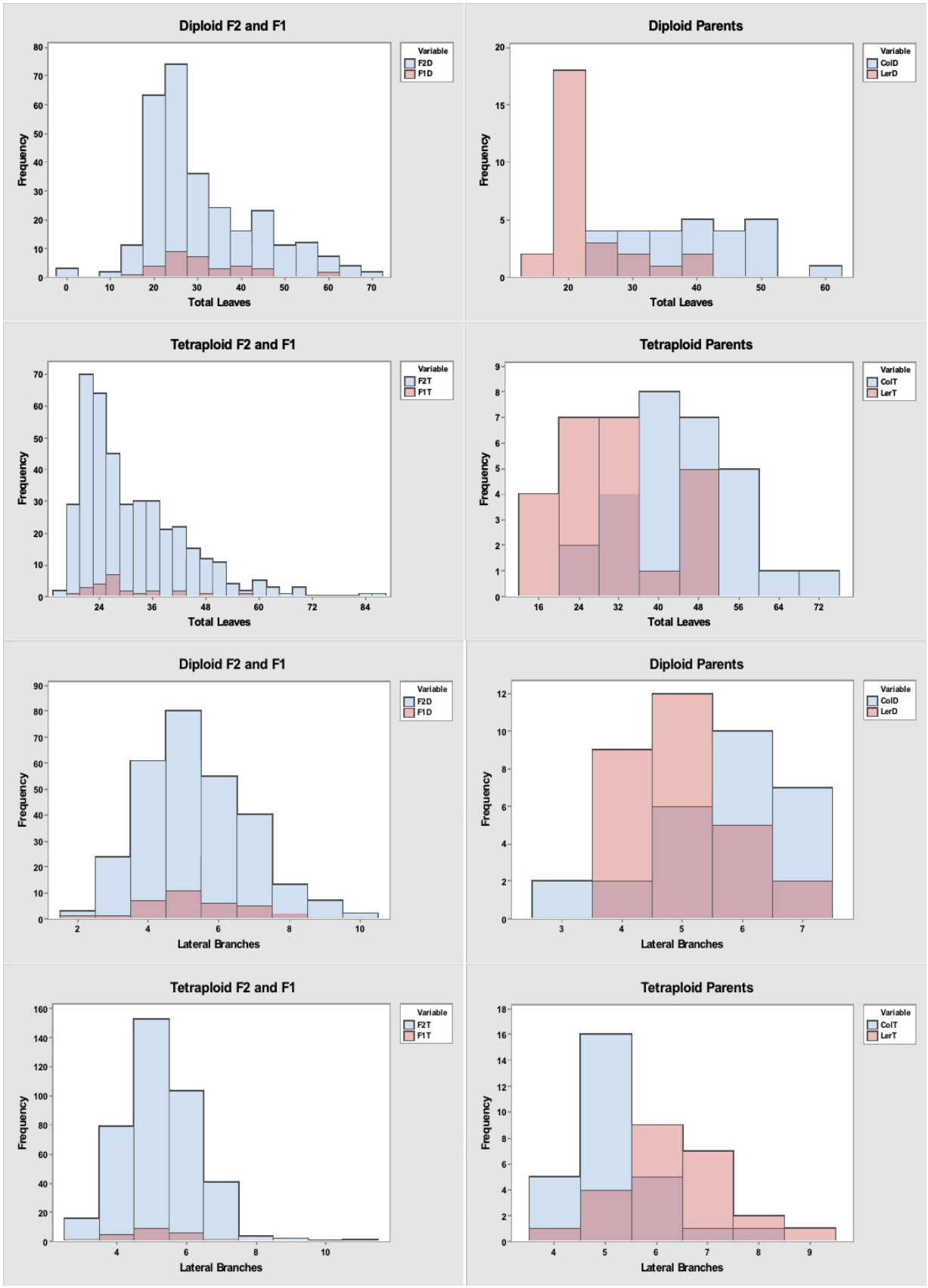

Tetraploid Parents
Variable
ColT
LerT
Frequency
Rosette leaves

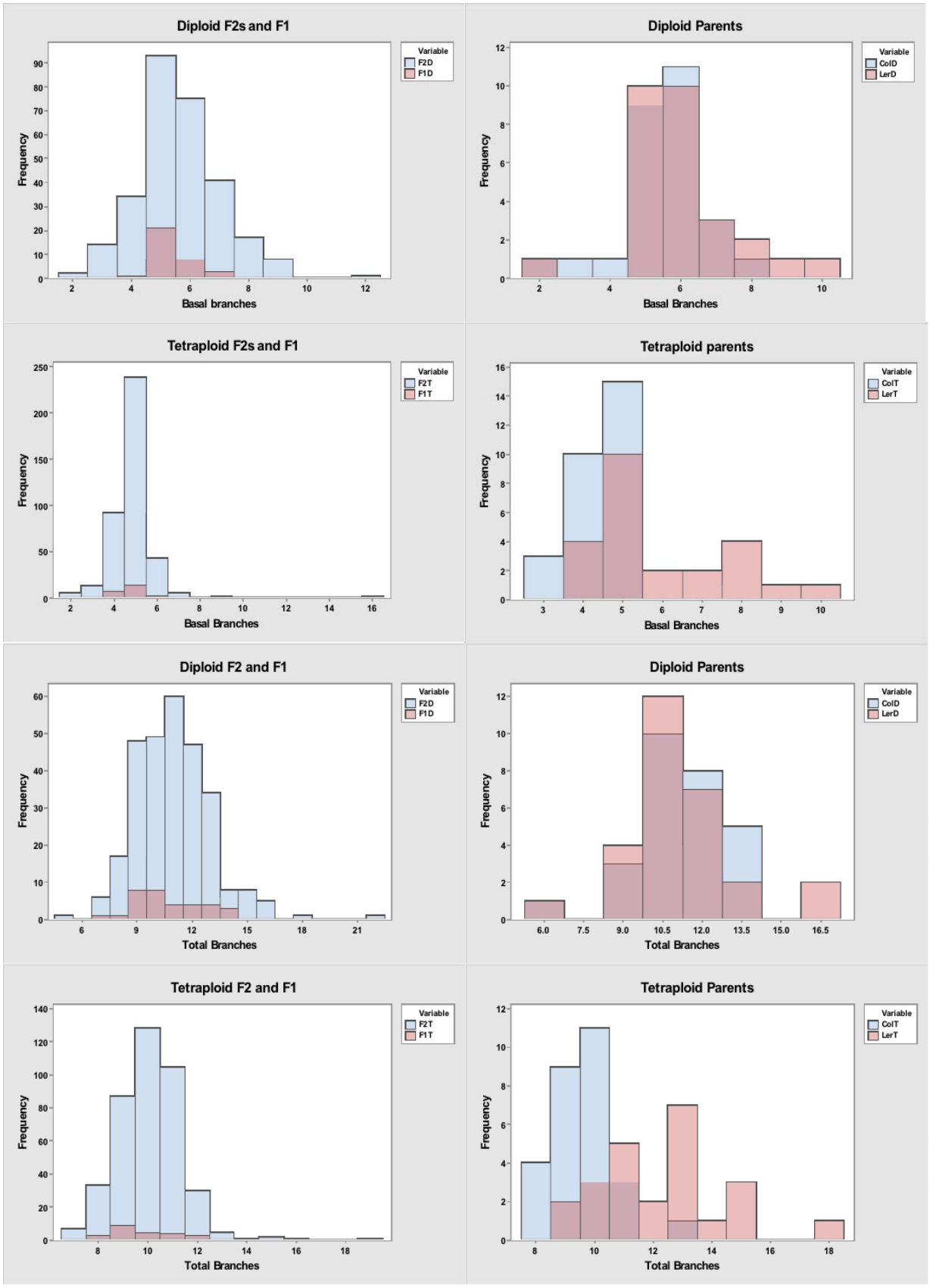

Diploid F2 and F1
Variable
F2D
F1D
Frequency
Cauline Leaves

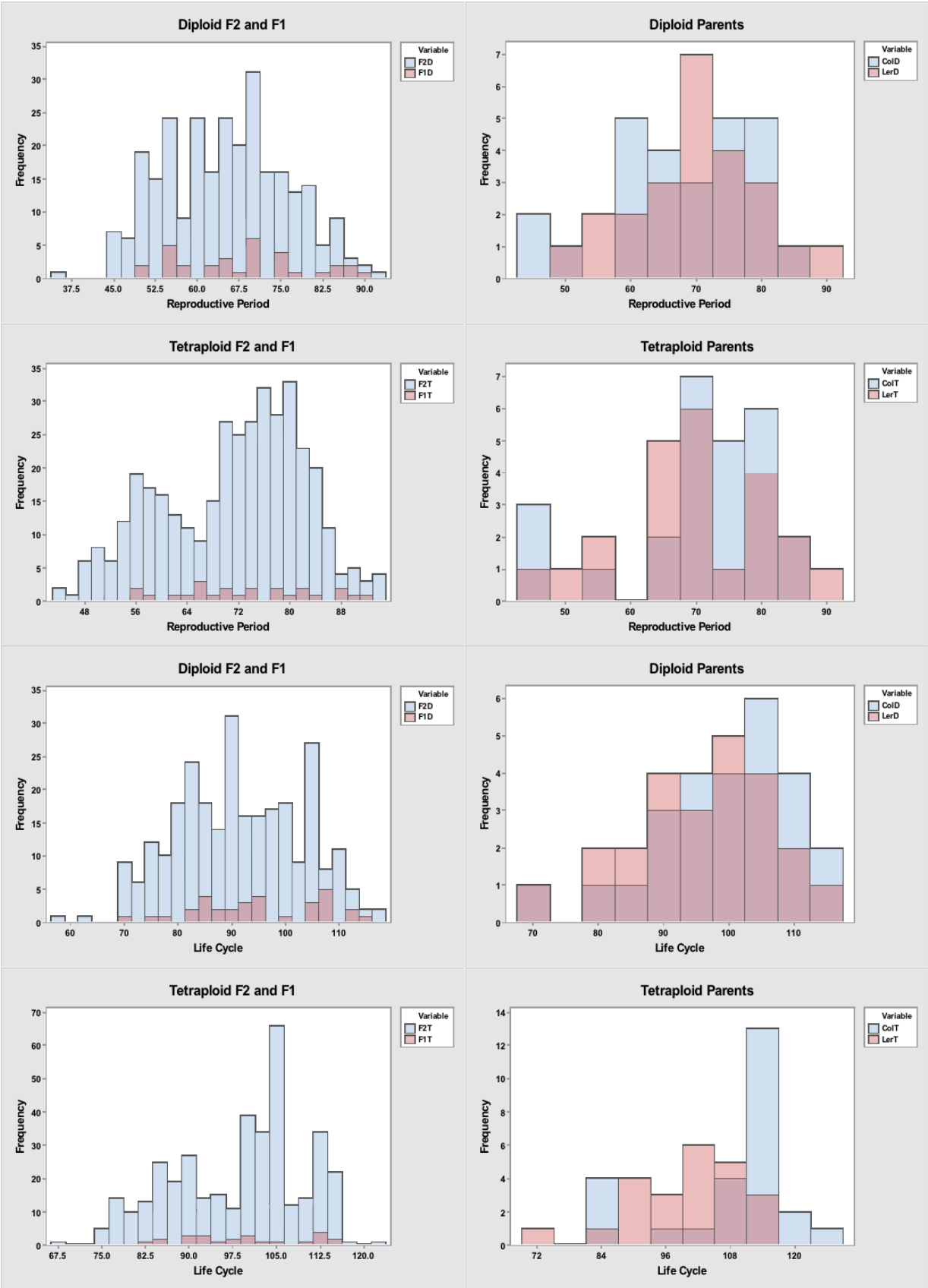

Diploid Parents
Variable
ColD
LerD
Frequency
Cauline Leaves

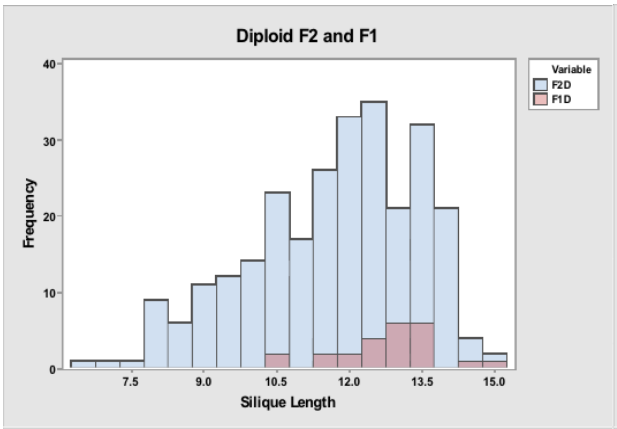

Tetraploid F2 and F1
Variable
F2T
F1T
Frequency
Cauline Leaves

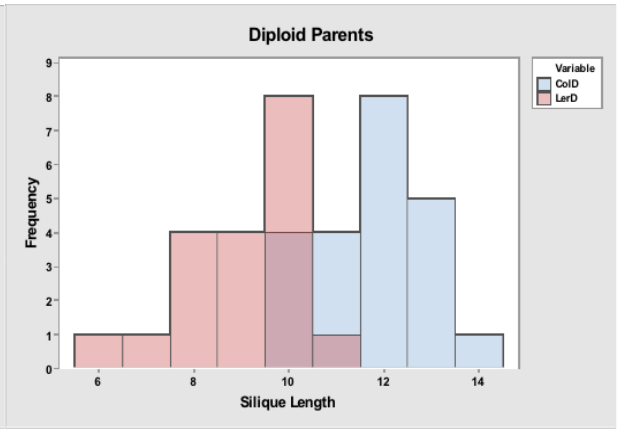

Tetraploid parents
Variable
ColT
LerT
Frequency
Cauline Leaves

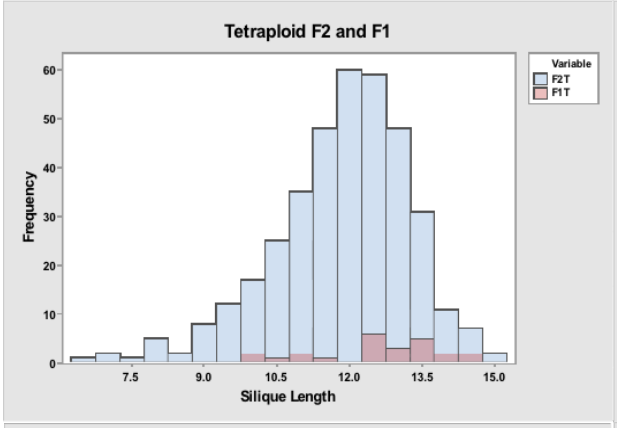

Diploid F2 and F1
Frequency
Total Leaves
Variable
F2D
F1D
Diploid Parents
Frequency
Total Leaves
Variable
ColD
LerD
Tetraploid F2 and F1
Frequency
Total Leaves
Variable
F2T
F1T
Tetraploid Parents
Frequency
Total Leaves
Variable
ColT
LerT
Diploid F2 and F1
Frequency
Lateral Branches
Variable
F2D
F1D
Diploid Parents
Frequency
Lateral Branches
Variable
ColD
LerD
Tetraploid F2 and F1
Frequency
Lateral Branches
Variable
F2T
F1T
Tetraploid Parents
Frequency
Lateral Branches
Variable
ColT
LerT

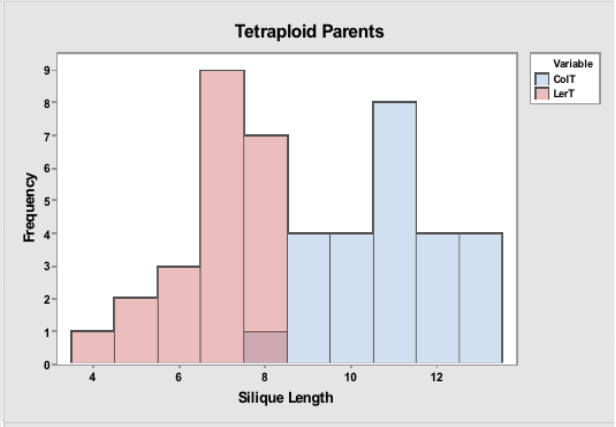

Diploid F2s and F1
Variable
F2D
F1D
Frequency
Basal branches
Diploid Parents
Variable
ColD
LerD
Frequency
Basal Branches
Tetraploid F2s and F1
Variable
F2T
F1T
Frequency
Basal Branches
Tetraploid parents
Variable
ColT
LerT
Frequency
Basal Branches
Diploid F2 and F1
Variable
F2D
F1D
Frequency
Total Branches
Diploid Parents
Variable
ColD
LerD
Frequency
Total Branches
Tetraploid F2 and F1
Variable
F2T
F1T
Frequency
Total Branches
Tetraploid Parents
Variable
ColT
LerT
Frequency
Total Branches

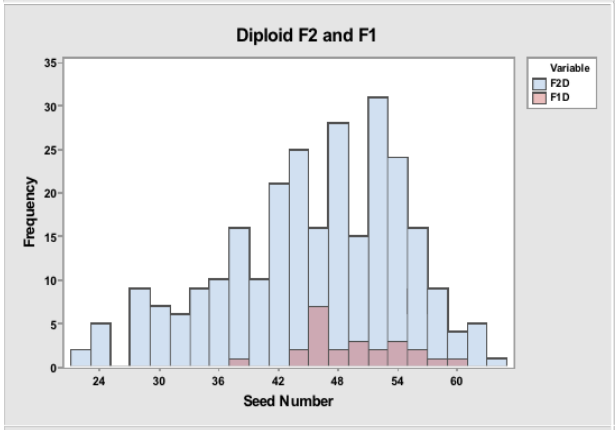

Diploid F2 and F1
Frequency
Reproductive Period
Variable
F2D
F1D
Diploid Parents
ColD
LerD
Tetraploid F2 and F1
F2T
F1T
Tetraploid Parents
ColT
LerT
Life Cycle

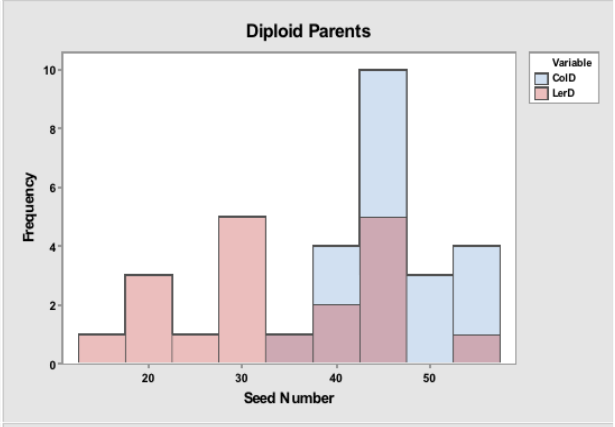

Diploid F2 and F1
Frequency
Silique Length
Variable
F2D
F1D

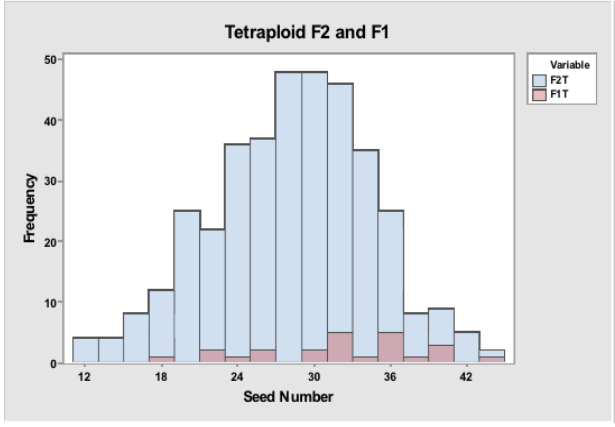

Diploid Parents
Frequency
Silique Length
Variable
ColD
LerD


Tetraploid F2 and F1
Frequency
Silique Length
Variable
F2T
F1T


Tetraploid Parents
Frequency
Silique Length
Variable
ColT
LerT


Diploid F2 and F1
Frequency
Seed Number
Variable
F2D
F1D


Diploid Parents
Frequency
Seed Number
Variable
ColD
LerD


Tetraploid F2 and F1
Frequency
Seed Number
Variable
F2T
F1T

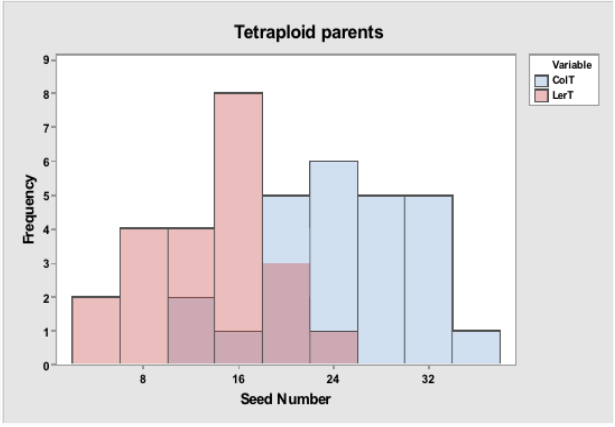

Tetraploid parents
Frequency
Seed Number
Variable
ColT
LerT

Figure 4-4 Histograms showing distribution of diploid and tetraploid F2 with F1 and the parental lines for different traits.

F2D and F2T represent diploid and tetraploid F2, ColD and ColT represent diploid and tetraploid Columbia parent, LerD and LerT are diploid and tetraploid Landsberg parent, F1D and F1T are diploid and tetraploid F1 respectively.

4.2.2 Exploratory data analysis for different traits

Exploratory data analysis involves looking at the differences in mean, variances, range, etc. between different lines. A summary table showing different exploratory data statistics for different traits in different lines is shown in **Table 4-1**. The coefficient of variation for the tetraploid F2 is lower than the diploid F2 for all the traits except the number of seeds, indicating less dispersion around the mean.

	TRAITS											
	DTF1			DTF2			DTF3			DTG		
variety	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var
F2D	286	26.27 ± 5.54	21.11	286	33.07 ± 5.81	17.56	286	36.24 ± 5.53	15.27	310	8.794 ± 5.89	67.00
F2T	401	27.89 ± 5.38	19.28	401	34.78 ± 4.68	13.44	401	38.07 ± 4.91	12.90	401	5.399 ± 1.82	33.73
ColD	27	31.85 ± 6.22	19.52	27	38.33 ± 6.46	16.86	27	41.67 ± 6.34	15.22	27	5.037 ± 2.14	42.47
ColT	28	37.75 ± 6.23	16.50	28	43.89 ± 6.78	15.44	28	46.86 ± 7.13	15.21	28	5.75 ± 4.04	70.31
LerD	28	25.82 ± 3.67	14.22	28	32.96 ± 3.97	12.04	28	36.5 ± 3.51	9.62	28	5.857 ± 2.65	45.22
LerT	24	29.04 ± 6.86	23.61	24	38.08 ± 7.38	19.39	24	41.46 ± 7.37	17.78	24	7.75 ± 4.94	63.70
F1D	33	26.55 ± 4.22	15.90	33	32.97 ± 4.12	12.49	33	36.36 ± 4.14	11.40	33	7.121 ± 5.86	82.32
F1T	24	26.88 ± 5.45	20.27	24	33.58 ± 5.52	16.43	24	38.38 ± 7.77	20.25	28	6.5 ± 5.31	81.68

	TRAITS											
	RLN			CLN			TLN			RP		
variety	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var
F2D	286	25.97 ± 11.60	44.65	285	5.291 ± 1.58	29.94	285	30.98 ± 12.78	41.26	275	65.01 ± 11.57	17.79
F2T	401	26.09 ± 11.14	42.68	400	5.208 ± 1.05	20.20	400	31.26 ± 11.62	37.19	377	70.63 ± 10.94	15.49
ColD	27	33.15 ± 8.47	25.57	27	5.556 ± 1.09	19.55	27	38.7 ± 9.10	23.52	26	67.88 ± 10.76	15.85
ColT	28	38.5 ±10.40	27.01	28	5.214 ± 0.92	17.59	28	43.71 ± 10.52	24.07	26	69.96 ± 11.36	16.24
LerD	28	18 ± 5.67	31.50	28	5 ± 0.90	18.05	28	23 ± 6.26	27.22	24	70.08 ± 9.93	14.16
LerT	24	22.83 ± 9.66	42.30	24	6.375 ± 1.35	21.10	24	29.21 ± 10.20	34.91	23	70 ± 11.48	16.40
F1D	33	26.24 ± 9.59	36.55	33	5.424 ± 1.52	28.05	33	31.67 ± 10.58	33.40	32	68.44 ± 11.35	16.58
F1T	24	25.04 ± 8.51	33.98	24	5.042 ± 1.04	20.66	24	30.08 ± 9.10	30.26	24	72.83 ± 10.95	15.03

	TRAITS											
	LB			BB			TB			LC		
variety	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var
F2D	285	5.33 ± 1.51	28.32	285	5.628 ± 1.41	25.05	285	10.96 ± 2.06	18.78	275	91.15 ± 12.34	13.54
F2T	400	5.252 ± 1.11	21.09	400	4.835 ± 0.97	20.16	400	10.09 ± 1.35	13.42	377	98.52 ± 11.08	11.25
ColD	27	5.667 ± 1.18	20.76	27	5.519 ± 1.22	22.12	27	11.19 ± 1.73	15.49	26	99.77 ± 10.46	10.49
ColT	28	5.179 ± 0.90	17.47	28	4.429 ± 0.69	15.58	28	9.607 ± 1.10	11.45	26	108 ± 11.59	10.73
LerD	28	5 ± 0.90	18.05	28	6 ± 1.52	25.26	28	11 ± 2.16	19.64	24	95.46 ± 10.96	11.48
LerT	24	6.333 ± 1.13	17.83	24	5.958 ± 1.73	29.06	24	12.29 ± 2.16	17.55	23	99.35 ± 10.30	10.37
F1D	33	5.303 ± 1.38	26.03	33	5.394 ± 0.70	13.06	33	10.7 ± 1.85	17.25	32	94.78 ± 11.70	12.34
F1T	24	5.042 ± 1.08	21.47	24	4.75 ± 0.61	12.80	24	9.792 ± 1.25	12.77	24	99.71 ± 10.36	10.40

	TRAITS					
	FERT 1			FERT 2		
variety	count(n)	mean ± sd	Coeff of var	count(n)	mean ± sd	Coeff of var
F2D	269	11.69 ± 1.72	14.71	269	45.44 ± 8.99	19.78
F2T	374	11.84 ± 1.41	11.91	374	28.27 ± 6.22	22
ColD	22	11.68 ± 1.08	9.24	22	45.81 ± 5.54	12.1
ColT	25	10.86 ± 1.3	11.97	25	24.38 ± 6.06	24.86
LerD	19	8.97 ± 1.4	15.61	19	33.81 ± 11.49	33.98
LerT	22	6.98 ± 1.11	15.9	22	13.9 ± 4.92	35.4
F1D	24	12.75 ± 1.01	7.92	24	49.52 ± 5.08	10.26
F1T	24	12.59 ± 1.34	10.64	24	32.06 ± 6.68	20.84

Table 4-1 A summary table showing mean, standard deviation and coefficient of variance of different traits for the 8 varieties.

F2D and F2T represent diploid and tetraploid F2, ColD and ColT represent diploid and tetraploid Columbia parent, LerD and LerT are diploid and tetraploid Landsberg parent, F1D and F1T represent diploid and tetraploid F1.

Based on the exploratory data analysis, it can be seen that both the F1 and F2 show heterosis over their respective parents for fertility traits, both in diploids as well as in tetraploids. For example, it can be seen in **Table 4-1** that the mean length of siliques and mean seed number for the F1 is more than their respective parents, both in diploids and tetraploids. This can be seen in **Figure 4-8** as well.

4.2.3 Normality testing and testing for significance

Data collected for different traits of diploid and tetraploid plants was tested to see if it followed a normal distribution using the Ryan Joiner test (equivalent of the Shapiro wilk test). It is a robust test sensitive to skewness and tails in the data. The normality assumption was found to be violated for all traits for both diploids and tetraploids (p-value < 0.01), except cauline leaves in tetraploids and lateral branches in diploids with p-value greater than 0.01.

The data distribution shows non-normality for most of the traits, hence the parametric statistical test analysis is unsuitable. Broad comparison between all diploids and all tetraploids for different traits, shows a significant difference (Mann Whitney test) between the two for 9 out of 14 traits, DTG (W = 214289, p-value = 0.000), DTF1 (W = 143715.50, p-value = 0.000), DTF2 (W = 141026.50, p-value = 0.000), DTF3 (W = 141049, p-value = 0.000), BB (W = 192688.50, p-value = 0.000), TB (W = 181363.50, p-value = 0.000), RP (W = 123947.50, p-value = 0.000), LC (W = 119191, p-value = 0.000) and FERT 2 (W = 197266.5, p-value = 0.000).

A further statistical analysis of different traits for the eight genotypes was undertaken to explore the differences between them using non-parametric tests and data shown using box plots. A significant difference (p-value < 0.05) was found between them for *all* the traits, which were then further probed using post hoc tests to exactly find out the differences.

Days to Germination- time from the seeds sown till the first cotyledons appear.

Kruskal-Wallis test (an equivalent of one way ANOVA) shows a significant difference in DTG between different lines ($\chi^2_{df=7,}$= 198.24, p-value < 2.2e-16). Post hoc Dunn test with Bonferroni correction produced a statistical significant (p-value < 0.05) difference between the lines as shown in the **Figure 4-5**, with tetraploids germinating earlier than diploids (**Figure 4-4**).

Flowering Time- DTF1, DTF2 and DTF3

Kruskal-Wallis test shows a significant difference in DTF1 (Days until floral meristem starts to divide, $\chi^2_{df=7,}$= 87.72, p-value = 3.633e-16), DTF2 (Days until floral stalk is 1cm in length, $\chi^2_{df=7,}$= 94.29, p-value < 2.2e-16) and DTF3 (days until first flower opens, $\chi^2_{df=7,}$= 92.07 p-value < 2.2e-16) between different lines. Post hoc Dunn test with Bonferroni correction produced a statistical significant difference between the lines at p-value < 0.05 (**Figure 4-5**), with tetraploids flowering later than diploids (**Figure 4-4**) with the exception of tetraploid F1s, which flowered earlier than the parents.

Number of leaves- RLN, CLN and TLN

Kruskal-Wallis test shows a significant difference in RLN (Rosette leaves number at flowering stage 1, $\chi^2_{df=7,}$= 73.29, p-value = 3.187e-13), CLN (Cauline leaves number on main stem, $\chi^2_{df=7,}$= 22.24, p-value =.002307) and TLN (sum of RLN and CLN, $\chi^2_{df=7,}$= 67.59, p-value = 4.59e-12) between different lines. Post hoc Dunn test with Bonferroni correction produced a statistical significant difference (p-value < 0.05) between the lines as shown in **Figure 4-6**, with ColT having the highest RLN and LerT showing the highest CLN. ColT produced the highest number of total leaves (TLN).

Number of branches- LB, BB and TB

Kruskal-Wallis test shows a significant difference in LB (number of lateral branches on main stalk, $\chi^2_{df=7,}$= 23.83, p-value = .00122), BB (Number of basal branches, $\chi^2_{df=7,}$= 126.4, p-value= 2.2e-16) and TB (total branches, sum of lateral and basal branches, $\chi^2_{df=7,}$= 78.79 p-value = 2.429e-14) between different lines. Post hoc Dunn test with Bonferroni correction produced a statistical significant difference (p-value < 0.05) between the lines as shown in **Figure 4-7**, with LerT having the highest number of lateral branches and differing significantly from LerD. However, tetraploids show fewer basal branches as well as fewer total branches than the diploids, as seen in **Figure 4-4**. It can be noted that Landsberg tetraploids showed more cauline leaves and more lateral branches, which can be correlated as lateral branches arise in the axil of cauline leaves.

RP- **Reproductive Period**- days between DTF1 and complete senescence

Kruskal-Wallis test shows a significant difference in RP between different lines ($\chi^2_{df=7,}$= 45.13, p-value = 1.291e-07). Post hoc Dunn test with Bonferroni correction produced a statistical significant (p-value < 0.05) difference between the diploid and tetraploid F2 as shown in **Figure 4-6**, with tetraploid F2 having a longer reproductive period (**Figure 4-4**).

LC- **Life Cycle**- days between germination and complete senescence

Kruskal-Wallis test shows a significant difference in LC between different lines ($\chi^2_{df=7,}$= 88.68, p-value = 2.307e-16). Post hoc Dunn test with Bonferroni correction produced a statistical significant (p-value < 0.05) difference between diploids and tetraploids as shown in **Figure 4-7**, with tetraploids having a longer life cycle than diploids (**Figure 4-4**).

FERT 1- **Silique length**- the average length of 10 siliques on the plant

Kruskal-Wallis test shows a significant difference in silique length between different lines ($\chi^2_{df=7,}$= 125.16, p value < 2.2e-16). Post hoc Dunn test with Bonferroni correction produced a statistical significant difference (p-value < 0.05) between diploids and tetraploids (**Figure 4-8**), with tetraploids having comparatively smaller siliques than diploids. F1 and F2 show heterosis.

FERT 2- **Seed numbers**- the average number of seed in 10 siliques

Kruskal-Wallis test shows a significant difference in seed numbers between different lines ($\chi^2_{df=7,}$= 462.92, p value < 2.2e-16). Post hoc Dunn test with Bonferroni correction produced a statistical

significant difference (p-value < .05) between diploids and tetraploids as shown in **Figure 4-8**, with tetraploids showing lower seed set in their siliques than diploids (**Figure 4-4**). F1 and F2 show heterosis.

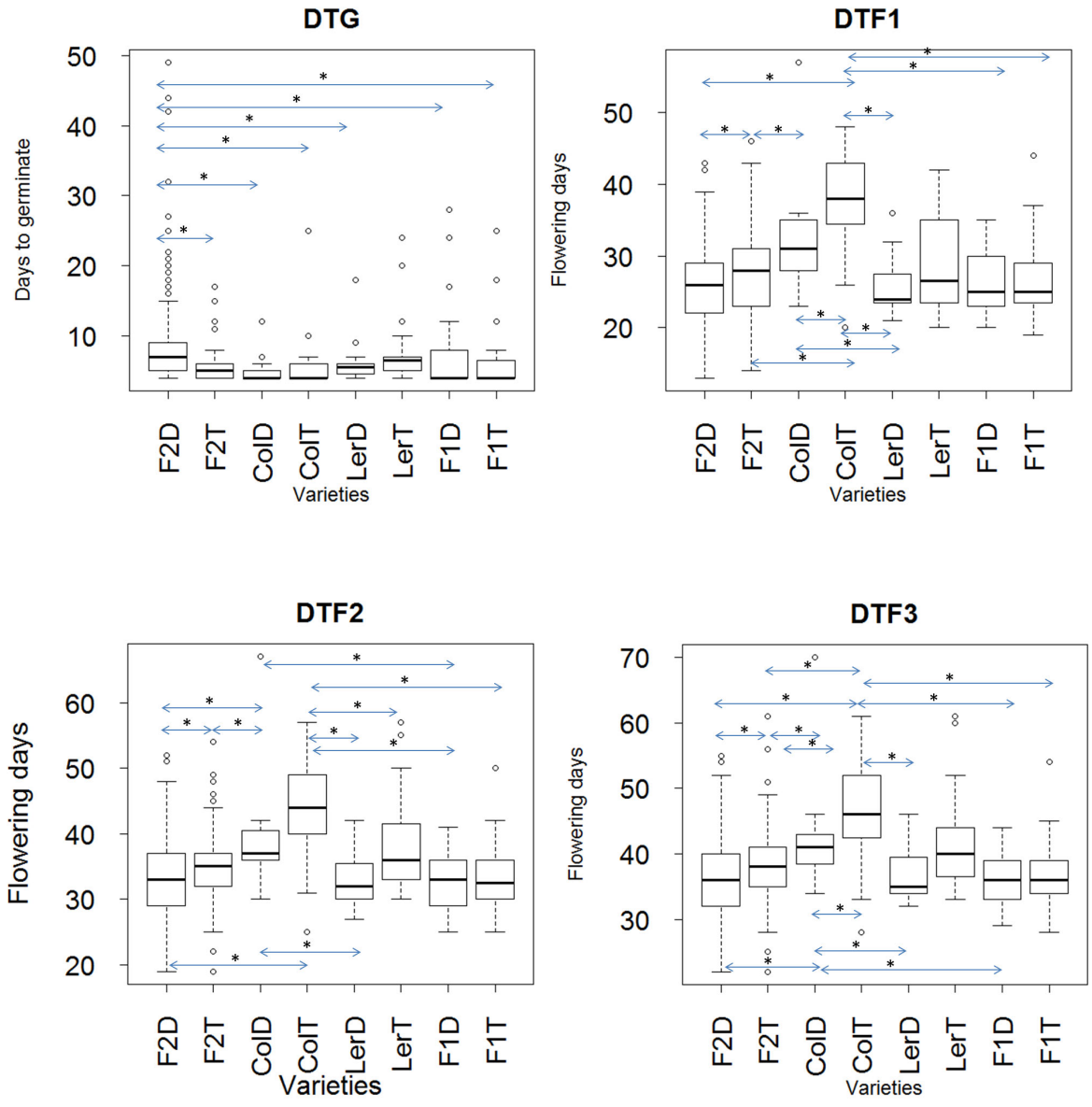


Figure 4-5 Boxplots showing distribution and significant differences between different varieties for four different traits – three flowering (DTF1, DTF2, DTF3) and Days to Germination (DTG).

* indicates significance at p-value < 0.05.

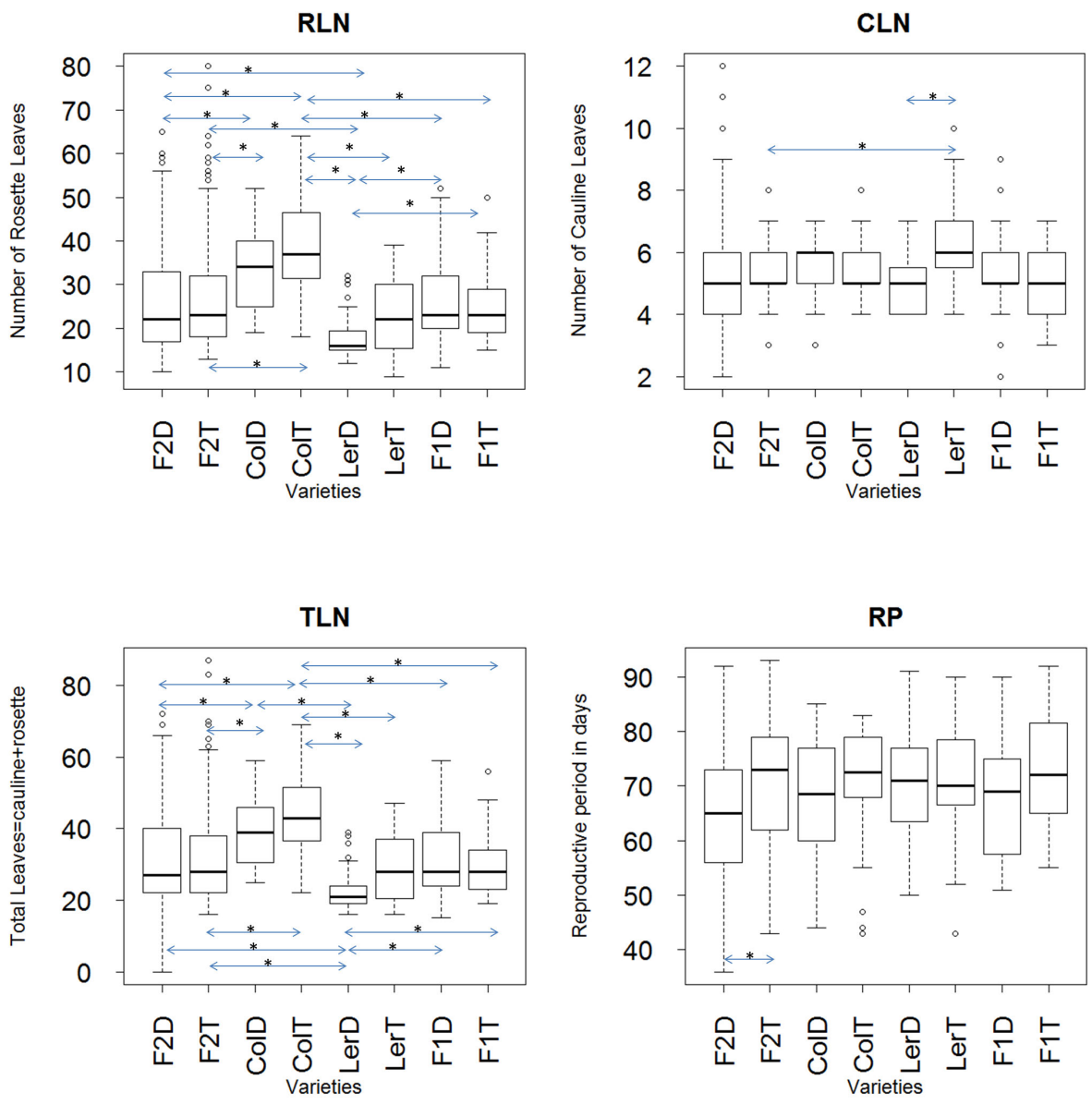


Figure 4-6 Boxplot showing distribution and significant differences between different varieties for four different traits – three leaf traits (RLN, CLN, TLN) and Reproductive period RP.

* indicates significance at p-value < 0.05.

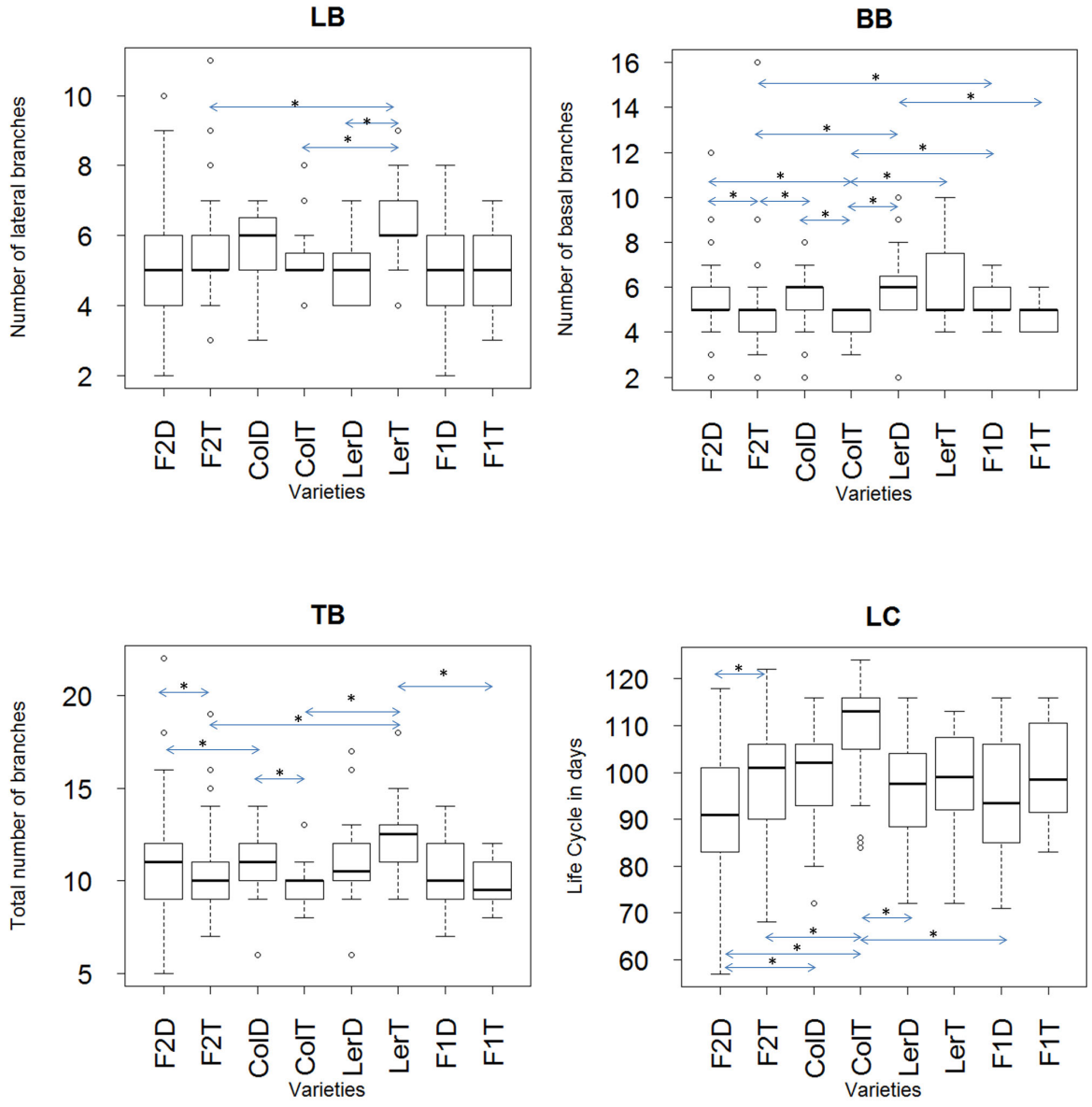


Figure 4-7 Boxplot showing distribution and significant differences between different varieties for four different traits – three branches (TB, LB, BB), and Life Cycle LC.

* indicates significance at p-value < 0.05.

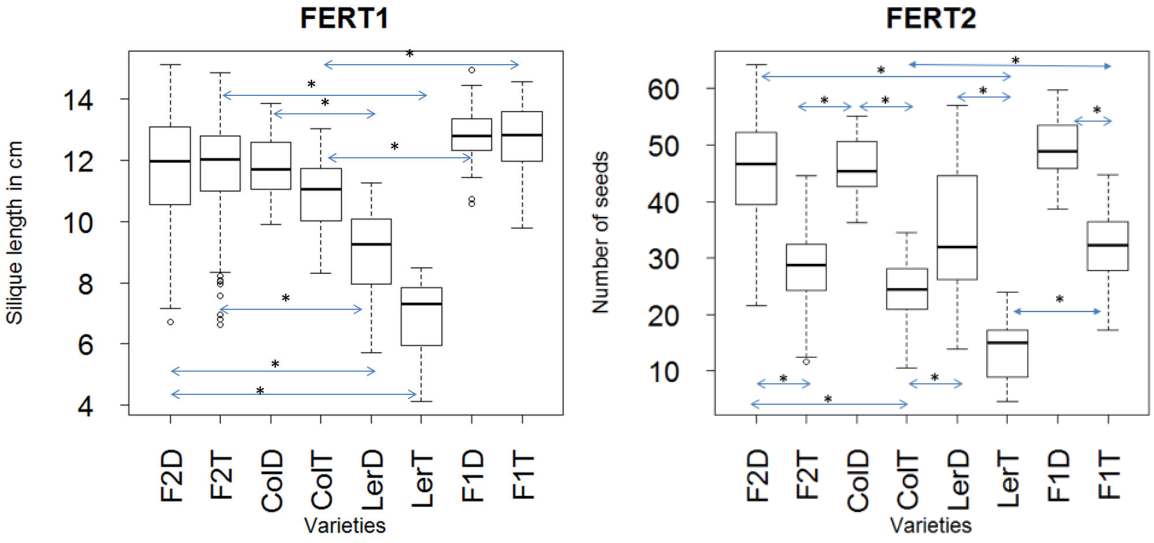


Figure 4-8 Boxplot showing distribution and significant differences between different varieties for two different fertility traits - silique length and seed numbers.

* indicates significance at p-value < 0.05.

To conclude, a significant difference ($p < 0.01$) has been found between diploids and tetraploids for at-least nine out of fourteen phenotype traits. In summary, tetraploids germinate earlier, flower later without producing more rosette leaves, and have less branches and lower seed set than diploids.

Correlation analysis was carried out for all traits in diploids as well as tetraploids. Spearman rho (considering most of the trait data distribution was not normally distributed) correlation coefficients were calculated using Minitab. **Table 4-2** and **Table 4-3** only show the significant data for Spearman rho.

In diploids, the three flowering traits in **Table 4-2** show highly significant (p-value < 0.001) and high values of the correlation coefficient (> 80%), which is expected as it is essentially measuring the time taken from the start of flower appearance to the first open flower. It is logical to expect that the earlier the flowering primordia appears in a plant, the earlier it will have an open flower, unless the environmental conditions are not favourable. Rosette leaf number also show a good correlation (> 60%) with the DTF1. Rosette leaves represent the vegetative growth phase, while DTF1 represents the start of reproductive phase. Hence, the more rosette leaves form, the longer it takes for DTF1 to occur. Total leaf number also shows a good correlation with DTF1 because it is the sum of RLN and CLN. Lateral branches and cauline leaves are highly correlated (> 90%) because lateral branches in *Arabidopsis* arise in the axil of the cauline leaves. Life cycle is highly correlated with reproductive period. Silique length and seed numbers (FERT1 and FERT2) are also highly correlated.

In tetraploids, the three flowering traits in **Table 4-3** are highly significantly correlated as also observed in diploids. Rosette leaves show a moderate to high correlation with all the three flowering traits, which is substantially higher than that of diploids. Cauline leaves also are moderately but significantly correlated with flowering traits and therefore TLN shows a high correlation with all the flowering traits. Similarly to diploids, there is a high correlation between lateral branches and cauline leaves, and between life cycle and reproductive period. The correlation between the silique length and seed numbers is also higher than that of diploids. The

relationship for selected traits, which show a significant and high correlation is shown in **Figure 4-9**, which indicate a near linear relationship.

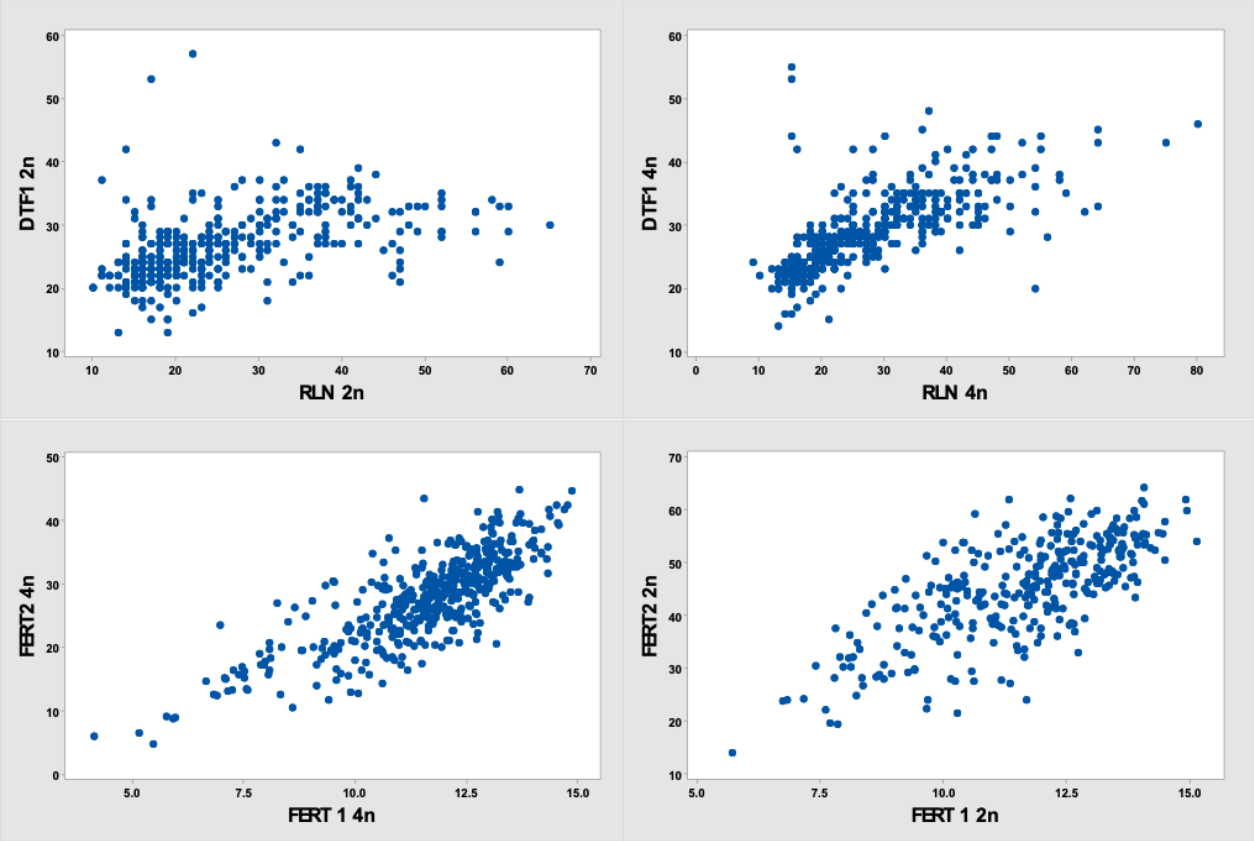


Figure 4-9 Scatter plots between different traits.

Figure shows plots between flowering trait DTF1 against rosette leaves and Fert 1 (silique length) with Fert 2 (seed number) traits in diploids and tetraploids.

	DTG	DTF1	DTF2	DTF3	RLN	CLN	TLN	LB	BB	TB	RP	LC	FERT1
DTF1	-0.156 **												
DTF2	-0.219 ***	0.874 ***											
DTF3	-0.188 ***	0.874 ***	0.958 ***										
RLN	0.152 **	0.606 ***	0.417 ***	0.438 ***									
CLN	0.120 *	0.349 ***	0.391 ***	0.423 ***	0.516 ***								
TLN	0.156 **	0.594 ***	0.433 ***	0.459 ***	0.988 ***	0.625 ***							
LB	0.130 *	0.346 ***	0.382 ***	0.418 ***	0.514 ***	0.955 ***	0.621 ***						
BB	ns	-0.114 *	ns	ns	ns	ns	ns	ns					
TB	ns	0.154 **	0.195 ***	0.231 ***	0.361 ***	0.680 ***	0.442 ***	0.723 ***	0.665 ***				
RP	ns	ns	ns	ns	-0.152 **	ns	-0.156 **	ns	ns	ns			
LC	-0.153 **	0.337 ***	0.350 ***	0.307 ***	ns	ns	ns	ns	ns	ns	0.886 ***		
FERT1	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	
FERT2	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.695 ***

Table 4-2 Correlation coefficient between different traits in diploids.

The Spearman rho correlation coefficient is shown. Text Highlights: Dark Green -Correlation coefficients above 80%, light green- above 60% and yellow – above 50%. *** indicate very highly significant, p-value ≤ 0.001, ** highly significant, p-value ≤ 0.01, * significant, p-value ≤ 0.05, ns denotes not significant.

	DTG	DTF1	DTF2	DTF3	RLN	CLN	TLN	LB	BB	TB	RP	LC	FERT1
DTF1	ns												
DTF2	ns	0.883 ***											
DTF3	-0.098 *	0.863 ***	0.957 ***										
RLN	ns	0.817 ***	0.674 ***	0.667 ***									
CLN	0.115 *	0.422 ***	0.499 ***	0.502 ***	0.404 ***								
TLN	0.096 *	0.820 ***	0.695 ***	0.690 ***	0.992 ***	0.498 ***							
LB	0.126 **	0.409 ***	0.491 ***	0.499 ***	0.394 ***	0.931 ***	0.479 ***						
BB	ns	-0.151 **	-0.128 **	-0.122 **	-0.201 ***	ns	-0.199 ***	-0.097 *					
TB	0.129 **	0.260 ***	0.345 ***	0.355 ***	0.224 ***	0.727 ***	0.295 ***	0.778 ***	0.502 ***				
RP	-0.102 *	-0.224 ***	-0.111 *	-0.114 *	-0.229 ***	-0.118 *	-0.228 ***	-0.102 *	ns	ns			
LC	-0.152 ***	0.300 ***	0.347 ***	0.339 ***	0.181 ***	0.092 *	0.183 ***	0.104 *	ns	ns	0.830 ***		
FERT1	ns	ns	ns	ns	ns	-0.130 **	ns	-0.110 *	ns	-0.104 *	ns	ns	
FERT2	ns	ns	ns	ns	ns	-0.120 *	ns	ns	ns	ns	ns	ns	0.776 ***

Table 4-3 Correlation coefficients between different traits in tetraploids.

The Spearman rho correlation coefficient is shown. Text Highlights: Dark Green -Correlation coefficients above 80%, light green - above 60% and yellow – above 50%. *** indicate very highly significant, p-value ≤ 0.001, ** highly significant, p-value ≤ 0.01, * significant, p-value ≤ 0.05, ns denotes not significant.

4.2.4 Heritability estimates

Broad sense heritability estimates of diploid and tetraploid F2 was carried out for different traits. Variation in phenotype among different individuals in a population is defined as the sum total of genetic and environmental variations between them.

The heritabilities estimated using both the methods as described in methodology (**2.10**) is given below in **Table 4-4**. It shows the heritability estimate using the variation in the non-segregating generations (P1, P2, F1) as an estimate of environmental variance (basic method) and the trait segregation analysis for estimating the major gene effects to analyse the variance components (computational method).

In using the basic method, where the Bartlett test showed a significant difference ($p < 0.05$) in variances between the F1 and both the parental lines, the heritability estimates were calculated using the F1 variance as the environmental variance. Otherwise, where there was no significant difference between the variances of the F1 and parental lines, a weighted average of all three variances was used as an estimate of the environmental variance.

Considering the above factors, for the basic method, it can be concluded that the flowering traits in the diploid F2 (**Table 4-4**) are 40-50% heritable, meaning that the differences in the flowering times in different F2 individuals can be attributed 40-50% to the genotype differences between them. Leaf numbers in the F2 show an almost 30% heritability, while branches show a broad range of heritability estimates between 13-75%, with basal branches

having the highest heritability. Reproductive period and life cycle show heritability estimates of 14 and 20% respectively, while silique length and seed numbers show 53% and 68% respectively.

Trait	Basic Method H^2		Computational Method H^2 (h^2)	
	Diploid	Tetraploid	Diploid	Tetraploid
DTG	0.01	-5.92	0.824 (0.549)	0.806 (0.456)
DTF1	0.42	-0.04	0.426 (0.284)	0.53 (0.324)
DTF2	0.5	-0.39	0.22 (0.147)	0.441 (0.335)
DTF3	0.44	-0.28	0.246 (0.164)	0.674 (0.444)
RLN	0.31	0.26	na	na
CLN	0.08	0.01	na	na
TLN	0.31	0.27	na	na
LB	0.4	0.11	na	na
BB	0.75	0.6	na	na
TB	0.13	0.15	na	na
RP	0.14	-0.0002	0.727 (0.727)	0.839 (0.801)
LC	0.2	0.06	0.696 (0.694)	0.898 (0.844)
FERT1	0.53	0.21	0.566 (0.378)	0.548 (0.362)
FERT2	0.68	0.09	0.543 (0.364)	0.815 (0.809)

Table 4-4 Heritability estimates of diploid and tetraploid F2s using different methods.

Text Highlights: Blue – high negative value, Yellow – high heritability estimates with little or no difference in H^2 and h^2. na denotes not applicable since this method can only be applied to quantitative traits.

For traits such as germination, flowering and reproductive period the estimates of heritability in tetraploids are negative. This is because the variance of F1 or collective variance of P1, P2 and F1 is larger than that of F2 for those traits. Considering these negative values to be 0, it

can be considered that any phenotypic variance observed in germination, flowering and reproductive period traits in tetraploid plants is purely environmental. Using instead the computational method, where the genetic variance could be dissected into its individual additive and dominance effects, narrow sense heritability for the traits could be calculated (**Table 4-4**). Using this method, most of the traits except silique length show higher heritability in tetraploids than diploids. Heritability for seed traits is more than 80% in tetraploids, compared with 36% in diploids, indicating more additive genetic control of trait variation in tetraploids than in diploids.

4.3 Discussion

World population is rising alarmingly and according to the United Nations it is expected to reach more than 9.7 billion by 2050. Food demand is rising rapidly, while agricultural output is not able to keep up the pace. With changes in the climatic conditions and more advent of droughts and floods alike in several places, agricultural practices need to be changed. More production and variety is required within the limited available space to not only feed the growing world, but also for growing uses of bioenergy and other industrial uses. Breeding high yield crops, adapted not only to climatic conditions but also providing more nutrition is the need of the hour (McKersie, 2015). Identifying the quantitative traits and genetic basis of those important traits can help in meeting the aforementioned goals.

Phenotype analysis to identify trait differences between different lines is a straightforward process, which can give a good idea about the variances and heritability of different lines. Several high throughput phenotyping assays have been developed to identify subtle traits, which can be utilized for breeding programmes. For example, a high throughput rice phenotyping facility was developed by integrating a rice automatic plant phenotyping device and a yield trait scorer to analyse 15 traits for which associated loci were identified using GWAS (Yang *et al.*, 2014). They, however, are slow to set up and expensive. Here, we have carried out traditional phenotyping on different *Arabidopsis* lines, which has given useful insights into the differences between the diploid and tetraploid plants. Tetraploids germinated earlier, flowered later (though did not produce more leaves during vegetative

growth), produced fewer basal branches, had a longer reproductive period as well as life cycle, and produced fewer seeds than the diploids, suggesting a reduction in fertility.

Phenotypic differences between diploid and polyploid plants have been studied previously. Tetraploid populations of the grass species *Lolium* were found to have longer and wider leaves than the diploids. It was found that the elongated leaf phenotype was due to an increased elongation rate in tetraploids compared with diploids (Sugiyama, 2005). In *A. thaliana* it has been seen that the differences in organ sizes in polyploids occur in a tissue dependent manner, with flowers and roots being larger in polyploids while leaves though having larger cell size had fewer cells leading to the same final size as in diploids (Del Pozo and Ramirez-Parra, 2014). The study also observed a late flowering phenotype in polyploid plants compared to diploids, with no increase in number of rosette leaves, just as we have seen here. To explain the phenomenon, an increase in the transcript level of Flowering Locus C (*FLC*), a flowering suppressor, and a decrease in the expression levels of floral activator *FT* was found. A similar late flowering phenotype has been observed in maize tetraploids and hexaploids, with reduced fertility as compared with diploids (Yao *et al.*, 2011). Though Del Pozo and Ramirez-Parra (2014) found the polyploid seeds to be dormant and germinate later than the diploid, we observed the opposite, where polyploid seeds germinated significantly earlier. This difference however, can be due to the difference in the time and microenvironment of the F1 and F2 seed production. Diploid F2 population was created earlier (2 years ago) than the tetraploid F2 and there could have been a reduction in their germination efficiency.

In the trial, the tetraploids flowered later and had longer reproductive period, but the seed set was lower, indicating that the fertility for tetraploids is reduced as compared to the diploid plants. This indicates difficulty in meiosis in autotetraploid *Arabidopsis* lines and therefore the time period required to complete the reproductive period increases. These difficulties could be during prophase I, when homologous chromosomes align and recombine. This can lead to improper homologous segregation and hence non-viable gamete formation, which is manifested as reduced seed number. There may well be differences between the diploid and tetraploid flowering time genes. QTL analysis in a cross between early and late flowering diploid accessions in *A. thaliana* identified seven QTLs, with five main QTLs accounting for 62% of the variance, with physiological evidences for interaction between them (Kuittinen, Sillanpää and Savolainen, 1997). This example shows that flowering is a very coordinated process involving the interaction between various genes and the environment, and this process may well be affected by the ploidy of the plant.

Natural genetic variation in *A. thaliana* autotetraploids has been found to affect gene expression changes after autotetraploid formation. It was found that Col 0 but not Ler 0 showed significant alteration in gene expression after tetraploid formation. This alteration was dependent on the developmental stage of the plant and changes in the methylation pattern of the DNA (Yu *et al.*, 2010). Polyploidy has been shown to confer stress resistance to plants and therefore they occur more commonly in extreme environmental conditions (Madlung, 2013). *A. thaliana* tetraploids have been shown to be salt and drought resistant (Del Pozo and Ramirez-Parra, 2014). Tetraploid Rangpur lime rootstock grafted with Valencia Delta sweet orange were found to be more resistant to drought than the diploids. There was

a differential gene expression for drought related genes in roots of the tetraploid plants and the leaves showed lower stomatal conductance (Allario *et al.*, 2013).

The highly significant correlation between leaves and flowers, silique length and seed number in both diploid and tetraploid varieties need to be analysed further at the molecular level to find out genetic component which could indicate the pleiotropy, or the presence of more than one QTL for the genes affecting these traits. For diploids, the correlation ranged from 34 to 60% between different leaves and flowering traits, and 69.5% between silique length and seed numbers, while for tetraploids it ranged from 40 to 81% for leaves with flowering time and 77.6% between silique length and seed numbers, indicating a genetic link between them. The correlation between traits can be used for indirect selection by trying to improve the second trait indirectly by improving the first trait (Lorencetti *et al.*, 2006).

Heritability estimates are a good first indication of the possibility of additive gene effect, which can be useful for artificial selection in breeding programmes. It has been suggested that the heritability of the morphological traits is generally higher than the life history traits (Falconer and Mackay, 1996). Genetic architecture for different traits was analysed in *Brassica napus* where different phenotypes were recorded for double haploid lines. The heritability for the plant architecture related trait was found to be higher than the plant yield traits. Several QTLs were identified and their analysis showed that there was a crosstalk between plant architecture, plant architecture related traits and plant yield traits (Cai *et al.*, 2016). However, in evening primrose *Oenothera biennis,* it was found that both morphological and life history traits were equally heritable (Johnson *et al.*, 2009). In our study, life history traits such as

reproductive period and fertility traits had higher heritability than the morphological traits, such as flowering.

Several horticulturally important traits such as petal numbers have been found to be highly heritable in rose, whereas other morphological traits such as side shoots had low heritability (Gitonga *et al.*, 2014). Heritability of fruit traits in *Capsicum annuum*, the cultivated pepper has been found to be high for various fruit characteristics such as shape and pericarp thickness (Naegele, Mitchell and Hausbeck, 2016). Here, I have identified higher heritability estimates for tetraploids than diploids for most of the traits using a major gene segregation effect analysis method (Chen *et al.*, 2018). This indicates that there are additive gene effects for those traits. In a diploid individual, there may be up to two alleles contributing to the phenotype of a trait. In contrast, there may be up to four different alleles in an autotetraploid genotype. A monogenic effect can be defined for each allele, but there are also multiple levels of interaction between two, three or even all four alleles at a locus (Chen *et al.*, 2018).

Negative heritability can be seen for a few traits in tetraploids including germination and flowering traits when using the basic method. The method assumes that the variance in the segregating generation (F2) will exceed the variance in the non-segregation generation(s), (either F1 or parents) due to the influence of genetic variation. However, we observed that variance in the F1 or the combined variance of the parents and F1 was larger than the variance in the F2 for a few traits. The differences in the variances between the parental and F1 lines can be due to several factors, which could not be controlled experimentally. The plants could not be moved around so they were at one specific location in the growth room throughout

the second trial. This can be the cause of large variances because there can be differences in the amount of water supply, amount of light available and also in the degree of cooling provided by the air vents. Another statistical reason for differences in variances can be the differences in the number of plants for each variety. While 401 F2 seeds were sown, there were only 33 F1 and 28 parental lines. Another important reason can be maternal effect for the plants; F1s used in the experiment are from different mothers whereas the F2s are all from the same F1 family. Maternal effects are defined as "the causal influence of the maternal genotype or phenotype on the offspring phenotype" (Wolf and Wade, 2009). Maternal effects on seeds and their germination is one of the most studied traits in plants. Maternal environmental effect on seed dormancy and germination has been shown in *A. thaliana*. Using recombinant inbred lines (RILs) derived from the cross between two natural accessions of *A. thaliana*, one from Sweden and another from Italy, it was shown that the genetic architecture of dormancy was affected by the maternal environment in which the seeds were produced (Postma and Ågren, 2015). In yet another experiment, maternal effect on the final lipid content in the seeds and subsequent seedling growth was found in *Arabidopsis* by analyzing starch turnover mutants (Andriotis *et al.*, 2012). Maternal effect has also been found to be affecting the gene expression of various genes in hybrid populations of *Arabidopsis lyrata* (Videvall *et al.*, 2015).

The heritability estimates (broad sense) obtained from the method based on F2 segregation analysis (Chen *et al.*, 2018) to estimate the major gene effects are different from those using the basic method. It can be seen (**Table 4-4**) that there is a difference between the diploids and tetraploids, with tetraploids showing a higher heritability for most of the traits except

DTG and LB. The heritability estimated using this method is generally higher than the basic method. This can be due to the initial assumption that there is one major gene present with no linkage disequilibrium, used in model development (Chen *et al.*, 2018). Thus, even if several small effect genes are present, they are highly overlapping giving the overall effect of one major gene. This will cause an overestimation of the genetic variance component, hence a higher estimate for heritability. Nevertheless, heritability estimation is a good initial indication prior to following up on these traits to identify the underlying QTLs and to analyse the differences between the diploids and tetraploids.

In my plant trial, the traits that are significantly different between the diploids and tetraploids will be assessed through sequencing and marker analysis to identify the genetic basis of this difference between the diploids and the tetraploids. DNA extraction has been carried out for more than 200 diploid and 200 tetraploid F2 leaves samples (**Appendix C**). We have received the sequencing raw data for 12 each of the diploid and tetraploid samples which will be analysed in future.

4.4 References

Allario, T. *et al.* (2013) 'Tetraploid Rangpur lime rootstock increases drought tolerance via enhanced constitutive root abscisic acid production', *Plant, Cell and Environment*, 36(4), pp. 856–868. doi: 10.1111/pce.12021.

Alonso-Blanco, C. *et al.* (2009) 'What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation?', *THE PLANT CELL ONLINE*, 21(7), pp. 1877–1896. doi: 10.1105/tpc.109.068114.

Andriotis, V. M. E. *et al.* (2012) 'Altered starch turnover in the maternal plant has major effects on Arabidopsis fruit growth and seed composition.', *Plant physiology*, 160(3), pp. 1175–86. doi: 10.1104/pp.112.205062.

Bevan, M. and Walsh, S. (2005) 'The Arabidopsis genome : A foundation for plant research The Arabidopsis genome : A foundation for plant research', *Genome Research*, 15(12), pp. 1632–1642. doi: 10.1101/gr.3723405.

Cai, G. *et al.* (2016) 'Genetic dissection of plant architecture and yield-related traits in Brassica napus', *Scientific Reports*, 6(1), pp. 21625. doi: 10.1038/srep21625.

Chen, J. *et al.* (2018) 'Orthogonal contrast based models for quantitative genetic analysis in autotetraploid species', *New Phytologist*, 220(1), pp 332-346. doi: 10.1111/nph.15284.

Falconer, D. S. and Mackay, T. F. C. (1996) *Introduction to Quantitative Genetics (4th Edition)*, *Trends in Genetics*. Available at: http://www.amazon.com/Introduction-Quantitative-Genetics-Douglas-Falconer/dp/0582243025.

Gitonga, V. W. *et al.* (2014) 'Genetic variation, heritability and genotype by environment interaction of morphological traits in a tetraploid rose population', *BMC Genetics*, 15(1), pp. 146. doi: 10.1186/s12863-014-0146-z.

Holland, J. B. (2007) 'Genetic architecture of complex traits in plants', *Current Opinion in Plant Biology*, 10(2) pp. 156–161. doi: 10.1016/j.pbi.2007.01.003.

Johnson, M. T. J. *et al.* (2009) 'Heritability, covariation and natural selection on 24 traits of common evening primrose (Oenothera biennis) from a field experiment', *Journal of Evolutionary Biology*, 22(6), pp. 1296–1307. doi: 10.1111/j.1420-9101.2009.01747.x.

Juenger, T., Purugganan, M. D. and Mackay, T. F. C. (2000) 'Quantitative trait loci for floral morphology in Arabidopsis thaliana', *Genetics*, 156(3), pp. 1379–1392. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461322&tool=pmcentrez&rendertype=abstract.

Kuittinen, H., Sillanpää, M. J. and Savolainen, O. (1997) 'Genetic basis of adaptation: Flowering time in Arabidopsis thaliana', *Theoretical and Applied Genetics*, 95(4), pp. 573–583. doi: 10.1007/s001220050598.

Lorencetti, C. *et al.* (2006) 'Aplicability of phenotypic and canonic correlations and path coefficients in the selection of oat genotypes', *Scientia Agricola*, 63(1), pp. 11–19. doi: 10.1590/S0103-90162006000100003.

Madlung, A. (2013) 'Polyploidy and its effect on evolutionary success: old questions revisited with new tools', *Heredity*, 110(2), pp. 99–104. doi: 10.1038/hdy.2012.79.

McKersie, B. (2015) 'Planning for food security in a changing climate', *Journal of Experimental Botany*, 66(12), pp. 3435–3450. doi: 10.1093/jxb/eru547.

Naegele, R. P., Mitchell, J. and Hausbeck, M. K. (2016) 'Genetic diversity, population structure, and heritability of fruit traits in Capsicum annuum', *PLoS ONE*, 11(7). doi: 10.1371/journal.pone.0156969.

Postma, F. M. and Ågren, J. (2015) 'Maternal environment affects the genetic basis of seed dormancy in Arabidopsis thaliana', *Molecular Ecology*, 24(4), pp. 785–797. doi: 10.1111/mec.13061.

Del Pozo, J. C. and Ramirez-Parra, E. (2014) 'Deciphering the molecular bases for drought tolerance in Arabidopsis autotetraploids', *Plant, Cell & Environment*, 37(12), pp. 2722–2737. doi: 10.1111/pce.12344.

Salomé, P. A. *et al.* (2011) 'Genetic architecture of flowering-time variation in Arabidopsis thaliana', *Genetics*, 188(2), pp. 421–433. doi: 10.1534/genetics.111.126607.

Sugiyama, S. I. (2005) 'Polyploidy and cellular mechanisms changing leaf size: Comparison of diploid and autotetraploid populations in two species of Lolium', *Annals of Botany*, 96(5), pp. 931–938. doi: 10.1093/aob/mci245.

Videvall, E. *et al.* (2015) 'Strong maternal effects on gene expression in Arabidopsis lyrata hybrids', *Molecular Biology and Evolution*, 33(4), pp. 984-994. doi: 10.1093/molbev/msv342.

Wolf, J. B. and Wade, M. J. (2009) 'What are maternal effects (and what are they not)?', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1520), pp. 1107–1115. doi: 10.1098/rstb.2008.0238.

Yang, W. *et al.* (2014) 'Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice.', *Nature communications*, 5(1), p. 5087. doi: 10.1038/ncomms6087.

Yao, H. *et al.* (2011) 'Phenotypic and gene expression analyses of a ploidy series of maize

inbred Oh43', *Plant Molecular Biology*, 75(3), pp. 237–251. doi: 10.1007/s11103-010-9722-4.

Yu, Z. *et al.* (2010) 'Impact of natural genetic variation on the transcriptome of autotetraploid Arabidopsis thaliana', *Proceedings of the National Academy of Sciences of the United States of America*, 107(41), pp. 17809–17814. doi: Doi 10.1073/Pnas.1000852107.

CHAPTER 5

COMPARATIVE CHIASMA ANALYSIS IN DIPLOID VS TETRAPLOID *ARABIDOPSIS THALIANA* F2

5 Comparative chiasma analysis in diploid vs tetraploid *Arabidopsis thaliana* F2

5.1 Introduction

Buds from the plants grown in the second trial were collected and F2 buds were analysed for counting the number of chromosomes to establish the stability of the autotetraploids as well as analyse chiasma frequency for different F2s. FISH using 5S and 45S probes to count the number of crossovers has worked well in *Arabidopsis thaliana* (hereafter referred to as *A. thaliana* or *Arabidopsis*). The method was used to analyse the crossovers in parental lines as described in Chapter 3 (**3.2.2.3**). Similar analysis was carried out on F2s, where the bivalent or the multivalent shape along with the probe position in metaphase I (M1) helped identify the chiasma number and position (short or long arm) on the chromosome. Higgins *et al.* (2012) analysed the number and position of chiasma in barley using a similar kind of analysis. It has been utilised for comparing different *Arabidopsis* lines as well as comparing wild type and mutant (Sanchez-Moran *et al.*, 2002). This method can also help to identify the proximal or the distal location of crossovers, though, this kind of analysis has not been undertaken in this study.

5.2 Results

A chromosome count by preparing chromosomal spreads and DAPI staining was carried out for 50 4n F2 plants to confirm their genome stability. Clear mitotic cells in pre metaphase I or M1 stage or meiotic cells in anaphase I or metaphase II (M2) stages were used for counting. 39 plants were found to give a full count of 20 in at-least 10 cells. The distribution of chromosomal counts can be seen in **Figure 5-1**.

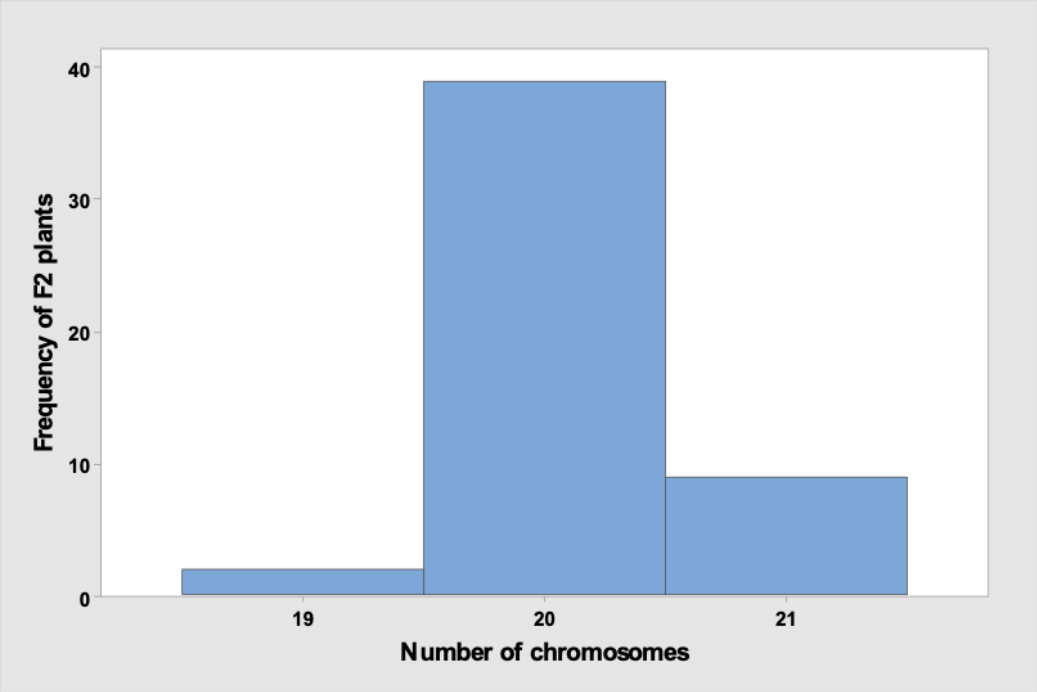


Figure 5-1 Arabidopsis thaliana F2 plant chromosomal count distribution.

After establishing 20 chromosomes in 78% (39 out of 50) of the 50 tetraploid F2s counted using DAPI staining, FISH was carried out on ten of those plants where clear M1s were obtained. The number of M1 cells analysed for different plants ranged from 8 to 45. Though more M1s were obtained, only those cells where a difference between bivalents and quadrivalents could be ascertained were selected for the analysis. A few M1s were also lost in the washing process of the FISH protocol. Out of the ten plants, probes did not work in one

plant indicating failure of the process. In the end, there were 9 plants for which a complete analysis could be carried out. Chromosome spreads were made for diploid F2s for comparison and two plants provided a good number of M1s. They were analysed for chiasma frequency using FISH probes to compare and contrast the difference in cross overs between the diploid and tetraploid F2s.

5.2.1 Chiasma analysis in tetraploid F2s

Chiasma analysis in 4n F2 plant 168

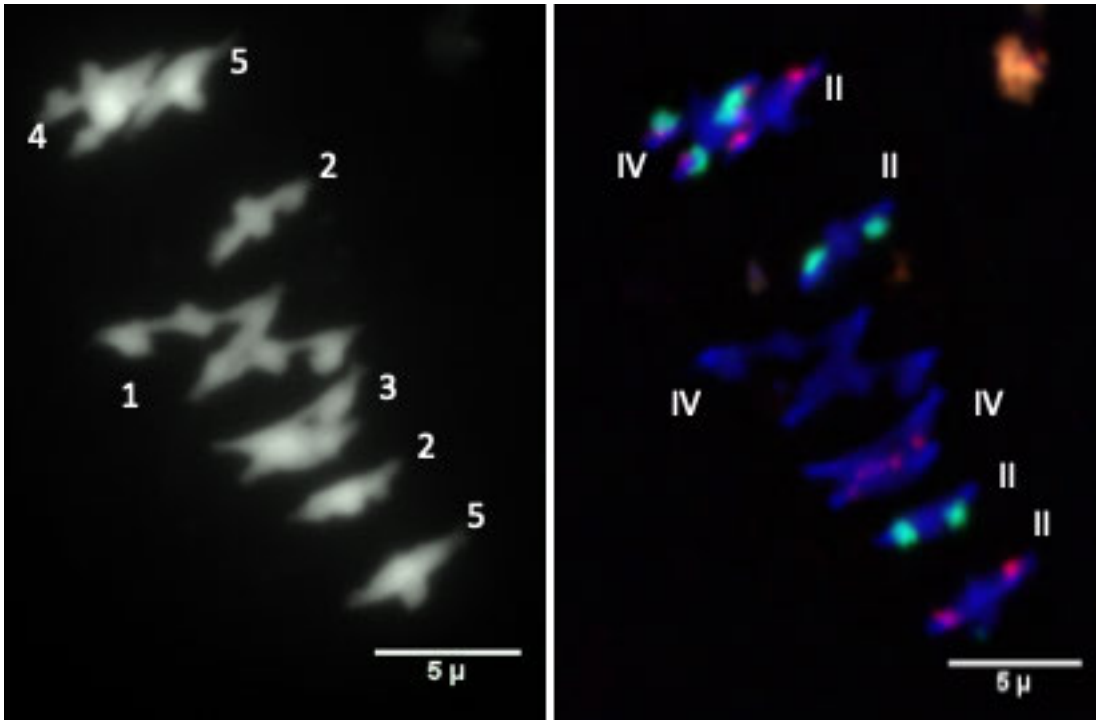


Figure 5-2 Comparison between cells in M1 in tetraploid F2 168.

Figure shows a DAPI stained M1 cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: II indicates bivalent, and IV indicates quadrivalent. Scale bar is 5 μm.

Figure 5-2 represents M1 in one of the F2s (168). A mix of bivalents and quadrivalents is seen, as was also seen in the tetraploid parental lines (**3.2.2.3**). Chromosome 1 as visible in the left

and right side of the figure, occurs as a chain quadrivalent with the two bivalents attached by the short arms. Chromosome 2 shows two separate rod bivalents, each with a chiasma in its longer arm. Chromosome 3 is a ring quadrivalent having at-least 4 chiasmata, one in each arm. Chromosome 4 is a chain quadrivalent with 2 chiasmata in the long arm and one chiasma in the short arms through which the bivalents are attached. Chromosome 5 shows 2 ring bivalents, with one chiasma in short and two in long arms for both bivalents. Thus, the cell has a total of 18 chiasmata. For this F2 (individual 168), 13 cells were counted and the mean chiasma number was found to be 15.92. Chiasma count was lower in the short arms than the long arms for all of the chromosomes, with the difference between the arms being smallest for chromosome 1 and largest for chromosome 2 (**Table 5-1**). Multivalents, mostly in the form of rings, chains or trivalents were present in all the chromosomes, as can be seen in **Table 5-2** and **Table 5-3**. Chromosome 5 showed the highest number of multivalents followed by chromosomes 1, 4, 3 and 2 respectively. The chromosomes 2 and 3 showed the most frequent univalents, indicating loss of the obligatory crossover.

Chiasma analysis in 4n F2 plant 412

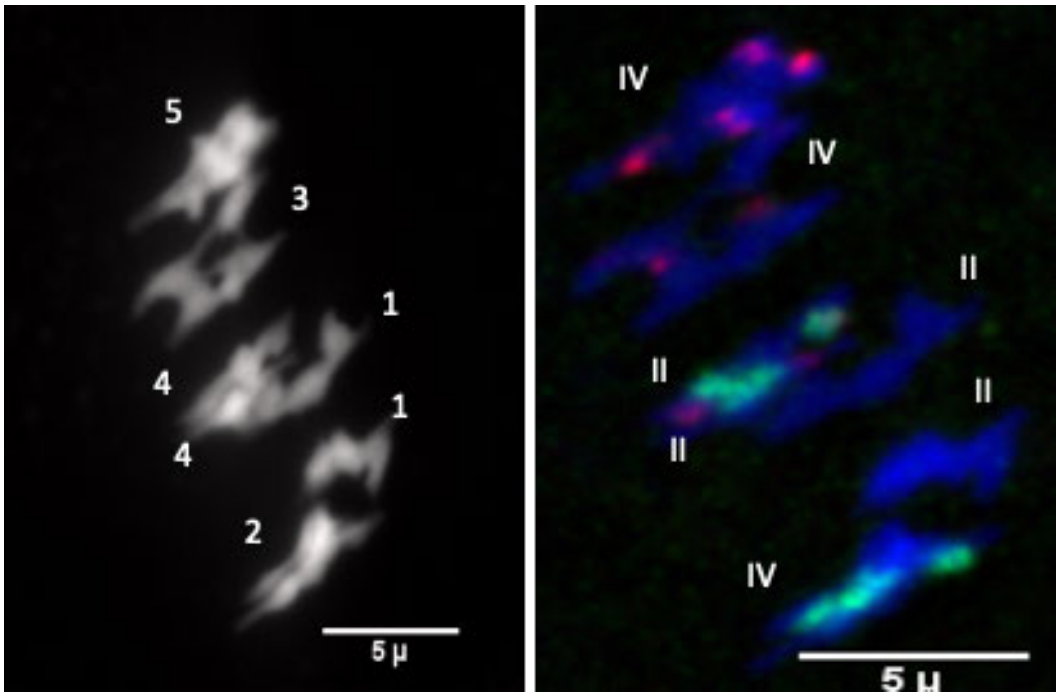


Figure 5-3 Comparison between cells in M1 in tetraploid F2 412.

The figure shows a DAPI stained cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: II indicates bivalent and IV indicates quadrivalent. Scale bar is 5 μm.

Figure 5-3 shows one representative M1 cell in another F2 (412) plant. A mix of bivalents and quadrivalents is present here, similar to plant 168. There are two bivalents of chromosome 1, one is a ring and the other is a rod having 2 and 1 chiasma respectively. Chromosome 2 shows a chain quadrivalent with 3 chiasmata, with the two bivalents attached by the short arms. Chromosome 3 is a ring quadrivalent having at-least 4 chiasmata, one in each arm. Chromosome 4 has 2 bivalents, one being a rod bound in the long arms and the other a ring bound in both long and short arms. Chromosome 5 again is a ring quadrivalent with at-least 4 chiasmata. Thus, the cell has a total of 17 chiasmata. For this F2 (individual 412), 41 cells were counted and the mean chiasma number was found to be 15.9. As seen in **Table 5-1**,

chromosome 1 has the highest mean chiasma and chromosome 4 has the fewest. Here again, the longer arms have more chiasmata than the smaller arm of the chromosomes and the difference is smallest for chromosome 1 and highest for chromosome 4. The number and percentage of multivalents formed can be seen in **Table 5-2**. Chromosomes 1 and 2 had the highest number of multivalents followed by chromosomes 3, 5 and 4 respectively. This is unusual for chromosome 2, considering it is one of the smaller acrocentric chromosome. Quadrivalents mostly occurred as rings or chains (**Table 5-3**).

Chiasma analysis in 4n F2 plant 466

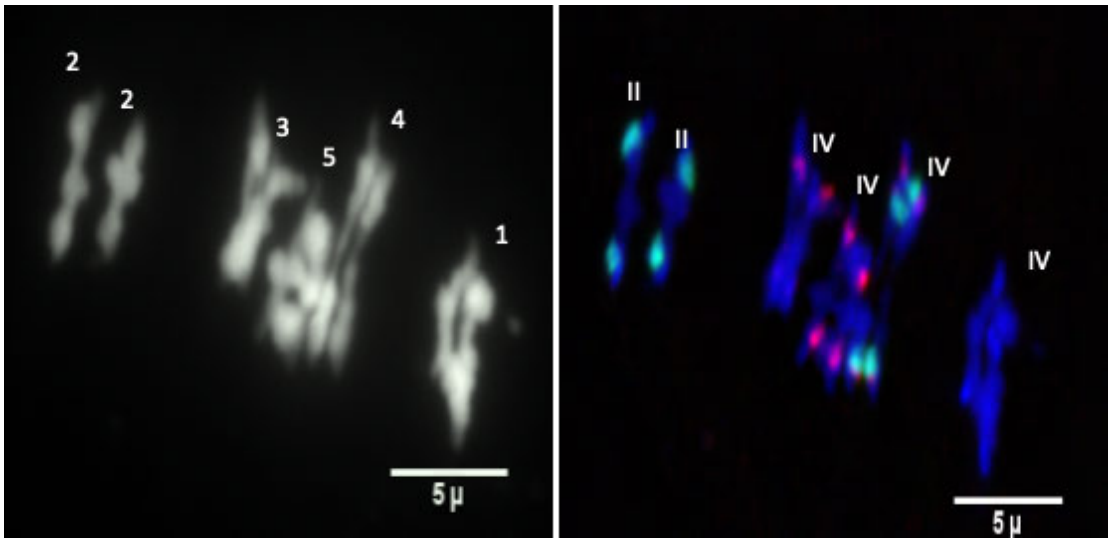


Figure 5-4 Comparison between cells in M1 in tetraploid F2 466.

The figure shows a DAPI stained cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: II indicates bivalent and IV indicates quadrivalent. Scale bar is 5 μm.

Figure 5-4 represents an M1 cell from another tetraploid F2 plant 466. Chromosome 1 presents a ring quadrivalent configuration with 2 chiasmata each in long and short arms. There are two rod bivalents for chromosome 2, each with one chiasma in the long arm.

Chromosomes 3 and 4 occur as chain quadrivalents with 3 chiasmata each, and chromosome 5 presents a ring quadrivalent. There are a total of 16 chiasmata for this cell. For F2 466, only 7 cells could be analysed, which showed a mean chiasma frequency of 16.57. As with other F2 plants (168 and 412), this plant also had more chiasmata in the longer arm than the smaller arm for all the chromosomes, with the lowest difference being in chromosome 1. Individual chromosome chiasma frequency can be seen in **Table 5-1**. Chromosome 1 has the highest chiasma frequency followed by chromosomes 5, then 3, 4 and smallest is for chromosome 2. Multivalents in the form of quadrivalents and trivalents occurred for all the five chromosomes (**Table 5-2** and **Table 5-3**). Chromosomes 1 and 5 had the same number of quadrivalents, with no univalents. Chromosome 2 had the next highest number of multivalents, but the lowest chiasma frequency. This was followed by 3 multivalents each for chromosomes 3 and 4, though for chromosome 3, all 3 were quadrivalents, while in chromosome 4, one cell had trivalent - univalent combination. Chromosome 2 had more multivalents than the sub-metacentric chromosome 3.

Chiasma analysis in 4n F2 plant 468

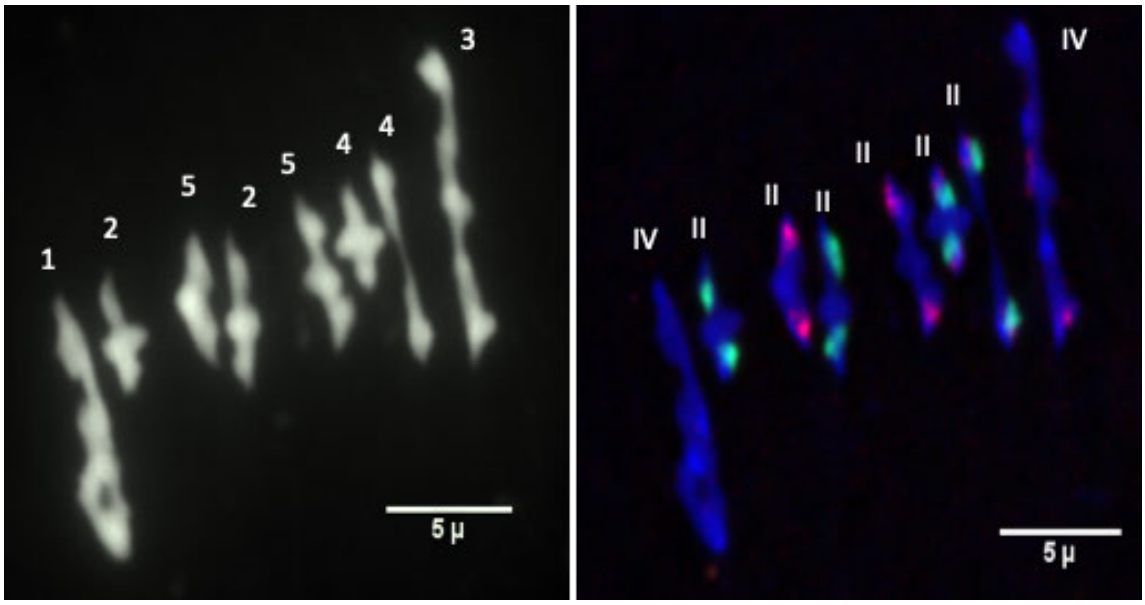


Figure 5-5 Comparison between cells in M1 in tetraploid F2 468.

The figure shows a DAPI stained cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: II indicates bivalent and IV indicates quadrivalent. Scale bar is 5 μm.

Another tetraploid F2 plant 468, was analysed and a representative M1 can be seen in **Figure 5-5.** It shows chromosomes 1 and 3 as spoon and chain quadrivalents with 4 and 3 chiasmata respectively, and all the other chromosomes as bivalents, mostly rods, while only one bivalent of chromosome 5 is a ring, with a chiasma in each arm. 17 M1 cells were analysed for plant 468, which gave a mean chiasma frequency of 15.71 (**Table 5-1**). Chromosome 1 shows the highest chiasma frequency followed by chromosomes 5, 3, 2 and 4. The difference between the chiasma frequency in long and short arms was smallest for chromosome 5 and highest for chromosome 4. For this plant chromosome 5 showed the highest number of multivalents followed by chromosomes 1, 4, 3 and 2 and chromosome 2 had the most univalents (**Table 5-2**). Quadrivalents occurred as rings, chain and spoon (**Table 5-3**).

Chiasma analysis in 4n F2 plant 471

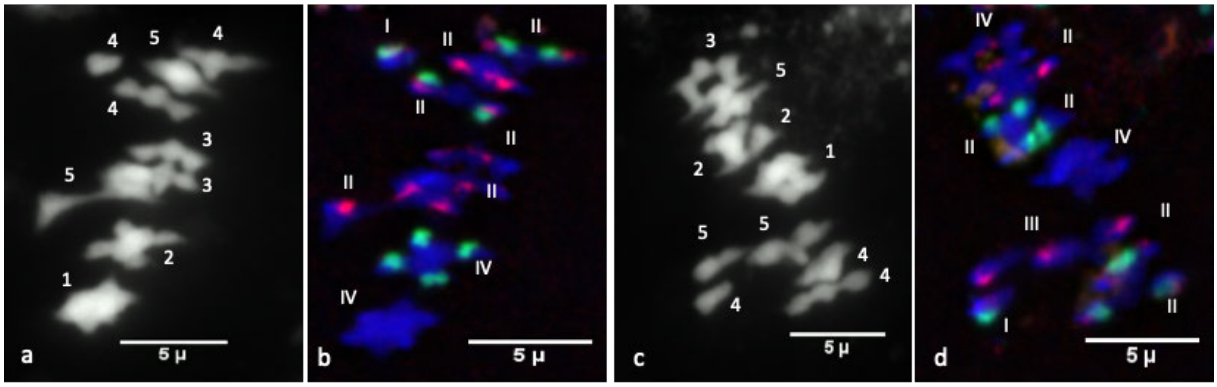


Figure 5-6 Comparison between 2 M1 cells in tetraploid F2 471.

The figure shows DAPI stained cells in a and c and the same cells also showing 5S (red) and 45S (green) FISH probes in b and d. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: II indicates bivalent and IV indicates quadrivalent. Scale bar is 5 µm.

Two cells in M1 are shown in **Figure 5-6** for plant 471. In panels a and b, chromosomes 1 and 2 are present as ring quadrivalents. Chromosome 3 has 2 rod bivalents and chromosome 5 has one rod and one ring bivalent. Chromosome 4 however, shows two rod bivalents and one univalent. This indicates the presence of 21 chromosomes leading to 4n+1 aneuploidy. In panels c and d, another M1 cell is represented, where again chromosome 1 is a ring quadrivalent, chromosome 2 has two bivalents, chromosome 3 is a ring quadrivalent, but chromosome 4 has a univalent in addition to 2 bivalents. In this cell, even chromosome 5 shows one trivalent and one bivalent. This indicates 4n+2 aneuploidy with two different chromosomes occurring in sets of 2 + 1 or 5 chromosomes each. Several cells for this plant showed this kind of behaviour wherein a few cells had an extra chromosome 4 while others had extra copies of both chromosomes 4 and 5.

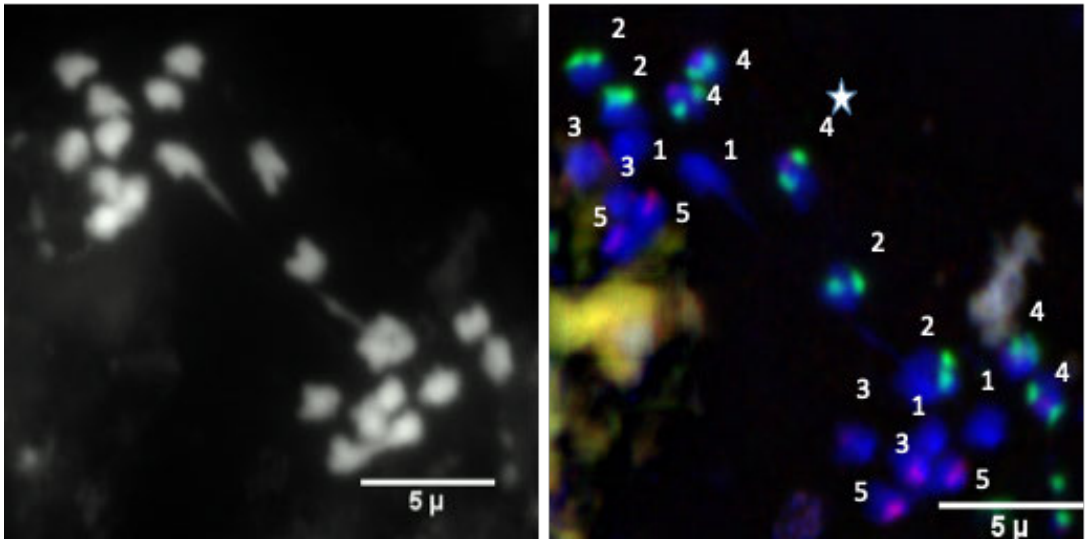


Figure 5-7 Comparison between cells in Anaphase I (A1) in tetraploid F2 471.

The figure shows a DAPI stained A1 cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. Numbers represent chromosome numbers. .* indicates extra copy of chromosome 4. Scale bar is 5 µm.

Another cell in anaphase I (A1) stage was checked to confirm what kind of aneuploidy was present. In **Figure 5-7**, homologues separating from each other in A1 stage can be seen. Apart from the usual chromosome complement, an extra chromosome 4 is visible indicated by a star. This indicates 4n+1 aneuploidy with an extra chromosome 4. To further confirm this behaviour, one of the M2 cells in the plant was analysed. It can be seen in **Figure 5-8**, that the cells have segregated as 11 and 11 instead of 10 and 10, and while on one side there are 3 copies of chromosome 4, on the other side there are 3 copies of chromosome 5. This indicates presence of one extra chromosome 4 as well as one extra chromosome 5, and 4n+2 aneuploidy. This plant was not used further for chiasma analysis as there was confusion about the presence of either one or two extra chromosomes.

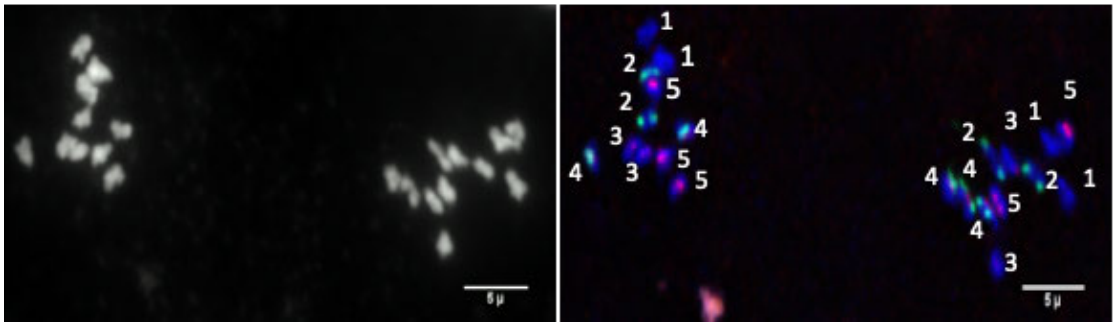


Figure 5-8 Comparison between cells in Metaphase II (M2) in tetraploid F2 471.

The figure shows a DAPI stained cell (left) and the same cell (right) in M2 also showing 5S (red) and 45S (green) FISH probes. Numbers represent chromosome numbers. Scale bar is 5 μm.

Chiasma analysis in 4n F2 plant 473

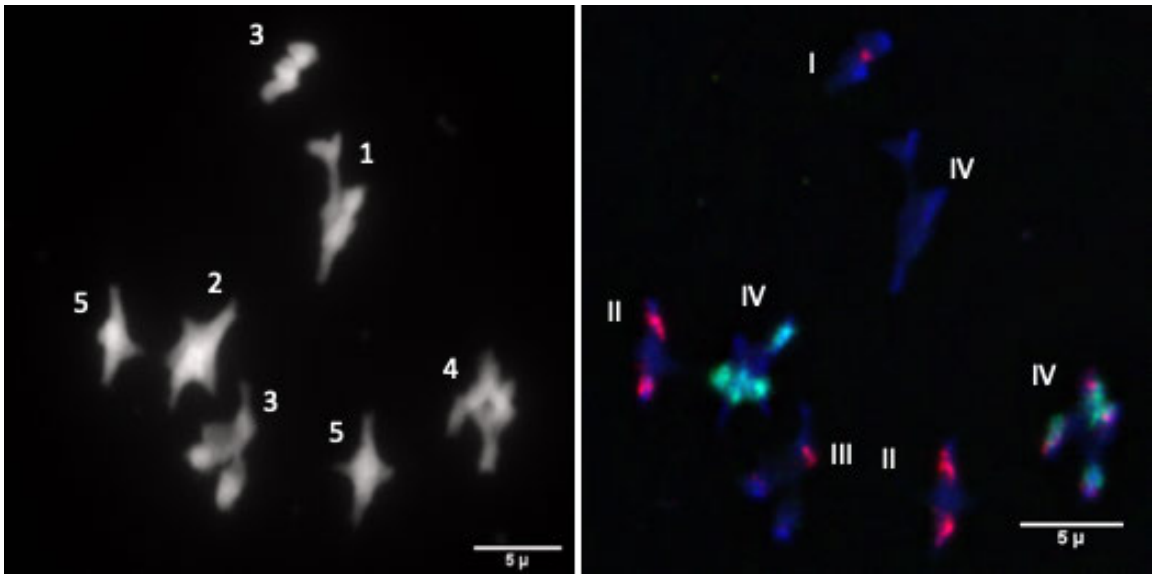


Figure 5-9 Comparison between cell in M1 in tetraploid F2 473.

The figure shows a DAPI stained M1 cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: I indicates univalent, II indicates bivalent, III indicates trivalent and IV indicates quadrivalent. Scale bar is 5 μm.

Figure 5-9 shows one M1 cell from yet another F2 (473). Chromosome 1 is a chain quadrivalent with 3 chiasmata, chromosome 2 is a ring quadrivalent with 4 chiasmata, while chromosome 3 shows a trivalent and a univalent indicating difficulties during meiosis. Chromosome 4 shows

a ring quadrivalent, while chromosome 5 has 2 ring bivalents. This cell thus has 16 chiasmata in total. For this plant, most of the metaphasic cells were lost during the FISH process. Only 8 cells could be analysed, which produced a mean chiasma of 16.13. Again, the smaller arm of the chromosomes had lower chiasma frequency than the longer arms, and the difference was lowest for chromosome 5 and highest for chromosome 2. Multivalent formation occurred for all of the chromosomes. In this plant, chromosome 1 presented quadrivalents in all 8 cells. Chromosome 3 produced the next highest number, with 5 multivalents (including 2 cells which formed a trivalent configuration). Chromosomes 2, 4 and 5 all had 4 quadrivalents (**Table 5-2**).

Chiasma analysis in 4n F2 plant 956

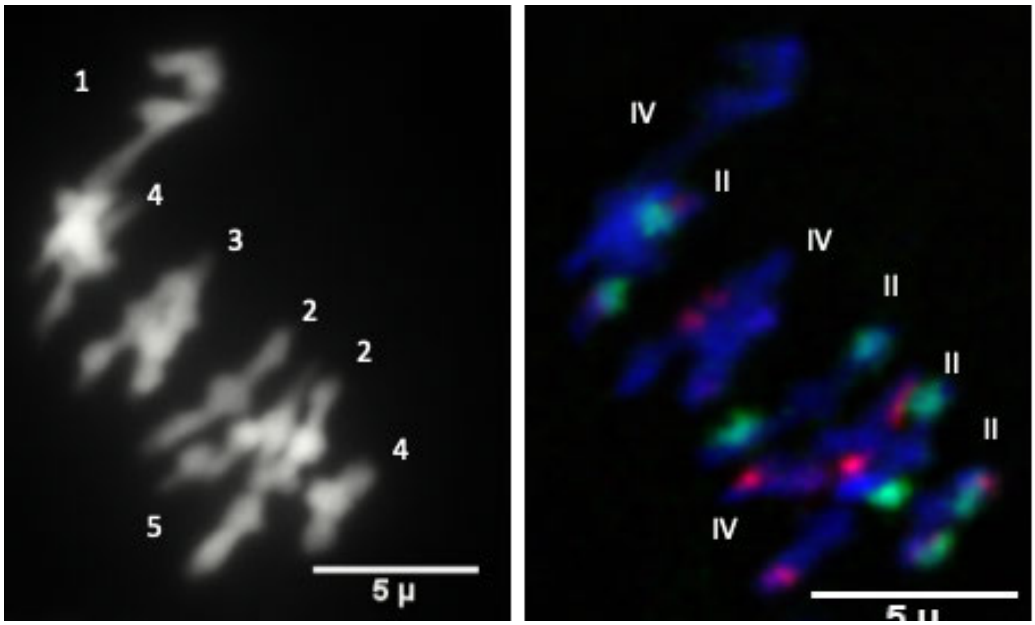


Figure 5-10 Comparison between a cell in M1 in tetraploid F2 956.

The figure shows a DAPI stained M1 cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: II indicates bivalent and IV indicates quadrivalent. Scale bar is 5 μm.

Figure 5-10 shows another M1 representative from another F2 plant (956). Chromosome 1 appears as a chain quadrivalent and chromosome 2 shows two rod bivalents with a chiasma each in the longer arm. Chromosome 3 is also a clear chain quadrivalent with 3 chiasmata and attached by the short arms. Chromosome 4 has 2 rod bivalents bound in the long arms. Chromosome 5 shows a chain quadrivalent with 3 chiasmata. For F2 956, 43 M1 cells were counted and the mean chiasma number was found to be 15.79, with chromosome 1 having the highest frequency and chromosome 4 the lowest (**Table 5-1**). The difference in the chiasma frequencies between the long and short arms was lowest for chromosome 1 and highest for chromosome 2. Multivalents occurred in all chromosomes, with chromosome 1

having the highest number followed by chromosomes 5, 2, 4 and 3 (**Table 5-2**). Chromosome 4 showed the highest number of trivalents.

Chiasma analysis in 4n F2 plant 958

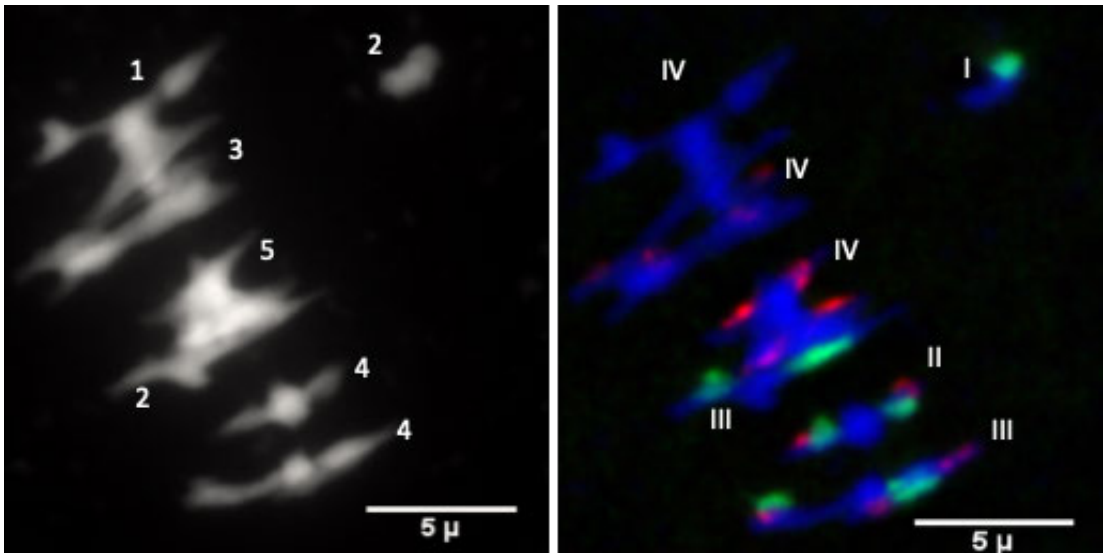


Figure 5-11 Comparison between a cell in M1 in tetraploid F2 958.

The figure shows a DAPI stained M1 cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: I indicates univalent, II indicates bivalent, III indicates trivalent and IV indicates quadrivalent. Scale bar is 5 µm.

Figure 5-11 represents an M1 cell from another F2, 958. Chromosome 1 is a ring quadrivalent with at-least 4 chiasmata. Chromosome 2 shows a trivalent and a univalent, chromosome 3 is ring quadrivalent again and chromosome 5 is also a ring quadrivalent. Chromosome 4 here shows a bivalent as well as a trivalent, indicating the presence of an extra chromosome 4 indicating a trisomic autotetraploid 4n+1. Several cells which could be analysed for M1 in plant 958 showed this kind of 4n+1 behaviour for chromosome 4, except for two cells where either two bivalents, one quadrivalent or a trivalent and a univalent were recorded. An example of

such a cell is shown in **Figure 5-12**, where chromosome 4 appears as a chain quadrivalent connected by short arms and all the chromosome complements look complete.

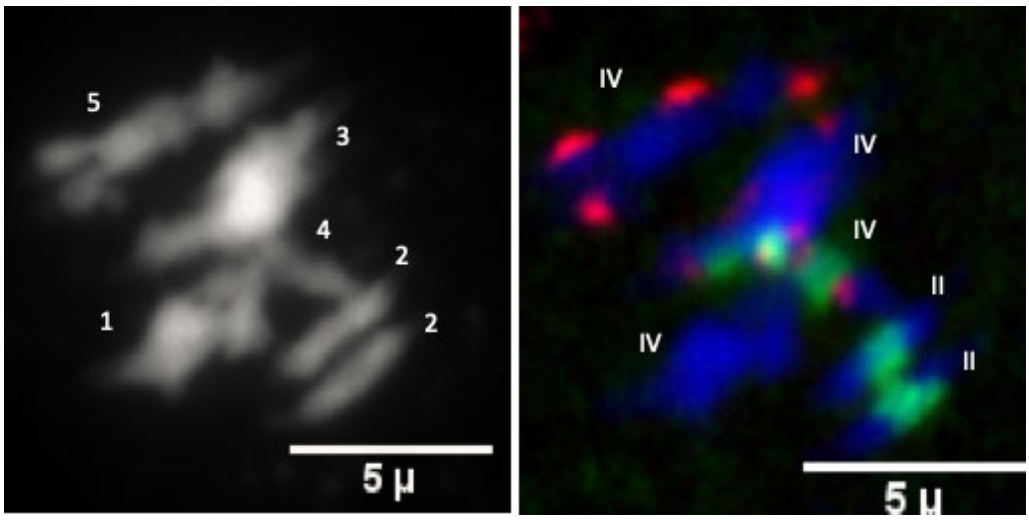


Figure 5-12 Comparison between a cell in M1 in tetraploid F2 958.

The figure shows a DAPI stained M1 cell and same cell also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: II indicates bivalent and IV indicates quadrivalent. Scale bar is 5 μm.

To confirm what was happening in the plant, mitotic metaphase cells which could be counted after FISH were analysed. One such cell is shown in **Figure 5-13**, where one extra chromosome 4 could be counted. In yet another mitotic cell however, five copies of chromosome 4 but only three of chromosome 3 can be identified, indicating compensation where the chromosome complement of 20 is maintained even though the right set of chromosomes may not be present. It is shown in **Figure 5-14,** where chromosome 4 also indicated by stars can be counted to be five in number whereas chromosome 3 also indicated by an asterisk has only count of three.

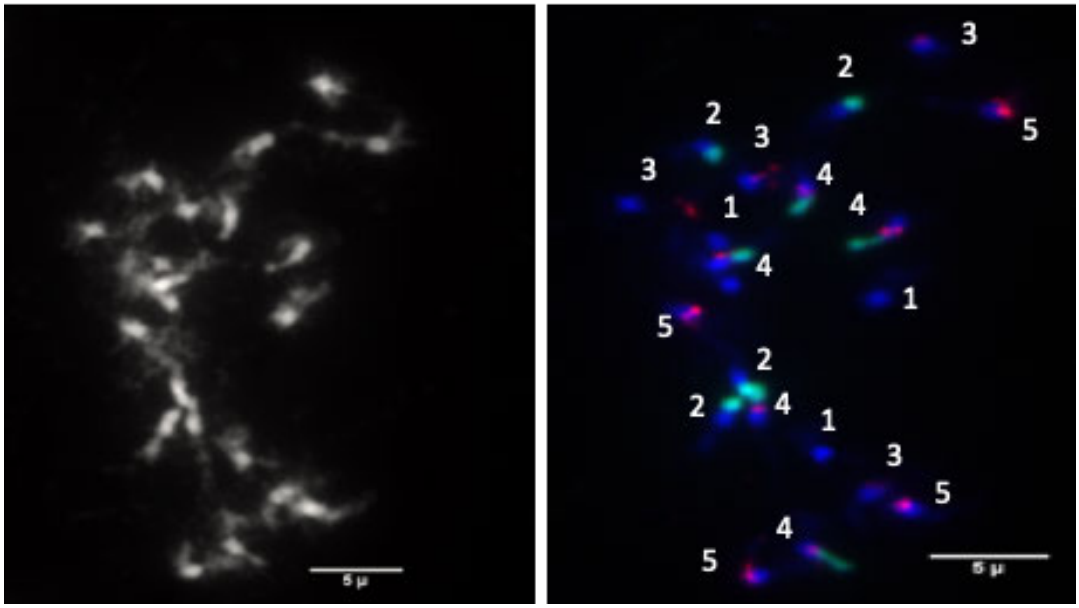


Figure 5-13 Comparison between a mitotic cell in tetraploid F2 958.

The figure shows DAPI stained mitotic cell and the same cell also showing 5S (red) and 45S (green) FISH probes. The numbers indicate chromosome number. Scale bar is 5 µm.

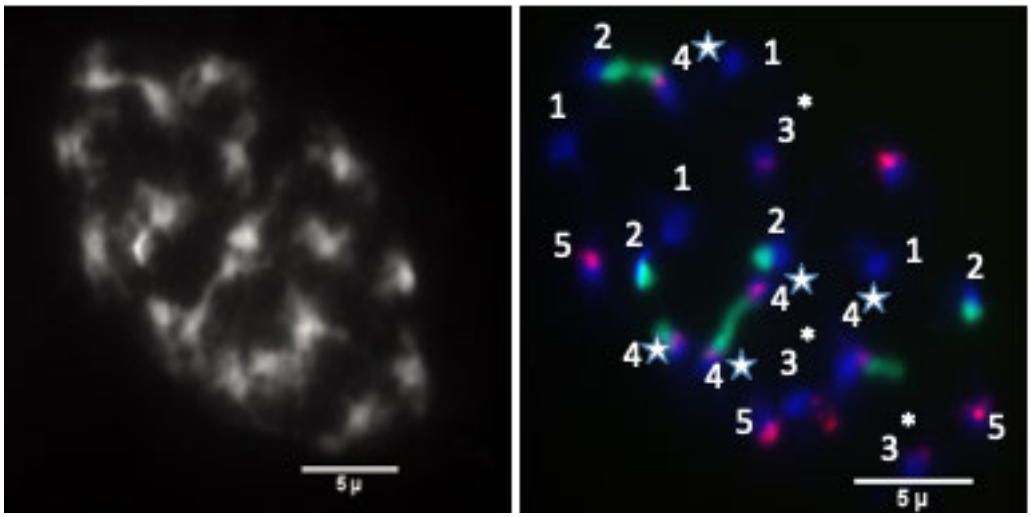


Figure 5-14 Comparison between another mitotic cell in tetraploid F2 958.

The figure shows a DAPI stained mitotic cell and the same cell also showing 5S (red) and 45S (green) FISH probes. The numbers are chromosome numbers. Star indicates chromosome 4 and * indicates chromosome 3. Scale bar is 5 µm.

In yet another mitotic cell, a full complement of 5 chromosomes could be counted as shown in **Figure 5-15**. In this mitotic cell, all the chromosomes have four homologues where correct

signals are observed. A clear conclusion over whether one extra chromosome 4 was present could not be reached. Considering these anomalies, the chiasma count analysis was not included from this plant.

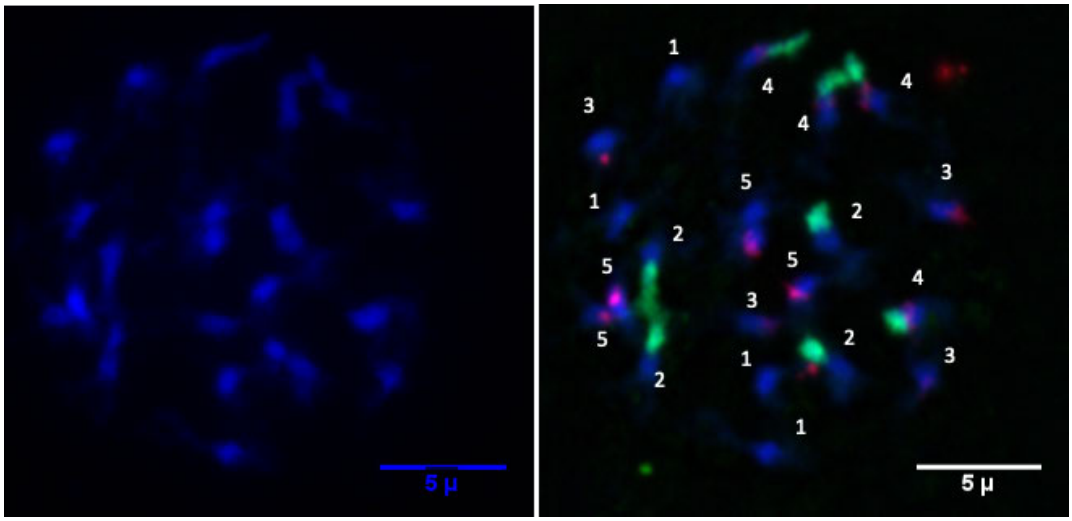


Figure 5-15 Comparison between mitotic cell in tetraploid F2 958 showing correct set of homologues.

The figure shows a DAPI stained mitotic cell and the same cell also showing 5S and 45S FISH probes. Numbers indicate chromosome number. Scale bar is 5 μm.

Chiasma analysis in plant 964

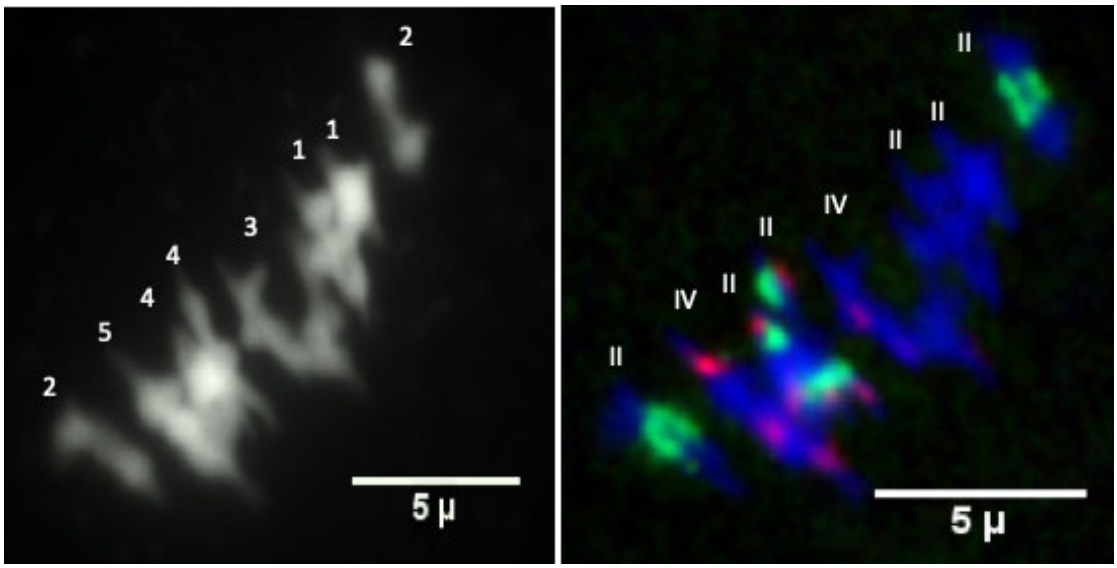


Figure 5-16 Comparison between a cell in M1 in tetraploid F2 964.

The figure shows a DAPI stained M1 cell (left) and the same cell (right) also showing 5S (red) and 45S (green) FISH probes. The numbers in the left represent the chromosome numbers while the Roman numerals on the right indicate the configuration: II indicates bivalent and IV indicates quadrivalent. Scale bar is 5 μm.

Figure 5-16 shows an M1 cell from another F2 plant 964. Two ring bivalents with at-least 1 chiasma each in the small and the long arm can be seen for chromosome 1. Chromosome 2 also shows two ring bivalents with chiasmata in both the arms for both the bivalents. Chromosome 3 represents a chain quadrivalent with 2 chiasmata in the long arms and 1 chiasma in the short arms. Chromosome 4 shows two rod bivalents each with a chiasma in the long arms. Chromosome 5 shows a ring quadrivalent with at-least 2 chiasmata each in small as well as long arms. For this F2 (individual 964), 12 cells were analysed and the mean chiasma number for 12 cells was equal to 15.25. Chromosome 1 has the highest chiasma frequency and chromosome 2 has the lowest (**Table 5-1**). The difference in the chiasma frequency between the long and short arms was similar among different chromosomes, though lowest for chromosomes 1 and 5 and highest for chromosome 2. Multivalents formed in all the

chromosomes, with chromosome 1 being the highest scorer and chromosome 2 scoring the lowest number (**Table 5-2**).

Chiasma scoring for each chromosome in different tetraploid F2 lines is presented in Table **5-1**. It is a conservative scoring as was done for the parental lines for different number of cells as indicated below the plant number. The percentage of quadrivalents, trivalents and univalents scored in each tetraploid F2 can be seen in **Table 5-2**. The highest percentage of quadrivalent formation is mostly seen for one of the two bigger chromosomes (1 or 5) with a low level of trivalent or univalent formation in all of the F2 lines analysed.

All the seven tetraploid F2s analysed showed a total mean chiasma of 15.89 ± 1.54, which lies in between the mean chiasma count of tetraploid parents Landsberg 15.35 ± 1.75 and Columbia 16.74 ± 1.35.

	Chr I	Chr II	Chr III	Chr IV	Chr V	Total
168 4n (n=13)	3.85 (1.77 2.1)	2.38 (0.54 1.85)	3.15 (1.23 1.92)	2.85 (1 1.85)	3.69 (1.54 2.15)	15.92
412 4n (n=41)	3.71 (1.61 2.1)	3 (1.1 1.93)	3.24 (1.27 1.98)	2.8 (0.8 2)	3.15 (1.17 1.98)	15.90
466 4n (n=7)	3.86 (1.71 2.14)	2.71 (0.86 1.86)	3.43 (1.43 2)	2.86 (1 1.86)	3.71 (1.57 2.14)	16.57
468 4n (n=17)	3.65 (1.58 2.06)	2.82 (1.06 1.76)	3.11 (1.18. 1.94)	2.65 (0.71 1.94)	3.47 (1.52 1.94)	15.71
473 4n (n=8)	3.38 (1.38 2)	2.75 (0.75 2)	2.88 (1.13 1.75)	3.25 (1.25 2)	3.88 (1.75 2.13)	16.13
956 4n (n=43)	3.56 (1.6 1.95)	2.93 (0.93 2)	3.14 (1.19 1.95)	2.72 (0.81 1.91)	3.44 (1.49 1.95)	15.79
964 4n (n=12)	3.75 (1.75 2)	2.5 (0.67 1.83)	3.17 (1.17 2)	2.42 (0.67 1.75)	3.41 (1.58 1.83)	15.25
4n Avg	3.68 (1.63 2.05)	2.73 (0.84 1.89)	3.16 (1.23 1.93)	2.79 (0.89 1.9)	3.54 (1.52 2.02)	15.89
205 2n (n=98)	1.97 (0.98 0.99)	1.38 (0.41 0.97)	1.71 (0.71 1)	1.31 (0.33 0.98)	1.79 (0.74 1.04)	8.15
977 2n (n=17)	1.94 (0.94 1)	1.29 (0.41 0.88)	1.53 (0.59 0.94)	1.17 (0.18 1)	1.88 (0.76 1.12)	7.82
2n Avg	1.95 (0.96 0.99)	1.34 (0.41 0.92)	1.62 (0.65 0.97)	1.24 (0.26 0.99)	1.84 (0.75 1.08)	7.99

Table 5-1 Mean chiasma frequency for each chromosome in Metaphase I stage in A. thaliana F2s.

The table shows the mean chiasma count for nine different tetraploid F2s. Numbers below the mean count in brackets is mean chiasma number in short arm and long arm respectively. n is the number of cells analysed.

		168 (n=13)	412 (n=41)	466 (n=7)	468 (n=17)	473 (n=8)	956 (n=43)	964 (n=12)	Overall (n=141)
Chr 1	**2II**	6 (46.2)	15 (36.6)	2 (28.6)	7 (41.2)	0	12 (37.9)	4 (33.3)	46 (32.6)
	IIV	7 (53.8)	26 (63.4)	5 (71.4)	8 (47.1)	8 (100)	29 (67.4)	8 (66.7)	92 (65.2)
	1III + 1I	0	0	0	1 (5.8)	0	2 (4.6)	0	3 (2.1)
	1II + 2I	0	0	0	0	0	0	0	0
Chr 2	**2II**	9 (69.2)	14 (34.1)	3 (43)	10 (58.8)	4 (50)	23 (53.5)	8 (66.7)	71 (50.4)
	IIV	2 (15.4)	24 (58.5)	3 (43)	5 (29.4)	4 (50)	20 (46.5)	2 (16.7)	60 (42.6)
	1III + 1I	1 (7.7)	2 (4.8)	1 (14)	1 (5.9)	0	0	2 (16.7)	7 (4.9)
	1II + 2I	1 (7.7)	1 (2.4)	0	1 (5.9)	0	0	0	3 (2.1)
Chr 3	**2II**	7 (53.8)	16 (39.0)	4 (57.1)	10 (58.8)	3 (37.5)	27 (62.8)	6 (50)	73 (51.8)
	IIV	5 (38.4)	24 (58.5)	3 (42.9)	6 (35.3)	3 (37.5)	14 (32.6)	6 (50)	61 (43.3)
	1III + 1I	0	0	0	1 (5.9)	2 (25)	2 (4.7)	0	5 (3.5)
	1II + 2I	1 (7.7)	1 (2.4)	0	0	0	0	0	2 (1.4)
Chr 4	**2II**	7 (53.8)	22 (53.7)	4 (57.1)	9 (52.9)	4 (50)	24 (55.8)	7 (58.3)	77 (54.6)
	IIV	4 (30.8)	17 (41.5)	2 (28.6)	8 (47.1)	4 (50)	15 (34.9)	3 (25)	53 (37.6)
	1III + 1I	2 (15.4)	2 (5)	1 (14.3)	0	0	3 (7)	2 (16.7)	10 (7.1)
	1II + 2I	0	0	0	0	0	1 (2.3)	0	1 (0.7)
Chr 5	**2II**	4 (30.8)	19 (46.3)	2 (28.6)	5 (29.4)	4 (50)	14 (32.6)	4 (33.3)	52 (36.8)
	IIV	9 (69.2)	21 (51.2)	5 (71.4)	12 (70.6)	4 (50)	27 (62.8)	6 (50)	84 (59.6)
	1III + 1I	0	1 (2.4)	0	0	0	1 (2.3)	1 (8.3)	3 (2.1)
	1II + 2I	0	0	0	0	0	1 (2.3)	1 (8.3)	2 (1.4)
Total	**Biv**	35 (53.8)	87 (42.4)	15 (42.9)	42 (49.4)	15 (37.5)	102 (47.4)	30 (50)	(46.2)
	Mult	30 (46.2)	117 (58.5)	20 (57.1)	43 (50.6)	25 (62.5)	113 (52.6)	30 (50)	(53.9)

Table 5-2 Chromosome M1 configurations for different tetraploid F2 plants in Arabidopsis thaliana.

Numbers in parenthesis indicate the percentage. II represents bivalent, IV represents quadrivalent, III represents trivalent and I represent univalent. Overall represents average bivalent, multivalent and univalents and respective percentages in parenthesis across all samples. Total represents total number of bivalents and multivalents across all chromosomes in individual F2s.

4n F2	IV configuration	Chr 1 n (%)	Chr 2 n (%)	Chr 3 n (%)	Chr 4 n (%)	Chr 5 n (%)
168 (N=13)	IV Ring	4 (57.1)	1 (50)	5 (100)	1 (25)	4 (44.4)
n_1=7,n_2=2,n_3=5,n_4=4,n_5=9	IV Chain	3 (42.9)	1 (50)	0 (0)	3 (75)	5 (55.6)
	IV Others	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
412 (N=41)	IV Ring	17 (65.4)	8 (33.3)	18 (75)	6 (35.3)	9 (42.9)
n_1=26,n_2=24,n_3=24,n_4=17,n_5=21	IV Chain	9 (34.6)	16 (66.7)	6 (25)	8 (47.1)	12 (57.1)
	IV Others	0 (0)	0 (0)	0 (0)	3 (17.6)	0 (0)
466 (N=7)	IV Ring	4 (80)	1 (33.3)	2 (66.7)	0 (0)	4 (80)
n_1=5,n_2=3,n_3=3, n_4=2,n_5=5	IV Chain	1 (20)	2 (66.7)	1 (33.3)	2 (100)	1 (20)
	IV Others	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
468 (N=17)	IV Ring	4 (44.4)	2 (40)	4 (66.7)	1 (12.5)	9 (75)
n_1=9,n_2=5,n_3=6, n_4=8,n_5=12	IV Chain	4 (44.4)	3 (60)	2 (33.3)	7 (87.5)	2 (16.7)
	IV Others	1 (11.1)	0 (0)	0 (0)	0 (0)	1 (8.3)
473 (N=8)	IV Ring	3 (37.5)	1 (25)	2 (66.7)	3 75)	4 (100)
n_1=8,n_2=4,n_3=3, n_4=4,n_5=4	IV Chain	5 (62.5)	3 (75)	1 (33.3)	1 (25)	0 (0)
	IV Others	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
956 (N=43)	IV Ring	16 (55.2)	9 ((45)	6 (42.9)	8 (53.3)	19 (70.4)
n_1=29,n_2=20,n_3=14,n_4=15,n_5=27	IV Chain	13 (44.8)	11 (55)	8 (57.1)	7 (46.7)	8 (29.6)
	IV Others	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
964 (N=12)	IV Ring	8 (100)	1 (50)	3 (50)	0 (0)	5 (83.3)
n_1=8,n_2=2,n_3=6, n_4=3,n_5=6	IV Chain	0 (0)	1 (50)	3 (50)	2 (66.7)	1 (16.7)
	IV Others	0 (0)	0 (0)	0 (0)	1 (33.3)	0 (0)

Table 5-3 Quadrivalent configurations in different tetraploid F2s.

N indicate the total number of cells analysed. n_1, n_2, n_3, n_4, n_5 indicate number of quadrivalents in chromosomes 1, 2 3, 4 and 5 respectively. Number in parenthesis indicate percentage of the quadrivalent configuration.

Four plants had highest quadrivalent formation in chromosome 1 followed by chromosome 5 and three plants had highest quadrivalent formation in chromosome 5 followed by chromosome 1. In a few plants (for example in 473 and 956), chromosomes 2 and 4 showed more quadrivalents than chromosome 3. Overall, chromosome 1 had the highest frequency of multivalent formation and chromosome 4 had the lowest. More closed configurations such as ring quadrivalents occurred for chromosomes 1 and 5 followed by chromosome 3, while more open configurations occurred for chromosomes 2, and 4. Other configurations such as X, spoon, and Y were occasionally seen for one or more chromosomes (**Table 5-3**).

Multivalent formation was more frequent than bivalent formation across all the chromosomes for different plants except for 168, where bivalents were more frequent and 964, where bivalents and multivalents were present in equal number. A χ^2 goodness to fit test was carried out to find out if chromosomes in different F2 lines followed the random end pairing model of 66% : 34% multivalent to bivalent configuration in autotetraploids (**Table 5-4**) (Sybenga, 1975). Overall, there is a trend of reduced multivalent formation (i.e. diploidisation) compared to what is expected from the random end pairing model. The reduction is highly significant in chromosomes 2, 3 and 4, whereas of the longer chromosomes, chromosome 1 shows a slight excess of multivalents which is not significant and 5 shows a slight reduction which is also not significant. It can also be ascertained from the table that multivalent formation is line and chromosome dependent. For example, chromosome 2 in F2 168 show 22.1% multivalent formation (**Table 5-2**), which highly significantly deviates from 66% multivalent formation (**Table 5-4**), while the same chromosome in F2 412, shows 63.3% multivalent formation, which is not a significant reduction compared with the random pairing model. Chromosomes 2, 3 and 4 consistently show a trend of reduced multivalent formation across all different F2s, whereas chromosome 1 and 5 also show increased multivalent formation in some F2s, though most of these are not significant.

A two sample proportion test between chromosomes, overall across all samples, showed a significant difference between a few chromosomes for multivalent formation after accounting for multiple testing by Bonferroni correction at p = 0.05. (**Table 5-5**). This suggests that multivalent formation is chromosome dependent. Though there are differences for individual chromosomes, overall across all chromosomes, the level of multivalent formation was not

significantly different between different F2 lines based on 2 sample proportion test (p-value > 0.05 for all the F2 pairs).

4n F2	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5
168	<	<	<	<	>
	0.86 ^{ns}	10.7 ***	4.39 *	2.3 ^{ns}	0.06 ^{ns}
412	<	<	<	<	<
	0.12 ^{ns}	0.12 ^{ns}	1.02 ^{ns}	7.06 *	2.78 ^{ns}
466	>	<	<	<	>
	0.1 ^{ns}	0.24 ^{ns}	1.67 ^{ns}	1.67 ^{ns}	0.1 ^{ns}
468	<	<	<	<	>
	0.39 ^{ns}	7.14 **	4.7 *	2.7 ^{ns}	0.16 ^{ns}
473	>	<	<	<	<
	4.1 *	0.9 ^{ns}	0.04 ^{ns}	0.9 ^{ns}	0.9 ^{ns}
956	>	<	<	<	<
	0.7 ^{ns}	7.3 **	15.9 ***	11.2 ***	0.01 ^{ns}
964	>	<	<	<	<
	0.002	5.7 *	1.4 ^{ns}	3.2 ^{ns}	0.31 ^{ns}
Across all F2s	>	<	<	<	<
	0.12 ^{ns}	21.46 ***	23.14 ***	28.56 ***	1.16 ^{ns}

Table 5-4 χ^2 goodness of fit to test deviation from random end pairing model.

The values in the table are χ^2 values and symbols indicate deviation direction, where < and > indicate less than or more than expected multivalents to bivalent ratio (66% multivalents : 34% bivalents) * indicates p-value ≤ 0.05, ** indicates p-value ≤ 0.01, *** p-value ≤ 0.001, ns indicates non-significant.

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5
Chr 1	-				
Chr 2	**3.37 (0.0007)**	-			
Chr 3	**3.49 (0.0005**)	0.12 (0.9050)	-		
Chr 4	**3.84 (0.0001)**	0.48 (0.6328)	0.36 (0.7199)	-	
Chr 5	1.00 (0.3193)	-2.39 (0.0167)	-2.51 (0.0120)	**-2.86 (0.0042)**	-

Table 5-5 Results of two sample proportion test between chromosomes across all F2 samples for multivalent formation.

The values reported are Z value followed by p-value.

5.2.2 Chiasma analysis in diploid F2s

Chiasma analysis in plant 205

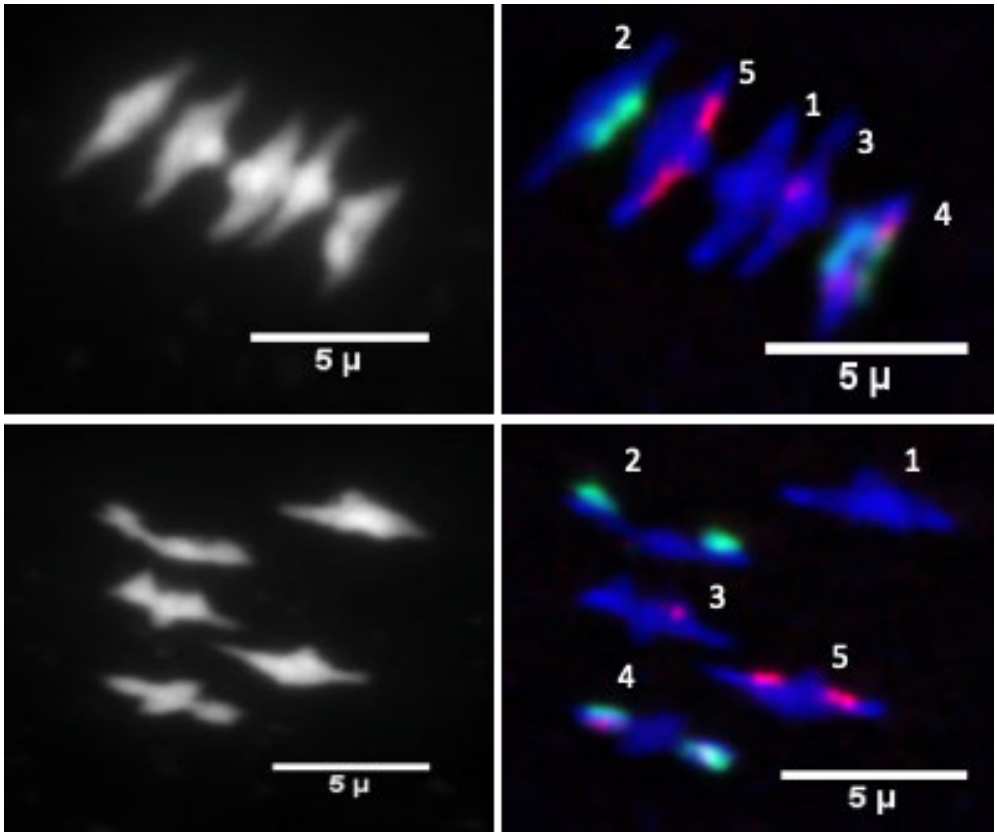


Figure 5-17 Comparison between two cells in M1 in diploid F2 205.

The figure shows two DAPI stained M1 cells (left) and the same cells (right) also showing 5S (red) and 45S (green) FISH probes. The numbers represent the chromosome numbers. Scale bar is 5 µm.

98 cells in diploid F2 plant 205 were probed using 5S and 45S FISH probes, two of which can be seen in **Figure 5-17**. The top panel show M1 stage where all the five chromosomes are in a ring configuration. They have at-least one chiasma in each of the short and long arms. Therefore, there are a total of 10 chiasmata in this cell. For another cell in the bottom panel, chromosomes 1 and 5 are ring bivalents having one chiasma in each of the short and long arms, chromosomes 2 and 4 are rod bivalents with one chiasma each in their long arm, while chromosome 3 is a rod bivalent with 2 chiasmata in the long arm, as a clear knob is visible.

Therefore, there are a total of 8 chiasmata in this cell. For the 98 cells analysed, the mean chiasma frequency was 8.15, with chromosome 1 having the highest frequency followed by chromosomes 5, 3, 2 and 4 in the order respectively (**Table 5-1**). Here, a decreasing pattern of crossover frequency in accordance with the decreasing size of the chromosomes is visible. Chromosome 1 is the biggest and chromosome 4 is one of the smallest.

Chiasma analysis in plant 977

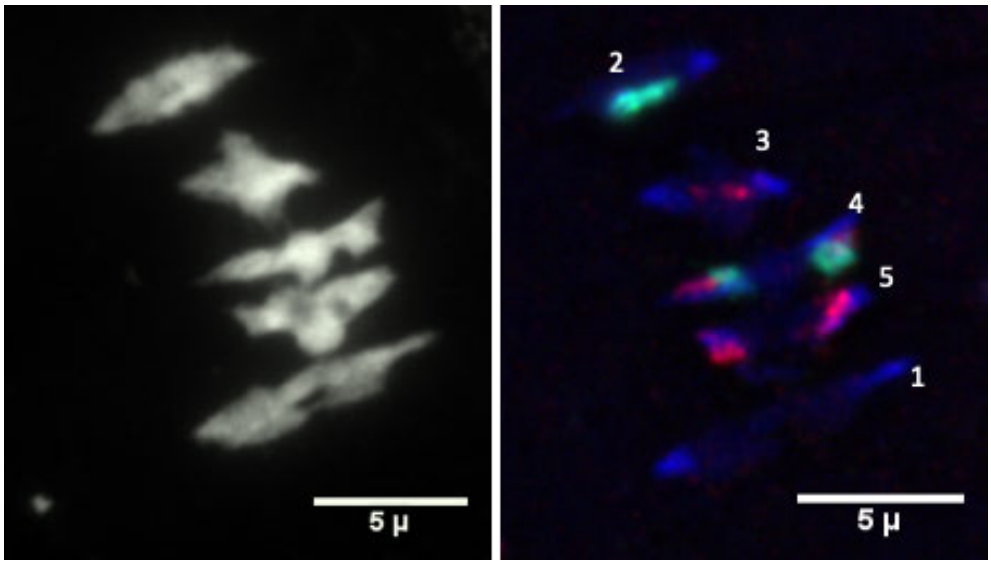


Figure 5-18 Comparison between a cell in M1 in diploid F2 977.

The figure shows a DAPI stained M1 cell (left) and the same cell (right) also showing 5S and 45S FISH probes. Numbers are chromosome numbers. Scale bar is 5 µm.

Another diploid F2 plant, 977, was analysed for chiasma frequency. A representative of the M1 cells, which were used in the analysis can be seen in **Figure 5-18**. Chromosomes 1, 2, 3 and 5 present ring configuration in the bivalents, where one chiasma is present in each arm, totalling at-least 2 chiasmata for each chromosome respectively. Chromosome 4 is a rod bivalent with one chiasma in the long arm. Thus, there are total of 9 chiasmata in this cell. For plant 977, 17 cells were analysed, which produced a mean chiasma frequency of 7.82 (**Table 5-1**). Here again, there is a decreasing chiasma frequency as the chromosome size decreases.

The average mean chiasma count for the 2n F2s was found to be 7.99 ± 0.23, which lies a little lower in the range between the mean chiasma count of the diploid Columbia and Landsberg parents, with their mean total scores being 8.3 ± 1.1 and 7.96 ± 0.82 respectively. A ring configuration, binding both the arms of the chromosome was predominantly seen for chromosomes 1, 5 and 3, while rods were predominant in chromosomes 2 and 4, as also seen in the parents.

5.2.3 Comparing chiasma count frequency between diploid and tetraploid F2s

The proportion of cells showing different numbers of chiasma was compared for different chromosomes in tetraploid and diploid F2s. **Figure 5-19** illustrates the differences between different F2s for number of chiasma in each chromosome. In tetraploids, cells showing one chiasma is not present except for a rare few in chromosomes 4 and 5 in plant 956. Cells with two chiasmata is much lower for chromosomes 1, and 5 compared to chromosomes 2, 3 and 4 in tetraploids. Cells with three chiasmata are distributed across all chromosomes. Cells with four chiasmata are present in higher number in chromosomes 1, 3 and 5 than chromosomes 2 and 4. Five chiasmata is only seen in a few plants in chromosome 1 in tetraploids and six chiasmata only in one tetraploid, 168 in chromosome 5. In contrast, there is a higher proportion of cells with 1 and 2 chiasmata in diploid F2s. More cells with 2 chiasmata occur in chromosomes 1, 3 and 5, while more cells with one chiasma are present in chromosomes 2 and 4. Only a few cells in diploid 977 have three chiasmata in chromosomes 1 and 5.

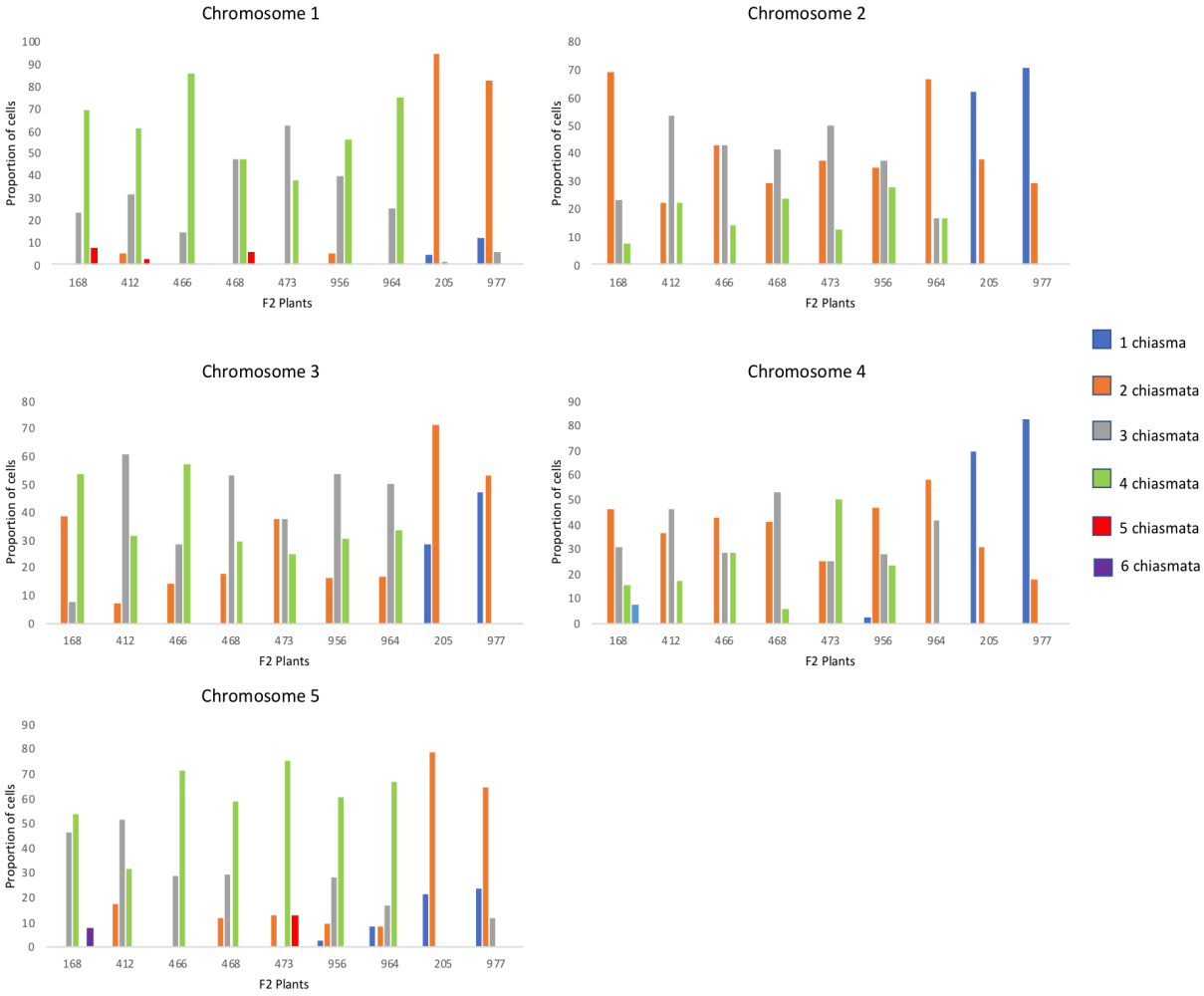


Figure 5-19 Proportion of meiotic cells with different number of chiasmata in five different chromosomes in A. thaliana F2s.

The figure shows chiasma distribution in cells in different tetraploid and diploid F2s. The coloured bars show the number of chiasmata in the respective F2 plants on horizontal axis.

A Kruskal-Wallis test showed a significant difference between different F2 plants ($\chi^2_{df=8}$ = 193.75, p-value < 2.26-16). This was followed by post hoc Dunn test with Bonferroni correction, which showed that tetraploid F2 plants differed significantly from diploid F2 plants (p-value << .01), however, they did not differ from each other within each ploidy group.

A per bivalent chiasma comparison between different F2 plants was carried out by extracting cells in tetraploid F2s, which had two bivalents. **Table 5-6** shows chiasma count comparison

between different F2s, when cells with two bivalents and multivalents have been separated in tetraploid F2s. A per bivalent comparison could now be performed by comparing the diploids chiasma count with per bivalent count in tetraploids (dividing the chiasma count by 2). **Figure 5-20** shows graphs of five different chromosomes comparing proportion of bivalent only meiotic cells in tetraploids with diploids. There are no cells with one chiasma in tetraploids and cells with two chiasmata are present in higher proportion in chromosomes 2 and 4 followed by chromosomes 3 and 5 with none in chromosome 1 in tetraploids indicating more rod bivalents in chromosome 2 and 4. A higher proportion of cells with three chiasmata can be seen in chromosome 3. Chromosomes 1 and 5 showed the highest proportion of cells with four chiasmata across different tetraploid F2s indicating more cells with ring bivalents. Similar to the total proportion of cells as seen in **Figure 5-19**, only a few cells in chromosome 1 and 5, and only in a few tetraploids had five and six chiasmata respectively. In comparison, diploids have more proportion of cells with one and two chiasmata across all chromosomes, with chromosomes 2 and 4 having more cells with one chiasma.

A Kruskal-Wallis test showed a significant difference in the per bivalent chiasma count between different F2s only for chromosomes 1 ($\chi^2_{df=8}$ =22.583, p-value =0.002014) and 5 ($\chi^2_{df=8}$ =25.816, p-value = 0.001129). Post hoc Dunn test showed this difference to be between 205 and 964 in chromosome 1, and 205 and 412 (p-value = 0.0045) and 977 and 412 (p-value 0.05) for chromosome 5. A significant Kruskal-Wallis was shown for chromosome 3 as well, however post hoc Dunn test did not give a significant difference between any variety. The per bivalent count for these chromosomes in tetraploids is less than the per bivalent count in

diploids. This indicates that instead of an increase, a reduction in chiasma count was seen in tetraploids.

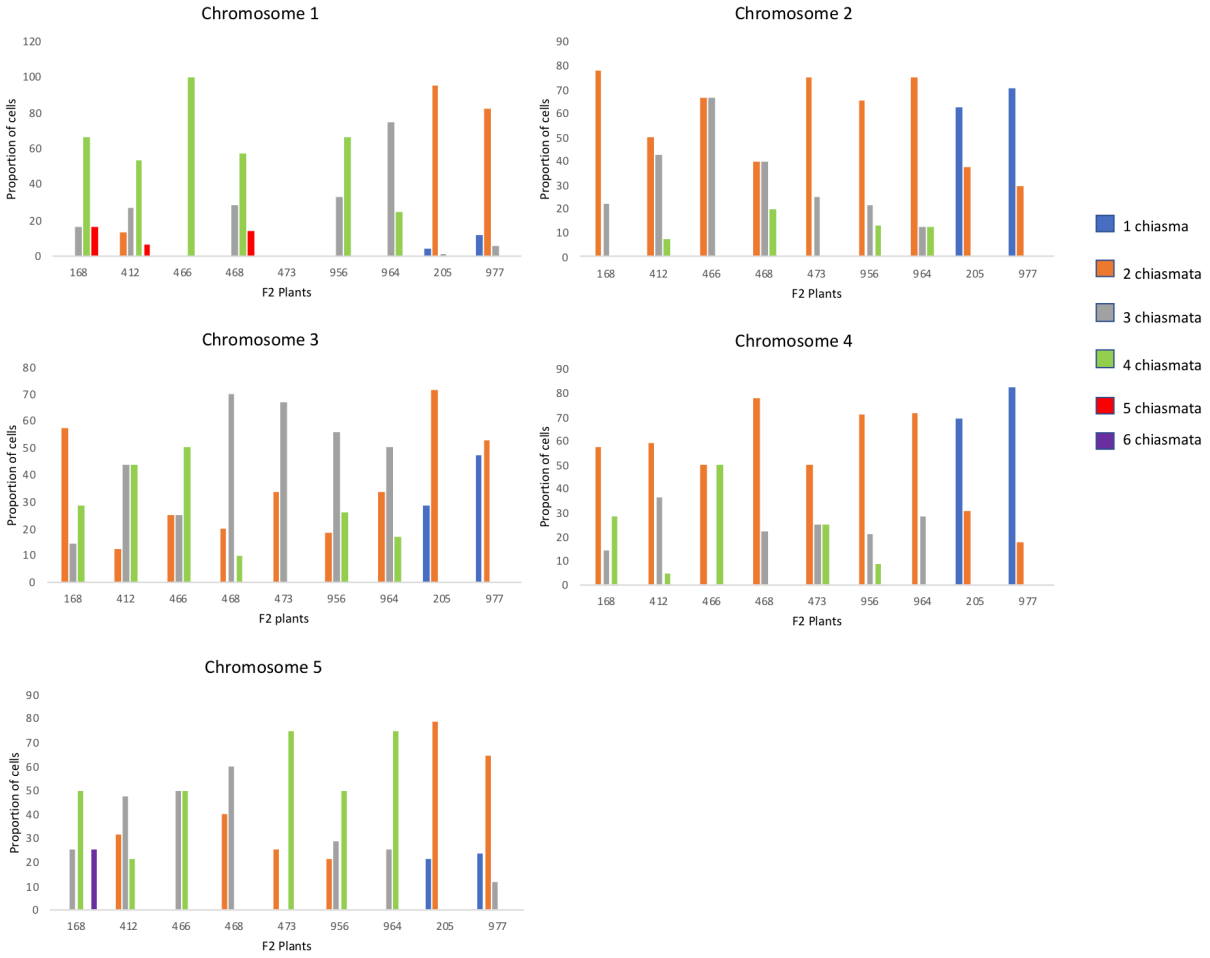


Figure 5-20 Proportion of meiotic cells showing only bivalents with different number of chiasmata across the five chromosomes in A. thaliana F2s.

The figure shows chiasma distribution in bivalent only cells in different tetraploid and diploid F2s. The coloured bars show the number of chiasma in the respective F2 plants on horizontal axis.

	Chr I	Chr II	Chr III	Chr IV	Chr V
168 4n (biv)	4 (n_1=6)	2.22 (n_2=9)	2.71 (n_3=7)	2.71 (n_4=7)	4.25 (n_5=4)
(N=13)	(1.83 2.17)	(0.22 2)	(0.71 2)	(0.71 2)	(1.75 2.5)
168.4n (mult)	3.71 (n_1=7)	3 (n_2=3)	4 (n_3=5)	3 (n_4=6)	3.4 (n_4=9)
(N=13)	(1.71. 2)	(1.3 1.7)	(2 2)	(1.3 1.7)	(1.4 2)
412 4n (biv)	3.53 (n_1=15)	2.57 (n_2=14)	3.31 (n_3=16)	2.45 (n_4=22)	2.89 (n_5=19)
(N=41)	(1.46 2.1)	(0.57 2)	(1.31 2)	(0.45 2)	(0.89 2)
412 4n (mult)	3.65 (n_1=26)	3.23 (n_2=26)	3.25 (n_3=24)	3.2 (n_4=19)	3.36 (n_5=22)
(N=41)	(1.65 2)	(1.31 1.92)	(1.25 2)	(1.2 2)	(1.41 1.95)
466 4n (biv)	4 (n_1=2)	2.33 (n_2=3)	3.25 (n_3=4)	3 (n_4=4)	3.5 (n_5=2)
(N=7)	(1.5 2.5)	(0.33 2)	(1.25 2)	(1 2)	(1 2.5)
466 4n (mult)	3.8 (n_1=5)	3 (n_2=4)	3.7 (n_3=3)	2.7 (n_4=3)	3.8 (n_5=5)
(N=7)	(1.8 2)	(1.25 1.75)	(1.7 2)	(1 1.7)	(1.8 2)
468 4n (biv)	3.86 (n_1=7)	2.8(n_2=10)	2.9 (n_3=10)	2.22 (n_4=9)	2.6 (n_5=5)
(N=17)	(1.71 2.14)	(1 1.8)	(0.9 2)	(0.33 1.89)	(0.8 1.8)
468 4n (mult)	3.5 (n_1=10)	3.2 (n_2=6)	3.43 (n_3=7)	3.13 (n_4=8)	3.83 (n_5=12)
(N=17)	(1.5 2)	(1.3 1.8)	(1.57 1.86)	(1.13 2)	(1.83 2)
473 4n (biv)	0 (n_1=0)	2.25 (n_1=4)	2.67 (n_3=3)	2.75 (n_4=4)	3.5 (n_5=4)
(N=8)	(0 0)	(0.25 2)	(0.67 2)	(0.75 2)	(1.5 2)
473 4n (mult)	3.4 (n_1=8)	3.25 (n_2=4)	3 (n_3=5)	3.75 (n_4=4)	4.25 (n_5=4)
(N=8)	(1.4 2)	(1.25 2)	(1.4 1.6)	(1.75 2)	(2 2.25)
956 4n (biv)	3.67 (n_1=12)	2.48 (n_2=23)	3.07 (n_3=27)	2.38 (n_4=24)	3.29 (n_5=14)
(N=43)	(1.67 2)	(0.48 2)	(1.07 2)	(0.38 2)	(1.29 2)
956 4n (mult)	3.41 (n_1=31)	3.45 (n_2=20)	3.3 (n_3=16)	3.3 (n_4=18)	3.64 (n_5=28)
(N=43)	(1.48. 1.93)	(1.45 2)	(1.4 1.9)	(1.4 4 1.83)	(1.68 1.96)
964 4n (biv)	3.25 (n_1=4)	2.38 (n_2=8)	2.83 (n_3=6)	2.29 (n_4=7)	3.75 (n_5=4)
(N=12)	(1.25 2)	(0.38 2)	(0.83 2)	(0.29 2)	(1.75 2)
964 4n (mult)	4 (n_1=8)	2.75 (n_2=4)	3.5 (n_3=6)	2.6 (n_4=5)	3.57 (n_5=7)
(N=12)	(2 2)	(1.25 1.5)	(1.5 2)	(1 1.6)	(1.71 1.86)
205 2n	1.97	1.38	1.71	1.31	1.79
(N=98)	(0.98 0.99)	(0.41 0.97)	(0.71 1)	(0.33 0.98)	(0.74 1.04)
977 2n	1.94	1.29	1.53	1.17	1.88
(N=17)	(0.94 1)	(0.41 0.88)	(0.59 0.94)	(0.18 1)	(0.76 1.12)
2n Avg	1.95	1.34	1.62	1.24	1.84
	(0.96 0.99)	(0.41 0.92)	(0.65 0.97)	(0.26 0.99)	(0.75 1.08)

Table 5-6 Mean chiasma count in diploids and tetraploids cells with bivalent only and multivalents only chromosome configurations in A. thaliana F2s. Numbers in parenthesis indicate chiasma in short and long arms.

The table shows mean chiasma count for each chromosome in different tetraploid and diploid F2s after extracting cells with 2 bivalents and multivalents separately in tetraploids. N indicates the total number of cells analysed, n1, n2, n3, n4 and n5 indicate the number of cells analysed in chromosomes 1,2 3,4 and 5 respectively.

5.3 Discussion

F2 plants, which were grown from the seeds of the F1 plants were created by crossing diploid Columbia and Landsberg parents, and autotetraploid Columbia and Landsberg parental lines to yield diploid and tetraploid F2s. Inflorescence were collected from the F2 plants and fixed, from which cytological chromosomal spreads were prepared. A good chromosomal stability (78%) was established for the plants. Those tetraploid F2s with M1s, for which an initial chromosome count of 20 chromosomes was established, were used to carry out FISH and to analyse the chiasma frequency. Though there were 39 such F2 plants, M1s enabling FISH analysis were only available in 9 of them. The exceptions were Plant 471 and 958, where a few mitotic cells counted 20, and a few others counted 21 and/or 22. Having seen few cells with 20 counts and with a good number of M1s, they were probed using 5S and 45S FISH probes. On the basis of the signals in M1, a few cells again looked to contain a full normal complement of all the chromosomes, whereas a few cells showed one extra chromosome 4 in plant 958, and sometimes one extra 4 as well as one extra chromosome 5 signal in plant 471. There can be two reasons for this; either one extra chromosome 4 is present, or there is a chromatid separation already occurring for one of the chromosomes involved in the bivalent or the quadrivalent, early on in M1 itself.

In *A. thaliana,* several proteins are involved in maintaining the cohesion among the sister chromatids during prophase I. The REC8 (sister chromatids cohesion maintenance protein in yeast) orthologue SYN1/DIF1 is important in maintaining the sister chromatid cohesion in prophase 1. Male meiocytes in *syn1* mutants show several defects including arm separation

of sister chromatids in leptotene (Cai *et al.*, 2003). A similar phenotype is present in other organisms as well with mutations in *REC8*, such as in mice where premature sister chromatid separation occurred in REC8 deficient mice (Xu *et al.*, 2005).

Reducing the amount of proteins such as SCC3, the homologue of yeast Scc3, cohesion subunit protein, again presented premature sister chromatid separation in male meiocytes in *A. thaliana* (Chelysheva *et al.*, 2005). Similarly, SMC-like gene *SWITCH* (*SWI1*) or *DYAD* is known to maintain the cohesion and meiotic chromosome structure in *Arabidopsis*. Mutation in the gene leads to sister chromatid separation as early as prometaphase 1 in male meiocytes (Mercier *et al.*, 2001). In a tetraploid plant, many of the genes encoding important proteins undergo changes. In this study, we have also hybridised two autotetraploid varieties to create an F2 population. There might be problems with one of the cohesion subunits leading to an early chromatid separation in metaphase for chromosome 4 or 5 in these plants. The spindle forces which are responsible for keeping the chromosomes aligned may be disturbed in these F2s.

Aneuploidy is a well know phenomenon in polyploids, which may occur due to failure of equal chromosomal segregation in meiosis (Ramsey and Schemske, 2002). Aneuploidy is less tolerated in the animal kingdom and only a few cases of viable trisomics are known, especially in humans (Williams and Amon, 2009). It is well tolerated in plants and viable trisomics of each chromosome with phenotypes associated with each trisomy has been described in Datura (Blakeslee, 1922). Trisomics have also been characterised for *A. thaliana* (Koornneef and Van der Veen, 1983) and the plant is capable of tolerating high levels of aneuploidy (Henry *et al.*,

2009). Established tetraploids such as *Arabidopsis arenosa* are also known to produce aneuploid gametes after meiosis at times (Morgan, 2016). The F2s in our study are the segregating generation for the chromosomes of Columbia and Landsberg, in which the meiosis is grappling with both tetraploidy and hybridisation, which can lead to such kind of differences in karyotype among different F2s. However, out of the nine plants analysed, only two plants showed abnormal karyotypes, indicating overall good stability in maintaining the right set of 20 chromosomes. Based on FISH analysis of M1 chromosomes, it was difficult to ascertain if the extra 5S and 45S signals visible in plants 471 and 958 were precocious chromatid segregation in one of the bivalents in chromosome 4 and 5, or if five sets of homologues of chromosome 4 and/or 5 were present in the plant. However, some irregularity was present in the plants, and therefore they were not included in the chiasma count analysis. The vegetative growth of the plants was good and the seed set was not different from the average seed set in the tetraploid F2s. It would be interesting to compare this cytological result with sequencing analysis from these plants.

Different kind of chromosomal combinations are possible in a tetraploid F2 (**Figure 5-21**).Theoretically, quadrivalent formation can occur between the four chromosomes in all types of the combinations as shown. However, in type 1, which represents two chromosomes of each parent, there may be a preference in pairing between the same parental chromosomes over non-parental chromosomes. These chromosome sets may produce only bivalents during meiosis. In type 2, two chromosomes are from one parent, one is from another parent and the fourth one is a recombinant, with most part being of second parent. In this type in meiosis, a quadrivalent formation may occur if the homologous parental

chromosomes synapse with each other and then the small blue part of the recombinant chromosome synapses with the blue homologue. Similarly, in type 3 and type 4 there are more chances of quadrivalent formation. The recombinant chromosomes in type 3 have an equal probability of synapsing with any of the other 3 chromosomes, while in type 4, 3 chromosomes are similar and the fourth one is different, which again can pair up with any of the three, only 2 chromosomes pairing at a time and may lead to quadrivalent formation (Sybenga, 1996). Presence of multivalents in F2s in our work indicate one or more of the four types of combinations occurring during meiosis.

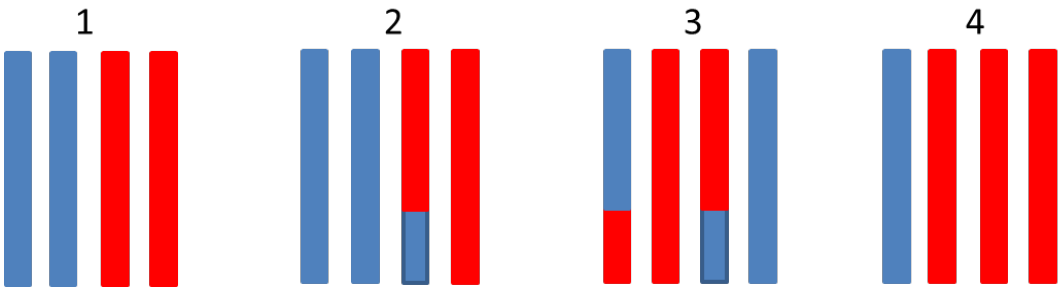


Figure 5-21 Representation of the possible chromosome combinations in an F2 generated from a hybrid F1. Figure adapted from Sybenga, 1996.

The mean chiasma frequency for the seven analysed tetraploid F2s was found to be 15.8 ± 1.54, which was within the range of the mean parental chiasma frequency of 15.35 ± 1.75 for Landsberg and 16.74 ± 1.35 for Columbia. A similar relationship was observed for other phenotypic traits analysed in phenotype comparison in **Results 4.2.2**, where the F2 mean trait value was either between the range of two parents or exceeded the parental trait value. The mean multivalent formation for all tetraploid F2 individual chromosomes was within the range of the parental multivalents, except for chromosome 4 which was higher than both the parental lines. This indicates heterosis in the F2 for multivalent formation in chromosome 4.

Chromosome 1, the biggest chromosome, showed the highest number of multivalents most of the time, followed by chromosome 5, which is the second biggest in size. Average multivalent formation for all the analysed F2s followed the chromosome size order. This kind of relationship has previously been seen in *Arabidopsis* (Moran *et al.*, 2001).

The multivalent frequency for each plant deviated from the random end pairing model for chromosomes 2, 3 and 4, being less than 66.6%, indicating diploidisation of these chromosomes or preferential pairing, which is possible in an F2 generation, as the chromosomes are still very heterogenous. Though, for chromosomes 1 and 5 it does not deviate much from 66%, indicating presence of more than one autonomous pairing sites (APS) leading to more frequent multivalent formation. Rapid cytological diploidisation in autotetraploids, where smaller acrocentric chromosomes achieved partial diploidisation rapidly within a few generations, along with the random pairing of homologues with the presence of more than one autonomous pairing sites for the longer metacentric chromosomes has been described before in *Arabidopsis* (Santos *et al.*, 2003). The exception here was chromosome 3, which also achieved diploidisation comparable to chromosomes 2 and 4 in F2s. There is an inversion present in the small arm of chromosome 4, and the long arm of chromosome 3 in Landsberg, which may be responsible for diploidisation of these chromosomes in F2 (Zapata *et al.*, 2016) by reducing crossover formation in the heterogenous homologues.

The formation of multivalents and bivalents differed between F2 lines analysed and overall chromosome 2 and 4 showed more bivalents than multivalents, the difference between the

varieties was not found to be significant. Overall across all chromosomes, the F2 multivalents followed the chromosomal size order, there was a significant reduction in multivalent formation in chromosomes 2, 3 and 4 compared to chromosome 1 and 5. This indicates that the frequency of multivalent formation is chromosome and line dependent, here chromosomes in individual F2 plants do show different trends for multivalent formation, as was also shown in Santos *et al.* (2003). However, the multivalent frequency will be higher at pachytene as a few associations open up before entering M1 stage and only few persist to M1 (Sybenga, 1975). The chiasma frequency for both the acrocentric chromosomes 2 and 4 was comparable and lower than other chromosomes in all F2s indicating several factors acting in cohesion to stabilise tetraploidy meiosis.

Univalents were present in low quantity, generally present with a trivalent formation. For chromosome 2, there were more univalents than for any other chromosome. It has been discussed that evolved autotetraploids stabilise themselves by reduction in chiasma formation by increased interference. This works along the four homologues that come together in zygotene, so that the two homologues involved in one crossover may not get involved again with a third or fourth homologue (Bomblies *et al.*, 2016). If indeed that is the case in *A. thaliana* F2s, then chromosomes 2 and 4 have to maintain a balance between the right level of interference so as not to lose any obligatory crossover. Fertility in terms of the number of seeds in F2s was found to be better than the parental lines and significantly higher than Landsberg parent (p-value < 0.01), indicating that meiosis may be stabilised in this successive generation, and heterosis occurred perhaps due to hybrid effects .

Analysis of zygotene, pachytene and diplotene was not carried out in this study. However, a few pachytenes observed showed loops which might indicate inversion or other structural differences in the pairing chromosomes. Such loop forming inversions have been described in detail by Sybenga (1975). To confirm such structures, electron microscopy can be useful. It will be interesting to compare this chiasma analysis with the sequencing results for these plants and for the F2 in general. For comparing and contrasting chiasma frequency, chromosomal preparations and FISH analysis was carried out in diploid F2s. A per bivalent chiasma comparison between tetraploid and diploid F2s showed a reduction in chiasma count in tetraploids which was significant in chromosomes 1 and 5. This may point to an increased interference in the chromosomes to stabilise meiosis in autotetraploids. Our result is in contrast with a previous study in *A. thaliana,* where an increase in chiasma frequency was seen in autotetraploids over diploids (Pecinka *et al.*, 2011). However, there might have been a change in distribution of the chiasma, which will be very interesting to see by analyzing the sequencing data from the F2 populations. A recent study on diploid and autotetraploid Col/Ler hybrids of *A. thaliana*, also found no substantial increase in chiasma frequency in tetraploids (Parra-nunez, Pradillo and Santos, 2019). They also found chromosome specific mechanisms controlling preferences for pairing between either identical or homologous (not identical) chromosomes in hybrid autotetraploid genomes.

Any change in chiasma formation, in number and/or distribution can be utilized by the breeders breeding for various useful traits like improved nutrition, pest resistance and so on by desirable alleles combination. A chiasma frequency assessment was successfully carried out on different F2s in this study. In future, the cytological images collected in the current

work could be reassessed to evaluate proximal vs distal location of chiasma to analyse the pairing behaviour. However, this kind of study can be best undertaken in hybrids, so more chromosomal spreads will have to be made from F1s. Sequencing will be carried out to generate molecular markers from different F2 and parental lines. The data thus generated will be analysed in future for molecular marker differences between the diploids and tetraploids, and analysing the rate and distribution of meiotic recombination.

5.4 References

Blakeslee, A. F. (1922) 'Variations in Datura Due to Changes in Chromosome Number', *The American Naturalist*. [University of Chicago Press, American Society of Naturalists], 56(642), pp. 16–31. Available at: http://www.jstor.org/stable/2556297.

Bomblies, K. *et al.* (2016) 'The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role?', *Chromosoma*, 125(2), pp. 287–300. doi: 10.1007/s00412-015-0571-4.

Cai, X. *et al.* (2003) 'The Arabidopsis SYN1 cohesin protein is required for sister chromatid arm cohesion and homologous chromosome pairing.', *Journal of cell science*, 116(Pt 14), pp. 2999–3007. doi: 10.1242/jcs.00601.

Chelysheva, L. *et al.* (2005) 'AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis.', *Journal of cell science*, 118(Pt 20), pp. 4621–32. doi: 10.1242/jcs.02583.

Henry, I. M. *et al.* (2009) 'Dosage and parent-of-origin effects shaping aneuploid swarms in A. thaliana', *Heredity*, 103(6), pp. 458–468. doi: 10.1038/hdy.2009.81.

Higgins, J. D. *et al.* (2012) 'Spatiotemporal Asymmetry of the Meiotic Program Underlies the Predominantly Distal Distribution of Meiotic Crossovers in Barley', *The Plant Cell*, 24(10), pp. 4096–4109. doi: 10.1105/tpc.112.102483.

Koornneef, M. and Van der Veen, J. H. (1983) 'Trisomics in Arabidopsis thaliana and the location of linkage groups', *Genetica*, 61(1), pp. 41–46. doi: 10.1007/BF00563230.

Mercier, R. *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(14), pp. 1859–1871. doi: 10.1101/gad.203201.

Moran, E. S. *et al.* (2001) 'Chiasma formation in Arabidopsis thaliana accession Wassileskija and in two meiotic mutants', *Chromosome Research*, 9(2), pp. 121-8. doi: 10.1023/A:1009278902994.

Morgan, C. (2016) *Coordination of meiotic recombination in diploid and tetraploid Arabidopsis (PhD thesis)*.

Parra-nunez, P., Pradillo, M. and Santos, J. L. (2019) 'Competition for Chiasma Formation Between Identical and Homologous (But Not Identical) Chromosomes in Synthetic Autotetraploids of Arabidopsis thaliana', *Frontiers in Plant Science*, 9(January), p. 1924. doi: 10.3389/fpls.2018.01924.

Pecinka, A. *et al.* (2011) 'Polyploidization increases meiotic recombination frequency in Arabidopsis', *BMC Biology*, 9(1), p. 24. doi: 10.1186/1741-7007-9-24.

Ramsey, J. and Schemske, D. W. (2002) 'Neopolyploidy in Flowering Plants', *Annual Review of Ecology and Systematics*. Annual Reviews, 33(1), pp. 589–639. doi: 10.1146/annurev.ecolsys.33.010802.150437.

Sanchez-Moran, E. *et al.* (2002) 'Variation in chiasma frequency among eight accessions of Arabidopsis thaliana', *Genetics*, 162(3), pp. 1415–1422. doi: 10.1007/bf00292272.

Santos, J. L. *et al.* (2003) 'Partial Diploidization of Meiosis in Autotetraploid', *Genetics*, 165(3), pp. 1533–1540.

Sybenga, J. (1975) 'The Analysis of Crossing-over', in *Meiotic Configurations: A Source of Information for Estimating Genetic Parameters*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 13–133. doi: 10.1007/978-3-642-80960-6_2.

Sybenga, J. (1996) 'Chromosome pairing affinity and quadrivalent formation in polyploids: do segmental allopolyploids exist?', *Genome*, 39(6), pp. 1176–1184. doi: 10.1139/g96-148.

Williams, B. R. and Amon, A. (2009) 'Aneuploidy: Cancer's fatal flaw?', *Cancer Research*, pp. 5289–5291. doi: 10.1158/0008-5472.CAN-09-0944.

Xu, H. *et al.* (2005) 'Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis', *Developmental Cell*, 8(6), pp. 949–961. doi: 10.1016/j.devcel.2005.03.018.

Zapata, L. *et al.* (2016) 'Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms', *Proceedings of the National Academy of Sciences*, 113(28), pp. E4052–E4060. doi: 10.1073/pnas.1607532113.

CHAPTER 6

GENOTYPING BY SEQUENCING RAD-SEQUENCING

6 Genotyping by sequencing – RAD-sequencing

6.1 Introduction

Ever since the molecular structure of DNA was elucidated by Watson and Crick, a range of technologies have developed to study not only the DNA bases, but also its products. Historically protein sequencing was the first, when Sanger identified amino acid residues of insulin in early 1950, which was followed by RNA sequencing when alanine tRNA sequence was identified (reviewed in Shendure *et al.*, 2017). DNA sequencing was later developed by Sanger and Nicklen, (1977) and Maxam and Gilbert, (1977) around same time using chain terminating dinucleotide or enzyme cleavage followed by gel resolution of bases. The process was slow, though it generated good quality long reads. Since then, next generation sequencing technologies have been developing at a fast rate generating a large amount of data in far less time, which poses its own challenges. Though many new cost effective next generation technologies have evolved, it is still expensive to carry out whole genome sequencing analysis on a population. An alternative is to genotype using high throughput sequencing by synthesis technologies, where a fraction of the whole genome is sequenced but it enables genome wide marker analysis and great reduction in costs. One such genotyping by sequencing method is Restriction site Associated DNA sequencing or RAD-seq (Davey *et al.*, 2011).

Before RAD-seq was developed, RAD markers were used in microarray techniques, for example in threespine stickleback, to identify a large number of markers for either individual or many samples together (Miller *et al.*, 2007). It was taken to the next level when RAD tag

library was sequenced using next generation Illumina sequencing platform, generating a large number of polymorphic markers from pooled F2 individuals to carry out genetic mapping in threespine stickleback (Baird *et al.*, 2008). In essence, restriction enzymes (RE) are used to chop the DNA of various individuals. The digested pieces of DNA are ligated to an adaptor, through the complementary overhangs of the RE digested site, which contains a forward primer and sample specific bar code. The samples are then pooled and sheared and a second adaptor is ligated to the ends containing the reverse primers. Only those RAD tags which have both the adapters on either side are size selected and amplified by PCR. The library can then be sequenced. The samples can finally be demultiplexed on the basis of the individual sample unique barcode. Thus, the short regions around the restriction sites across the whole genome are sequenced (Baird *et al.*, 2008). The original process of RAD-seq library preparation and sequencing can be seen in **Figure 6-1**.

Subsequently, RAD-seq has had several project specific adaptations and customisations for a wide range of applications. For example, DNA digestion using double restriction enzymes, one frequent cutter and one rare cutter was used for generation of reference sequence as well as linkage maps in *Arabidopsis thaliana* and lettuce (Truong *et al.*, 2012). Another adaptation (ez RAD) involved using TruSeq Illumina adaptors after restriction enzyme DNA digestion by two isoschizomer enzymes, to generate the fragments of desired size. This enabled cost reduction as the adaptors used were not custom designed. Several non-model organisms were sequenced by the process for which no references were available (Toonen *et al.*, 2013).

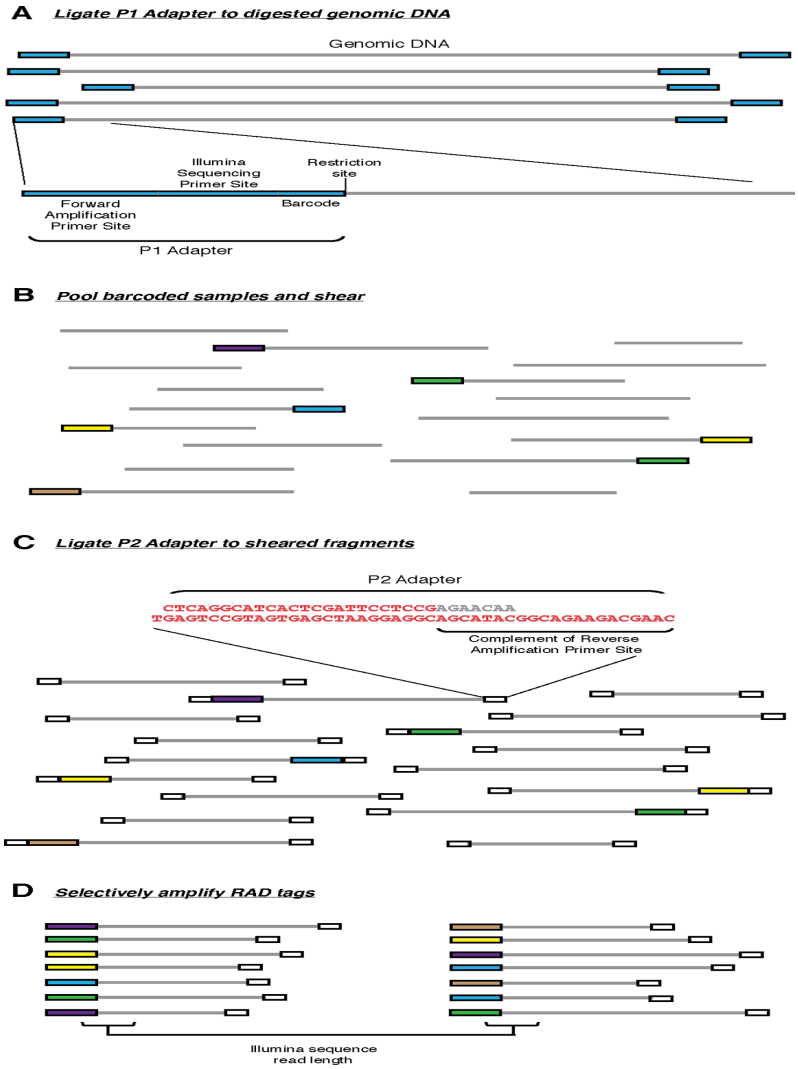


Figure 6-1 Representation of RAD-seq marker generation

Figure reproduced from Baird *et al.* (2008). Permission obtained, CC BY 4.0.

The modifications to the original protocol aim to improve the process of marker generation for robust downstream analysis and reduce the cost of sequencing further. One such modification has been applied in RAD-seq approach, which has been utilised in this project, where it was optimised to reduce the sequencing reads from chloroplast and rRNA genes. Three restriction enzymes, HindIII, EcoRI, and MspI were used to initially digest the genomic sequences followed by ligation with barcoded adaptors. The samples were pooled and size selected, followed by a second round of digestion, again using three enzymes SnaBI, StuI and

TfiI to remove the chloroplast and rRNA genes sequences. The pooled samples were PCR amplified and again size selected to ensure recovery of required DNA fragments and removal of primer dimers. The pooled samples were sequenced on an Illumina MiSeq platform generating 2 x 150 bp paired-end reads (Jiang *et al.*, 2016).

Sequencing genomes generate molecular markers to enable several downstream analyses such as QTL linkage mapping, phylogeny creation, and understanding genetic diversity. Variant and genotype calling after the initial bioinformatic analysis on the sequences, is an important step in downstream analysis. In plants, while the existing bioinformatic tools exist for variant and genotype calling with good certainty in diploids, only a few and scattered tools are known for polyploids, especially autopolyploids. Since many important crop plants are autopolyploids, it is important to have proper tools for variant calling, genotype assignment and dosage calling from next generation sequencing datasets.

6.1.1 Variant calling in Autopolyploids

SNP calling through various methods has been attempted in several autopolyploids. A few successful methods as reviewed in Clevenger *et al.* (2015) and described elsewhere (Jiang *et al.*, 2016; McCallum *et al.*, 2016) in different plants is presented in **Table 6-1**.

Species	Type of Sequencing	Mapping and Alignment	SNP calling software	No of SNPs called
Alfalfa	Illumina, RNA Seq	GSNAP	Alpheus pipeline	872,384
Arabidopsis thaliana	Illumina, RRL	Bowtie2	SAMtools	28,649
Potato	Illumina, sequence capture	BWA	Freebayes	42,625
Potato	Illumina, RRL	Bowtie2	SAMtools	125,291
Switchgrass	Illumina, RRL	panati	panati	149,502
Switchgrass	Illumina, RRL	UNEAK	UNEAK	1,242,869
Blueberry	Illumina, GBS	UNEAK	UNEAK	109,044

Table 6-1 Platform of sequencing and SNP calling in a few autotetraploids.

It can be seen that Illumina technology has been the choice of sequencing for autopolyploids, mainly because of cheaper sequencing costs, which is due to its ability to generate more reads per run. This is useful as higher coverage depth is required for quality variant calling in polyploids (Clevenger *et al.*, 2015).

6.1.2 Genotype Dosage Assignment in Polyploids

In a diploid cross between two contrasting allelic parents AA x aa, three genotypes, two homozygous AA, aa and one heterozygous Aa are possible in the F2 generation. However, in an autotetraploid for a biallelic cross between a nulliplex aaaa and tetraplex AAAA, five genotypes can be expected in an F2 population, which can be two homozygous and three heterozygotes AAAa (triplex), AAaa (duplex), Aaaa (simplex). Using the tools for diploid genotype calling, for example bcftools, in polyploids, while it will be possible to find if the genotype is homozygous or heterozygous, dosage of allele will not be identified. Therefore, either the existing tools need to be optimised to work with polyploids or new tools should be developed.

Reference	Tool/ Method	Summary
Voorrips, Gort and Vosman (2011).	fitTetra now updated to fitPoly	R package developed originally for genotype assignment in potato genotypes generated from SNP arrays.
Depristo *et al.* (2011)	GATK, Haplotype Caller	Most widely used genotype caller, can now provide polyploid genotyping.
Serang, Mollinari and Garcia (2012).	SuperMASSA, web based software	Statistical software tool for dosage calling in polyploids for dataset obtained from SNP arrays. Illustrated in potato and sugarcane.
Garrison and Marth (2012).	Freebayes	Haplotype based tool for allele dosage assignment from sequencing data.
Hackett, McLean and Bryan (2013).	Illumina, Infinium proprietary method	Originally used in potato for constructing linkage map from the SNP dosage information generated from SNP array.
Schmitz Carley *et al.* (2017).	ClusterCall	R package developed for autotetraploid dosage assignment for genotype data generated from SNP array with application shown in potato data.
Gerard *et al.* (2018).	updog, empirical Bayes approach	R package for genotyping from messy polyploid sequencing data, illustrated in hexaploid sweet potato.
Pereira, Garcia and Margarido (2018).	VCF2SM	Python based tool, which makes use of modified TASSEL-GBS and SuperMASSA softwares for polyploid genotype calling making use of the read depths of the alleles called.
Blischak, Kubatko and Wolfe (2018).	EBG	A C ++ programme, which uses genotype likelihood to call alleles in polyploids.
Clark, Lipka and Sacks (2019).	polyRAD	R based package developed for genotype calling using the allelic read depth of the variants called from low or uneven read depth sequencing such as RAD-seq.

Table 6-2 Tools available for genotype and dosage assignment in polyploids.

Most of the polyploid dosage assignment tools have been developed for microarray based SNP array genotyping (Bourke *et al.*, 2018). However, a few other tools have been developed in the recent past to identify allele dosage in an autotetraploid NGS dataset and have been described in **Table 6-2**. It is by no means an exhaustive list of tools that can be utilised, however it does point out the fact that the diploid genotype calling tools cannot provide complete information about polyploids for downstream analysis, and therefore better, flexible and easy to use tools need to be developed for analysing NGS polyploid datasets.

6.1.3 Rationale for sequencing

Genotyping by sequencing will generate data, which can be mined for sequence polymorphisms distinguishing plant ecotypes, that can be used to infer the occurrence of meiotic recombination on a genome-wide basis. Thus, genetic markers scoring the occurrence of meiotic recombination can be combined with experimental data obtained from more traditional cytogenetic methods for investigating meiotic chromosome behaviour and answering the fundamental question of whether ploidy affects a change in the frequency or distribution of recombination during meiosis.

6.2 Bioinformatics pipeline for analysis of RAD-seq data

6.2.1 Quality checks

Triple digest RAD-seq was carried out on the two parental lines, 2 F1s and 8 F2s for each diploid and tetraploid *Arabidopsis thaliana* population leaf samples collected during second plant trial (**2.3.2**), as described (**6.1**) in China, which provided us with dataset in FASTQ format (Jiang *et al.*, 2016). The quality of sequencing data for all the samples was checked using FASTQC (http://www.bioinformatics.babraham.ac.uk). Cutadapt version 1.16 (Martin, 2011) was used to trim 9 bases from the start of the R1 read and 3 bases from the start of the R2 read, which included barcodes and restriction enzyme site overhangs.

6.2.2 Read Alignment

The short paired reads were aligned to the reference sequence using Bowtie2 version 2.3.4.1 aligner allowing at-least 1 mismatch (Langmead and Salzberg, 2012). The aligned output from Bowtie aligner was straight converted into sorted bam files using samtools view, with

parameter -q 20 to only allow reads with mapping quality of 20 or more, and samtools sort command of SAMtools version 1.8 (Li *et al.*, 2009). Uniquely mapped reads were filtered by using the "XS:i" tag on the aligned and sorted BAM reads. The unique reads for individual samples from different libraries were then merged.

6.2.3 Variant Calling

Samtools mpileup and bcftools call was used to call variants, SNPs and INDELs with minimum base quality and mapping quality of 20 (Phred quality score). They were further filtered for genotype quality > 10 and sample depth > 4 for each sample used. Though I completely realise that bcftools may not be an ideal tool for tetraploid genotype calling (see section **6.1.2**), due to lack of time and ease of comparison, I have used it for the same.

6.3 Results

Six libraries were constructed and were sequenced in three rounds of 2 libraries each (**Table 6-3**) in China. In the first round, 8 samples each of diploid and tetraploid F2 were sequenced. The libraries were called AT190101, which consisted of diploid F2s and AT190102, which consisted of tetraploid F2s. For the second round of sequencing, libraries AT190201 for diploids and AT190202 for tetraploids were used which consisted of 4 parents (P1, P2 and 2 F1s) in addition to the same 8 F2 samples as used in first round. In the third round of sequencing, libraries used were AT190201b and AT190202b, which consisted of exactly same samples as in the second round.

Library	AT190101 (2n)	AT190102 (4n)	AT190201 (2n)	AT190202 (4n)	AT190201b (2n)	AT190202b (4n)
Samples	13 (F2_1)	87 (F2_1)	25 (P1 Col)	76 (P1 Col)	25 (P1 Col)	76 (P1 Col)
	98 (F2_2)	412 (F2_2)	437 (P2 Ler)	845 (P2 Ler)	437 (P2 Ler)	845 (P2 Ler)
	141 (F2_3)	466 (F2_3)	927 (F1_1)	501 (F1_1)	927 (F1_1)	501 (F1_1)
	205 (F2_4)	468 (F2_4)	930 (F1_2)	703 (F1_2)	930 (F1_2)	703 (F1_2)
	226 (F2_5)	503 (F2_5)	13 (F2_1)	87 (F2_1)	13 (F2_1)	87 (F2_1)
	369 (F2_6)	956 (F2_6)	98 (F2_2)	412 (F2_2)	98 (F2_2)	412 (F2_2)
	435 (F2_7)	958 (F2_7)	141 (F2_3)	466 (F2_3)	141 (F2_3)	466 (F2_3)
	649 (F2_8)	964 (F2_8)	205 (F2_4)	468 (F2_4)	205 (F2_4)	468 (F2_4)
			226 (F2_5)	503 (F2_5)	226 (F2_5)	503 (F2_5)
			369 (F2_6)	956 (F2_6)	369 (F2_6)	956 (F2_6)
			435 (F2_7)	958 (F2_7)	435 (F2_7)	958 (F2_7)
			649 (F2_8)	964 (F2_8)	649 (F2_8)	964 (F2_8)

Table 6-3 RAD-Seq Libraries and the samples used in each library.

P1 Col indicates Columbia parent and P2 Ler indicates Landsberg parent. Notation in parenthesis indicate the number assigned to the sample. 2n and 4n represents diploid and tetraploid respectively.

Number of paired end reads in millions can be seen in **Figure 6-2.** For parental samples, the number of reads in tetraploids in higher than those in diploids. It is varying for the F2 samples, however the average for both diploids and tetraploids is little more than 6M reads.

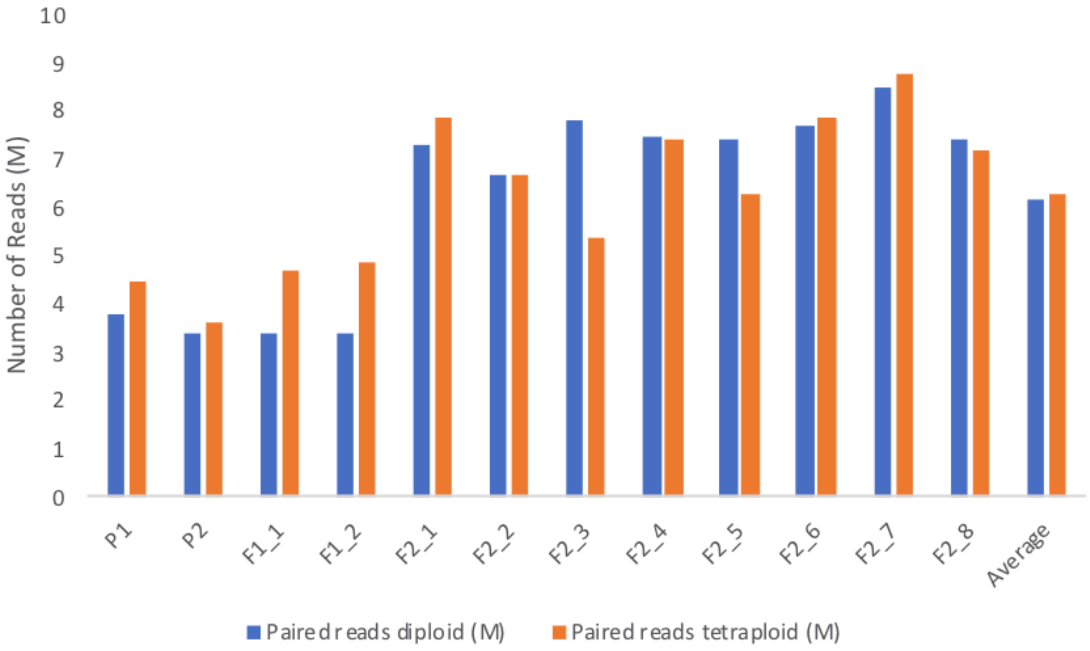


Figure 6-2 Number of paired end reads in million for diploid and tetraploid Arabidopsis thaliana samples.

6.3.1 Quality check

A quality check was performed on all the samples from all the rounds using FASTQC. An example of one of the samples for reads from one end of the paired read is shown in **Figure 6-3**. FASTQC produces a series of graphs, which can be interpreted for quality check. **Figure 6-3** illustrates number of sequences for read one of the paired read for tetraploid Columbia parent. A total of 1.8M paired 150 bp reads have been sequenced. The average sequence quality is 40, which is considered good. It means that only 1 in 10,000 called bases can be wrong. A FASTQC report was produced for all the 24 samples sequenced in this manner and the quality of all the sequenced samples was acceptable for further analysis. It was recognised that the sample barcodes and restriction site overhangs were overrepresented in all the sequenced samples. Cutadapt was then used to trim the first 9 bases from the read one of the pair and first three bases off the read two of the pair.

Basic Statistics

Measure	Value
Filename	76_R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1801445
Sequences flagged as poor quality	0
Sequence length	150
%GC	40

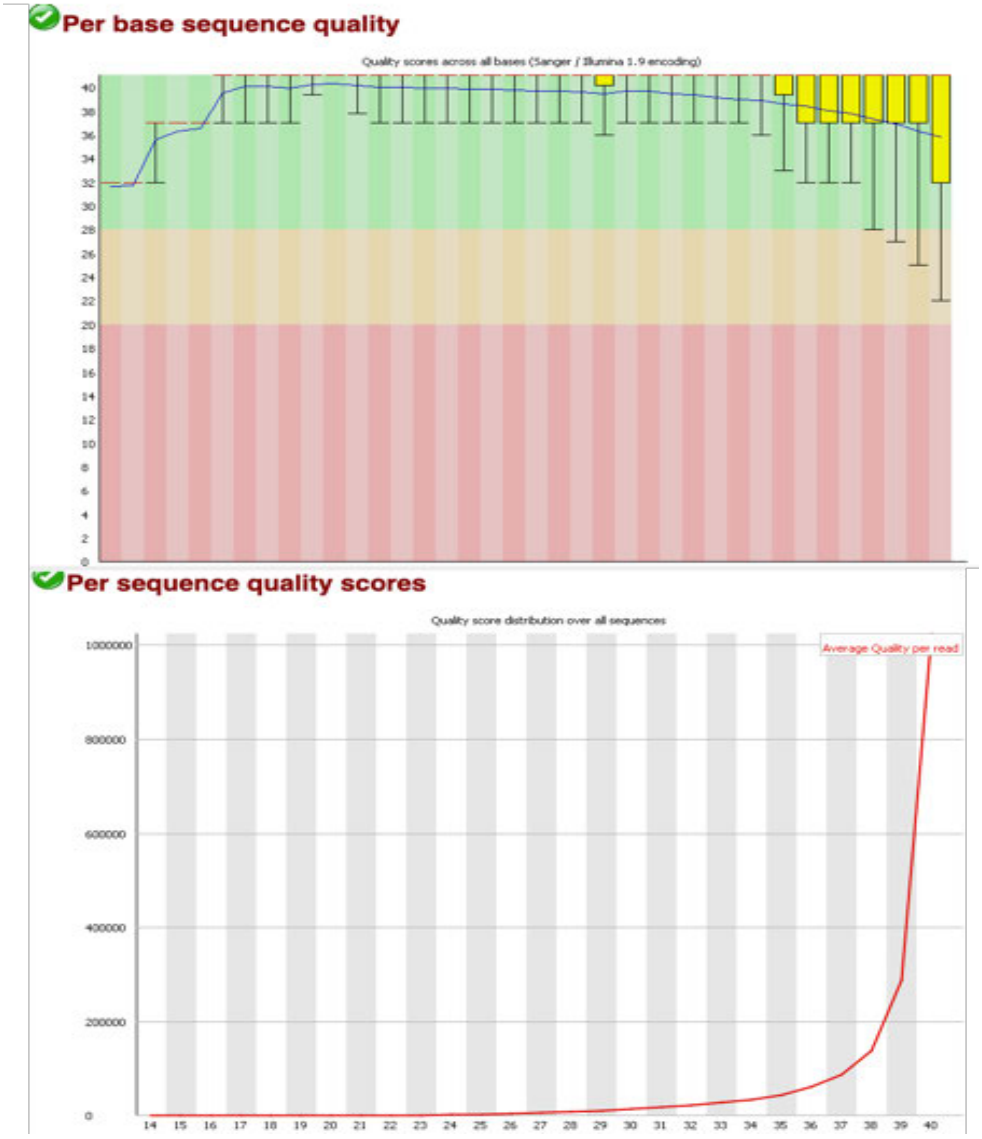


Figure 6-3 FASTQC output of one of the sequenced sample, 4n Columbia parent.

6.3.2 Mapping and Alignments

The sequenced and trimmed reads were aligned with the Bowtie2 aligner. The total number of mapped or aligned reads along with reads aligned after filtering for mapping quality 20 and uniquely aligned reads that were found can be seen in **Table 6-4.** It represents a good overall alignment for all the samples.

Sample	% overall alignment (2n)	% aligned reads (2n)	% unique alignments (2n)	% overall alignment (4n)	% aligned reads (4n)	% unique alignments (4n)
P1	92.4	59.5	52.2	93.6	57.4	50.4
P2	88.1	49.4	46.2	90.0	49.0	45.8
F1_1	90.6	56.9	51.5	93.8	57.1	51.5
F1_2	92.6	60.1	54.5	93.9	56.0	50.6
F2_1	89.9	54.0	49.8	91.9	54.7	49.5
F2_2	89.4	53.3	49.2	91.9	55.9	51.1
F2_3	91.9	57.9	52.4	93.9	60.4	55.0
F2_4	89.8	54.1	50.5	91.9	56.2	51.3
F2_5	90.5	53.4	48.7	91.8	55.8	50.7
F2_6	90.9	53.6	49.0	92.8	55.8	50.8
F2_7	91.7	55.9	51.5	93.6	56.4	50.9
F2_8	90.0	54.3	50.4	91.7	55.0	49.8

Table 6-4 Percentage of mapped reads.

% overall alignment indicates the output of bowtie 2 alignment,% aligned reads indicate output of bam files filtered for mapping quality 20 and% unique alignments are the alignments matched exactly at one place from aligned reads with mapping quality greater than 20.

6.3.3 Variants called

The uniquely aligned reads of all the 24 samples were extracted and variant calling was performed only on the unique reads with mean base quality ≥ 20 and mapping quality ≥ 20, which produced variants as shown in **Table 6-5**. Total number of variants called were large which reduced to 7217 SNPs and 506 indels in diploids, and 6761 SNPs and 609 indels in tetraploids, on filtering the variants based on the genotype quality and sample depth.

	Diploid		Tetraploid	
	Total	After filtering per sample	Total	After filtering per sample
Candidate Variants	56,976	7723	53,728	7370
SNPs	49,866	7217	46,646	6761
INDELs	7110	506	7082	609

Table 6-5 The number of genetic variants detected from Arabidopsis thaliana diploid and tetraploid RAD-seq datasets.

The average read depths of variant sites in different samples across all variants is shown in

Table 6-6.

Sample	Average Read depth (2n)	Average Read depth (4n)
P1	24.9	29.1
P2	23.8	26.3
F1_1	25.9	35.5
F1_2	27.5	35.7
F2_1	44.8	48.7
F2_2	41.1	44.3
F2_3	48.2	40.7
F2_4	47.2	47.8
F2_5	42.4	41.4
F2_6	45.5	50.6
F2_7	51.1	53.1
F2_8	45.7	45.7

Table 6-6 Average read depth across the variant sites in diploid and tetraploid samples.

6.4 Discussion

A variant calling pipeline was developed for the sequenced diploid samples, which yielded a high number of variants after quality filtering. The same pipeline was used here for tetraploids as well, to give a useful comparison. Bowtie 2 aligner was preferred over BWA-Mem for alignment of the sequenced bases to the reference genome. Bowtie 2 is relatively faster for shorter read sequences (Langmead and Salzberg, 2012), such as our 2 x 150 bp paired end sequence reads. Only unique alignments were used for variant calling in both diploids and tetraploid samples, which is expected to increase the quality of downstream analysis.

Further analysis needs to be performed on the 24 samples sequenced to provide the preliminary differences in meiotic recombination between the diploid and the tetraploid F2 populations. A more appropriate variant caller and genotype dosage assignment tool as described in **6.1.2** can be used to call genotypes from the tetraploid samples. The variants so called can be further validated using the annotated polymorphism data between the Col and Ler lines from the TAIR database, as was also done for *A. thaliana* diploid and tetraploid parental lines in Jiang *et al*. (2016).

DNA from more than 400 diploid and tetraploid F2 has been extracted, which will be sequenced in future (**Appendix C**). Further, collected leaf samples (161 diploids and 237 tetraploids) will be processed to enable RAD-seq genotyping of the full *A. thaliana* populations created. Molecular marker analysis will be then carried out to analyse differences in the frequency and distribution of meiotic recombination in diploids vs tetraploids, along with

studying meiotic recombination patterns. Comparative QTL analysis of various quantitative traits in the diploid and autotetraploid F2 population will also be conducted. This will help to dissect the genetic architecture of the differentiating traits between the diploids and tetraploids.

6.5 References

Baird, N. A. *et al.* (2008) 'Rapid SNP discovery and genetic mapping using sequenced RAD markers.', *PloS one*. doi: 10.1371/journal.pone.0003376.

Blischak, P. D., Kubatko, L. S. and Wolfe, A. D. (2018) 'SNP genotyping and parameter estimation in polyploids using low-coverage sequencing data', *Bioinformatics*. doi: 10.1093/bioinformatics/btx587.

Bourke, P. M. *et al.* (2018) 'Tools for Genetic Studies in Experimental Populations of Polyploids', *Frontiers in Plant Science*. doi: 10.3389/fpls.2018.00513.

Clark, L. V., Lipka, A. E. and Sacks, E. J. (2019) 'polyRAD: Genotype Calling with Uncertainty from Sequencing Data in Polyploids and Diploids', *G3: Genes|Genomes|Genetics*. doi: 10.1534/g3.118.200913.

Clevenger, J. *et al.* (2015) 'Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations', *Molecular Plant*. doi: 10.1016/j.molp.2015.02.002.

Davey, J. *et al.* (2011) 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', *Nature Reviews Genetics*, 12(7), pp. 499–510. doi: 10.1038/nrg3012.

Depristo, M. A. *et al.* (2011) 'A framework for variation discovery and genotyping using next-generation DNA sequencing data', *Nature Genetics*. doi: 10.1038/ng.806.

Garrison, E. and Marth, G. (2012) 'Haplotype-based variant detection from short-read sequencing', pp. 1–9. Available at: http://arxiv.org/abs/1207.3907.

Gerard, D. *et al.* (2018) 'Genotyping polyploids from messy sequencing data', *Genetics*. doi: 10.1534/genetics.118.301468.

Hackett, C. A., McLean, K. and Bryan, G. J. (2013) 'Linkage Analysis and QTL Mapping Using SNP Dosage Data in a Tetraploid Potato Mapping Population', *PLoS ONE*. doi: 10.1371/journal.pone.0063939.

Jiang, N. *et al.* (2016) 'A highly robust and optimized sequence-based approach for genetic polymorphism discovery and genotyping in large plant populations', *Theoretical and Applied Genetics*. doi: 10.1007/s00122-016-2736-9.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2.', *Nature methods*. doi: 10.1038/nmeth.1923.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*. doi: 10.14806/ej.17.1.200.

Maxam, A. and Gilbert, W. (1977) 'A new method for sequencing DNA. 1977.', *Proc. Natl. Acad. Sci*. doi: 10.1073/pnas.74.2.560.

McCallum, S. *et al.* (2016) 'Construction of a SNP and SSR linkage map in autotetraploid blueberry using genotyping by sequencing', *Molecular Breeding*. doi: 10.1007/s11032-016-0443-5.

Miller, M. R. *et al.* (2007) 'Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers', *Genome Research*. doi: 10.1101/gr.5681207.

Pereira, G. S., Garcia, A. A. F. and Margarido, G. R. A. (2018) 'A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids', *BMC Bioinformatics*. doi: 10.1186/s12859-018-2433-6.

Sanger, F. and Nicklen, S. (1977) 'DNA sequencing with chain-terminating', *Proc. Nati. Acad. Sci. USA*.

Schmitz Carley, C. A. *et al.* (2017) 'Automated tetraploid genotype calling by hierarchical clustering', *Theoretical and Applied Genetics*. doi: 10.1007/s00122-016-2845-5.

Serang, O., Mollinari, M. and Garcia, A. A. F. (2012) 'Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids', *PLoS ONE*. doi: 10.1371/journal.pone.0030906.

Shendure, J. *et al.* (2017) 'DNA sequencing at 40: Past, present and future', *Nature*. doi: 10.1038/nature24286.

Toonen, R. J. *et al.* (2013) 'ezRAD: a simplified method for genomic genotyping in non-model organisms', *PeerJ*. doi: 10.7717/peerj.203.

Truong, H. T. *et al.* (2012) 'Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations', *PLoS ONE*. doi: 10.1371/journal.pone.0037565.

Voorrips, R. E., Gort, G. and Vosman, B. (2011) 'Genotype calling in tetraploid species from bi-allelic marker data using mixture models', *BMC Bioinformatics*. doi: 10.1186/1471-2105-12-172.

CHAPTER 7

MEIOTIC CHROMOSOME BEHAVIOUR IN *SOLANUM TUBEROSUM*

7 Meiotic chromosome behaviour in *Solanum tuberosum*

7.1 Introduction

The cultivated potato, *Solanum tuberosum* belongs to family Solanaceae, section petota, which also includes tomato, peppers, brinjal, petunia and tobacco. Tubers, the modified stems, are the most important part of the plant and are used both as food and as the source of propagation. Potato is the third most important food crop after rice and wheat (Gaiero *et al.*, 2016), and provides a rich source of vitamins, especially vitamin C and minerals (Camire, 2016). It originated in South America almost 10,000 years ago, from where it spread to the whole world, now being grown in 160 countries and having 4000 cultivars (Camire, 2016). Considering its importance as an important global food crop, the Food and Agricultural Organisation of the United Nations (FAO) declared 2008 as the International Year of the potato. Though some potato varieties are capable of producing true seeds borne in the fruit pods, they are propagated asexually through tubers.

As discussed in Chapter 1 (**1.2**), many important crop varieties are polyploids, where they have more than two sets of chromosomes. These polyploid plants may face a reduction in fertility due to difficulties in meiosis, which serves as a bottleneck in their propagation, especially at an early stage of polyploid development. In the long term, they may develop ways to overcome this problem or they may be maintained through asexual propagation. Most cultivated potatoes are considered autotetraploid (2n=4x=48) having 4 sets of homologous chromosomes. However, diploid, triploid, pentaploid and hexaploid varieties also exist

(Gavrilenko, 2007). The potatoes also have high heterozygosity due to their outcrossing behaviour.

Figure 7-1 Solanum tuberosum, variety Sante plant with flowers.

In **Figure 7-1**, a Sante potato plant with flowers in bloom can be seen. The flowers occur in an inflorescence and have fused 5 petals. There are five anthers and a pistil coming out of those anthers. A good amount of pollen could be collected by gently agitating the flower from the rear. In **Figure 7-2**, tubers from tetraploid and diploid varieties can be seen and an immediate difference in the size is visible. Two different colours of flowers in the panel 'd' and 'e' for varieties Maris Peer and Cara can also be seen in **Figure 7-2**. The frame in panel 'f' in **Figure 7-2** shows a berry growing on a Cara plant in the glass house. Berry formation is rare in some varieties, while more common in others. One of the reasons could be the difficulties in meiosis due to the presence of 4 sets of each chromosome leading to mis-segregation during anaphase.

Figure 7-2 Tubers, Flower and Berry from different varieties.

a and d show the tubers and flower of tetraploid variety Maris Peer. b, e and f show the tubers, flower and berry in the frame of tetraploid variety Cara. c shows tubers of diploid variety Mayan Gold.

Being autotetraploid, potato has complex genetics owing to a polysomic pattern of inheritance. A polysomic pattern of inheritance occurs when there is more than one homologous partner to pair with during meiosis. In polyploids, two kinds of pairing behaviour are possible, namely random pairing and preferential pairing. Random pairing, where all the homologues are free to pair with each other may lead to the formation of multivalents. In a tetraploid, there may be up to four distinct alleles present at each locus. This may lead to various combinations of alleles in gametes during bivalent pairing, which are more complex to predict than in diploids. There may be a combination of bivalent and multivalent segregation in meiosis. Double reduction is a direct effect from multivalent formation, where

identical alleles carried on sister chromatids may end up in the same gamete (Wu *et al.*, 2001). On the other hand, preferential pairing where two copies of closely related chromosomes pair, out of the four chromosome copies present, mostly occurs in allopolyploids leading to normal bivalent pairing with little to no chance of double reduction (Sybenga, 1975).

Potato is an important autotetraploid crop plant and an insight into its cytogenetics can help improve the breeding efforts. Traditionally, karyotyping and cytogenetic studies have been difficult in potato species due to small mitotic metaphase chromosomes and a low level of divergence among the species, which made it difficult to differentiate between different types on a cytogenetic basis (Gavrilenko, 2007). Initial analysis on somatic chromosomes was made by using aceto-carmine, DNA binding dyes, followed by Giemsa C-banding techniques to analyse each chromosome based on the differences in the distribution of highly repetitive DNA sequences (Gavrilenko, 2007).

More recently, FISH signals from RFLP tagged bacterial artificial chromosomes (BACs) were developed and used as chromosome specific cytogenetic DNA markers, which enabled identification of all 12 somatic metaphase chromosomes in a haploid potato line (2n=2x=24) (Dong *et al.*, 2000). They were able to identify and map the 5S rRNA, 45S rRNA and potato late blight resistance gene to specific locations on chromosomes 1, 2 and 8 respectively. A pachytene cytogenetic map (meiotic), which identified the arms of 12 potato bivalents was developed using a set of 60 FISH BAC clones, 5 FISH signals for each chromosome, in a diploid line (Tang *et al.*, 2009). These FISH BAC probes have been utilised in other studies to find the collinearity between the wild cultivars and the cultivated *S. tuberosum,* to enable

introgression of various useful wild type genes into the cultivated variety (Gaiero *et al.*, 2016). These techniques can be applied to polyploid varieties directly to study the differences in meiosis in polyploid plants from the diploids.

Most recently, oligo probes have been developed and used to identify the different mitotic chromosomes in potato (Braz *et al.*, 2018). The usefulness and cost effectiveness of oligo probes over BAC probes was also discussed, where it was suggested that oligo probes can be developed for any plant with sequence availability and can be designed for use in different varieties (Braz *et al.*, 2018). BAC probes on the other hand, though useful, are difficult to generate for each chromosome and may produce unspecific signals in FISH in plants with large genomes having large amounts of repetitive DNA (Braz *et al.*, 2018). Oligo probes have also been recently used in potato for pairing analysis of chromosomes 2, 4, 7 and 11 during prophase I and metaphase I of meiosis (He *et al.*, 2018). The probes could successfully identify the individual chromosomes in mitotic cells as well as the pairing pattern in pachytene, diakinesis and early M1s in diploid, tetraploid and hexaploid potato. These are exciting new developments in the field of molecular cytology in cultivated potato.

Here, a cytological analysis has been carried out on chromosomes 1 and 2 (bearing 5S and 45S rDNA), in different potato varieties, building up on the knowledge and skills gained in *Arabidopsis thaliana* (hereafter referred to as *A. thaliana* or *Arabidopsis*).

7.2 Potato material

Four tetraploid and two diploid varieties of cultivated potato have been used for this work. The information about them can be found on Agriculture and Horticulture Development Board (AHDB, a levy board in UK working alongside farmers offering consultation and strategic advice about farming) website http://varieties.ahdb.org.uk/varieties. The tetraploid varieties belong to the association panel of 300 varieties as used in a recent study (Sharma *et al.*, 2018) and are as follows:

Sante: This variety originated in Netherlands and was released in 1983. It can have an early, intermediate and late maturity. It has multipurpose use in crisps, flour, and french fries with low after cooking darkening. According to AHDB Sante has good resistance to Potato virus Y^O, dry rot, powdery scab and late blight on tubers and is therefore grown commonly. Sante produces white flowers but no berries.

Sarpo Mira: This variety originated in Hungary and has a very late maturity. It has multipurpose use and can produce good french fries and flour. According to AHDB, Sarpo Mira has good resistance to blackleg, late blight on foliage, late blight on tubers and Potato virus Y^O. It produces purplish flowers and few berries.

Cara: This variety originated in Ireland and was released in 1973. It can have late to very late maturity and can be used as flour or salad, though it does not produce good crisps. According to AHDB, Cara is a robust and high yielding variety with resistance to late blight on tubers, Potato Virus Y^O and also to cyst nematode *Globodera rostochiensis* Ro 1. It produces white flowers and medium frequencies of berries. In our work, Cara was the only plant that produced two berries.

Maris Peer: This variety originated in UK and was released in 1962. It has an early maturity and boils good. According to AHDB, Maris Peer has good resistance to powdery scab, gangrene damage and bruising. It produces purplish flowers with rare production of berries.

The diploid varieties used in the study are as follows:

Mayan Gold: This diploid variety originated in UK. It has very late maturity with tall plants. According to AHDB, Mayan Gold has high resistance to common and powdery scab and good for chip making. It produces purplish flowers and have frequent berry formation though there was no berry production in our glasshouse trial.

Scapa: This diploid variety originated from a cross between Mayan Gold and Mayan Twilight. According to AHDB, Scapa has high resistance to common scab. It produces purplish flowers and has a late maturity.

7.3 Results

Four tetraploid and two diploid European varieties of *S. tuberosum* were grown in a glasshouse trial in the summer of 2017 and 2018. The buds collected from these varieties were processed for cytology. Variety Sarpo Mira does not have any cytology result as most of its buds were used as recipient for pollen from variety Sante. The remaining few buds were fixed but a comprehensive data could not be collected from them. Cytological results of the remaining tetraploid varieties, Sante, Maris Peer, Cara and diploid varieties Mayan Gold and Scapa is presented here.

7.3.1 Checking pollen viability

Crossing between two tetraploid varieties, Sante and Sarpo Mira was tried where Sante pollen was used to pollinate Sarpo Mira buds or vice versa, but it was not successful. Pollen viability check to ensure they were not sterile was performed for Sante using Alexander staining.

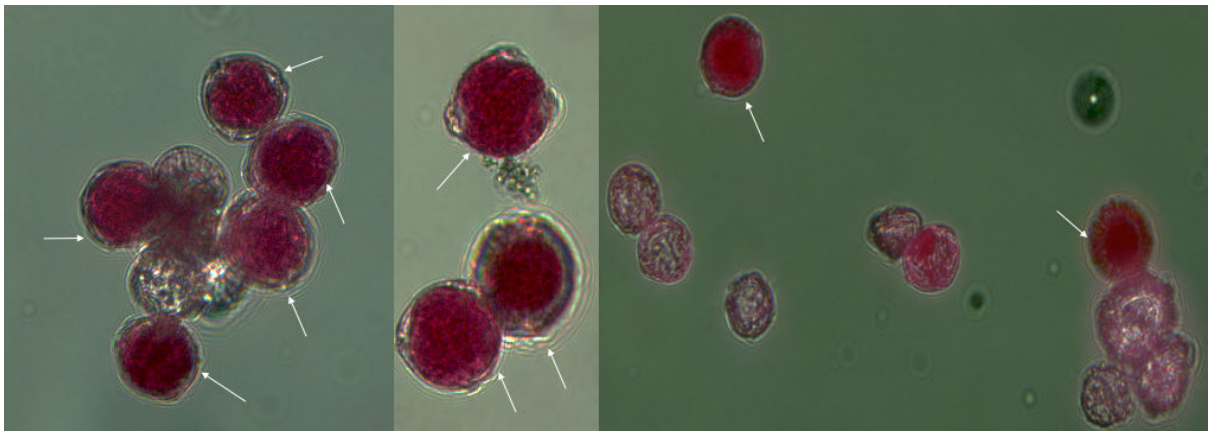


Figure 7-3 Alexander staining to check pollen viability in Sante.

Viable pollen are dark red and non-viable are pale pinkish. White arrows indicate the viable pollen.

The viable pollen takes up the dark red stain as indicated by white arrows in the **Figure 7-3**, whereas the non-viable pollen is very lightly stained. The pollen which appears pale pink and not marked by arrows are non-viable pollen. Though a count was not performed, just by looking at the picture above, the ratio of viable to non-viable pollen is roughly 1:1. This indicates a reduced fertility but not sterility. The plant did not produce any berries and the cross pollination using Sante pollen onto a different variety Sarpo Mira was unsuccessful. This indicates that there are some other factors involved, which do not allow the crossing to be successful. Flower dropping two to three days after pollination was observed. This could be one of the major reasons for the crossing to be unsuccessful.

7.3.2 Production of Meiotic Atlas

After the buds from different tetraploid and diploid varieties of *S. tuberosum* were collected and fixed, the first step was to find out and correlate anther sizes with different meiotic stages. It was found that all the anthers in a single bud were not of the same size and could be in different stages of meiosis. This worked as a boon and bane at the same time. Boon, because it increased the chances of finding more stages in the anthers from one bud itself and bane because it meant that only one cell with a particular stage may be present in a pool of cells at a different, but same meiotic stage. For example, a single M1 cell could be present among a pool of pachytenes. It was also found that a few stages would be present in a range of anther sizes. The range of sizes of anthers and the corresponding stage can be seen in **Table 7-1**. It can be seen that a range of meiotic stages could be present in the same anther or they may in present in a range of anthers. Therefore, a large number of slides were required to capture most of the stages. This table was made considering anther sizes - meiotic stage relationship

in at-least two different varieties. Only a small difference can be identified between the diploid and tetraploid varieties in terms of meiotic stages to be found in a particular anther size. For example, in anther range size 1.7-1.85, leptotene to pachytene could be found in diploid, while only zygotene to pachytene stages were found in tetraploids.

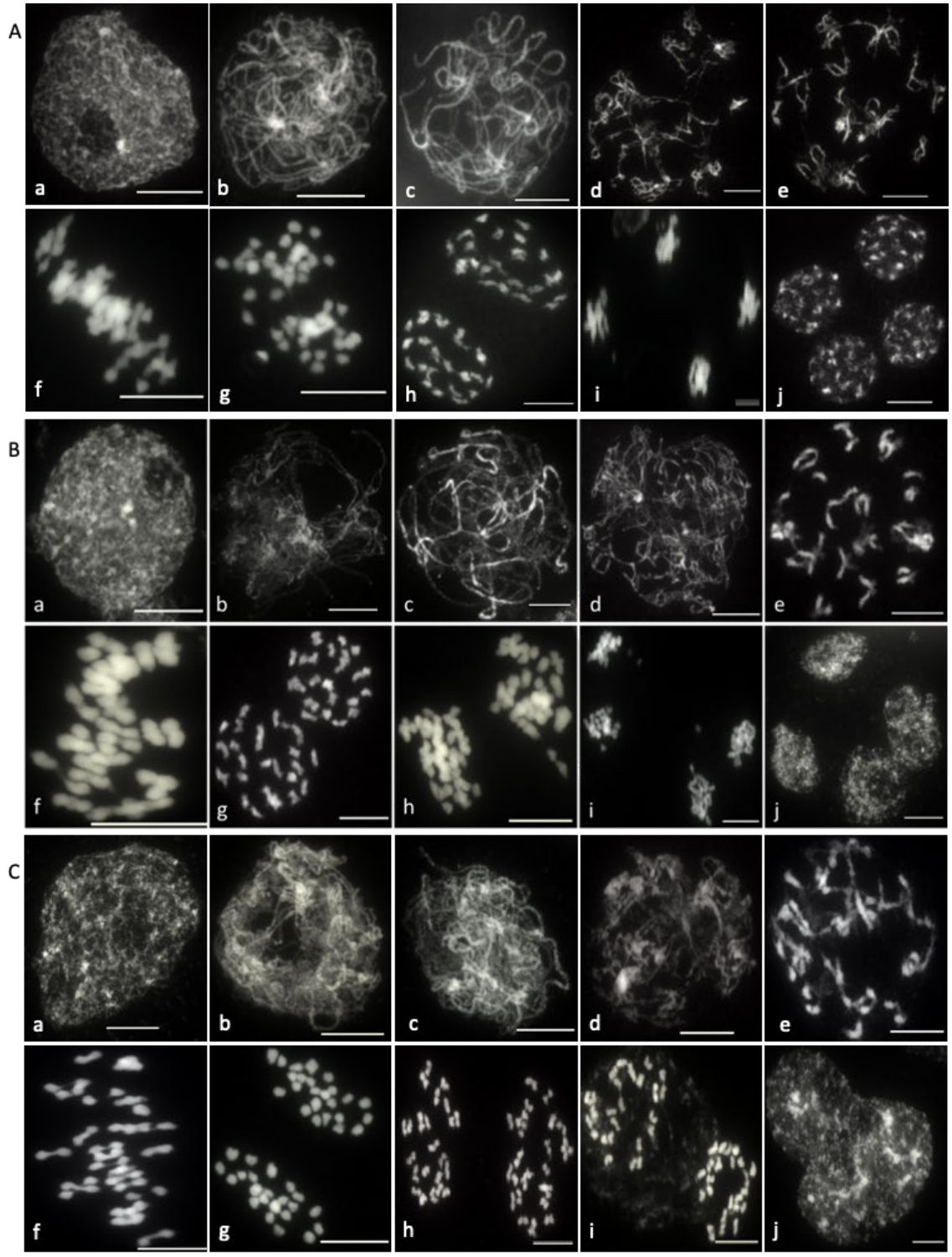
Anther Size (mm)	**Diploid (meiotic stage)**	**Tetraploid (meiotic stage)**
1.5	G2 and leptotene	G2
1.6	Leptotene-zygotene	Leptotene-zygotene
1.65-1.7	Leptotene-zygotene-pachytene	Leptotene, zygotene
1.7-1.85	Leptotene-zygotene-pachytene	Zygotene-pachytene
1.9-2.0	Diakinesis-metaphase I-tetrad	Diakinesis-metaphase I
2.0-2.1	Metaphase II-tetrad	Metaphase I-tetrad

Table 7-1 Anther sizes and meiotic stages in diploid and tetraploid Solanum tuberosum.

A time course analysis of different stages was not done, though it is clear that most time was spent in prophase I of meiosis, as it was comparatively easier to find cells in different stages of prophase I than in any other stages. Most of the time, though not a rule, when the chromosome spread was prepared, late zygotenes, diplotenes and late diakinesis would be found in abundance, and there would be paucity of later stages. This indicates that homologous identification, pairing and chiasma formation takes up most of the time. In tetraploids particularly, correction in pairing and interlock resolution may take more time due to formation of multivalents. M1 was a difficult find, both in diploids and tetraploids. In tetraploids however, when found they occurred in good numbers, but they were elusive in

diploids. This indicates that meiosis goes away quickly after pairing and homologous recombination has been sorted in prophase I.

An atlas of different meiotic stages is presented in **Figure 7-4**. Panel A, B and C represent tetraploid varieties Maris Peer, Cara and Sante, respectively. Most of the stages could be identified in these varieties. Leptotene to M1 can be seen in all three of them. A clear anaphase I was not seen in Cara, while a clear metaphase II was not seen in Maris Peer. A clear presence of 24 bivalents in M1 in tetraploids was only seen in a few cells and the spindle arrangement looked overcrowded. A clear 24 bivalents can be counted in variety Sante in Panel C of the figure in metaphase I, and 24 homologues could be counted in the dyad stage or metaphase II stage in all of them. In panels D and E, diploid varieties Mayan Gold and Scapa atlases can be seen. Prophase I stages can be clearly seen in both the diploid varieties. For Mayan Gold, only pre-M1 cells were found as can be seen in 'f' in panel D in **Figure 7-4**, they too were few and far between. However, a clear 12 bivalents are visible in both the varieties during this stage. This is in contrast to the tetraploids, where 24 bivalents were not always clearly visible. Similarly, 12 homologues could also be counted during the metaphase II stage.


A
a
b
c
d
e
f
g
h
i
j
B
a
b
c
d
e
f
g
h
i
j
C
a
b
c
d
e
f
g
h
i
j

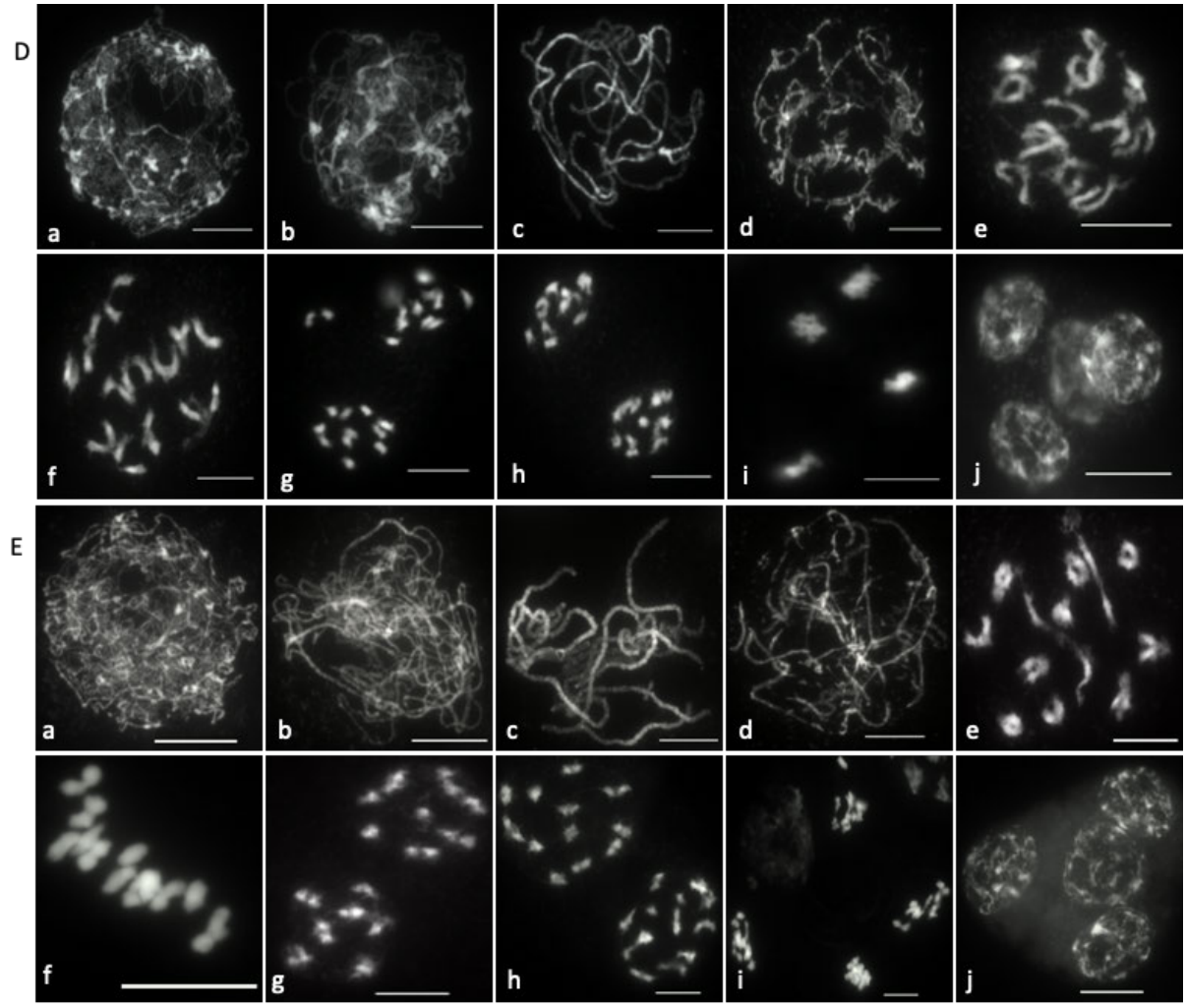


Figure 7-4 Meiotic Atlas of different varieties of potato.

The figure shows tetraploid varieties Maris Peer, Cara and Sante in A, B and C panels respectively and diploid varieties Mayan Gold and Scapa in D and E respectively. Scale bar is 10 μm. a. Leptotene, b. Zygotene, c. Pachytene, d. Diplotene, e. Diakinesis, f. Metaphase I, g. Anaphase I, h. Metaphase II, i. Anaphase II, j. Tetrad.

7.3.3 Identification and Immunolocalisation of ASY1 and ZYP1 proteins

ASY1 and ZYP1 are two important structural proteins, coordinating essential loading and interaction of various precursors and maintenance proteins for the successful axis modulation and synaptonemal complex formation. In *Arabidopsis*, ASY1 is a HORMA domain protein, which is known to interact with chromatin, leading to a failure in synapsis in the event of its mutation (Armstrong, 2002). Similarly, ZYP1 in *Arabidopsis* was identified as the SC

(synaptonemal complex) protein essential for its formation. Defects in the SC protein lead to a delay in the progression of meiosis and in the meiotic recombination fidelity, leading to multivalent formation (Higgins *et al.*, 2005). A keyword search for ASY1, asynaptic 1 in the spud-db potato database (http://solanaceae.plantbiology.msu.edu/cgi-bin/annotation_search.cgi) under functional annotation search gave a result for Meiotic asynaptic mutant 1 labelled PGSC0003DMP400050690, the protein sequence of which matched with *A. thaliana* AT1G67370.1, the ASY1 protein available on the TAIR 10 database. The potato ASY1 showed a 76.7% similarity with *Arabidopsis* ASY1 in the spud-db database. A query for potato predicted protein sequence with *Arabidopsis* sequence as subject was carried out using Blast, which showed 62% identity and 76% similarity, as can be seen in **Figure 7-5**. Percent identity indicates the exact amino acid match and similarity indicates a substitution in amino acid with similar physiochemical properties (Madden, 2002). The first 180 amino acids matched the HORMA domain superfamily on NCBI website (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

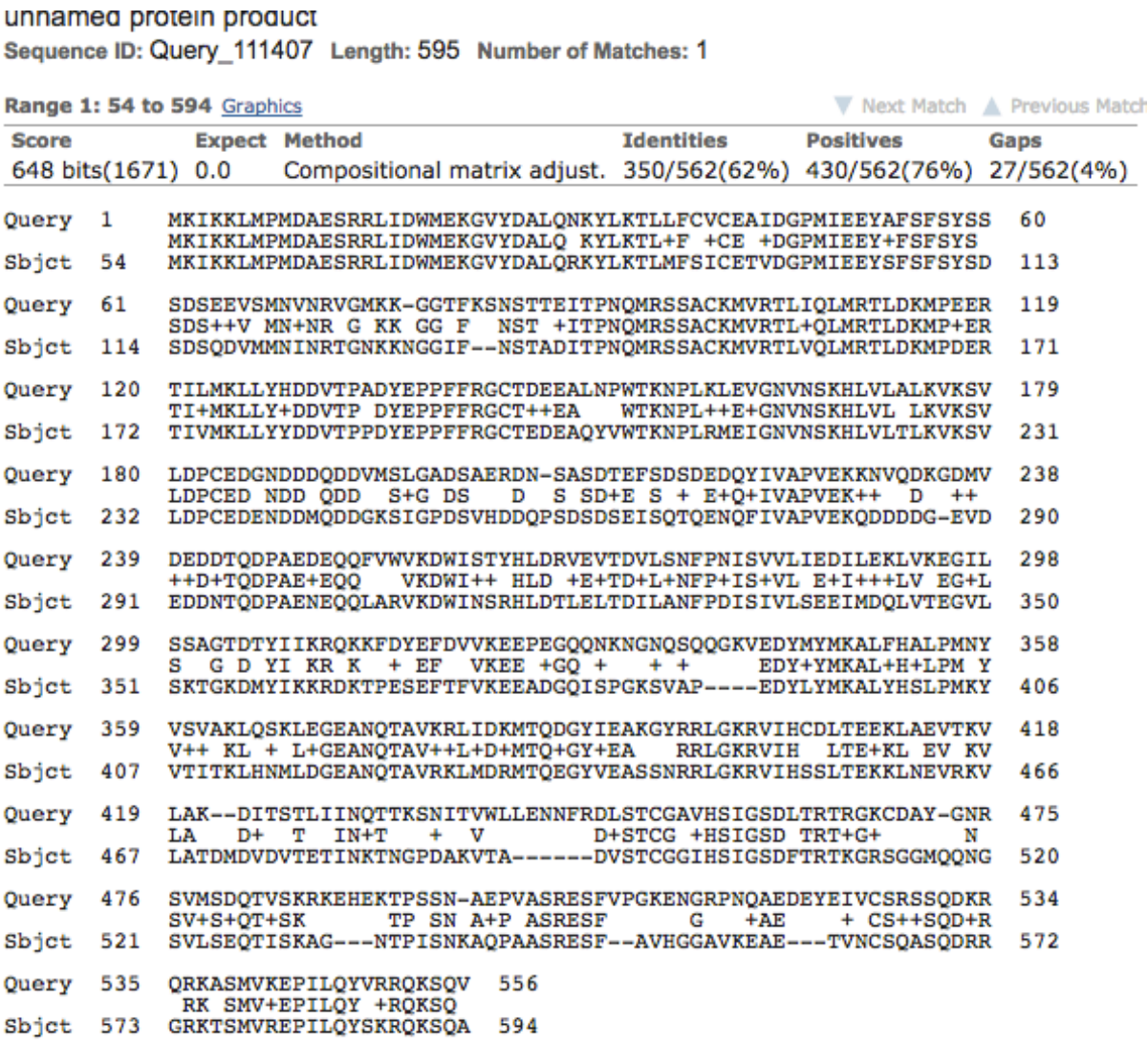


Figure 7-5 Blast output showing differences and similarities in axis protein ASY1 *between Arabidopsis thaliana and Solanum tuberosum.*

To search for ZYP1, an alternative approach was employed. Typing in synaptonemal complex protein, zipper protein or transverse filament protein in the spud-db keyword functional annotation search did not give any conclusive results, and instead produced more than 20,000 hits. Hence, the TAIR annotations for *Arabidopsis* ZYP1a and ZYP1b were used to find out their protein sequences from the TAIR website. Those sequences were blastp in the spud-db database, which resulted in a ribosome binding protein PGSC0003DMP400011376, which

gave 56% similarity with ZYP1a and 67% with ZYP1b. The AtZYP1a protein sequence was blasted in NCBI against *Solanum tuberosum* for alignments. The aligned protein sequences are shown in **Figure 7-6** showing 48% identity and 68% similarity. The top alignment for this protein in NCBI blastp was *S. tuberosum* predicted synaptonemal complex protein 1-like. The amino acids from place 42 to 684 matched the Smc superfamily on NCBI website, the proteins which are involved in cell cycle control.

Score	Expect	Method	Identities	Positives	Gaps
699 bits(1805)	0.0	Compositional matrix adjust.	414/869(48%)	595/869(68%)	37/869(4%)

```
Query  1    MKGLDHFKSLSGSGVGAAKTMSIPTRMSSDMVSTGSFANLKLTAEKLVKEQASAKTDLQL  60
            MK LD  +SLSGS    A   S   R   D VS+GSF+NLKLTAEKLVK+QA+ +TDL+L
Sbjct  9    MKSLDKPRSLSGS----ANMYSFSNRKPPDSVSSGSFSNLKLTAEKLVKDQAAMRTDLEL  64

Query  61   ASSKLKKLTEEVHMLEEKLKNAYNENAKLKVKQKEDEKLWKGLESKFSSTKTLCDQLTET  120
            A+ KLKK  E V+ LEEKL+NA+NENAKL+V++KEDEKLW+GLESKFSSTKTLCDQLTET
Sbjct  65   ANCKLKKSMEHVYALEEKLQNAFNENAKLRVRKKEDEKLWRGLESKFSSTKTLCDQLTET  124

Query  121  LQHLAGVVQDAEKDKAFFEDRQSATSVVVDNLQDDLKSLSLRLESSEETVRNCKRELNEL  180
            LQHLA  VQDAEKDK FFE + S +S   +D+L    ++ +SLRL++++E + +   +EL EL
Sbjct  125  LQHLASQVQDAEKDKGFFETKFSTSSEAIDSLNQQMRDMSLRLDAAKEEITSRDKELEEL  184

Query  181  GIEKEKMKNCFMVEQSKCTSVIGEKDAMIKEFEATVAVNGLAVENLKNKLEELHLESRLK  240
             +EK++ +   +  E+     S+I +KDA+I + EA+ A    L +ENL ++LE++HLE   K
Sbjct  185  KLEKQQKEMFYQTERCGTASLIEKKDAVITKLEASAAERKLNIENLNSQLEKVHLELTTK  244

Query  241  EDKLEDLRTAKKNAEKEKSDL-VSKNSEFAKQLDTSLQEIKNLNEFVNEMVVKVTDLDSQ  299
            ED+++DL + ++  EKEK+ + +S ++ F K L +S QE+K L+E V  +V ++T+LD +
Sbjct  245  EDEVKDLVSIQEKLEKEKTSVQLSADNCFEK-LVSSEQEVKKLDELVQYLVAELTELDKK  303

Query  300  SLAFAEKIIQLTALFDSGFEMMRERRELAAQHAQQKFGKLHDQYTSITSEKNALLLANKD  359
            +L F EK  +L+ L+D+   ++++ R+LA   AQ+ F  L  +   + + K AL  A  +
Sbjct  304  NLTFKEKFDKLSGLYDTHIMLLQKDRDLALDRAQRSFDNLQGELFRVAATKEALESAGNE  363

Query  360  LKDKVSALQKEQEHAMVQHAQESLLAEDQIRKLDSEVELLLSKKKEMELLISKLQENIVT  419
            L +K+  LQ ++E  + Q +          I KL+SE + L+SK  + E  IS+L+E + T
Sbjct  364  LNEKIVELQNDKESLISQLSGLRCSTSQTIDKLESEAKGLVSKHADAESAISQLKEEMET  423

Query  420  LSDSSKLSENEMQNLSLKLSEMETENKDHIGKLQSDMQKKEDEIHVLLKEIDNYTETVDS  479
            L +S K SE++ Q LSLKLS +E E+K+   KLQ+D Q++ +E+  L KE +++     D
Sbjct  424  LLESVKTSEDKKQELSLKLSSLEMESKEKCEKLQADAQRQVEELETLQKESESHQLQADL  483

Query  480  LEKHVTEINNKLEEKDQLVQELQDKEKQLEAEREKIQASLLAAESKLTESKKQYDQMLES  539
            L K V ++   +EEK  ++ +  + EKQL  +  K +  L  AE+KL E+KKQYD MLES
Sbjct  484  LAKEVNQLQTVIEEKGHVILQCNENEKQLNQQIIKDKELLATAETKLAEAKKQYDLMLES  543

Query  540  KQLELSRHLKELSQKNDQAINDIRRRYDLEKLESVNLEKEKAEKIVGEMEKNCELKLSEC  599
            KQLELSRHLKELSQ+NDQAIN+IRR+YD+EK E +N EK+K EKI+ ++     + +LS+C
Sbjct  544  KQLELSRHLKELSQRNDQAINEIRRKYDVEKHEIINSEKDKVEKIIKDLSNKFDKELSDC  603

Query  600  REESKQNLKHVQEEHANLVCQIQQAHSRKEMSLVASHNEELKRSRFQYENELREKTNSMR  659
            +EESK+ L  +QEEH++L+  +++ H  KE++L A +++EL++S+ Q ENEL+E+  +++
Sbjct  604  KEESKRQLLTIQEEHSSLILSLREEHESKELNLKAKYDQELRQSQIQAENELKERITALK  663

Query  660  NEHEAQLRALRLELEDDSRRLQEELYMQKSKEEKQRALLQLQWKVMGDN-PEEEEVTSKK  718
            +EH+AQL+A + + EDD ++LQEEL +Q+ KEE+QRAL+QLQWKVM DN PEE+EV S K
Sbjct  664  SEHDAQLKAFKCQYEDDCKKLQEELDLQRKKEERQRALVQLQWKVMSDNPPEEQEVNSNK  723

Query  719  NYSCSVTKR--NQPDSGKPPVRAEAKDV-DSHYLVGNQIPVSNLLRKVEQVNSGSL----  771
            NYS S   R      S    VR++  +V DS ++   + PVS +L+K + VN+GS+
Sbjct  724  NYSISKDSRLGGSKRSEHIRVRSDNDNVQDSPFVKAKETPVSKILKKAQNVNAGSVLSIP  783

Query  772  -PQHSRKVTHHEYEVETTNSRIITKRRKTKSTVMFD----------------VVKKG---  811
             P+H  KVTH EYEVET N R +TKRRKT++T MF+                 + KG
Sbjct  784  NPKHHSKVTHREYEVETNNGR-VTKRRKTRNTTMFEEPQRRRTRATPKLTPQSIAKGTGM  842

Query  812  --HPKPANIGDLFSEGSLNPYADDPYAFD  838
              H + ANIGDLFSEGSLNPYADDPYAFD
Sbjct  843  TSHARSANIGDLFSEGSLNPYADDPYAFD  871
```

Figure 7-6 Blast output showing differences and similarities in synaptonemal complex protein ZYP1 *between Arabidopsis thaliana and Solanum tuberosum.*

After establishing the identity of ASY1 and ZYP1 in potato, immunolocalisation of the two proteins was carried out at meiotic prophase I using antibodies derived from *A. thaliana,* as good protein level similarities existed between the two in blastp. Using fresh anthers as well as pre-prepared chromosomal spread slides as the material, the *Arabidopsis* antibodies against ASY1 and ZYP1 were found to be effective in potato.

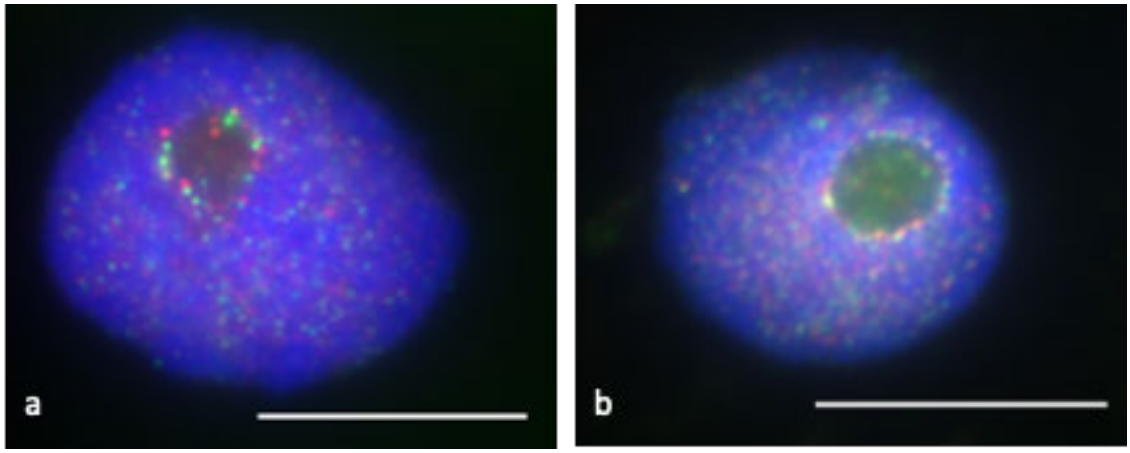


Figure 7-7 Nucleus in G2 stage showing ASY1 *(green) and* ZYP1 *(red) foci signals.*

a. Sante and b. Scapa showing clear nucleolus in the cell. Scale bar is 10 μm.

In **Figure 7-7**, a nucleus in the G2 stage is visible in both a tetraploid variety Sante and a diploid variety Scapa. The nucleolus is visible in the upper left corner in tetraploid Sante and the middle right in diploid Scapa. Both ASY1 and ZYP1 foci are conspicuous in the cell in the chromatin as well as on the nucleolus periphery. The ASY1 green signals look more frequent than the ZYP1. This may indicate that loading of ASY1 and ZYP1 occurs simultaneously in potato, though at separate sites to ensure proper SC formation between homologues. However, the ZYP1 only elongates or form linear tracts during later stages of prophase I at places where ASY1 is absent or not visible, as can be seen later in **Figure 7-8**. Other meiotic proteins such as AtMLH1 and SWI1 have been found to be associated with the nucleolus

before in *Arabidopsis* (Jackson *et al.*, 2006). The nucleolus has been considered to keep a check on cell cycle regulating proteins by sequestration (Visintin and Amon, 2000). A similar sequestration might be the cause of the occurrence of the ASY1 and ZYP1 signals on the nucleolus. In contrast, the bright spots around the nucleolus may also be an artifact, as the polyclonal antibodies used are specific for ASY1 and ZYP1 in *Arabidopsis,* but they may bind other epitopes in other organisms including potato.

It can be seen in **Figure 7-8**, that the green signal indicating ASY1 is present predominantly in the cell, indicating late zygotene/early pachytene stages of prophase I. There are continuous linear red signals as well along with the green, indicating presence of ZYP1. The presence of continuous red ZYP1 signals indicates the complete formation of synaptonemal complex protein between the homologues.

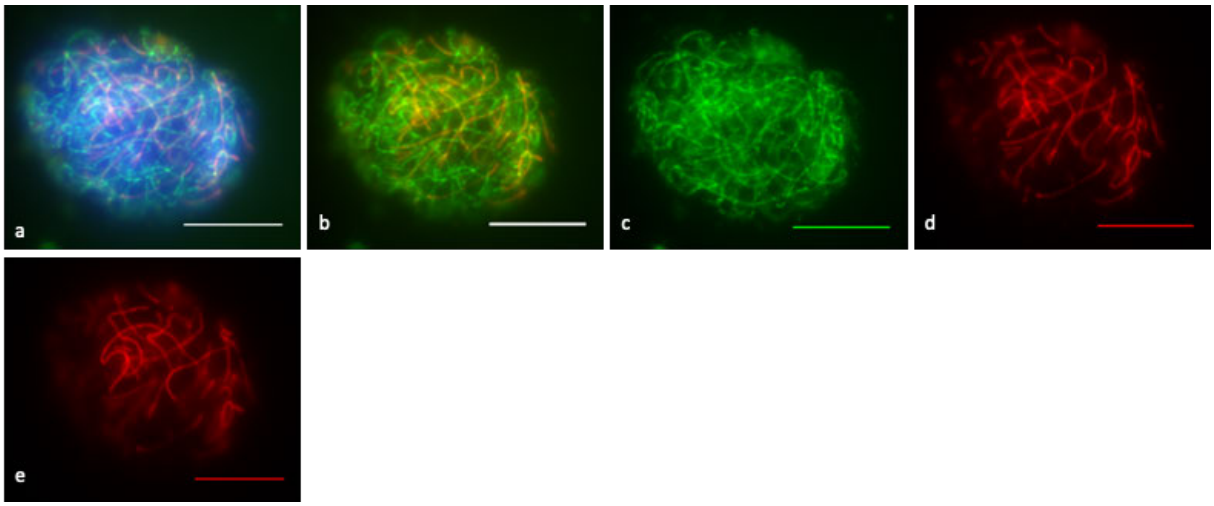


Figure 7-8 Immunolocalisation of ASY1 *(green) and* ZYP1 *(red) in Solanum tuberosum meiotic Prophase I in variety Sante.*

a. Merged DAPI, ASY1 and ZYP1 colour coded blue, green and red, b. Merged ASY1 and ZYP1, c. ASY1, d. and e. ZYP1 as seen in different planes. Scale bar is 10 μm. The antibodies used were raised against *Arabidopsis thaliana* proteins.

In *Arabidopsis*, the *ASY1* gene has been found to play an important role in chromosome synapsis (Sanchez-Moran *et al.,* 2007). The ASY1 signal has been found to be continuous in zygotene, and to disappear in the late diplotene when the homologues start to separate out. It was found to be associated with the lateral elements of the chromosome axis, indicating an important structural maintenance role in synapsis formation (Armstrong, 2002). Here, in potato we can observe continuous ASY1 signals along with ZYP1 at certain places, indicating a similar morphology as in *A. thaliana*. The ASY1 signals appear to be in patches of higher and lower intensity, while the ZYP1 signal is more continuous indicating reduction in intensity of ASY1 in synapsing regions. This pattern of ASY1 has been reported before in *A. thaliana* (Lambing *et al.*, 2015).

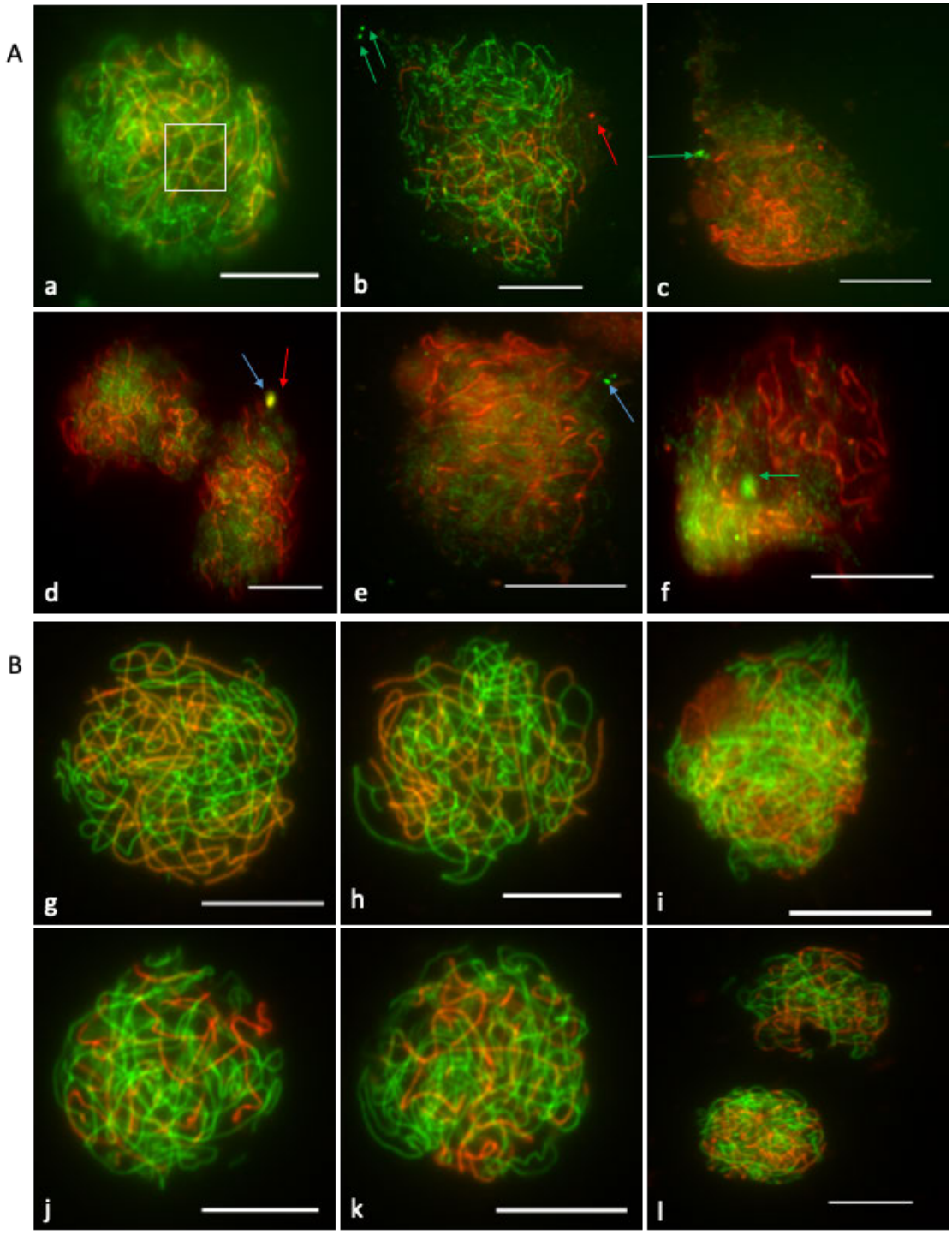


Figure 7-9 Immunolocalisation of ASY1 *(green) and* ZYP1 *(red) in Solanum tuberosum prophase I cells.*

Panel A shows prophase I cells in Sante and B in Scapa. The frame in first cell, panel a of A indicates the possible partner switch. Green and red arrows indicate ASY1 and ZYP1 big foci in other cells of Sante in panel A compared to normal localization of the same proteins in diploid variety Scapa in panel B. Scale bar is 10 μm.

In panel A of **Figure 7-9**, seven different cells of tetraploid Sante with various degrees of ASY1 and ZYP1 localisation at zygotene-pachytene stage can be seen. As the ZYP1 polymerises, an absence of the ASY1 is conspicuous, with ASY1 showing discontinuous signals with interspersed ZYP1. Aggregates of the ZYP1 and the ASY1 proteins are also observed, visible as bright dots indicated by the arrows (green arrows for ASY1 and red arrows for ZYP1 foci in panels b, c, d, e and f of A). Loop formation and chromosome entanglement is also visible, which can either be resolved before pachytene or may persist till metaphase, leading to mis-segregation of the chromosomes. There are sites such as those represented within the square in the panel 'a of A' in **Figure 7-9**, which might indicate pairing partner switching between the homologues. Panel B shows the zygotene-pachytene stage in a diploid variety Scapa. In contrast to the tetraploids, no aberrations, foci and entanglements are visible, indicating the proper loading of the meiotic proteins. The domain arrangement of ASY1 is also not as enhanced in diploids as seen in tetraploids, which might indicate delayed synapsis in tetraploids, as was also seen in *pch2* mutant in *Arabidopsis* (Lambing *et al.*, 2015).

Loop formation, entanglement and possible partner switching can again be seen indicated by arrows in **Figure 7-10**, in different tetraploid varieties. The arrows in panel A in variety Sante point possibly towards aligned chromosomes running together, indicating a possible synaptic partner switching (SPS) or an interlock. Synaptic partner switch occurs when one chromosome synapses with more than one partner at the same time. In panel B, variety Cara is represented. The arrows in the left and right indicates the possible point of initiation of the synaptonemal complex formation. Here it seems, that two parallel running homologues pair, come together, synapse and open up again. This might indicate that SC formation is not complete between

these two homologues and they are free to pair up with other homologues, leading to SPS multivalents. Similarly, the arrows in panel C in Maris Peer also points towards the possible partner switch occurring between the homologues. Two parallel synapsed strands seem to exchange the partners at the rightmost arrow point of their meeting. The middle and the left arrow indicates the exchanged partners again synapsing. This can be represented diagrammatically as shown in **Figure 7-11**.

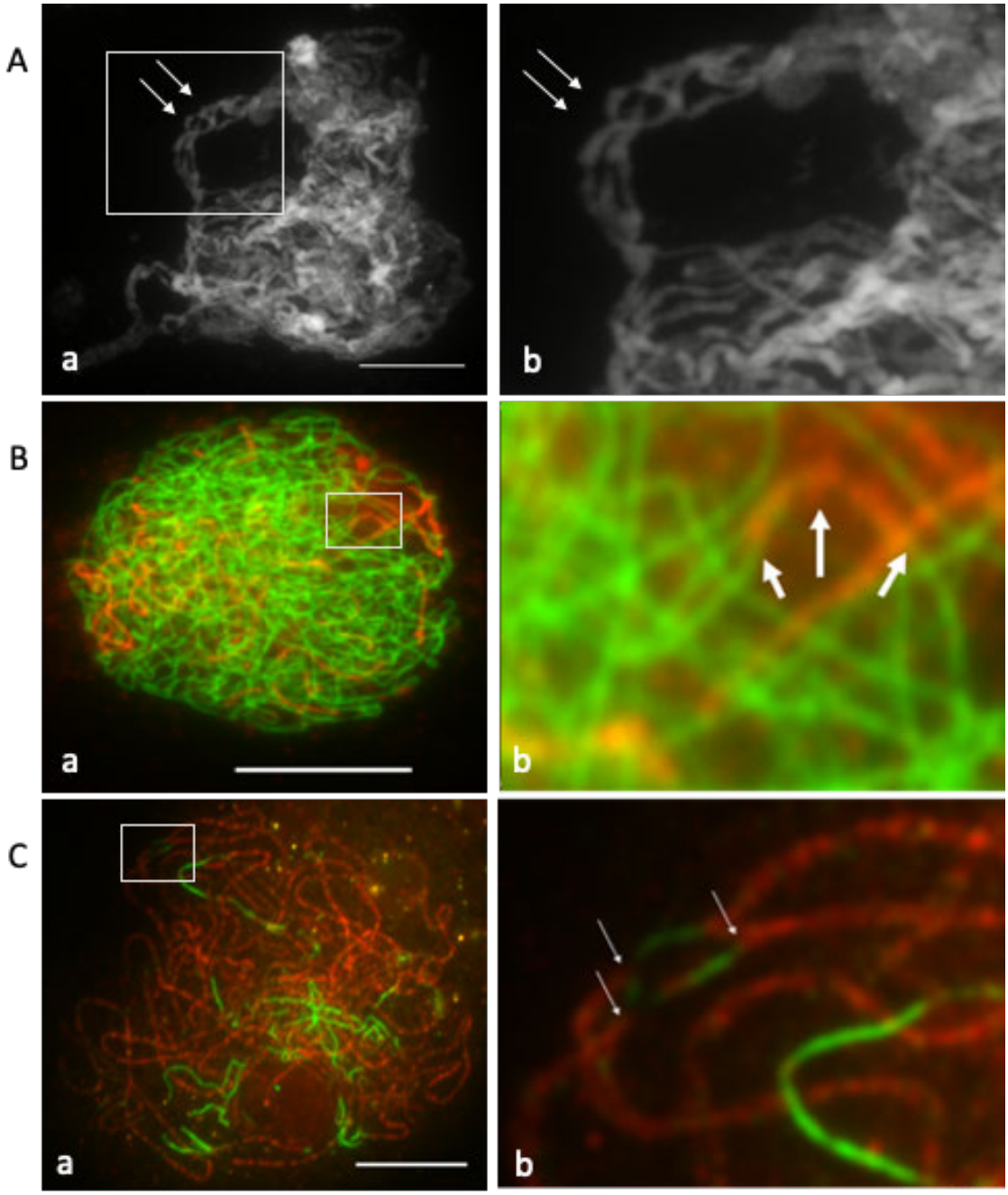


Figure 7-10 Zygotene and Pachytene stages shown in three different tetraploid varieties.

Panel A represents Sante, Panel B represents Cara and Panel C represents Maris Peer. Picture on the right in b in all three panels is the magnification of the part visible in the box in a. The white arrows point towards the possible synaptic partner switch in A and B. Panel A shows a zygotene stained with DAPI. In panels B and C dual immunolocalisation with axis protein ASY1 (green) and SC protein ZYP1 (red) can be seen. Scale bar is 10 μm.

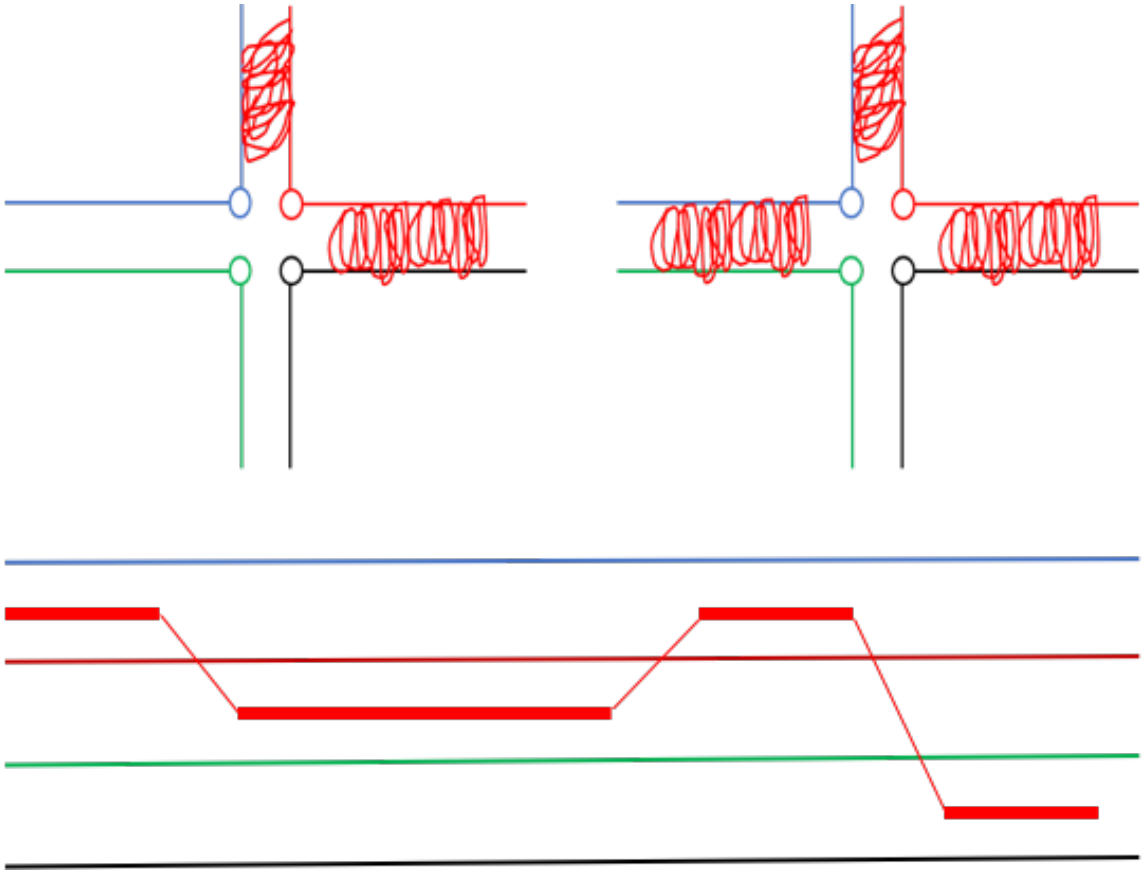


Figure 7-11 Diagrammatic representation of synapsis between different homologues leading to multivalent formation in a tetraploid.

Figure 7-11 (upper left), shows synapsis formation between three homologues, where it can lead to a trivalent formation along with the univalent for the unsynapsed chromosome or this partner switch may be resolved before entering M1. On the right, all four homologues are involved, where a blue chromosome can be seen synapsing with red as well as green chromosomes. Similarly, the red chromosome can be seen synapsing with blue as well as black chromosomes. This can lead to a quadrivalent formation, which can persist until M1 or resolve before that, leading to bivalent formation. The presence of quadrivalents in M1s, which we will see later in **Figure 7-12**, indicates the presence of synaptic pair switching where potentially all four homologues are bound. This needs to be analysed further using immunolocalisation of different proteins along with super resolution microscopy to have a clear understanding.

7.3.4 Chiasma Analysis using FISH probes in tetraploid and diploid varieties

FISH analysis using the 45S and 5S probes used in *A. thaliana* was carried out on the chromosomal spreads. Dong *et al.* (2000) have earlier shown that 5S rDNA repeats are present in the short arm of chromosome 1, close to the centromere, while 45S rDNA repeats are present in the NOR regions of chromosome 2 in the short arms again. Based on this information, the chiasma analysis of the two chromosomes in three different tetraploid varieties and one diploid variety of *S. tuberosum* is presented here. Although four tetraploid and two diploid varieties were grown, a sufficient number of M1 cells for chiasma analysis was found only in three tetraploids, Sante, Maris Peer and Cara, and one diploid variety, Scapa. As expected, two chromosome pairs, chromosomes 1 and 2, produced the signals on hybridization with 5S and 45S ribosomal DNA probes.

Hypothesis: The null hypothesis will be that the ploidy of the variety should not have an effect on the chiasma frequency of the plant. Alternatively, the number of crossovers should be more than doubled in tetraploids compared with diploids.

7.3.4.1 Chiasma analysis in 4n Sante

Two representative M1 cells from tetraploid variety Sante is shown in **Figure 7-12**. 5S signals (red) can be seen in the two bivalents near the centromeric region in chromosome 1. The two bivalents in panels a and b represent two rods indicating at-least 1 chiasma in the long arm for each bivalent. 45S signals (green) can also be seen in both the bivalents of chromosome 2, which have assumed the shape of rods indicating one chiasma, here in the longer arm as chromosome 2 is subtelocentric. In panels c and d on the other hand, a clear ring quadrivalent

for the chromosome 1 as indicated by red 5S rDNA signals and two rod bivalents of chromosome 2 as indicated by 4 green 45S signals can be seen.

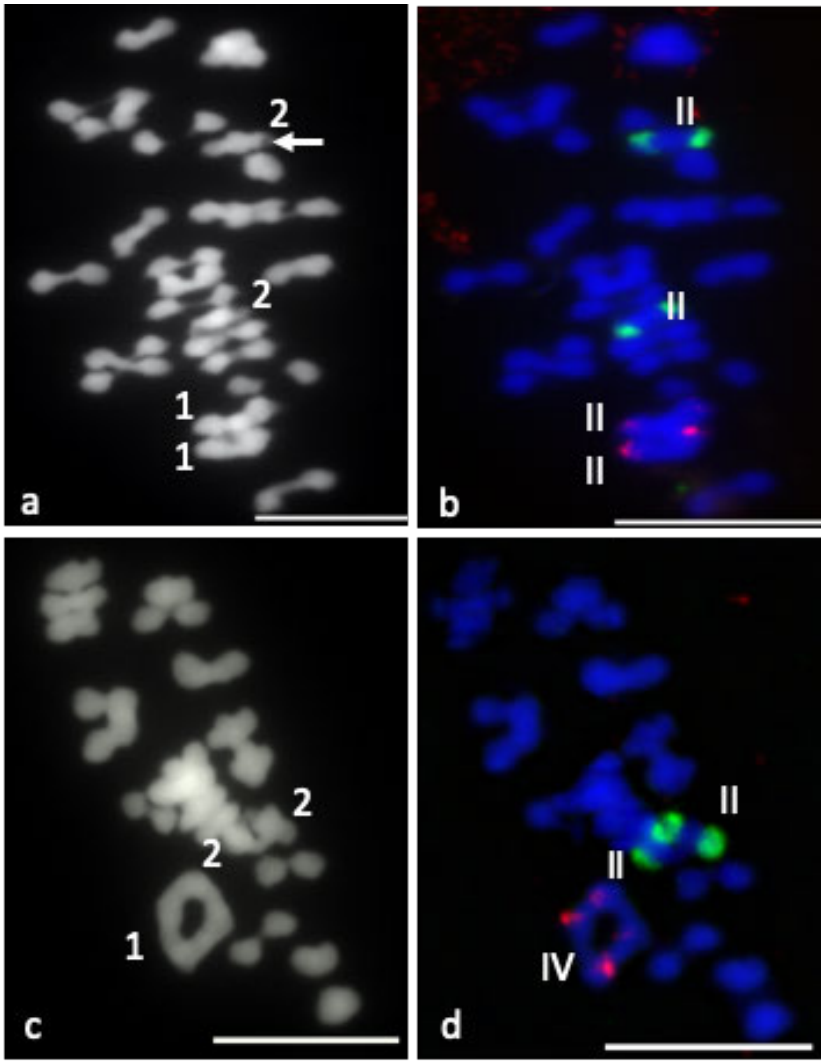


Figure 7-12 Comparison between cells in M1 in Solanum tuberosum, variety Sante, in chromosomes 1 and 2.

The figure shows two DAPI stained M1 cells (left) and the same cells (right) also showing 5S (red) and 45S (green) FISH probes. Numbers in the left indicate chromosome numbers and Roman numerals on the right indicate the chromosome configuration, II is bivalent, IV is quadrivalent. The arrow in panel a indicates the NOR region. Scale bar is 10 μm.

A chiasma count for chromosomes 1 and 2 based on meiotic M1 configurations was undertaken for 71 cells. The mean chiasma count was found out to be 2.59 for chromosome 1 and 2.13 for chromosome 2 (**Table 7-2**). It was found that the chiasma occurrence in short arms was far less than the long arms. For example, the chiasma count frequency in

chromosome 1 for short arms was 0.62, while it was 1.97 for the long arms. Chromosome 2 is a subtelocentric chromosome with short arm largely made up of a NOR region and a satellite. The chiasma count frequency was found to be lower for short arms in chromosome 2 than chromosome 1. A chi square test for association showed a significant association of the chiasma frequency in the short and long arms for both the chromosomes ($\chi^2_{df=1}$ = 15.00, p-value < 0.0001), indicating that the chiasma distribution is associated with the chromosome.

Multivalent formations, mostly in the form of ring quadrivalents, chain quadrivalents or trivalents also occurred for these chromosomes. The numbers and percentage of bivalents, quadrivalents, trivalents and univalents formed is shown in **Table 7-3**. The low level of multivalent formation for both the chromosomes (12 out of 71 in Chr 1 and 10 out of 71 in Chr 2) is significantly different from the expected 2:1 (66.66% multivalents) ratio according to the random end model ($\chi^2_{df=1}$ = 76.27, p-value < 0.0001 for chr 1 and $\chi^2_{df=1}$ = 85.28, p-value < 0.0001 for chr 2). The remaining chromosomes, which were not probed, predominantly occurred as bivalents, mostly rods with a few rings and few univalents as well (as visible in DAPI pictures). An overall chiasma analysis could not be performed for those chromosomes because of the overcrowding of the bivalents along the equatorial plate, which made it difficult to distinguish clearly between rods, rings or multivalents without a probe.

7.3.4.2 Chiasma analysis in 4n Maris Peer

Another tetraploid variety, Maris Peer was probed for chiasma frequency in chromosomes 1 and 2. **Figure 7-13** shows two representative M1 cells. Two rod bivalents each of chromosome 1 and 2 can be seen representing a chiasma in the long arm in panels a and b. In panels c and d, chromosome 1 can be seen as two bivalents; one ring showing 2 chiasmata, one in each arm, and one rod with a chiasma in the long arm. Chromosome 2 can be seen forming a chain quadrivalent where there are 2 chiasmata in the long arms and 1 chiasma in the short arms through which 2 bivalents seem to be attached. In the past, people have analysed chiasma frequency using cells in diplotene, diakinesis, early M1 and M1s. Post pachytene, chiasma can be visible in very clear diplotene preparations. However, it becomes clearer during late diakinesis or early M1s. Recently, diakinesis and pre M1s were used in potato to ascertain the percentage of quadrivalents and bivalents for few chromosomes using oligo probes (He *et al.*, 2018). The post prophase I stages, as discussed before in the results section **7.3.2**, are transient and pass away quickly. Therefore, it is more important to consider cells in late diakinesis and early M1s along with M1s for chiasma analysis.

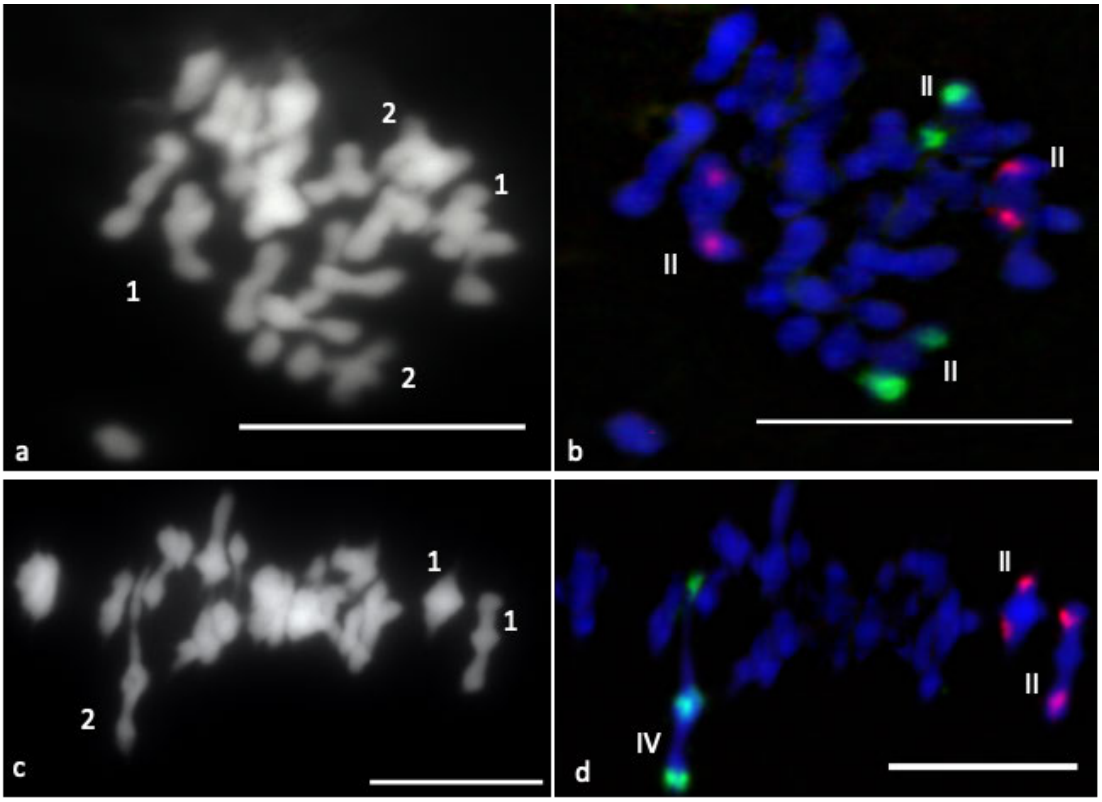


Figure 7-13 Comparison between cells in M1 in Solanum tuberosum, variety Maris Peer, in chromosomes 1 and 2.

The figure shows two DAPI stained M1 cells and the same cells also showing 5S (red) and 45S (green) FISH probes. Numbers in the left indicate chromosome numbers and Roman numerals on the right indicate the chromosome configuration, II is bivalent, IV is quadrivalent. Scale bar is 10 μm.

The chiasma analysis was carried out on 305 cells, including cells in late diakinesis, early M1s and M1. The mean chiasma count was found to be 2.87 for chromosome 1 and 2.32 for chromosome 2. As with the variety Sante, the mean chiasma count frequency in the short arms was far less than in the long arms for Maris Peer as well (**Table 7-2**). The distribution of the chiasma over short and long arms was found to be significantly associated with the chromosomes ($\chi^2_{df=1}$ = 16.52, p-value < 0.0001). Multivalents in the form of rings, chains, Y and trivalents, as seen in Sante, occurred in Maris Peer as well (**Table 7-3**). Chromosome 1 shows a higher percentage of quadrivalents than chromosome 2; however, the number of trivalents and univalents are higher in chromosome 2 in Sante and chromosome 1 in Maris

Peer. The multivalent formation for both the chromosomes (126 out of 305 in Chr 1 and 90 out of 305 in Chr 2) significantly deviated from the expected 2:1 (66.66% multivalents) ratio according to the random end model ($\chi^2_{df=1}$ = 82.85, p-value < 0.0001 for chr 1 and $\chi^2_{df=1}$ = 181.00, p-value < 0.0001 for chr 2).

To find out if there could be a difference in chiasma count frequency between the diakinesis and M1 cells, separate chiasma count of the 50 cells in diakinesis and 255 cells in M1 from the total of 305 cells was also carried out. Cells in diakinesis provided a lower chiasma frequency of 2.62 for chromosome 1 (0.74 and 1.88 for short and long arm) and 2.14 for chromosome 2 (0.3 and 1.84 for short and long arm) compared to the joint 305 cells. On the other hand, the cells in M1 gave a higher chiasma frequency of 2.91 for chromosome 1 (0.98 and 1.93 for short and long arm) and 2.4 for chromosome 2 (0.42 and 1.96 for short and long arm). However, these frequencies were not significantly different from each other (Mann Whitney test, p value > .05). Chiasma frequency was consistently lower in the short arm for both the chromosomes. Therefore, a combined analysis of the diakinesis and M1 cells can be undertaken. However, the multivalent frequency was found to be higher for both the chromosomes for cells in M1 than the cells in diakinesis. The diakinetic cell showed 28% (14 out of 50 cells) multivalent formation for chromosome 1 and 18% (9 out of 50 cells) for chromosome 2 while the M1 cells showed 44% (112 out of 255 cells) multivalents for chromosome 1 and 32% (81 out of 255 cells) for chromosome 2. There can be two reasons for this. First, the cell count in diakinesis is much lower than the cell count in M1s. Second, the bivalents in the M1 lay very close to each other owing to the large number of chromosomes, which can give an impression of multivalents. However, the multivalent formation is higher in

Maris Peer (even separately in diakinesis and M1) when compared with Sante. This indicate the genetic variation that exists among the varieties in potato.

7.3.4.3 Chiasma analysis in 4n Cara

27 cells in another tetraploid variety Cara, were analysed for chromosomes 1 and 2. A representative M1 cell can be seen in **Figure 7-14**. Chromosome 1 can be seen in a chain configuration having 2 chiasmata in the long arm and one in the short arm, while chromosome 2 has two rod bivalents, each having one chiasma in the long arm.

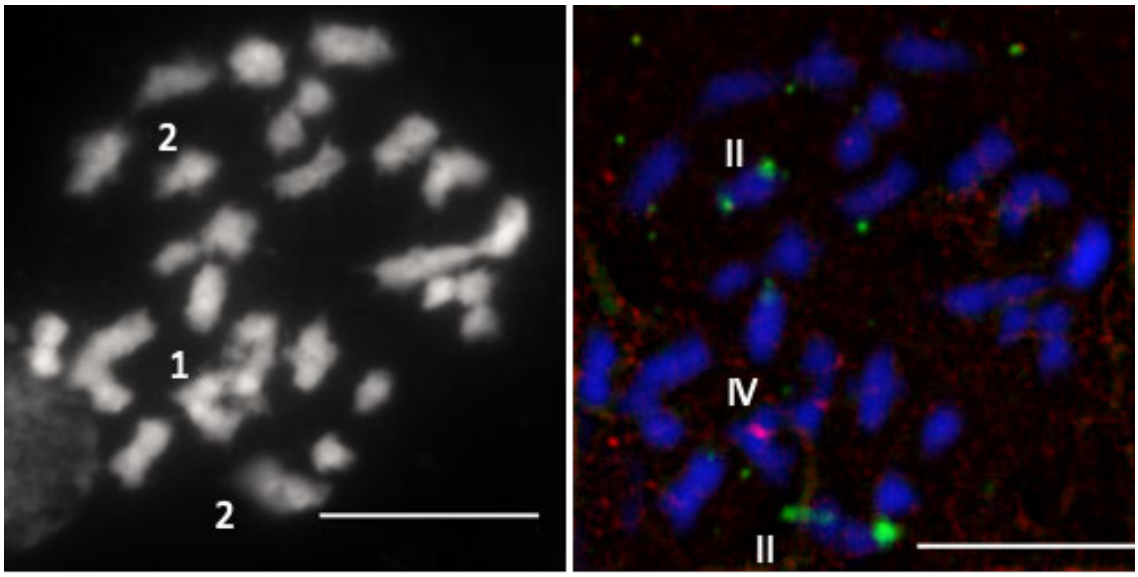


Figure 7-14 Comparison between a cell in M1 in Solanum tuberosum, variety Cara, in chromosomes 1 and 2.

The figure shows a DAPI stained M1 cell and the same cell also showing 5S (red) and 45S (green) FISH probes. Numbers in the left indicate chromosome numbers and Roman numerals on the right indicate the chromosome configuration, II is bivalent, IV is quadrivalent. Scale bar is 10 μm.

Chiasma analysis was carried out in 27 diakinetic or M1 cells and the mean chiasma count was found to be 3.07 (1.11 in short arm and 1.96 in long arm) for chromosome 1 and 2.48 (0.48 in short arm and 2 in long arm) for chromosome 2. Similar to the other two tetraploid varieties, the mean chiasma count frequency in the short arms was less than the long arms (**Table 7-2**). The distribution of the chiasma over short and long arms was found to be significantly

associated with the chromosomes ($\chi^2_{df=1}$ = 18.94, p-value < .0001). Multivalents, mostly in the form of ring and chain quadrivalents were found (**Table 7-3**). Multivalent formation was lower than the bivalent formation (13 out of 27 in Chr 1 and 7 out of 27 for Chr 2). Deviation from the random end model predicting at-least 66.66% multivalents is borderline significant for Chr 1 ($\chi^2_{df=1}$ = 3.83, p-value = 0.05) and significantly different for Chr 2 ($\chi^2_{df=1}$ = 19.32, p-value = 0.000).

Variety (Number of cells)	Chromosome 1 (4n/2n)			Chromosome 2 (4n/2n)		
	Short arm	Long arm	Total	Short arm	Long arm	Total
Cara (4n) (n=27)	1.11 (2.2)	1.96 (2.0)	3.07 (2.1)	0.48 (2.7)	2 (2.1)	2.48 (2.2)
n_{b1}=14, n_{b2}=20	0.71 (1.4)	1.93 (1.9)	2.63 (1.8)	0.2 (1.1)	2.0 (2.1)	2.2 (1.9)
n_{m1}=13, n_{m2}=7	1.54 (3.1)	2 (2.1)	3.54 (2.4)	1.29 (7.2)	2.0 (2.1)	3.29 (2.9)
Sante (4n) (n=71)	0.62 (1.2)	1.97 (2.0)	2.59 (1.8)	0.21 (1.2)	1.92 (2.0)	2.13 (1.9)
n_{b1}=59, n_{b2}=60	0.46 (0.9)	1.98 (2.0)	2.44 (1.7)	0.05 (0.3)	2 (2.1)	2.05 (1.9)
n_{m1}=12, n_{m2}=10	1.42 (2.8)	1.91 (1.9)	3.33 (2.3)	1.2 (2.2)	1.5 (1.6)	2.7 (2.4)
M. Peer (4n) (n=305)	0.94 (1.9)	1.92 (2.0)	2.87 (2.0)	0.40 (2.2)	1.93 (2.0)	2.32 (2.0)
n_{b1}=178, n_{b2}=208	0.64 (1.3)	1.94 (2.0)	2.58 (1.8)	0.09 (0.5)	1.94 (2.0)	2.02 (1.8)
n_{m1}=126, n_{m2}=90	1.35 (2.7)	1.93 (1.9)	3.28 (2.2)	1.13 (6.3)	1.98 (2.0)	3.11 (2.7)
Scapa (2n) (n=236)	0.50	0.97	1.47	0.18	0.97	1.15

Table 7-2 Mean chiasma frequency for chromosomes 1 and 2 in different varieties of Solanum tuberosum.

The first row shows the chiasma frequency for all the cells. The second row shows chiasma frequency for cells with bivalents only, in tetraploids. The third row shows chiasma frequency for cells with multivalent only, in tetraploids. n_{b1} and n_{b2} represent the number of cells with bivalents only, and n_{m1} and n_{m2} represent the number of cells with multivalents only for chromosomes 1 and 2 respectively in tetraploids.

Var	Chr	2II	1IV	1III + 1I	1 II + 2I	IV Ring	IV Chain	IV Other	n
		n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	
Cara	**1**	14 (51.9)	13 (48.1)	0 (0)	0 (0)	7 (53.8)	6 (46.2)	0	27
Cara	**2**	20 (74.1)	7 (25.9)	0 (0)	0 (0)	1 (14.3)	5 (71.4)	1 (14.3)	27
Sante	**1**	59 (83.1)	11 (15.5)	1 (1.4)	0 (0)	5 (45.5)	6 (54.5)	0	71
Sante	**2**	60 (84.5)	5 (7.0)	5 (7.0)	1 (1.4)	2 (40)	3 (60)	0	71
MP	**1**	178 (58.4)	116 (38.0)	10 (3.3)	1 (0.33)	42 (36.2)	70 (60.3)	4 (3.4)	305
MP	**2**	208 (68.2)	87 (28.5)	3 (1.0)	7 (2.3)	12 (13.8)	73 (83.9)	2 (2.3)	305

Table 7-3 Number of cells showing quadrivalents, trivalents and univalents for chromosomes 1 and 2.

II indicate bivalents, IV indicate quadrivalents, III indicate trivalents and I indicate univalents. The percentages are shown in the parenthesis. MP is tetraploid variety Maris Peer.

7.3.4.4 Chiasma analysis in 2n Scapa

In **Figure 7-15**, a cell in M1 stage of the diploid variety Scapa, when stained with DAPI as well as after carrying FISH with 5S and 45 S probes in panels a and b can be seen. 12 clear bivalents are visible. Chromosome 1 can be identified in panels a and b as a rod bivalent with 2 chiasmata in the long arm, as a clear knob is visible. Chromosome 2 can also be identified as a rod with single chiasma in the long arm. Panels c and d represent cells in late diakinesis/ early M1. Again, chromosomes 1 and 2 both can be identified showing rod conformations, each with a single chiasma in their long arms. 236 cells were used to count for the number of crossovers in Scapa. The mean chiasma count was found to be 1.47 for chromosome 1 and 1.15 for chromosome 2 (**Table 7-2**). The chiasma frequency was smaller in the short arm as compared to the long arm for both the chromosomes. This was similar to tetraploids.

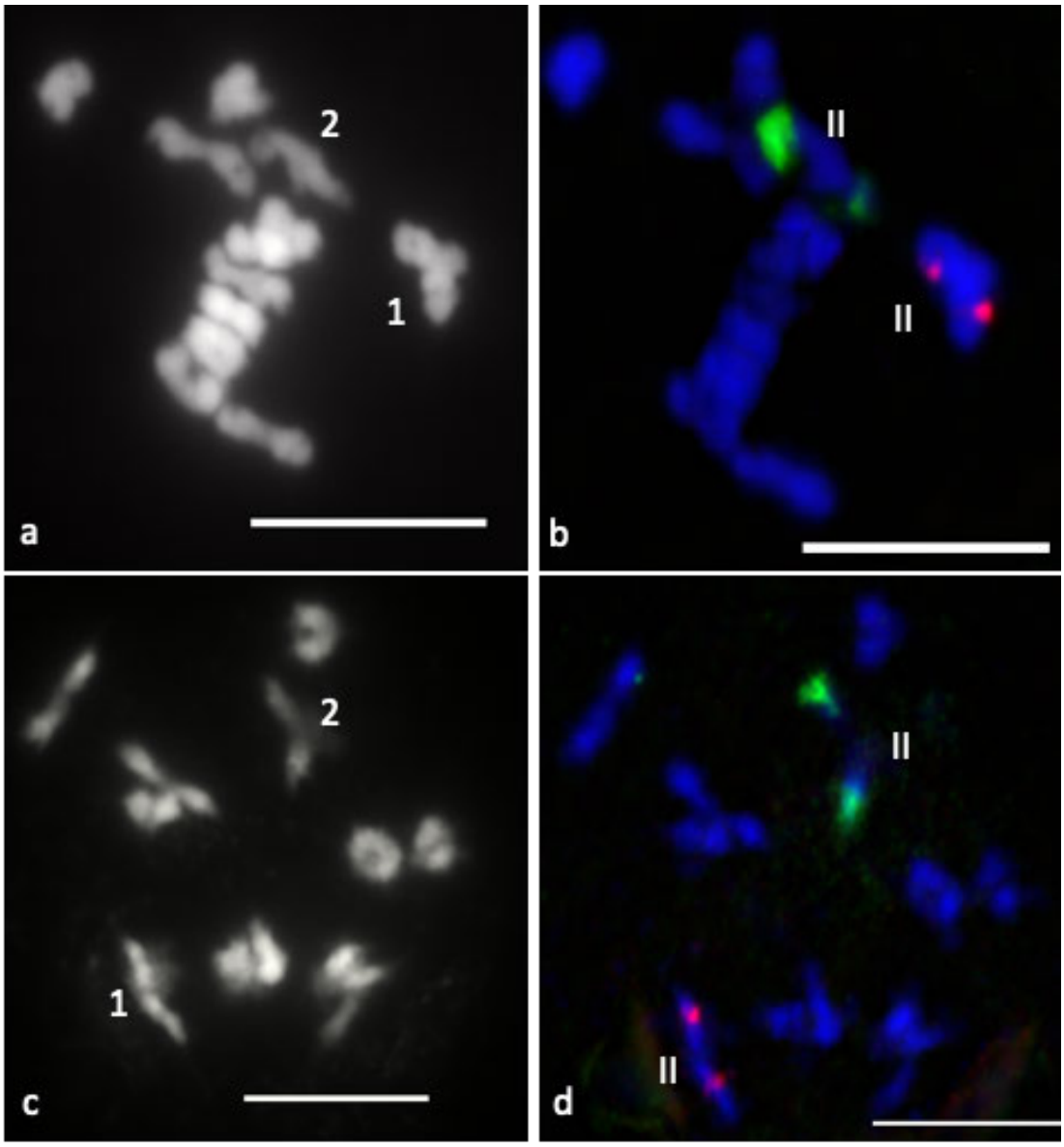


Figure 7-15 Comparison between meiotic cells in Solanum tuberosum, variety Scapa, in chromosomes 1 and 2.

The figure shows DAPI stained meiotic cells and same cells also showing 5S (red) and 45S (green) FISH probes. Panels a and b indicate M1 cells and panels c and d indicate late diakinesis/pre M1 cells. Numbers on the left are chromosome numbers while Roman numerals on the right indicate configurations. II indicate bivalents. Scale bar is 10 μm.

7.3.5 Varietal Variation in Configurations and Chiasma frequency

It was hypothesised that the chiasma frequency in tetraploids would be more than doubled than that in diploids, because the number of chromosomes in a tetraploid is double than that of the diploid and multivalent formation occurs. The chiasma counts for the two chromosomes between different varieties was compared using two different tests.

Frequency for different chiasma counts were calculated and varieties were compared with Kruskal-Wallis and Fisher exact test. Kruskal-Wallis test showed that four different varieties are significantly different from each other ($\chi^2_{df=3}$ = 443.67, p-value < 2.2e-16, chr 1: $\chi^2_{df=3}$ = 336.47, p-value < 2.2e-16, chr 2: $\chi^2_{df=3}$ = 403.96, p-value < 2.2e-16). Post hoc Dunn test with Bonferroni corrections was performed to ascertain which variety pair and for which chromosomes was different. It showed that diploid Scapa was highly significantly different from all three tetraploids (**Table 7-4**). A significant difference between the varieties was also found in Fisher test which gave a p value of 0.0004998.

Variety (total) **Chromosome1** **Chromosome 2**	**Cara**	**Maris Peer**	**Sante**
Maris Peer (total)	0.5957	-	-
Chr 1	1.00	-	-
Chr 2	1.00	-	-
Sante (total)	0.12	0.10	-
Chr 1	0.16	0.19	-
Chr 2	0.48	0.56	-
Scapa (total)	<2e-16	<2e-16	<2e-16
Chr1	<2e-16	<2e-16	<2e-16
Chr2	<2e-16	<2e-16	<2e-16

Table 7-4 Post hoc Dunn test p-values for total and individual chromosomes after Kruskal-Wallis.

The chiasma frequency in Scapa is not exactly half for chromosome 1 when compared to Sante and Maris Peer (**Table 7-2**). It is exactly half of Maris Peer, though not Sante for chromosome 2 and half of Cara for both chromosomes. Thus, it can be seen that there is no substantial increase in chiasma frequency in the tetraploids as compared with diploids. However, when treated separately for short and long arms, doubling in the chiasma frequency in long arms, but not in short arms was seen for Sante and Maris Peer as compared with the diploid Scapa. It was only doubled for short arm of chromosome 2 in Maris Peer. The doubling of Cara over Scapa was seen for both the short and the long arms in both chromosomes with most increase seen in the short arm in chromosome 2.

To compare the per bivalent chiasma formation between different varieties, cells with two bivalents and cells with multivalents, for both chromosomes 1 and 2 were separated for the tetraploid varieties and analysed (**Table 7-2)**. The per bivalent chiasma frequency for both the chromosomes, as calculated by dividing the chiasma frequency obtained for cells with two bivalents by 2 in tetraploids, is smaller than the diploid variety, Scapa.

Kruskal-Wallis test showed a significant difference in per-bivalent chiasma frequency between different varieties for both chromosomes (chr 1: $\chi^2_{df=3}$ =17.591, p-value = 0.0005341, chr 2: $\chi^2_{df=3}$ =23.31, p-value = 3.479e-05). Post hoc Dunn test with Bonferroni corrections showed that the chiasma frequency was significantly lower in Sante for chromosome 1 and significantly lower for chromosomes 1 and 2 in Maris Peer than Scapa (**Table 7-5**). When considering only the cells with multivalents in tetraploids, a 2.9 times and 2.7 times increase in CO in chromosome 2 in Cara and Maris Peer can be seen compared with the diploid Scapa.

When treated separately for short and long arms, the boost was higher in short arms of chromosome 2 (7.2 times in Cara and 6.3 times in Maris Peer).

Variety (Chr)	Cara	Maris Peer	Sante
Maris Peer (Chr1)	1.0000	-	-
Maris Peer (Chr 2)	0.130	-	-
		-	-
Sante (Chr 1	1.0000	1.0000	-
Sante (Chr 2)	0.403	1.0000	-
Scapa (Chr 1)	1.0000	**0.0053**	**0.0039**
Scapa (Chr2)	1.000	**5e-05**	0.079

Table 7-5 Post hoc Dunn test p-values for per bivalent chiasma frequency difference in chromosomes after Kruskal-Wallis.

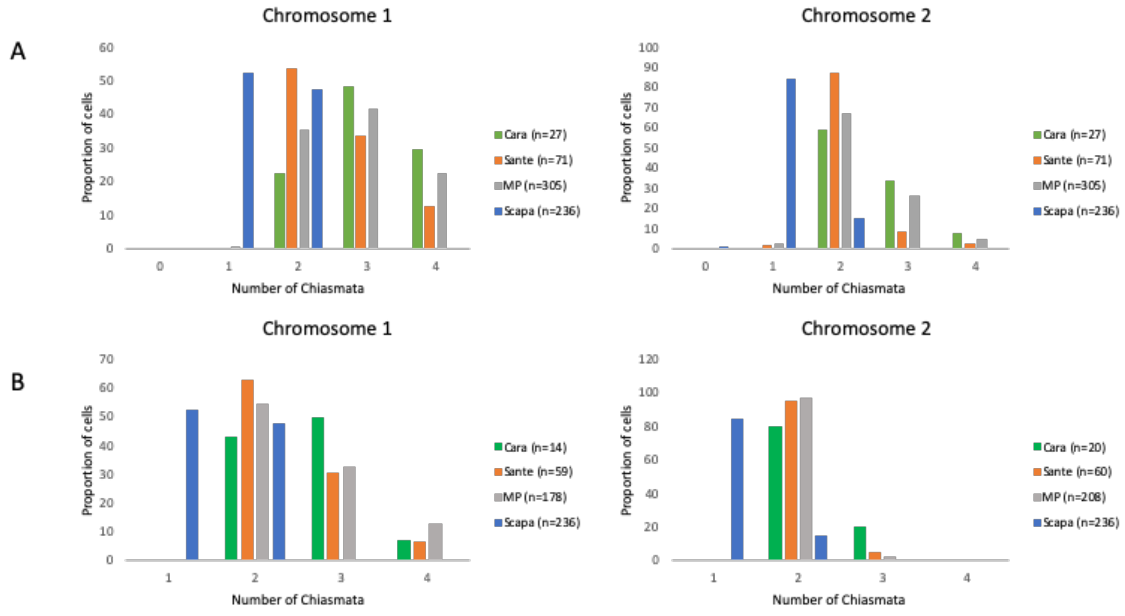


Figure 7-16 Different varieties showing the proportion of meiotic cells with different number of chiasmata in chromosomes 1 and 2.

Panel A represents chiasmata number in proportion of meiotic cells for all the cells. Panel B represents proportion of meiotic cells showing number of chiasmata in bivalents only. MP is Maris Peer.

The proportion of meiotic cells showing number of chiasmata across the two chromosomes 1 and 2, in different varieties of *S. tuberosum* can be seen in **Figure 7-16**. A high proportion of cells showing a lower number of chiasmata can be seen for diploid variety, Scapa, compared to tetraploid varieties Sante, Maris Peer and Cara, where a higher proportion of cells with a larger number of chiasmata can be seen for both chromosomes in panel A. Variety Sante, shows a higher proportion of cells with fewer crossovers among the tetraploid varieties, while variety Cara shows higher proportion of cells with 3 and 4 chiasmata than other tetraploid varieties.

The proportion of bivalent only meiotic cells in tetraploids, showing number of chiasmata in chromosomes 1 and 2 in comparison with the diploid variety can be seen in panel B of **Figure 7-16**. Proportion of cells with two chiasmata is higher than the proportion of cells with one, three and four chiasmata in tetraploids for both the chromosomes. Fewer cells with three chiasmata are present in chromosome 2 as compared to chromosome 1 and there is no cell with four chiasmata in chromosome 2 in tetraploids. Sante shows the highest proportion of cells with two chiasmata in chromosome 1 and comparable cells with Maris Peer in chromosome 2, while Cara shows the highest proportion of cells with 3 chiasmata in chromosome 1 as well as chromosome 2. However, diploid Scapa shows more cells with one chiasma for both chromosomes with a massive reduction in cells showing two chiasmata in chromosome 2. No cells with three or four chiasmata can be seen for diploids.

To compare and contrast the frequency of multivalent formation between the three tetraploid varieties, a 2 sample proportion test was carried out for each chromosome. The multivalent

formation was found to be significantly different from each other for Sante and Maris Peer in chromosome 1 at 95% confidence interval (Z = -3.84, p-value = 0.0001). Similarly, multivalent formation differed significantly between the two tetraploid varieties for chromosome 2 (Z = -2.65, p-value = 0.0081). The observed quadrivalent frequency in Sante was less than half the frequency observed in Maris Peer. It was also significantly lower than Cara for chromosome 1 (Z= 3.17, p-value = 0.002), though not for chromosome 2. Again, the observed quadrivalent frequency in Sante was less than half the frequency observed in Cara (**Table 7-3**). The multivalent formation was not found to be significantly different between Cara and Maris Peer for both chromosomes.

7.3.6 General observations in meiotic cells

A few more interesting observations were made in the cells in different varieties of *Solanum tuberosum* and where suitable, parallels have been drawn from the literature.

Possible alignment of four homologues

Linear and parallel chromosomes running together have been seen in autotetraploid *Arabidopsis arenosa* (Higgins *et al.*, 2014), where it was discussed if they could be the homologues running together. This can indicate an alignment of all the four homologues, which can then persist as multivalent if synapsis occurs between all of them, or can resolve before moving into further stages. A similar kind of alignment can be seen in **Figure 7-17**, which shows a zygotene stage in tetraploid variety Sante stained with DAPI and probed using anti-ZYP1 immunolocalisation with microwave method. Linear tracts of ZYP1 are visible, indicating synaptonemal complex formation between the homologues. The white arrow

indicates two parallel strands of aligned chromosomes running linearly together, which can be seen in the ZYP1 localised stage as well. Such parallel running homologues can also lead to synaptic pair switching (**Figure 7-10**), due to alignment of all four homologues.

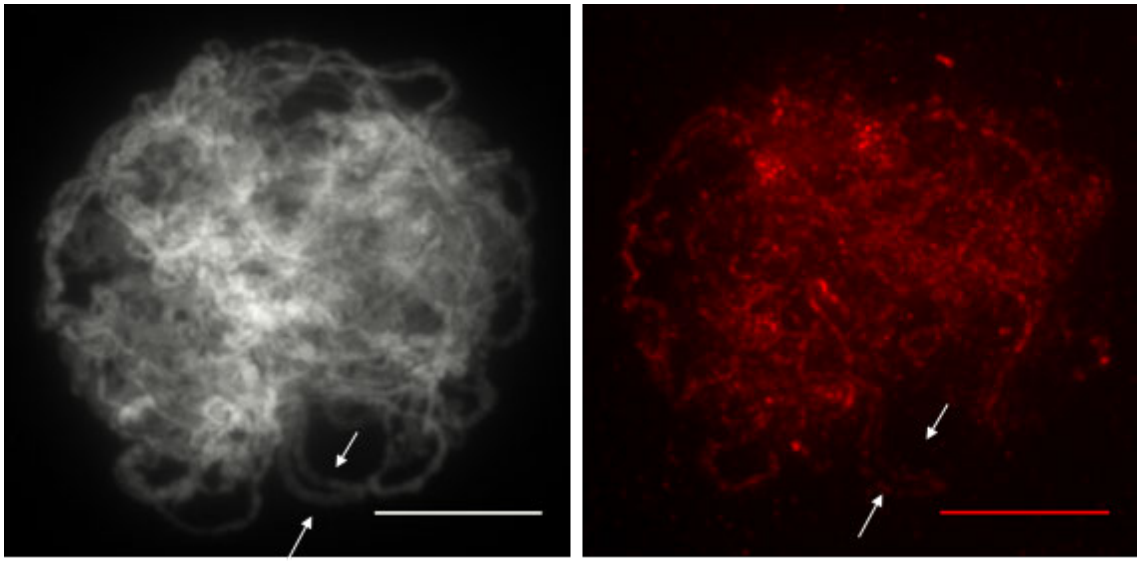


Figure 7-17 Zygotene stage with ZYP1 *immunolocalisation.*

The white arrows indicate parallel strands of homologous chromosomes running together in Sante. Scale bar is 10 μm.

DNA mismatch repair protein MLH1 localisation

The immunolocalisation of the mismatch repair protein, MLH1, has been used to identify the number of class I crossovers (CO) in tomato (Lhuissier *et al.*, 2007). Before, it has also been used in *Arabidopsis* to mark a subset of crossovers where it co-localises with another protein MLH3 in the pachytene stage of prophase I (Jackson *et al.*, 2006). To find out if the same could be utilized in potatoes, a search for MLH1 was made in the keyword search of spud-db database. It was identified as PGSC0003DMP400026441 which showed 89.7% similarity with *A. thaliana* MLH1 protein. The *Arabidopsis thaliana* MLH1 protein sequence was blasted in NCBI blastp against *Solanum tuberosum,* which is shown in **Figure 7-18.** It shows that the

identified MLH1 sequence in potato is 72% identical to that in *A. thaliana*. The antibodies raised against MLH1 in *Arabidopsis* should therefore be able to work well in potato.

PREDICTED: DNA mismatch repair protein MLH1 [Solanum tuberosum]
Sequence ID: XP_006338878.1 Length: 738 Number of Matches: 1

Range 1: 3 to 738 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
1101 bits(2848)	0.0	Compositional matrix adjust.	534/739(72%)	625/739(84%)	6/739(0%)

```
Query  2    IDDSSLTAEMEEEESPATTIVPREPPKIQRLEESVVNRIAAGEVIQRPVSAVKELVENSL  61
            I+D+ +  EME E  PA   +P+EPPKI RLEE VVNRIAAGEVIQRPVSAVKEL+ENSL
Sbjct  3    IEDTIIEVEMENEAIPAP--IPKEPPKILRLEECVVNRIAAGEVIQRPVSAVKELIENSL  60

Query  62   DADSSSISVVVKDGGLKLIQVSDDGHGIRREDLPILCERHTTSKLTKFEDLFSLSSMGFR  121
            DADS+SISVVVKDGGLKLIQVSDDGHGI  EDLPILCERHTTSKL+KFEDL S+ SMGFR
Sbjct  61   DADSTSISVVVKDGGLKLIQVSDDGHGICYEDLPILCERHTTSKLSKFEDLQSIRSMGFR  120

Query  122  GEALASMTYVAHVTVTTITKGQIHGYRVSYRDGVMEHEPKACAAVKGTQIMVENLFYNMI  181
            GEALASMTYV HVTVTTIT GQ+HGYR +YRDG+M  EPKACAAVKGTQIM+ENLFYNM
Sbjct  121  GEALASMTYVGHVTVTTITMGQLHGYRATYRDGLMVDEPKACAAVKGTQIMIENLFYNMA  180

Query  182  ARRKTLQNSADDYGKIVDLLSRMAIHYNNVSFSCRKHGAVKADVHSVVSPSRLDSIRSVY  241
            ARRKTLQNSADDY KIVDL+SR AIH+ +VSFSCRKHGA +ADVH++ + SRLD+IRSVY
Sbjct  181  ARRKTLQNSADDYPKIVDLISRFAIHHTHVSFSCRKHGAGRADVHTIATSSRLDAIRSVY  240

Query  242  GVSVAKNLMKVEVSSCDSSGCTFDMEGFISNSNYVAKKTILVLFINDRLVECSALKRAIE  301
            GVSVA+NLM +EVS        F M+GFISNSNY+AKK  +VLFINDRLV+C ALKRAIE
Sbjct  241  GVSVARNLMNIEVSDTGPLNSVFKMDGFISNSNYIAKKITMVLFINDRLVDCGALKRAIE  300

Query  302  IVYAATLPKASKPFVYMSINLPREHVDINIHPTKKEVSLLNQEIIIEMIQSEVEVKLRNA  361
            IVY ATLPKASKPF+YMSI LP EHVD+NIHPTK+EVSLLNQE +IE IQS V  KLR++
Sbjct  301  IVYTATLPKASKPFIYMSIILPPEHVDVNIHPTKREVSLLNQEFVIEKIQSVVGSKLRSS  360

Query  362  NDTRTFQEQKVEYIQSTLTSQKSDSPVSQKPSGQKTQKVPVNKMVRTDSSDPAGRLHAFL  421
            N++RTFQEQ +++  S+ T+   DS     PSG K+QKVP +KMVRTD+ DP+GRLHA++
Sbjct  361  NESRTFQEQTMDFSSSSPTATSKDSIKEPSPSGIKSQKVP-HKMVRTDTLDPSGRLHAYM  419

Query  422  QPKPQSLPDKVSSLSVVRSSVRQRRNPKETADLSSVQELIAGVDSCCHPGMLETVRNCTY  481
            Q KP    ++ S LS VRSS+RQRRNP ETADL+S+QEL+  +D+ CHPG+L+ VRNCTY
Sbjct  420  QMKPPGNSERGSCLSSVRSSIRQRRNPSETADLTSIQELVNEIDNDCHPGLLDIVRNCTY  479

Query  482  VGMADDVFALVQYNTHLYLANVVNLSKELMYQQTLRRFAHFNAIQLSDPAPLSELILLAL  541
            +GMAD++FAL+Q+NTHLYL NV+NLSKELMYQQ LRRFAHFNAIQLS+PA L EL++LAL
Sbjct  480  IGMADEIFALLQHNTHLYLVNVINLSKELMYQQVLRRFAHFNAIQLSEPASLPELVMLAL  539

Query  542  KEEDLDPGNDTKDDLKERIAEMNTELLKEKAEMLEEYFSVHIDSSANLSRLPVILDQYTP  601
            KEE  DP  +   +L+ +IAEMNTELLK+KA MLEEYFS+HIDS+ N+SRLPVILDQYTP
Sbjct  540  KEEGTDPEGNESKELRGKIAEMNTELLKQKAGMLEEYFSIHIDSNGNMSRLPVILDQYTP  599

Query  602  DMDRVPEFLLCLGNDVEWEDEKSCFQGVSAAIGNFYAMHPPLLPNPSGDGIQFYSKRGES  661
            DMDR+PEF+LCLGNDV+WEDEK CFQ ++AA+GNFYAMHPPLL NPSGDG++FY KR  S
Sbjct  600  DMDRIPEFILCLGNDVDWEDEKICFQTIAAALGNFYAMHPPLLRNPSGDGLKFYRKRVLS  659

Query  662  SQEKSDLEGNVD---MEDNLDQDLLSDAENAWAQREWSIQHVLFPSMRLFLKPPASMASN  718
            S  +     N+    ME   +++LL +AENAWAQREWSIQHVLFPS+RLF K P SMA+N
Sbjct  660  SGSEVTSTENIQNDTMEAEFEEELLLEAENAWAQREWSIQHVLFPSLRLFFKTPTSMATN  719

Query  719  GTFVKVASLEKLYKIFERC  737
            GTFV+VASLEKLY+IFERC
Sbjct  720  GTFVQVASLEKLYRIFERC  738
```

Figure 7-18 Blast output showing similarities and differences in the DNA mismatch repair protein MLH1 *between Arabidopsis thaliana and Solanum tuberosum.*

Query is *A. thaliana* sequence and subject is potato sequence.

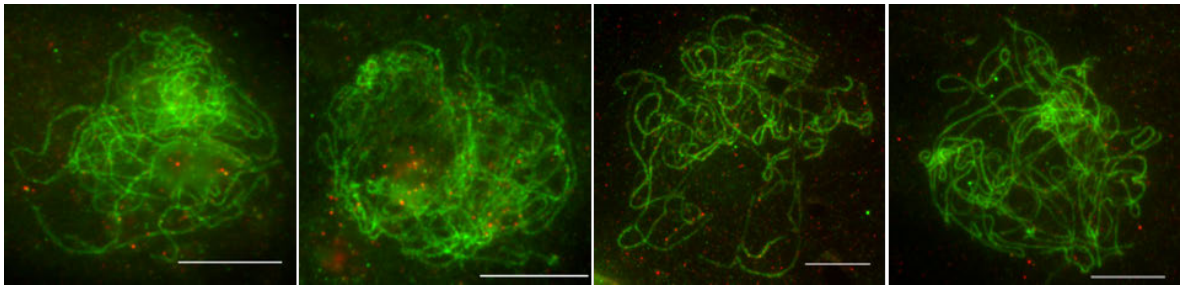


Figure 7-19 Immunolocalisation of ZYP1 *(green) and* MLH1 *(red) in 4n Maris Peer.*

Scale bar is 10 μm.

Based on the information from bioinformatics, dual immunolocalisation of ZYP1 and MLH1 was undertaken on the pre-prepared zygotene and pachytene slides of Maris Peer. ZYP1 should appear as the continuous signal in pachytene and MLH1 should appear as foci signals, indicating the crossover sites, on the continuous ZYP1. Continuous ZYP1 green signal is visible in **Figure 7-19**. However, the red MLH1 foci seems to be present not just on the chromatin, but outside as well, indicating the presence of background. This made it difficult to distinguish between the real signal and the background noise, therefore it was not utilized to count the total number of class I CO in the plant. Due to the lack of time and for saving material for FISH analysis, it was not possible to carry out more immunolocalisations using MLH1. In future, this technique can be utilized on the fresh material as it can give an overall estimate of the number of class I COs.

Heteromorphism in 5S signal in Sante

The 5S signal in chromosome 1 has been found to be close to the centromeric region in the 'p' or the upper arm of the submetacentric chromosome 1 (Tang *et al.*, 2009). In the chiasma analysis on meiotic M1s in variety Sante, the 5S signal appeared to be heteromorphic. The heteromorphism was seen in all the cells, with the 5S signal as depicted in **Figure 7-20**, where

the strength of the 5S signal observed was different in different cells in the homologous bivalents, indicating the difference in the number of 5S repeats.

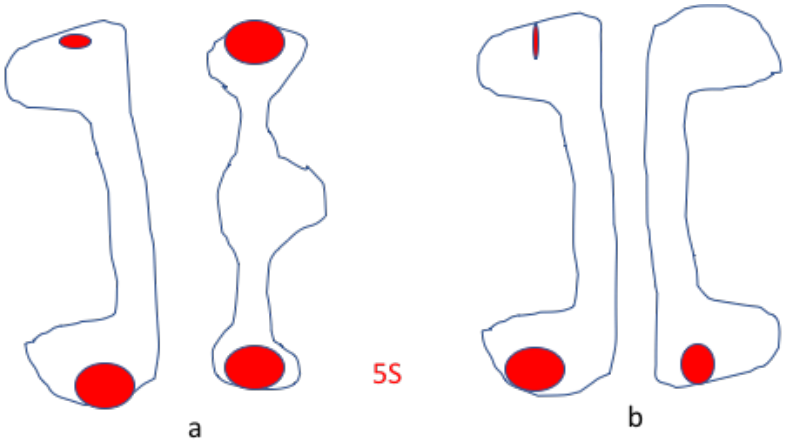


Figure 7-20 Diagrammatic representation of the presence of heteromorphism in 5S rDNA in chromosome 1 in Sante.

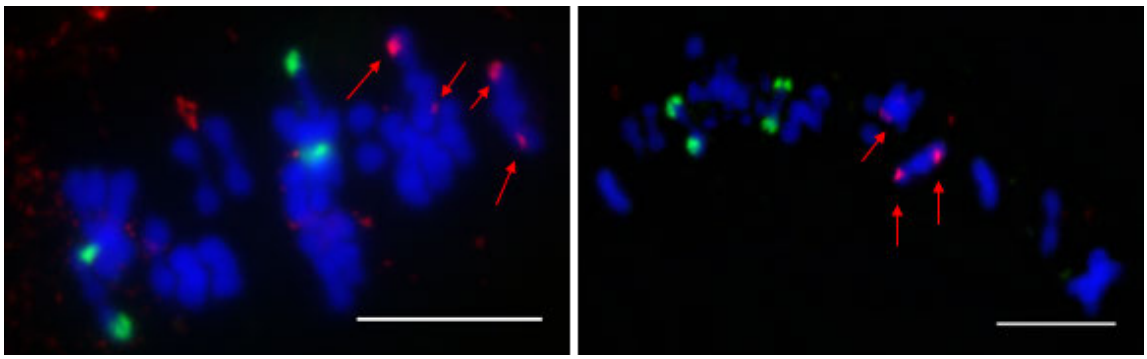


Figure 7-21 Two M1 cells showing different types of heteromorphic 5S rDNA in Sante.

The red arrows indicates the 5S rDNA sites in chromosome 1. Scale bar is 10 μm.

In **Figure 7-21**, two different cells showing type 'a' and type 'b' heteromorphism as depicted in **Figure 7-20**, in 5S signals in chromosome 1 is shown. In **Figure 7-21** on the left, 4 clear 5S rDNA sites are stained, with one bivalent showing two good signals, and the other bivalent showing one good and one weak signal. The figure on the right shows one bivalent with two good 5S signals, whereas another bivalent has only one signal. However, most cells showed

type ‘a’ heteromorphism with two bright and two very faint signals. The type ‘b’ heteromorphism, which is visible only in few cells, could also be an artifact where the 5S probe could not hybridise well during the procedure.

Mis-segregation of chromosomes in tetraploids

Stickiness of the bivalents on the equatorial plate was observed and is represented in circles in **Figure 7-22**. This kind of stickiness was also observed in other cells in different varieties. This can lead to mis-segregation of the bivalents in anaphase. Meiotic chromosome stickiness has been observed before in diploid potato (Buckseth and Saggoo, 2016).

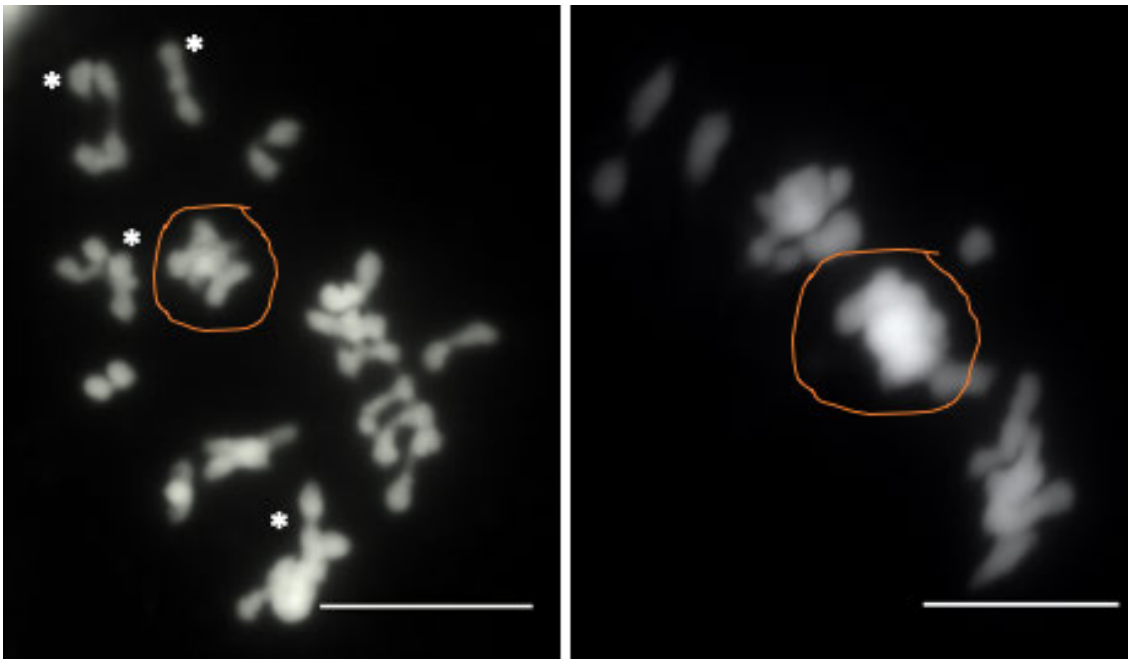


Figure 7-22 M1 cell showing non orientation and stickiness of bivalents in Sante and Maris Peer.

a) shows a cell in M1 in Sante, b) shows a cell in M1 in Maris Peer. The bivalents in circles indicate stickiness and asterisk depict non-orientation. Scale bar is 10 μm.

Non-orientation of bivalents on the M1 plate was also observed, which can be seen in **Figure 7-22** depicted with asterisk in panel a. The mis-orientation of rod bivalents has been observed earlier in diploid as well as tetraploid potato (Sangowawa, 1989).

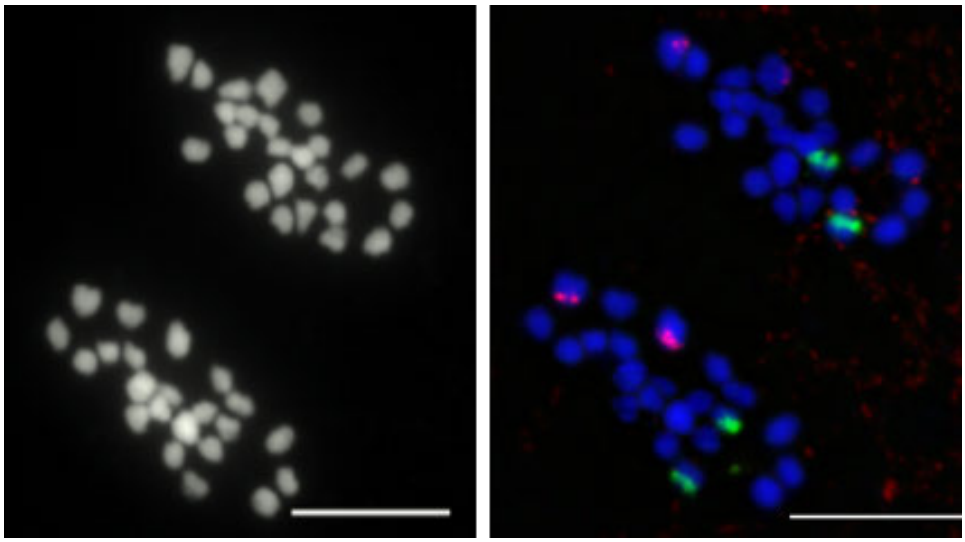


Figure 7-23 Cell in Anaphase I showing mis-segregation of the chromosomes in Sante.

Left panel shows a DAPI stained cell and right pane shows the same cell with 5S (red) and 45S (green) FISH probes. Scale bar is 10 μm.

In **Figure 7-23**, an anaphase I cell in tetraploid variety Sante shows 22 and 24 chromosomes separated along two poles, indicating loss of 2 chromosomes during segregation. These chromosomes could be lost due to non-orientation of the bivalents on the M1 plate or they could be univalents, which were slow to move to the anaphase pole and were lost in the process. 2 signals for chromosome 1 arm (5S red) and 2 signals for chromosome 2 (45S green) can be seen in the separated cells. Here again, the heteromorphic nature of the 5S rDNA is visible, with two chromosomes having a bright signal and other two chromosomes with one faint and another not so faint 5S rDNA signal.

7.4 Discussion

Cytological analysis of the cultivated potato, which is an important food crop, has not been extensively done owing to small somatic chromosomes. BAC FISH analysis has enabled to identify all the 12 sets of the diploid potato chromosomes (Dong *et al.*, 2000). However, little has been done in recent times to identify and map chiasma using meiotic metaphase chromosomes in cultivated potato. Here, we identify and show that the chromosomal spread preparation method used for *A. thaliana* can also be used in potato, with slight modifications, to prepare the meiotic spreads and tease out different information based on those spreads using FISH and immunolocalisation.

To start with, an atlas showing different meiotic stages of tetraploid varieties Sante, Maris Peer, Cara, and diploid varieties Scapa and Mayan Gold were produced and a relationship between the meiotic stage and anther size was established. As seen before in autotetraploid *Arabidosis arenosa*, more prophase I stages persisted in bigger anthers than in diploids (Higgins *et al.*, 2014) in *S. tuberosum*. This gave a positive headstart in cultivated potato cytology using present-day techniques. The varieties used belong to the European panel of varieties probed in a recent study, making them highly relevant to modern UK potato breeding. Immunolocalisation of two different structural proteins, ASY1 and ZYP1, was also successful after undertaking a similarity search between the *A. thaliana* and *S. tuberosum* for those proteins. ASY1 is the HORMA domain protein, known to be present in different plants such as *A. thaliana*, *Brassica*, rice, and plays an important role in the axial element formation (Osman *et al.*, 2018). In the absence of the protein, synapsis and crossover formation is

disturbed (Sanchez-Moran *et al.*, 2007). ZYP1 is the transverse filament protein required for the formation of proper synaptonemal complex and crossovers (Higgins *et al.*, 2005). Immunolocalisation using ASY1 and ZYP1 antibodies from *A. thaliana* gave us a good indication along with the bioinformatics search of the proteins in potato that these proteins share homology.

In the present study, a continuous ZYP1 signal was not seen in the cells. There may be two reasons; first, that a complete pachytene stage is not visible and therefore the complete synapsis is absent. The second reason can be an absence of complete synapsis, due to problems with the homology identification, as four homologues are present. Incomplete pairing during pachytene has been observed before in wild tetraploid variety *Solanum hjertingii* (Sangowawa, 1989). Synapsis formation can start to occur between two homologues and then switch over to the second and third homologue and switch again between first and fourth homologue, as shown in bottom panel in **Figure 7-11**. In that case, a complete synaptonemal complex between two homologues will not be clearly visible. Putative pairing partner switch (PPS) or synaptic pairing switch (SPS) was observed in the late zygotene-pachytene stages in tetraploid Maris Peer using dual immunolocalisation of ASY1 and ZYP1, which corroborates incomplete synapsis formation. This is the first instance of a PPS being shown in cultivated potato using immunolocalisation of axis proteins. An incomplete synapsis during pachytene has been traditionally observed in other tetraploids as well. For example in *Allium porrum,* where late zygotene, pachytene and early diplotene were grouped together for analyzing the prophase I configurations due to incomplete synapsis (Khazanehdari, Jones and Ford-Lloyd, 1995). Pairing partner switch was also visualised using electron microscopy in

the study, though it was restricted in number to being one or two, suggesting *Allium porrum* to be a weak segmental allotetraploid. Recently, presence of PPS has been visualised using powerful structured illumination microscopy (SIM) in an established autotetraploid *A. arenosa* (Morgan, 2016). This explained the presence of multivalents in the plant.

The various problems in prophase I seems to be resolved for a large proportion of cells by the metaphase I stage, because the cells observed in M1 showed more bivalents (82-85% in Sante and 58-70% in Maris Peer) and fewer multivalents than expected under the random pairing model for an autotetraploid. A high level of synapsis between all four homologues in tetraploid *S. tuberosum* variety Kathadin (65-78%), for chromosome 7 and 11 was observed in pachytene stage, which was reduced to almost half in late diakinesis and M1, indicating the resolution of the multivalent formation, by not necessarily forming a crossover between the synapsed regions. In contrast, the same chromosomes in hexaploid *Solanum demissum* paired predominantly as bivalents, indicating diploidisation of chromosomes and stabilization of meiosis (He *et al.*, 2018).

Bright foci (aggregates) formation for ASY1 and ZYP1 was seen in the cells. These resemble polycomplexes seen before. The aggregates for both ZIP1 and ASY1 proteins have been described before in yeast (Sym and Roeder, 1995) and wheat *ph1b* mutants (Boden *et al.*, 2009). The polycomplexes of ZIP1 were formed when the protein was overexpressed in yeast, indicating its role as a structural protein. They could be formed after synaptonemal complex (SC) dissolution or even before the SC formation, serving as a storehouse of SC proteins (Sym and Roeder, 1995). In wheat, the absence of *Ph1* led to an increased transcription of *TaASY1*

(Wheat Asy1), with formation of polycomplexes during prophase I, persisting from leptotene to diakinesis (Boden *et al.*, 2009). The aggregates have been found to occur in meiotic mutants of yeast unable to form proper SC (Henderson and Keeney, 2004). It has been suggested that the central filament proteins of SC can self-assemble when its proper assembly and polymerisation is disturbed due to various internal and external factors (De Carvalho and Colaiácovo, 2006). Aggregates of ASY1 and ZYP1 have also been observed in autotetraploid *A. arenosa* when grown at a temperature of 33 °C, disturbing the axis formation and downstream CO formation (Morgan, Zhang and Bomblies, 2017). This is a good example showing the effect of an external factor such as temperature on the meiotic process. In tetraploid potato, there are four homologues of each chromosome with the possibility of all trying to pair up and there may be an over-production of ZYP1 as well as ASY1 leading to such big foci formation. This indicates structural problems in the SC formation as explained above, where synapsis might be occurring at various different locations between the homologues leading to the formation of aggregates of both ASY1 and ZYP1 proteins. This situation can be transient until the completion of pachytene stage.

The presence of the ZYP1 foci along with the ASY1 foci in G2 is in contrast with *A. thaliana,* but similar to observations in rye (Mikhailova *et al.*, 2006). ASY1 foci appear initially during late G2, early leptotene in *Arabidopsis* and ZYP1 foci only appear when ASY1 becomes more continuous (Higgins *et al.*, 2005). However, in rye, linear tracts of ZYP1 were observed in leptotene (Mikhailova *et al.*, 2006), whereas only foci were observed for potato. The potato genome is bigger than *Arabidopsis* and it is possible that different loading mechanisms are in place to ensure synapsis and proper segregation. It was interesting to carry out the techniques

in the diploid potato line Scapa, to analyse the differences if any in the ASY1 and ZYP1 interaction during SC formation. However, a similar pattern was seen in Scapa as well. This indicates a difference from *Arabidopsis*, and it will be interesting to find out the roles of ZYP1 in potato during the earliest stages of prophase. However, it could be an artifact as well. *Arabidopsis* ZYP1 protein shows only 68% similarity with potato and therefore the anti ZYP1 antibody could be binding other epitopes in potato.

FISH analysis probing 5S and 45S rDNA on chromosomes 1 and 2 gave useful information about chiasma formation, occurrence of multivalents and univalents during M1, which may result in mis-segregation of chromosomes later. This provides the useful information and proof about the existence of quadrivalent pairing, which can lead to double reduction during gamete formation. A clear ring quadrivalent for chromosome 1 as indicated by the presence of the 5S rDNA sites (red stains) can be seen in **Figure 7-12** and a clear chain quadrivalent for chromosome 2, indicated by green 45S rDNA signals in **Figure 7-13**. The chromosomes can then segregate normally in regular two by two fashion or may mis-segregate during anaphase.

Chromosome 1 is the larger of the two chromosomes and therefore may be expected to show more quadrivalents compared with the smaller chromosome 2. Overall, the number of multivalents is low in Sante, with 18% and 14% for chromosomes 1 and 2 respectively, indicating the diploidisation of the genome. It is interesting to note that potato is grown vegetatively, so there should be no recent selection for fertility; even then, there is a diploidisation of the chromosomes. This indicates a more genetic control enabling even segregation of the chromosomes, which can ensure fertility of the plant in the longer term.

This variety however, is known not to bear fruit pods indicating different reasons rather than irregular meiosis to affect its fertility. The diploidisation of chromosomes was also visible in Cara and Maris Peer, as the multivalent formation was found to be 48% and 26% in Cara and 41% and 28% in Maris Peer for chromosomes 1 and 2 respectively. However, this was significantly higher than Sante indicating a genetic variation, which leads to different propensity to multivalent formation in different varieties of potato. This variation could be at the nucleotide level or at the level of chromatin modellers. Considering the cytological behaviour of the two analysed chromosomes the overall multivalent formation was low when compared with bivalent pairing.

Lower multivalent formation than the stipulated 66.66% from random pairing model, may also indicate partial preferential pairing of the chromosomes. This kind of pairing has recently been analysed in tetraploid rose, where it was found to be variable between chromosomes, genotypes identified and even between the chromosome arms, with the presence of quadrivalents as well as bivalents (Bourke *et al.*, 2017). This type of preferential pairing occurs in segmental allopolyploids. However, cytological behaviour alone cannot be the basis of implying segmental allopolyploidy in cultivated potato. Using high density marker linkage map analysis to quantify preferential pairing can be utilized, as was done in Bourke *et al*. (2017). The frequency of multivalent formation that affects the degree of double reduction (DR) was analysed in *S. tuberosum* using simulation. A high density linkage map was produced and the effects of the pairing behavior of the chromosomes on the map construction was analysed (Bourke *et al.*, 2015). The frequency of quadrivalent formation was found between 20% and 30% with no evidence for preferential pairing by comparing the predicted DR rates based on

simulation of a mapping population with the observed rates of DR based on the marker data of a linkage mapping population created by crossing two tetraploid lines.

The chromosome pairing behaviour and the rate of multivalent formation has been analysed in several polyploid species. Diploid, natural and induced tetraploids of Kiwifruit *Actinidia chinensis* were used to compare the multivalent frequency between them. Inheritance of the microsatellite alleles in tetraploid progeny from the cross between tetraploids *A. chinensis* and *A. arguta* was used to find out whether preferential or non-preferential chromosome pairing occurred. Non-preferential chromosome pairing in the tetraploids was found, along with a higher tendency of multivalent formation in induced tetraploids than the natural tetraploids. Based on the number of quadrivalents and bivalents formed, the multivalent frequency ranged from 8.97-12% in induced tetraploids and 7-8% in natural tetraploids (Wu *et al.*, 2014). In another hexaploid variety *Actinidia chineneis* var *deliciosa* (6x), a non-preferential chromosome pairing was found. This was based on the inheritance of microsatellite alleles in the tetraploid progeny of a cross between *A. chinensis* var *deliciosa* and the distantly related *Actinidia eriantha* Benth (2x). The quadrivalent frequency was found to be low, 1 in 20 cells or 5% only (Mertten *et al.*, 2012).

Tetrasomic inheritance with random pairing of chromosomes was also found in blueberries. The chromosomes mostly paired as bivalents in meiosis with less than 10% quadrivalent formation in commercial tetraploid *Vaccinium corymbosum* called Bluecrop as well as wild tetraploid variety called CEL. A synthetic hybrid variety called US75 derived from the cross between 2n gametes of diploid *Vaccinium darrowi* and Bluecrop also showed less than 25%

quadrivalent formation. Tetrasomic inheritance for US75 was established based on the inheritance pattern of the RAPD markers (present in the parental lines of US75) in a segregating population of a cross between US75 and *V. corymbosum*. Some genomic divergence with easy transfer of genes was found between different *Vaccinium* species (Qu, Hancock and Whallon, 1998). This can be favourable for breeders making transfer of desirable genes easier between ploidies.

Meiotic studies were carried out on natural and commercial diploid (2n=14) and natural tetraploid (2n=28) crested wheatgrasses in Turkey. Quadrivalent formation was found to be 28% in the natural tetraploid crested wheatgrass *Agropyron desertorum,* which is used in revegetating arid range lands (Deniz and Dogru, 2006). In another perennial autotetraploid grass *Pennisetum orientale,* 36-47% chromosomes occurred as quadrivalents during meiosis, with some cells also showing higher associations of hexavalents and octavalents, indicating chromosomal changes such as translocations and or inversions (Koul, Nagpal and Sharma, 1999). A reciprocal translocation may occur between different set of homologous chromosomes, which may lead to higher associations. This may eventually lead to the cytological diploidisation with regular 2 x 2 meiosis.

Cytological diploidisation can occur after autopolyploidization to ensure proper segregation of chromosomes. A range of multivalent frequencies per cell ranging from 1.58 to 4.80 has been cytologically identified in different commercial varieties of potato (Swaminathan, 1954). Recently, pairing configurations for chromosomes 2, 4, 7 and 11 in tetraploid and hexaploid varieties of *S. tuberosum* was analysed using oligo probes in late diakinesis and M1 (He *et al.*,

2018). Quadrivalent formation occurred for all the four chromosomes ranging from 21 to 42% in tetraploids and 0 to 16.8% in hexaploid potato. This multivalent frequency is comparable with our two chromosomes in three different tetraploid varieties. A range of multivalent frequencies in different chromosomes in same species and in different varieties may indicate genetic control over chromosome pairing configuration. A few chromosomes, may be because of size or other structural differences, may diploidise earlier than the other chromosomes in the same variety.

Chiasma analysis has been performed in the past cytologically on different *Solanum* varieties where an overall estimation was made. Mean chiasma frequencies ranging from 1.0 to 1.68 in diploids, 1.42-1.71 in tetraploids to 1.12 to 1.86 per bivalent in hexaploids has been found in various tuber bearing *Solanum* varieties (reviewed in Magoon, Ramanujam and Cooper, 1962). Though, this analysis was carried out for the whole set of chromosomes, it is comparable with our results for diploids and tetraploids where the per bivalent CO frequency ranged from 1.15 to 1.47 in diploids and 1.01 to 1.32 in different tetraploid varieties, which is lower than the diploid variety and the above example. It has to be kept in mind that the analysis has been done for two chromosomes and will be different once all the chromosomes are taken into account. For the two chromosomes in the tetraploids, the chiasma frequency distribution was associated with the chromosome arms and therefore may vary from chromosome to chromosome. Though no significant difference was found between the varieties, however subtle differences were seen in chiasma frequency.

The difference in chiasma frequencies amongst varieties has been reported before in other plants such as in *A. thaliana* (López *et al.*, 2012). The chiasma frequencies for the wild tetraploid potatoes has been found to be still lower than the commercial varieties. The overall average chiasma frequency was found to be 1.27 in wild tetraploid potato *Solanum hjertingii* (Sangowawa, 1989). This helps to put the chiasma frequency calculated for different varieties, though only for two chromosomes in context. It was useful to analyse chiasma and behaviour of metaphase chromosomes in diploid variety Scapa, where rod and rings were found in almost equal proportion for chromosome 1 and rod bivalents predominantly occurred for chromosome 2, indicating the presence of low chiasma for the smaller chromosome and still lower for the smaller arm. Occasional univalents for chromosome 2 was also found, indicating the loss of the obligate crossover.

Increased interference has been attributed with the low chiasma frequency and stabilization of meiosis in autotetraploid *A. arenosa* (Bomblies, Higgins and Yant, 2015). In our study, per bivalent crossover frequency calculated for chromosomes 1 and 2 in tetraploids was found to be lower than the diploid variety. It was significantly lower in Sante for chromosome 1 and for chromosomes 1 and 2 both, in Maris Peer. This might indicate presence of stronger interference in tetraploid varieties reducing the chiasma frequency to enable bivalent formation and proper segregation of chromosomes in meiosis.

The chromosome 1 in potato is roughly 80 Mb in length with smaller arm being 20 and longer being 60 Mb, while chromosome 2 is roughly 40 Mb with smaller arm being 5 and longer being 35 Mb in length (The Potato Genome Sequencing, 2011). The genetic map length was found

to be 93 cM for chromosome 1 and 76 cM for chromosome 2 in a segregating population in potato (Sharma *et al.*, 2013). This explains smaller number of crossovers in chromosome 2 compared with chromosome 1 and also in the smaller arm for both the chromosomes for all the varieties. When separating bivalents and multivalents in tetraploids, a boost in the chiasma frequency in smaller arm was seen in multivalents only cells in both chromosomes. The presence of the crossovers on the small arm in chromosome 2 is noteworthy. Chromosome 2 is considered subtelocentric with a NOR region and a satellite in its small arm (Dong *et al*., 2000). NOR and satellite regions are considered highly heterochromatic, and NOR may be responsible for pairing up of homologous chromosomes and remaining associated for longer time giving it more time to form crossovers, as has been observed in *A. thaliana* (Da Ines, Gallego and White, 2014). However, there must be some euchromatic regions present on the small arm which are capable of recombining during meiosis. Similar regions were identified in *A. thaliana* on chromosomes 2 and 4, which are NOR associated and were considered free of crossovers before the cytological analysis revealed the presence of crossovers on their small arms (Moran *et al.*, 2001). Short arm of chromosome 2, in varieties Cara and Maris Peer also showed the largest increase in CO rate when comparing only multivalent forming cells in tetraploids with diploid Scapa. However, the number of multivalents formed in chromosome 2 were lower than chromosome 1, implying earlier diploidisation of smaller chromosomes than the bigger chromosomes.

Only two chromosomes have been analysed in the present analysis. BAC FISH probes as well as oligo probes have been developed which can be used to identify all the 24 bivalents and perform chiasma analysis for the complete set (Braz *et al.*, 2018). These probes have been

useful in identifying a single set of chromosomes in an experiment (He *et al.*, 2018), though it will be useful to see if all 12 probes identifying 24 sets of chromosomes could be used together. Antibodies against MLH1 foci can be used to count the number as well as position of Class I crossovers (Anderson *et al.*, 2014). The anti-MLH1 was tried on the prepared slides and the foci could not be reliably counted due to the presence of large background. However, in future it can be used on the fresh material to analyse the sites of CO.

A very interesting feature that came across during FISH analysis was the heteromorphic 5S signal in Sante, though not in other varieties pointing to the genetic variation among the varieties. The intensity of signal was different in different homologous bivalents. It is known that 5S rDNA is present only on chromosome 1; hence, it is beyond doubt that there is heteromorphism in the 5S rDNA present in the tetraploid plant variety Sante. This kind of heteromorphism has been seen in other polyploid plants as well. Two stronger and two weaker signals of 5S have been identified in tetraploid *Fragaria corymbosa,* belonging to the strawberry genus, and was identified as an allotetraploid based on this result (Liu and Davis, 2011). The study also identified polymorphic 25S signals in various diploid accessions. The polymorphism indicated towards the dynamic nature of these loci in the diploids, which could also lead to the loss of few repeats, duplication or rearrangement. Heteromorphic 5S rDNA sites have also been reported in *Ceratozamia kuesteriana*, a plant belonging to cycad family (Kokubugata, Vovides and Kondo, 2004). Rearrangements, translocations, gene loss can accumulate in the non-offspring producing organisms, where there is no selection pressure. Here, potato is grown vegetatively and specifically the variety Sante does not produce berries, indicating a reproduction barrier. It is possible that either the autotetraploid Sante is actually

a segmental allotetraploid where two very close but differentiated polyploid accessions came together or the diploid ancestor was a hybrid between two very close accessions with copy number variation in 5S rDNA that doubled its genome.

Stickiness and non-orientation of chromosomes, as observed in M1 stages between the chromosomes in Sante and Maris Peer can lead to mis-segregation and loss of chromosomes. This kind of behaviour has been reported before in plants. A recessive sticky gene has been identified in maize, which in homozygous state produced abnormal meiosis having sticky chromosomes and led to sterility (Beadle, 1933). It was suggested that translocations occurred between chromosomes, which led to chromosomal associations leading to stickiness in these plants. A review of various plant studies established that interconnections between chromosomes could be present in M1 stages in meiosis, especially in hybrids and polyploids where a diffused stage between pachytene and diplotene was present in meiosis, which was responsible for the stickiness. This stickiness was similar to that seen by the X-ray irradiation of grasshopper chromosomes (Klasterska, 2009). The diffused stage occurs during early diplotene, where the nucleus resembles interphase and/or leptotene, with a network of desynapsing chromatin threads. It has been referred to as “tinsel-like structures” in cereals (Colas *et al.*, 2017). Chromosome stickiness was observed in hybrid *Panicum maximum,* which led to abnormal meiosis and sterility (Pessim *et al.*, 2015). Non orientation has been seen before, where an increased distance between the centromeres with reduced repulsion was suggested as the reason (Sangowawa, 1989). Therefore, genetic as well as environmental factors can be responsible for such kind of behaviour.

In our work, crossing between Sante and Sarpo Mira varieties did not produce berries. Reciprocal crossing was attempted, but the flower would fall off after a few days of pollination. Many different ways of supporting flowers after cross pollination have been described in the literature, which can be utilized to support the crosses (Howard, 1961). Several fertilization problems have been known to occur in cultivated potato plants, ranging from male sterility, low viability of pollen, dropping of flower buds and self-incompatibility. 4x *Solanum tuberosum* is considered to be the least fertile class of the potatoes (Muthoni *et al.*, 2012). In general, fertilization can be a problem in autotetraploids, where all four homologues can potentially recombine. In variety Cara, Sante and Maris Peer, the meiotic cells were checked, where a minority of quadrivalent formations occurred. This indicates that the fertility may be reduced, but they should not be sterile. The crosses between tetraploid *A. thaliana* lines worked (Chapter 3) and even the tetraploid parents set seeds, though the seed set was considerably lower than the diploids. It is possible that crossing between many more varieties need to be attempted to find out the cross compatibility and produce true seeds.

Comparison between diploid and tetraploid potato lines using molecular cytogenetics techniques along with powerful microscopy such as super resolution microscopy can give a better understanding about the meiotic processes. This can be useful in understanding about the stabilization of the polyploid genome as well as work out the phylogeny of the potatoes based on the molecular cytogenetic markers. This is also important to understand the variation present between the varieties which can be helpful in developing high yielding varieties. Potato genome has been sequenced and published (The Potato Genome Sequencing, 2011), which paves the way forward for molecular genetics along with

cytogenetics approach to understand its genome evolution. Being one of the important food crops, research into different aspects of potato biology can help accelerate breeding programmes, which can help overcome the imminent food insecurity, which we are facing due to multiple factors.

7.5 References

Anderson, L. K. *et al.* (2014) 'Combined fluorescent and electron microscopic imaging unveils the specific properties of two classes of meiotic crossovers', *Proceedings of the National Academy of Sciences*, 111(37), pp. 13415–13420. doi: 10.1073/pnas.1406846111.

Armstrong, S. J. (2002) 'Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica', *Journal of Cell Science*, 115(18), pp. 3645–3655. doi: 10.1242/jcs.00048.

Beadle, G. W. (1933) 'A gene for sticky chromosomes inZea mays', *Zeitschrift für Induktive Abstammungs- und Vererbungslehre*, 63(1), pp. 195–217. doi: 10.1007/BF01849089.

Boden, S. A. *et al.* (2009) 'TaASY1 promotes homologous chromosome interactions and is affected by deletion of Ph1', *Plant Journal*, 57(3), pp. 487–497. doi: 10.1111/j.1365-313X.2008.03701.x.

Bomblies, K., Higgins, J. D. and Yant, L. (2015) 'Meiosis evolves: adaptation to external and internal environments', *New Phytologist*, 208(2), pp. 306–323. doi: 10.1111/nph.13499.

Bourke, P. M. *et al.* (2015) 'The double-reduction landscape in tetraploid potato as revealed by a high-density linkage map', *Genetics*, 201(3), pp. 853–863. doi: 10.1534/genetics.115.181008.

Bourke, P. M. *et al.* (2017) 'Partial preferential chromosome pairing is genotype dependent in tetraploid rose', *Plant Journal*, 90(2), pp. 330–343. doi: 10.1111/tpj.13496.

Braz, G. T. *et al.* (2018) 'Comparative oligo-FISH mapping: An efficient and powerful methodology to reveal karyotypic and chromosomal evolution', *Genetics*, 208(2), pp. 513–523. doi: 10.1534/genetics.117.300344.

Buckseth, T. and Saggoo, M. I. S. (2016) 'Cytological investigations in diploid potato (*Solanum spp*.)', *Chromosome Botany*, 11(1), pp. 9–13. doi: 10.3199/iscb.11.9.

Camire, M. E. (2016) 'Potatoes and Human Health', in *Advances in Potato Chemistry and Technology: Second Edition*, pp. 685–704. doi: 10.1016/B978-0-12-800002-1.00023-6.

De Carvalho, C. E. and Colaiácovo, M. P. (2006) 'SUMO-mediated regulation of synaptonemal complex formation during meiosis', *Genes and Development*, pp. 1986–1992. doi: 10.1101/gad.1457806.

Colas, I. *et al.* (2017) 'Observation of Extensive Chromosome Axis Remodeling during the "Diffuse-Phase" of Meiosis in Large Genome Cereals', *Frontiers in Plant Science*. doi: 10.3389/fpls.2017.01235.

Deniz, B. and Dogru, U. (2006) 'Meiotic behaviour in natural diploid, tetraploid, and commercial diploid crested wheatgrass', *New Zealand Journal of Agricultural Research*, 49(4), pp. 405–409. doi: 10.1080/00288233.2006.9513731.

Dong, F. *et al.* (2000) 'Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato', *Theoretical and Applied Genetics*, 101(7), pp. 1001–1007. doi: 10.1007/s001220051573.

Gaiero, P. *et al.* (2016) 'Collinearity between potato (Solanum tuberosum L.) and wild relatives assessed by comparative cytogenetic mapping', *Genome*, 60(3), pp 228-240. doi: 10.1139/gen-2016-0150.

Gavrilenko, T. (2007) 'Potato cytogenetics', in *Potato Biology and Biotechnology: Advances and Perspectives*, pp. 203–216. doi: 10.1016/B978-044451018-1/50052-X.

He, L. *et al.* (2018) 'Chromosome painting in meiosis reveals pairing of specific chromosomes in polyploid Solanum species', *Chromosoma*, 127(4), pp 505-513. doi: 10.1007/s00412-018-0682-9.

Henderson, K. A. and Keeney, S. (2004) 'Tying synaptonemal complex initiation to the formation and programmed repair of DNA double-strand breaks', *Proceedings of the National Academy of Sciences*, 101(13), pp. 4519–4524. doi: 10.1073/pnas.0400843101.

Higgins, J. D. *et al.* (2005) 'The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over', *Genes and Development*, 19(20), pp. 2488–2500. doi: 10.1101/gad.354705.

Higgins, J. D. *et al.* (2014) 'Cytological techniques to analyze meiosis in Arabidopsis arenosa for investigating adaptation to polyploidy', *Frontiers in Plant Science*, 4(January), pp. 1–7. doi: 10.3389/fpls.2013.00546.

Howard, H. W. (1961) *Potato Cytology and Genetics 1952–1959*. Dordrecht: Springer Netherlands. doi: 10.1007/978-94-015-3518-2.

Da Ines, O., Gallego, M. E. and White, C. I. (2014) 'Recombination-independent mechanisms and pairing of homologous chromosomes during meiosis in plants', *Molecular Plant*, 7(3), pp 492-501. doi: 10.1093/mp/sst172.

Jackson, N. *et al.* (2006) 'Reduced meiotic crossovers and delayed prophase I progression in AtMLH3-deficient Arabidopsis', *EMBO Journal*, 25(6), pp. 1315–1323. doi: 10.1038/sj.emboj.7600992.

Khazanehdari, K. A., Jones, G. H. and Ford-Lloyd, B. V. (1995) 'Meiosis in the leek (Allium porrum L.) revisited I. Prophase I pairing', *Chromosome Research*. doi: 10.1007/BF00713894.

Klasterska, I. (2009) ‘A new look on the role of the diffuse stage in problems of plant and animal meiosis’, *Hereditas*, 82(2), pp. 193–203. doi: 10.1111/j.1601-5223.1976.tb01557.x.

Kokubugata, G., Vovides, A. P. and Kondo, K. (2004) ‘Mapping 5S ribosomal DNA on somatic chromosomes of four species of Ceratozamia and Stangeria eriopus (Cycadales)’, *Botanical Journal of the Linnean Society*, 145(4), pp. 499–504. doi: 10.1111/j.1095-8339.2004.00299.x.

Koul, K. K., Nagpal, R. and Sharma, A. (1999) ‘Peculiarities of chromosomal behaviour in the natural autotetraploid Pennisetum orientale L.C. Rich’, *Caryologia*, 52(3–4), pp. 185–195. doi: 10.1080/00087114.1998.10589172.

Lambing, C. *et al.* (2015) ‘Arabidopsis PCH2 Mediates Meiotic Chromosome Remodeling and Maturation of Crossovers’, *PLoS Genetics*, 11(7). doi: 10.1371/journal.pgen.1005372.

Lhuissier, F. G. P. *et al.* (2007) ‘The Mismatch Repair Protein MLH1 Marks a Subset of Strongly Interfering Crossovers in Tomato’, *THE PLANT CELL ONLINE*, 19(3), pp. 862–876. doi: 10.1105/tpc.106.049106.

Liu, B. and Davis, T. M. (2011) ‘Conservation and loss of ribosomal RNA gene sites in diploid and polyploid Fragaria (Rosaceae)’, *BMC Plant Biology*. BioMed Central Ltd, 11(1), pp. 157. doi: 10.1186/1471-2229-11-157.

López, E. *et al.* (2012) ‘Looking for natural variation in chiasma frequency in Arabidopsis thaliana’, *Journal of Experimental Botany*, 63(2), pp. 887–894. doi: 10.1093/jxb/err319.

Madden, T. (2002) ‘Chapter 16 : The BLAST Sequence Analysis Tool’, *The NCBI Handbook Internet National Library of (2002)*. doi: 9305 [pii].

Magoon, M. L., Ramanujam, S. and Cooper, D. C. (1962) ‘Cytogenetical Studies in Relation to the Origin and Differentiation of Species in the Genus Solanum L.’, *Caryologia*, 15(1), pp 151-252. doi: 10.1080/00087114.1962.10796057.

Mertten, D. *et al.* (2012) ‘Meiotic chromosome pairing in Actinidia chinensis var. deliciosa’, *Genetica*, 140(10–12), pp. 455–462. doi: 10.1007/s10709-012-9693-2.

Mikhailova, E. I. *et al.* (2006) ‘Molecular assembly of meiotic proteins Asy1 and Zyp1 and pairing promiscuity in rye (Secale cereale L.) and its synaptic mutant sy10’, *Genetics*, 174(3), pp. 1247–1258. doi: 10.1534/genetics.106.064105.

Moran, E. S. *et al.* (2001) ‘Chiasma formation in Arabidopsis thaliana accession Wassileskija and in two meiotic mutants’, *Chromosome Research*, 9(2), pp. 121–128. doi: 10.1023/A:1009278902994.

Morgan, C. (2016) *Coordination of meiotic recombination in diploid and tetraploid Arabidopsis (PhD thesis)*.

Morgan, C. H., Zhang, H. and Bomblies, K. (2017) 'Are the effects of elevated temperature on meiotic recombination and thermotolerance linked via the axis and synaptonemal complex?', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 372(1736), pp. 20160470. doi: 10.1098/rstb.2016.0470.

Muthoni, J. *et al.* (2012) 'Reproductive biology and early generation's selection in conventional potato breeding', *Australian Journal of Crop Science*, 6(3), pp. 488–497.

Osman, K. *et al.* (2018) 'Affinity proteomics reveals extensive phosphorylation of the Brassica chromosome axis protein ASY1 and a network of associated proteins at prophase I of meiosis', *Plant Journal*, 93(1), pp. 17–33. doi: 10.1111/tpj.13752.

Pessim, C. *et al.* (2015) 'Chromosome stickiness impairs meiosis and influences reproductive success in panicum maximum (Poaceae) hybrid plants', *Genetics and Molecular Research*, 14(2), pp 4195-4202. doi: 10.4238/2015.April.28.2.

Qu, L., Hancock, J. F. and Whallon, J. H. (1998) 'Evolution in an autopolyploid group displaying predominantly bivalent pairing at meiosis: Genomic similarity of diploid Vaccinium darrowi and autotetraploid V. corymbosum (Ericaceae)', *American Journal of Botany*, 85(5), pp. 698–703. doi: 10.2307/2446540.

Sanchez-Moran, E. *et al.* (2007) 'ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis', *Genes and Development*, 21(17), pp. 2220–2233. doi: 10.1101/gad.439007.

Sangowawa, B. G. (1989) 'Meiotic studies in a wild tetraploid potato (Solanum hjertingii Hawkes).', *CYTOLOGIA*, 54(4), pp. 617–626. doi: 10.1508/cytologia.54.617.

Sharma, S. K. *et al.* (2013) 'Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps', *G3: Genes|Genomes|Genetics*, 3(11), pp 2031-2047 doi: 10.1534/g3.113.007153.

Sharma, S. K. *et al.* (2018) 'Linkage Disequilibrium and Evaluation of Genome-Wide Association Mapping Models in Tetraploid Potato', *G3: Genes|Genomes|Genetics*, 8(10), pp 3185-3202. doi: 10.1534/g3.118.200377.

Swaminathan, M. S. (1954) 'Microsporogenesis in Some Commercial Potato Varieties', *Journal of Heredity*, 45(6), pp. 265–272. doi: 10.1093/oxfordjournals.jhered.a106489.

Sybenga, J. (1975) 'The Analysis of Chromosome Pairing', in *Meiotic Configurations: A Source of Information for Estimating Genetic Parameters*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 134–199. doi: 10.1007/978-3-642-80960-6_3.

Sym, M. and Roeder, G. S. (1995) 'Zip1-induced changes in synaptonemal complex structure

and polycomplex assembly.', *The Journal of Cell Biology*. Rockefeller University Press, 128(4), pp. 455–466. doi: 10.1083/jcb.128.4.455.

Tang, X. *et al.* (2009) 'Assignment of genetic linkage maps to diploid Solanum tuberosum pachytene chromosomes by BAC-FISH technology', *Chromosome Research*, 17(7), pp. 899–915. doi: 10.1007/s10577-009-9077-3.

The Potato Genome Sequencing (2011) 'Genome sequence and analysis of the tuber crop potato.', *Nature*, 475, pp. 189–195. doi: 10.1038/nature10158.

Visintin, R. and Amon, A. (2000) 'The nucleolus: The magician's hat for cell cycle tricks', *Current Opinion in Cell Biology*, pp. 372–377. doi: 10.1016/S0955-0674(00)00102-2.

Wu, J. H. *et al.* (2014) 'Meiotic chromosome pairing behaviour of natural tetraploids and induced autotetraploids of Actinidia chinensis.', *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 127(3), pp. 549–557. doi: 10.1007/s00122-013-2238-y.

Wu, S. S. *et al.* (2001) 'A multivalent pairing model of linkage analysis in autotetraploids', *Genetics*, 159(3), pp. 1339–1350.

CHAPTER 8

GENERAL DISCUSSION

8 General Discussion

8.1 Introduction

Polyploidy, which is considered as one of the many important factors shaping evolution of land plants, is present in most of the important crop plants today. Phenotypic and cytological comparison has been done in my thesis to compare and contrast the differences between diploid and polyploid plants, with emphasis on meiotic recombination in the model plant *Arabidopsis thaliana* and transferring the cytological techniques developed in this plant to the crop plant *Solanum tuberosum*. Though I started with an initial aim of characterising structural and functional differences between diploid and tetraploid *A. thaliana* using phenotypic data analysis along with sequencing data analysis in F2 lines created, backed up by cytological analysis; an exciting new aim of developing and analysing cytology in crop plant potato developed during the journey.

8.2 Comparative phenotypic and genotypic analysis between diploids and autotetraploid *A. thaliana*

Firstly, I was successful in creating the diploid and tetraploid F2 population in *A. thaliana* by crossing Columbia and Landsberg plants. This was in line with an initial aim of creating a mapping population, where recombination differences along with the genetic basis of any phenotypic variation could be analysed between the diploid and tetraploid varieties. A large number of polymorphic molecular markers along with autotetraploid lines are available for the varieties, therefore the two parents were chosen to create a mapping population. Owing to the time constraints of a PhD project, an F2 population was chosen as it requires only two

generations to develop. Though it is quick to develop, each plant is unique and therefore replications cannot be performed which can impose certain limitations. However, we created a population of more than 300 individuals, which increases the statistical power to detect patterns. A second plant trial was conducted after the first trial was unsuccessful owing to various environmental conditions in the glasshouse. Another major reason was the presence of greater variance in F1 than F2 for the phenotypic traits analysed. However, this discrepancy was not removed in the second trial as well and various reasons have been discussed in Chapter 4. It is plausible that subtle differences were missed during phenotyping. For example, the measurements were taken during the course of the day at different times starting from one end of the growth facility to the other, where the exact time difference could not be taken into account. A way to describe this would be that, one plant might have germinated in the early morning, whereas another plant could have germinated a few hours later in the morning. Both were scored as germinating on the same day. Here, sophisticated digital phenotyping tools could be more useful, which can make note of such subtle differences between individuals (reviewed in Das Choudhury, Samal and Awada, 2019).

Various traits were analysed phenotypically to find out the differences between the diploid and tetraploids, where a significant difference in distribution for numerous traits was found between the diploids and tetraploids. An ideal situation would have been to identify the genotypic differences between the two populations to account for the phenotypic differences observed through QTL mapping. However, only a handful of 24 samples could be sequenced in time for this project, and more will be sequenced in the future. Preliminary data analysis

reached the stage of variant calling, performed on the data for 24 samples, which arrived in April, 2019.

8.3 Cytological comparison between diploids and tetraploids in model plant *A. thaliana* and crop plant *S. tuberosum*.

Diploid and autotetraploid parental lines and F2s of *A. thaliana* were analysed cytologically to establish a difference in mean chiasma frequency between them. Cytogenetic analysis in polyploid plants is challenging, more so when the aim is to analyse the chiasma frequency along with the configuration behaviour of the chromosomes. The parental autotetraploids used in my project were the established autotetraploids, even then it was important to establish their true tetraploidy, to enable creation of the F2 population. Not only the autotetraploid parents were confirmed for their tetraploidy by counting the number of chromosomes, chiasma count analysis was also performed for Columbia and Landsberg diploids and tetraploids. Similar analysis was carried out for a few plants from the F2 population used in the plant trial. It was hypothesised that polyploidy can lead to an increase in the number of crossovers over and above what is expected (based on doubling of chromosome number) and therefore, can serve as a potential pathway of increasing meiotic recombination in the crop plants, considering many important crop plants are polyploids. Though a significant increase in the number of crossovers was found in tetraploids, it was not over and above the expected doubling that will happen in tetraploids owing to the doubling of the chromosomes. On the other hand, a per bivalent CO frequency was found to be less in tetraploid F2s than the diploid F2s indicating diploidisation of chromosomes. This result is in contradiction with a few studies in the past on autotetraploid *A. thaliana* where an increase in the recombination frequency in autotetraploids was found (Pecinka *et al.*, 2011).

As discussed in chapters 3 and 5, several configurations can be attained by the metaphase I chromosomes in an autotetraploid. It has been seen that neo-autopolyploids form high rate of multivalents and cytological diploidisation occurs over the successive generations (Santos *et al.*, 2003). Even after numerous generations, we observed a substantial rate of multivalent formation both in the autotetraploid parents as well as in the F2s in *A. thaliana*. The multivalent formation between the two *A. thaliana* varieties did not differ significantly from each other except for chromosome 4. The multivalents formed during M1 can segregate two by two or may mis-segregate leading to aneuploidy and fertility problems. A reduction in fertility as compared with diploids was indeed visible, but it was still not low enough to endanger the plant. This also highlighted the problems the process of meiosis in gametes was facing owing to increased chromosomes. It would be interesting in future to compare the differences in sequence and expression levels of core meiotic genes (e.g. *ASY1*), as was done in *Arabidopsis arenosa* (Yant *et al.*, 2013).

To understand chromosomal behaviour in a crop plant, cytological analysis was carried out in diploid and autotetraploid *Solanum tuberosum*. Methods were optimised to produce a clear atlas of different meiotic stages. Similar to *A. thaliana*, chiasma analysis was undertaken for two chromosomes, chromosome 1 which is the biggest and chromosome 2 which is one of the smaller chromosomes, both of which could be probed with 5S and 45S FISH probes developed from *A. thaliana*. The chiasma count between the diploid and tetraploid potato varieties differed significantly, but did not exceed doubling in a tetraploid, as was seen in *A. thaliana* for all the cells observed. Similar to *A. thaliana* F2s, the per bivalent CO frequency was lower in tetraploids than in the diploid variety. The reduction seen in per bivalent CO

frequency in tetraploid *A. thaliana* and crop plant *S. tuberosum* in our work, resonates with the reduction in chiasma frequency in *A. arenosa*, which has been reduced to one crossover per bivalent due to increased interference, which is thought to stabilise meiosis (Bomblies *et al.*, 2016).

Prophase I in diploid and tetraploid potato varieties was also probed using antibodies against axis proteins ASY1 and synaptonemal complex protein ZYP1 used in *A. thaliana*. This is the first instance of a probable pairing partner switch (PPS) shown in a tetraploid potato variety, which again indicates the problems even an established autotetraploid such as potato may face during meiosis. Foci resembling polycomplexes of ASY1 and ZYP1 was also seen in tetraploids, indicating possible issues with axis maintenance and synapsis formation. These observations help understand the challenges faced during meiosis in an autopolyploid crop plant directly. Though we did not delve deep in prophase I with these proteins, it opens up a whole new area of research in an autotetraploid crop plant to understand meiosis better. For example, the effect of interference could be studied by measuring the length of SC complex in tetraploids vs diploids using ZYP1 protein immunolocalisation during prophase I. Various important genes such as the S-RNase gene responsible for self-incompatibility in diploids is present on chromosome 1 in potatoes (Enciso-Rodriguez *et al.*, 2019). Similarly QTLs for yield have been identified on both chromosomes 1 and 2 (Manrique-Carpintero *et al.*, 2015). Understanding the factors affecting meiotic recombination can help improve these specific traits.

The pairing partner switch that occurs during zygotene-pachytene in meiosis, can lead to multivalent formation, which may lead to mis-segregation of chromosomes. Chromosome

pairing configuration was studied and multivalent formation was found to occur in all the tetraploid potato varieties. The rate of multivalent formation and diploidisation was found to be variety (though not significant) as well as chromosome dependent, ranging from 14-48% in different varieties. The deviation from the 66% multivalent formation according to random end pairing model indicated substantial diploidisation in both the chromosomes.

Potato has assumed random pairing behaviour and tetrasomic inheritance with different levels of diploidisation happening in different varieties. To put it into context the pairing behaviour can be compared and contrasted with other autotetraploid plants and animals. In the autotetraploid Leek *Allium porrum*, 71% quadrivalent formation was found during prophase I in meiosis, but only 40% of cells retained the multivalent configuration in metaphase I (Jones, Khazanehdari and Ford-Lloyd, 1996). Autotetraploid alfalfa shows predominant bivalent pairing of chromosomes in meiosis, but show tetrasomic pattern of inheritance (Quiros, 1982). All these examples indicate the genetic control of chromosome pairing and a tendency towards the cytological diploidisation of the chromosomes in various plant species.

In animals, polyploidy is not prevalent in all taxa, however it is present in many fishes, reptiles and amphibians (Comai, 2005). A few studies about pairing behaviour in these animals have been carried out. Preferential bivalent pairing of 86% of chromosomes was found in the spermatocytes of the induced tetraploid Pacific oyster *Crassostrea gigas* Thunberg. This was in contrast to the eggs in the tetraploid females, where chromosomes paired mostly as quadrivalents. Though they produced aneuploid gametes, fertility was comparable with

diploids (Zhang *et al.*, 2014). Predominant quadrivalent formation with equal segregation of chromosomes and normal fertility was also found in the male and female gonads of tetraploid South American frog *Odontophrynus americanus* (Beçak, Beçak and Rabello, 1966). This indicates that these animals have developed mechanisms to control fertility for species continuation along with the evolutionary advantages of polyploidy, which do not necessarily need to involve extensive chromosome diploidisation.

All these examples indicate that natural selection works to stabilise the fertility of an organism to ensure its survival. A few organisms already have mechanisms in place to deal with the polyploidisation, while others develop ways during the course of evolution to sustain themselves. In my experiments, the multivalent formation in *A. thaliana* ranged from 15-100% in different F2s across different chromosomes, while in parents its ranged from 13-70% with chromosome 1 always forming highest multivalents. Similarly in *S. tuberosum* it ranged from 14-48% across the two chromosomes and three varieties. While the variety Sante had lower quadrivalent formation for both the chromosomes compared to the other two varieties, it again indicates that diploidisation of chromosomes is variety and chromosome dependent. In future, a phenotypic and cytological comparison can be made between the *Arabidopsis* F2s and potato F1s to understand the differences and similarities in various traits, chromosome behaviour and chiasma frequency between an inbreeder and an outbreeder.

8.4 Cytogenetics, meiotic recombination and its role in sustainable crop breeding and improvement

The Earth's population is increasing at a stark rate, growing by 83 million people annually. It has been projected to reach 8.6 billion in 2030 and almost 10 billion in 2050 (United Nations Department of Economic and Social Affairs Population Division, 2017). Global climate change impacts agriculture at an unprecedented level with several regions experiencing drought and/or flood affecting the land's production ability. It has an adverse impact on biodiversity, soil and water resources, thus endangering food security. It is therefore imperative that breeders employ innovative solutions to increase productivity without increasing the use of the resources available (McKersie, 2015). Meiotic recombination or crossover formation in meiosis is at the heart of plant breeding and crop improvement, as the process can create novel allele combinations desired by breeders. Breeding novel varieties resilient to climate changes and sustainable in long term is the way forward. My project has contributed important knowledge towards this goal by testing if polyploidisation could cause an increase in the frequency of meiotic recombination. Meiotic recombination in tetraploids is clearly higher than in diploids in both *Arabidopsis thaliana* and potato, though the increase does not go beyond what is expected based on a doubling of the chromosome number. Any change either in the frequency or in the distribution of chiasma along a chromosome can lead to recombination in otherwise non recombining areas with important genes. For example, it is known in cereals that recombination occurs in the distal arms of the chromosomes and proximal areas which contain many important genes do not readily undergo recombination (Higgins *et al.*, 2012). Ways to redistribute recombination in important crops will be welcomed by breeders. In the future, the cytological data collected here along with RAD-sequencing of

the full set of diploid and tetraploid F2 populations, can be used to address whether polyploidisation may redistribute meiotic recombination along the chromosomes.

Cytogenetic analysis using FISH helped to understand the meiotic behaviour of chromosomes in polyploid *A. thaliana* and potato. Polyploid plants often show bigger flowers or leaves as compared to their diploid counterparts (Comai, 2005). This is especially important in garden plants and plants whose vegetative parts such as leaves or flowers are consumed. However, autopolyploids often have issues with fertility due to problems during meiosis. This is important to address, especially in the plants where seeds are the main yield. Considering many important crop plants are polyploids, it is important to understand the process of meiotic recombination and the ways it can be stabilised to ensure higher fertility.

Potatoes are important autopolyploid crops which are clonally propagated. While many ploidy levels exist, the most widely consumed *S. tuberosum* is a tetraploid. Conventional potato breeding involves crossing between two tetraploid cultivars, followed by growing F1 seedlings for tuber production from many crosses, which are grown in subsequent years for selection and finally, a cultivar is released. The process takes 13–14 years (Jansky and Spooner, 2018). The process requires large land and labour resources and lead to a narrow genetic base in the cultivated potatoes. With everchanging climate conditions and great demand from an ever-growing population, it has been realized by the farming and scientific community alike, a need to widen the genetic basis of potatoes by introgressing various useful traits from wild varieties without introducing the linkage drag of the undesirable traits. Since most wild varieties are diploids, it has been suggested that the way to move forward is to create diploid inbred lines

which can produce true potato seeds with higher tuber yields and disease resistance (Jansky and Spooner, 2018). These true breeding diploid cultivars can be selected for tetraploid production which may have superior qualities than the diploids.

Cytogenetics can play a role in sustainable potato breeding. Cytological techniques such as FISH and GISH (Genomic *in situ* Hybridisation) can help to identify the introgression of genetic material from wild variety into cultivated variety. Oligo probes as suggested in Braz *et al.* (2018) can be created for specific regions from the wild varieties to test the introduction of desired genes from them to the cultivated varieties. It was shown in my project that pairing behaviour in M1 can be studied using FISH probes in cultivated potato varieties. Oligo FISH probes could also be used for this purpose for all 12 chromosomes. While the chromosomes in the metaphase are highly condensed in meiosis, pachytene chromosomes can help understand the meiotic behaviour and aberrations in an autopolyploid. Using axis proteins such as ASY1 and ZYP1, as used in my project, along with FISH probes, detailed pairing mechanisms and recombination patterns can be identified in other cultivated potato varieties to help in the breeding programmes.

Anti-crossover pathways have been identified in *A. thaliana* and the orthologues of the genes identified have recently been manipulated in crop plants to increase crossovers (Mieulet *et al.*, 2018). Similarly, recombination rate was modulated in *A. thaliana* by enhancing pro-crossover *HEI10* gene dosage and mutating and reducing anti-crossover factors together (Serra *et al.*, 2018). Immunocytochemistry can be utilized to identify similar proteins in potato, building on the work presented here on ASY1 and ZYP1 proteins. A similar approach could

then be utilized in potatoes to manipulate recombination with a view to enhancing yield and disease resistance.

8.5 Conclusion

This project has successfully characterised the frequency and distribution of meiotic recombination in diploid and autotetraploid genotypes of both *Arabidopsis thaliana* and *Solanum tuberosum* (potato) species. In the future, RAD-seq genotyping of the full *A. thaliana* populations created needs to be completed to analyse the chromosomal distribution of recombination in more detail and enable a QTL mapping study to dissect the genetic basis of phenotypic variation in diploid versus autotetraploid populations.

8.6 References

Beçak, M. L., Beçak, W. and Rabello, M. N. (1966) 'Cytological evidence of constant tetraploidy in the bisexual South American frog Odontophrynus americanus', *Chromosoma*, 19(2), pp. 188–193. doi: 10.1007/BF00293683.

Bomblies, K. *et al.* (2016) 'The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role?', *Chromosoma*, 125(2), pp. 287–300. doi: 10.1007/s00412-015-0571-4.

Braz, G. T. *et al.* (2018) 'Comparative oligo-FISH mapping: An efficient and powerful methodology to reveal karyotypic and chromosomal evolution', *Genetics*, 208(2), pp. 513–523. doi: 10.1534/genetics.117.300344.

Das Choudhury, S., Samal, A. and Awada, T. (2019) 'Leveraging Image Analysis for High-Throughput Plant Phenotyping', *Frontiers in Plant Science*. doi: 10.3389/fpls.2019.00508.

Comai, L. (2005) 'The advantages and disadvantages of being polyploid.', *Nature reviews. Genetics*, 6(11), pp. 836–46. doi: 10.1038/nrg1711.

Enciso-Rodriguez, F. *et al.* (2019) 'Overcoming Self-Incompatibility in Diploid Potato Using CRISPR-Cas9', *Frontiers in Plant Science*. doi: 10.3389/fpls.2019.00376.

Higgins, J. D. *et al.* (2012) 'Spatiotemporal Asymmetry of the Meiotic Program Underlies the Predominantly Distal Distribution of Meiotic Crossovers in Barley', *The Plant Cell*, 24(10), pp. 4096–4109. doi: 10.1105/tpc.112.102483.

Jansky, S. H. and Spooner, D. M. (2018) 'The Evolution of Potato Breeding', in *Plant Breeding Reviews*. John Wiley & Sons, Ltd, pp. 169–214. doi: 10.1002/9781119414735.ch4.

Jones, G. H., Khazanehdari, K. A. and Ford-Lloyd, B. V. (1996) 'Meiosis in the leek (Allium porrum L.) revisited. II. Metaphase I observations', *Heredity*, 76(2), pp. 186–191. doi: 10.1038/hdy.1996.26.

Manrique-Carpintero, N. C. *et al.* (2015) 'Genetic map and QTL analysis of agronomic traits in a diploid potato population using single nucleotide polymorphism markers', *Crop Science*. doi: 10.2135/cropsci2014.10.0745.

McKersie, B. (2015) 'Planning for food security in a changing climate', *Journal of Experimental Botany*, pp. 3435–3450. doi: 10.1093/jxb/eru547.

Mieulet, D. *et al.* (2018) 'Unleashing meiotic crossovers in crops', *Nature Plants*. doi: 10.1038/s41477-018-0311-x.

Pecinka, A. *et al.* (2011) 'Polyploidization increases meiotic recombination frequency in Arabidopsis', *BMC Biology*, 9(1), p. 24. doi: 10.1186/1741-7007-9-24.

Quiros, C. F. (1982) 'Tetrasomic segregation for multiple alleles in alfalfa.', *Genetics*, 101(1), pp. 117–27.

Santos, J. L. *et al.* (2003) 'Partial Diploidization of Meiosis in Autotetraploid', *Genetics*, 165(3), pp. 1533–1540.

Serra, H. *et al.* (2018) 'Massive crossover elevation via combination of HEI10 and recq4a recq4b during Arabidopsis meiosis.', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1713071115.

United Nations Department of Economic and Social Affairs Population Division (2017) 'E02 World Population Prospects The 2017 Revision: Key Findings and Advance Tables', *World Population Prospects The 2017*. doi: 10.1017/CBO9781107415324.004.

Yant, L. *et al.* (2013) 'Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa', *Current Biology*. Elsevier Ltd, 23(21), pp. 2151–2156. doi: 10.1016/j.cub.2013.08.059.

Zhang, Z. *et al.* (2014) 'Preferential bivalent formation in tetraploid male of Pacific oyster Crassostrea gigas Thunberg', *Journal of Ocean University of China*, 13(2), pp. 297–302. doi: 10.1007/s11802-014-2319-9.

APPENDICES

Appendix A

Comparing variance – the measure of dispersion between the two *Arabidopsis thaliana* trials.

A plant trial with 920 plants consisting of a population of 391 diploid and tetraploid F2 and 23 each of parental lines was undertaken in summer 2015. Various phenotype traits as described and defined in 2.3 were collected. On analysing a few traits, it was found that the variance in the F1 for various traits was larger than the variance in the F2 as seen in **Table A-1** and **Table A-2**. Three traits DTG, DTF1 and CLN had F1 variance greater than the F2 variance for both diploid and tetraploid F2 population. For DTF3 F1 variance was larger than the F2 variance for tetraploid F2 population only, while LC and RP had larger F1 variance in diploid F2 population. Since F1 population is genetically similar, any difference in the phenotypic trait variance for F1 is considered as the environmental variance, which should be lower than F2 since the environment is maintained and therefore should be same for all the plants growing in the glass house.

	DTG	DTF1	DTF2	DTF3	RLN	CLN	TLN	LB
A (F2D)	0	1.8	2.82	3.73	0.98	0.27	1.63	0.27
B (F2T)	0.0026	0.8	4.03	1.28	0.67	0.22	1.06	0.22
C (ColD)	0	0.2	0.36	0.81	0.31	0.27	0.44	0.27
D (ColT)	0	1.27	0.99	1.08	0.95	0.26	0.98	0.26
E (LerD)	0	0	0.36	0.8	0.3	0.24	0.53	0.24
F (LerT)	0	1.81	1.1	2.9	0.54	0.26	1.18	0.26
G (F1D)	0.24	166.49	0.35	0.8	0.6	0.26	1.09	0.26
H (F1T)	0.2	1.85	1.66	3.54	0.4	0.13	0.66	0.13

Table A-1 Variance of various phenotypic traits collected during first A. thaliana trial in 2015.

Yellow highlight indicates the F1 variance to be greater than the respective F2 variance.

	BB	TB	LC	RP	PL	FERT1	FERT2
A (F2D)	0.76	1.23	59.62	60.9	102.6	2.24	26.04
B (F2T)	0.98	1.34	45.04	45.5	83.5	3.06	50.87
C (ColD)	0.61	1.15	36.45	38.2	6.03	NA	NA
D (ColT)	0.68	1.21	38.13	35.4	120.1	NA	NA
E (LerD)	0.08	0.26	30.43	30.43	3.92	NA	NA
F (LerT)	0.44	0.47	22.83	26.7	12.29	NA	NA
G (F1D)	0.39	0.44	93.2	98.1	26.32	0.21	0.85
H (F1T)	0.49	0.69	42.1	41.53	52.1	1.77	16.82

Table A-2 Variance of various phenotypic traits collected during first A. thaliana trial in 2015.

Yellow highlight indicates the F1 variance to be greater than the respective F2 variance.

There was a fly infestation in the glasshouse, which affected the plant growth and therefore an adequate amount of leaf samples could not be collected. Buds for cytology were not collected for the same reason. Therefore, a second trial was conducted in 2016 (Chapter 4). Though, the problem of larger F1 variance could not be solved entirely, but it was only seen for a fewer traits compared with the first trial and only for tetraploids as seen in **Table A-3** and **Table A-4**.

	DTG	DTF1	DTF2	DTF3	RLN	CLN	TLN	LB
F2D(A)	34.71	30.74	33.73	30.63	134.46	2.51	163.36	2.28
F2T(B)	3.32	28.91	21.86	24.13	124.02	1.11	135.14	1.23
ColD(C)	4.58	38.67	41.77	40.23	71.82	1.18	82.83	1.38
ColT(D)	16.34	38.79	45.95	50.79	108.11	0.84	110.66	0.82
LerD(E)	7.02	13.49	15.74	12.33	32.15	0.81	39.19	0.81
LerT(F)	24.37	47.00	54.51	54.35	93.28	1.81	104.00	1.28
F1D(G)	34.36	17.82	16.97	17.18	92.00	2.31	111.92	1.91
F1T(H)	28.19	29.68	30.43	60.42	72.39	1.09	82.86	1.17

Table A-3 Variance of various phenotypic traits collected during second A. thaliana trial in 2016.

Yellow highlight indicates the F1 variance to be greater than the respective F2 variance.

	BB	TB	RP	LC	FERT1	FERT2
F2D(A)	1.99	4.24	133.78	152.26	2.94	80.75
F2T(B)	0.95	1.83	119.77	122.87	1.99	38.74
ColD(C)	1.49	3.00	115.71	109.46	1.16	30.68
ColT(D)	0.48	1.21	129.16	134.36	1.96	132.05
LerD(E)	2.30	4.67	98.51	120.09	1.68	36.75
LerT(F)	3.00	4.65	131.82	106.06	1.24	24.2
F1D(G)	0.50	3.41	128.77	136.89	1.02	25.81
F1T(H)	0.37	1.56	119.80	107.43	1.79	44.69

Table A-4 Variance of various phenotypic traits collected during second A. thaliana trial in 2016.

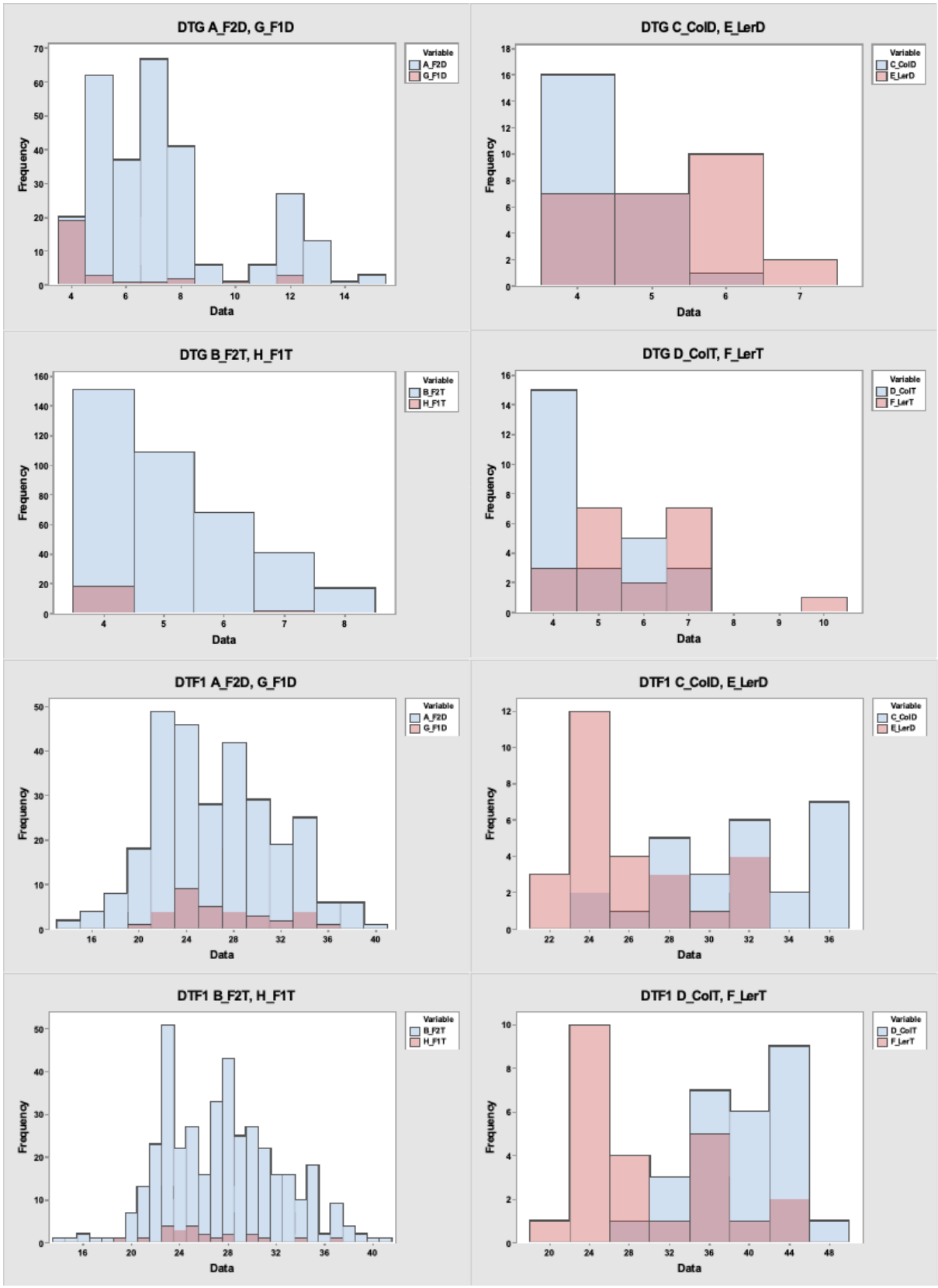
Yellow highlight indicates the F1 variance to be greater than the respective F2 variance.

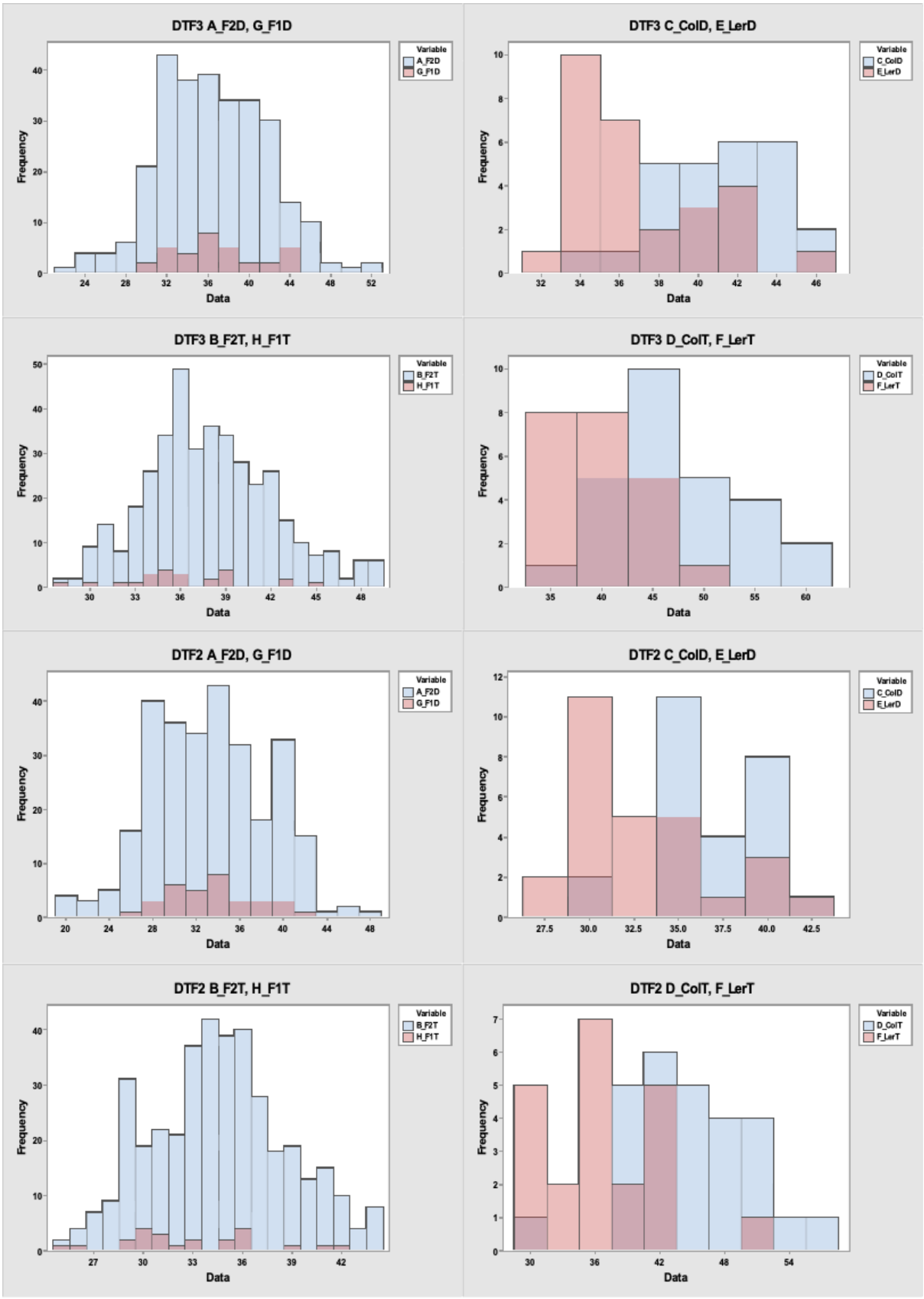
On comparing second trial with the first, it can be seen than the larger F1 variance than the F2 variance was now seen in DTG, DTF1, DTF2, DTF3 and RP for tetraploid F2 only, whereas it was there for more traits and for both diploid and tetraploid population in the first trial. The plants in the second trial were healthy and adequate leaf samples as well as buds could be collected for sequencing and cytological analysis.

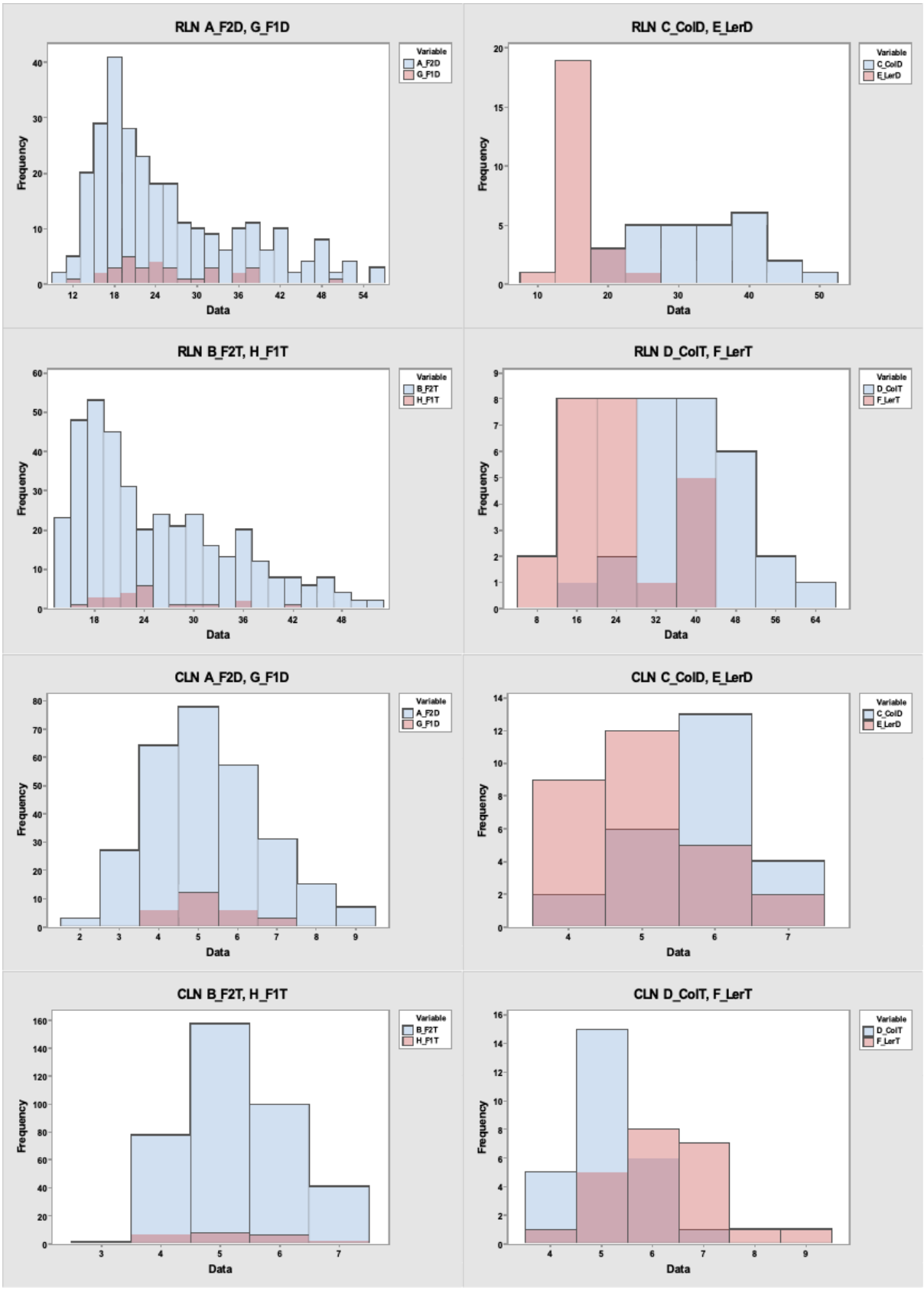
Appendix B

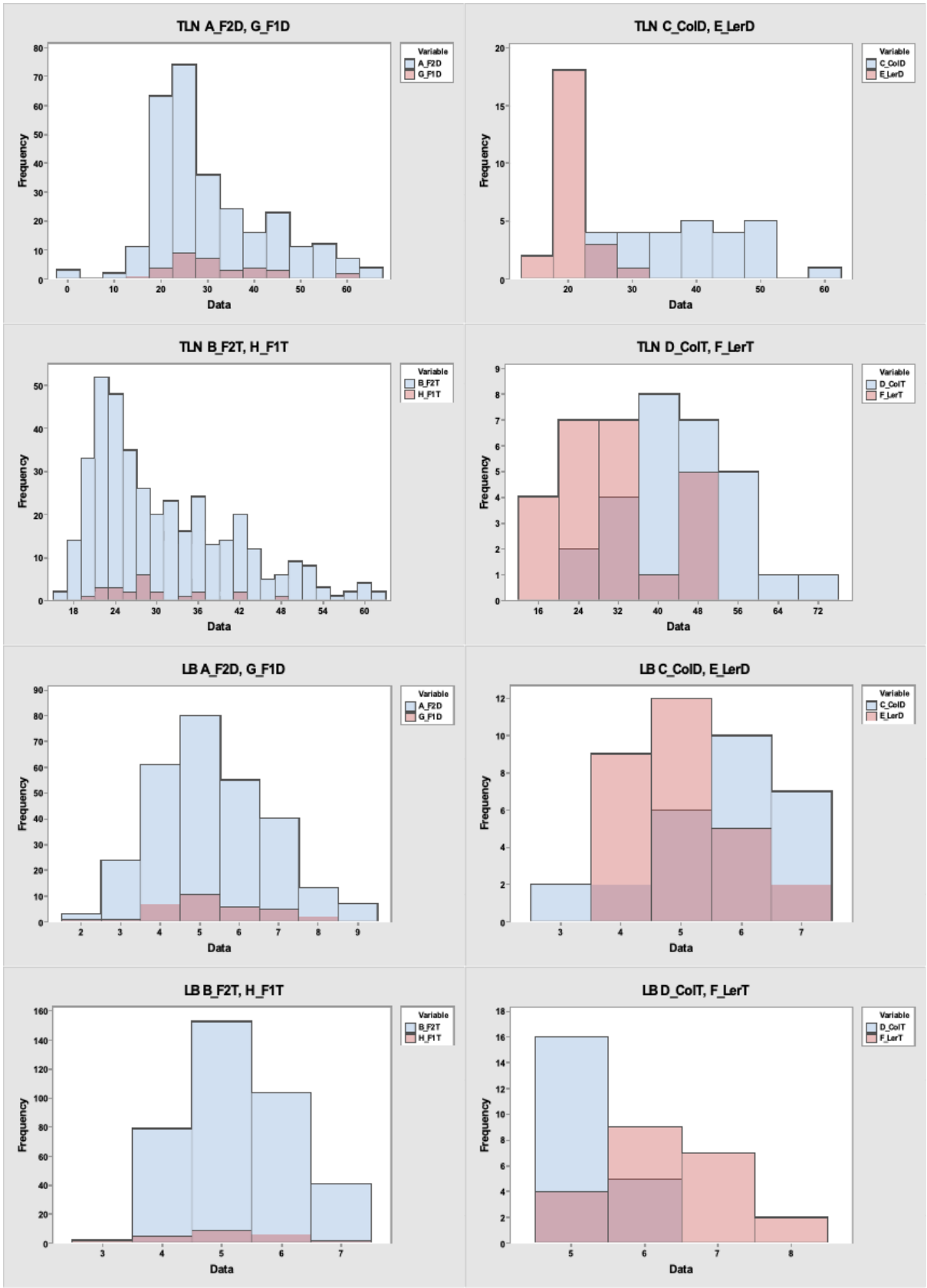
Distribution of trait data collected during second *Arabidopsis thaliana* trial after removing outliers.

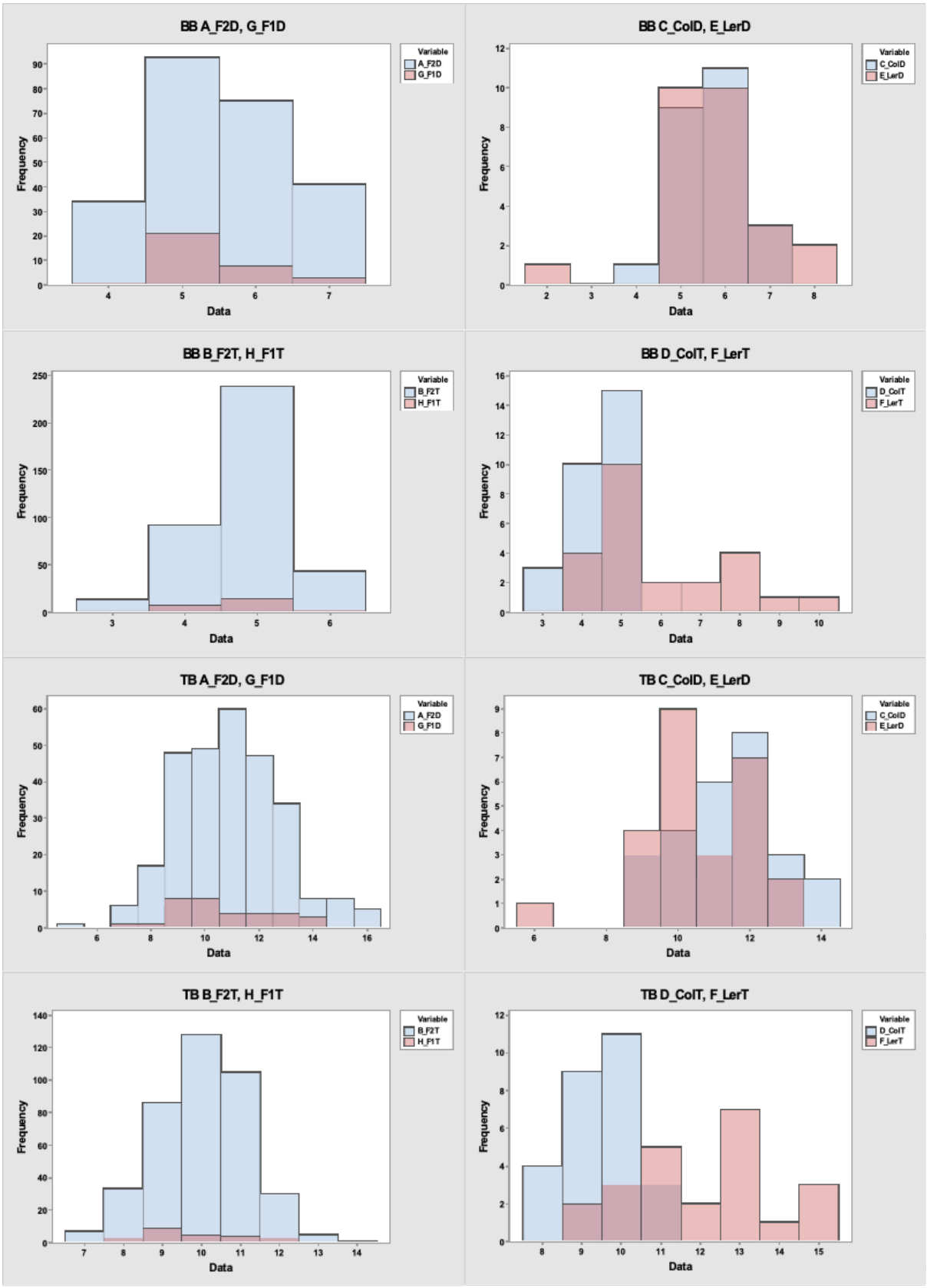
Distribution of each trait after removing outliers from the trait data collected during second *Arabidopsis thaliana* trial, was compared with the distribution of the traits with complete data. Histograms comparing diploid and tetraploid F2 with F1 and parental lines for different traits after removing outliers are shown in **Figure B-1.** As with complete data (**Figure 4 4**), here as well for days to germination (DTG) F2 are more variable than F1, and germination in Landsberg is more variable than Columbia in both diploids as well as tetraploids. The flowering distribution (DTF1, DTF2 and DTF3) all follow the same distribution pattern as with the complete data with F2 being more variable and Landsberg flowering earlier than Columbia. Similarly, the F2 leaf distribution for rosette leaf is right tailed, whereas it is symmetrical for cauline leaves. Overall, Columbia parents produced more leaves and more uniform distribution than the Landsberg parent as was seen in the complete data. Distribution across different F2 population and parental lines for reproductive period (RP), life cycle (LC), silique length (FERT1) and seed number (FERT2) follows the same pattern as the complete data set shown in **4.1.2.1.** The removal of outliers did not produce any change in the trait data distribution for all the traits. Therefore, it was considered unnecessary to remove the outliers and the phenotype data analysis was done with complete data set for all the traits.

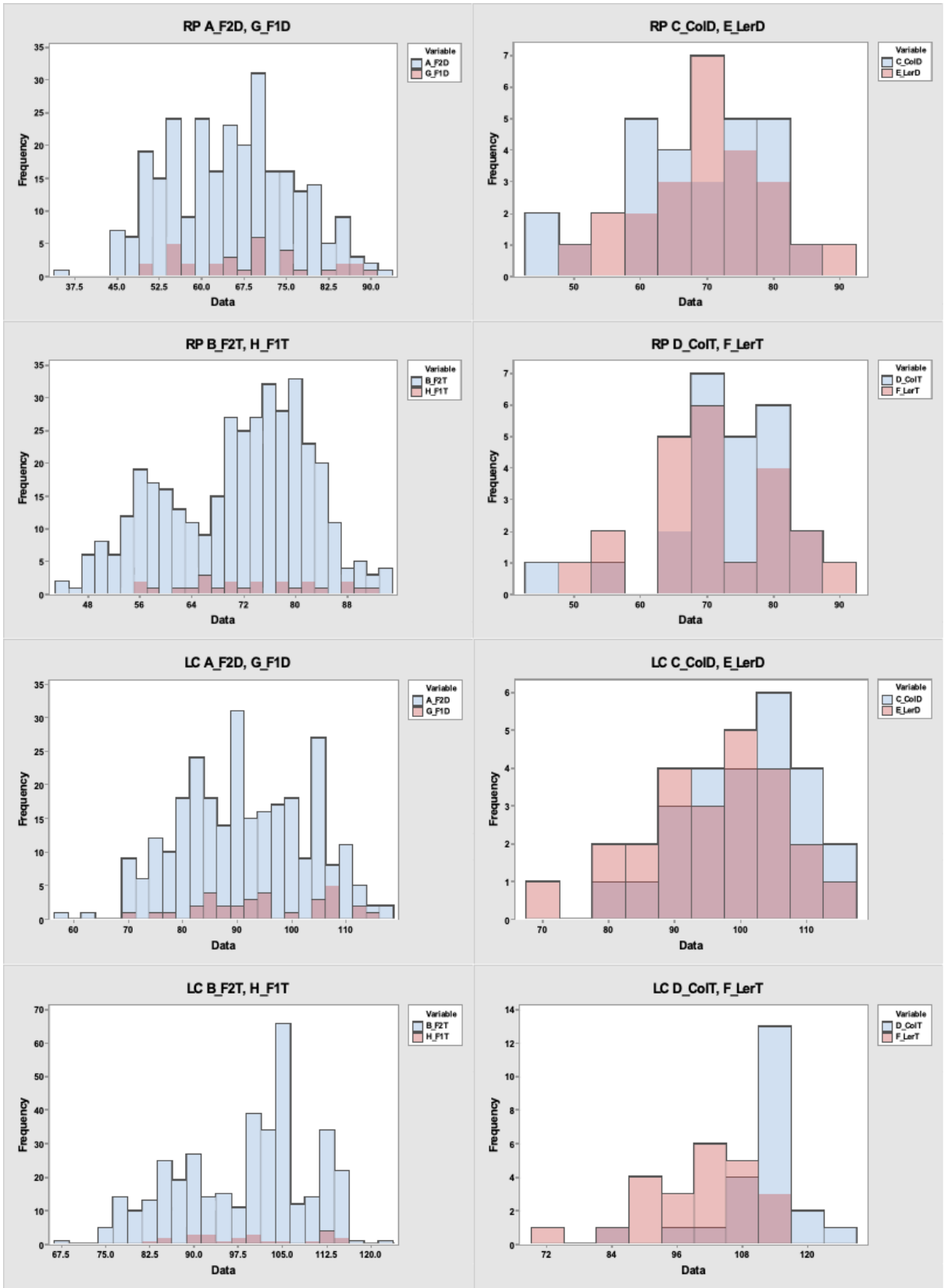

DTG A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
DTG C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
DTG B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
DTG D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data
DTF1 A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
DTF1 C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
DTF1 B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
DTF1 D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data


DTF3 A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
DTF3 C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
DTF3 B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
DTF3 D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data
DTF2 A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
DTF2 C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
DTF2 B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
DTF2 D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data


RLN A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
RLN C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
RLN B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
RLN D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data
CLN A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
CLN C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
CLN B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
CLN D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data


TLN A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
TLN C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
TLN B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
TLN D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data
LB A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
LB C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
LB B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
LB D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data


BB A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
BB C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
BB B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
BB D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data
TB A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
TB C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
TB B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
TB D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data


RP A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
RP C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
RP B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
RP D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data
LC A_F2D, G_F1D
Variable
A_F2D
G_F1D
Frequency
Data
LC C_ColD, E_LerD
Variable
C_ColD
E_LerD
Frequency
Data
LC B_F2T, H_F1T
Variable
B_F2T
H_F1T
Frequency
Data
LC D_ColT, F_LerT
Variable
D_ColT
F_LerT
Frequency
Data

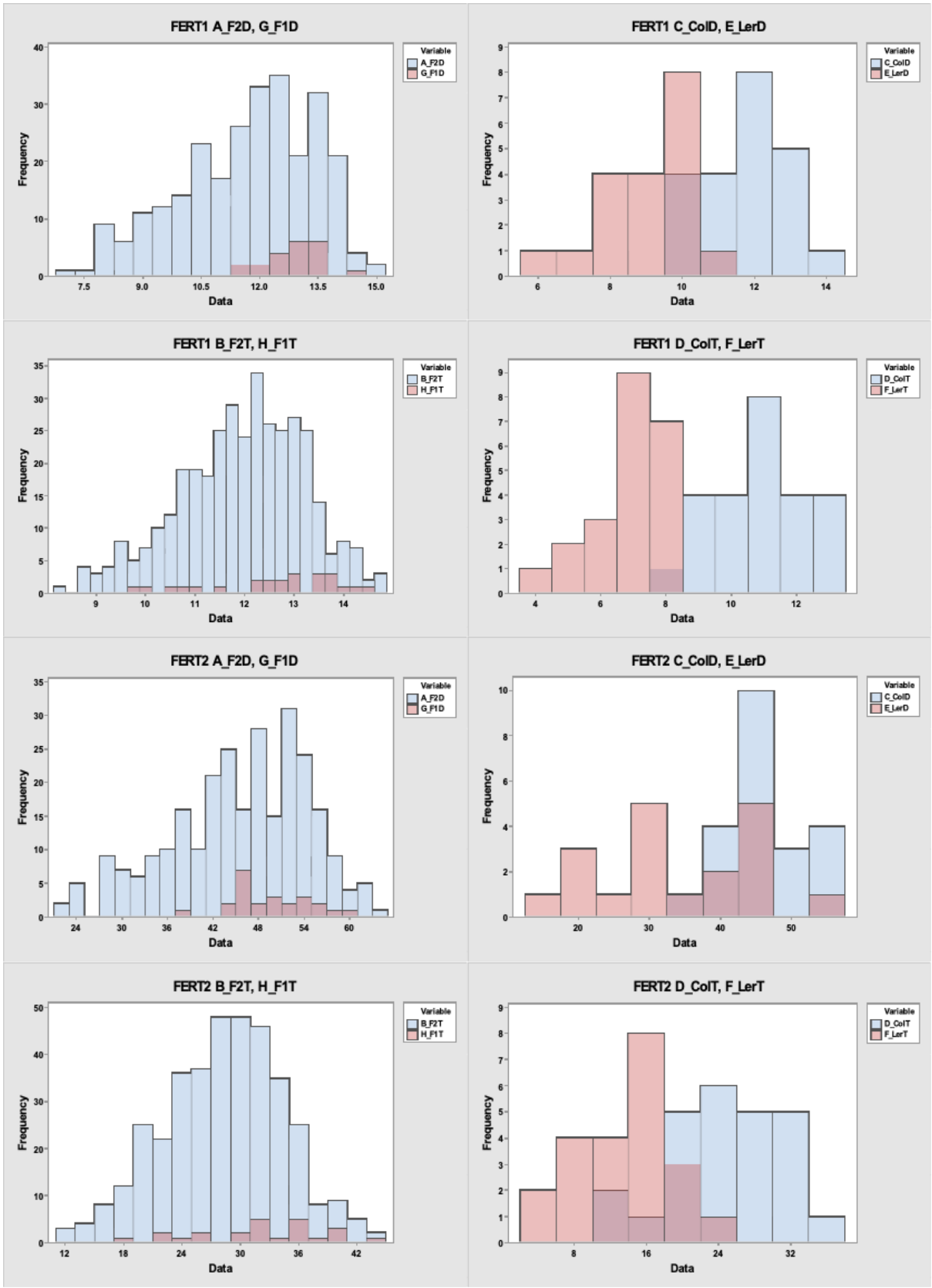


Figure B-1 Histograms showing distribution of diploid and tetraploid F2s with F1s and the parental lines for different traits after removing outliers.

A_F2D and B_F2T represent diploid and tetraploid F2, C_ColD and D_ColT represent diploid and tetraploid Columbia parent, E_LerD and F_LerT are diploid and tetraploid Landsberg parent, G_F1D and H_F1T are diploid and tetraploid F1 respectively.

Appendix C

Quantity and quality of the DNA extracted from leaf samples of *Arabidopsis thaliana* F2 and parental population, grown in second trial for RAD sequencing .

DNA for RAD-Seq was extracted from 209 tetraploid and 203 diploid F2 and 8 parental *Arabidopsis thaliana* leaf samples collected during 2nd trial in 2016 as shown in **Table C-1**, **Table C-2** and **Table C-3**.

4n F2	Amt (ng/ µl)	4n F2	Amt (ng/ µl)	4n F2	Amt (ng/ µl)	4n F2	Amt (ng/ µl)	4n F2	Am (ng/ µl)	4n F2	Amt (ng/ µl)	4n F2	Amt (ng/ µl)
1	58.8	99	58.6	321	22	438	17.9	518	25.4	593	39.4	748	48.8
7	41	106	18.68	322	17.72	441	46	519	34.8	594	19.74	750	52.8
11	40.4	109	23.2	329	30.4	452	14.1	523	32.6	595	14.56	761	46.4
18	44	124	44.4	330	31.2	456	13.74	524	20.4	596	59.2	771	52.6
19	36.2	127	48	332	52.4	462	40.6	525	35.8	605	29	775	27.6
22	38.4	135	39.4	349	21.2	463	12.42	527	32	612	31.4	777	47.2
24	40.8	136	12.48	352	26.4	466	19.58	530	34.6	614	32.6	779	51.4
30	49.2	138	27	359	27.6	467	15.62	532	30.6	624	35.6	780	36.6
31	48.6	139	28	364	66.2	468	25	533	31	630	28.2	784	37.8
31	48.6	150	30.6	373	42.4	469	53.6	542	34.4	631	29.4	791	34.2
38	43.6	153	30.6	374	36.2	470	23.2	543	34.2	632	20.4	793	33
40	29.4	157	31.2	377	37.8	471	42.2	545	34.6	636	39.8	794	63
43	41.4	165	30	384	36.4	473	9.8	547	31	645	28.8	803	16.66
44	35.4	172	28	389	39.6	476	53.4	549	26.4	650	25.4	837	23.6
46	49	174	40	392	58.4	477	28	550	36.6	666	25.6	844	26.2
49	49	175	20.6	393	35.8	481	38	554	33.2	670	45.8	856	32.8
51	36.8	178	70.6	395	35.2	483	41.8	557	43	673	33	859	39
54	50.2	228	44	404	41.6	487	39.6	559	22	674	24.4	862	38
55	39.6	229	30.6	408	29.4	490	30.8	560	19.3	675	33	869	42.2
56	63.4	252	24.4	410	36.2	492	32	561	29.2	681	39.2	874	34
62	50.6	267	38.4	411	23.8	492	32	563	16.1	682	32.2	877	29.6
63	31.4	275	37	412	19.06	496	30.2	564	14.92	698	49.6	882	27.4
69	34.4	285	33.4	413	28.4	497	41.6	565	23	701	32.2	891	34.4
79	33.4	288	25.2	416	24.6	500	27.2	566	41.6	712	56	893	25.6
80	31	289	55.6	420	25	503	11.08	573	28.2	716	38.4	895	33.6
81	27.8	291	19.82	423	33.8	507	38	579	20.2	719	20.6	899	28
85	26.6	294	17.52	426	29.2	511	35.8	581	29.2	725	21.8	902	31.2
91	45.8	301	49	429	15.4	513	27.4	583	18.24	727	45	911	69
94	55.6	303	18	434	32	516	25.4	586	25.8	729	30.4	914	34.8
96	33	317	26.2	436	11.2	517	30.2	590	23.8	735	65.4		

Table C-1 Sample number and the amount of DNA extracted (in ng/µl) from 4n F2 samples.

2n F2	Amt (ng/ µl)	2n F2	Amt (ng/ µl)	2n F2	Amt (ng/ µl)	2n F2	Amt (ng/ µl)	2n F2	Am (ng/ µl)	2n F2	Amt (ng/ µl)	2n F2	Amt (ng/ µl)
2	31	98	19.2	181	32.6	251	39.4	485	32	672	32.6	832	40
3	29.2	101	55.8	183	26.4	272	20.8	491	59.6	692	30.2	833	32.6
4	38.8	104	53.4	184	43.6	273	56	493	46.2	707	49.2	836	19.9
5	48.6	105	34.8	186	33.8	280	48.6	502	37.2	714	36.4	841	45.8
13	50	107	50.8	188	33	307	42.8	505	35.2	717	42	842	24.8
17	41.6	108	31.4	192	45.8	316	45.8	506	32.6	722	37.6	848	34.2
20	43.8	110	46.4	194	17.78	325	32.6	510	34.8	733	24	849	34.6
21	48.4	112	60	195	37.4	334	37.6	522	24.8	734	30.8	866	26.2
23	54.6	114	42.6	197	35.6	336	35.6	529	42.2	736	28.6	888	36.8
26	53.4	119	50.6	200	34.8	340	28	548	34.4	741	43	889	28.6
33	42.8	122	26.2	204	28.4	341	38.8	551	30.6	742	32	892	40
39	27.2	130	30.4	205	11.94	344	42.4	552	32.4	743	37	896	32.6
42	67	131	33.4	206	18.62	346	42.8	558	32.4	745	23	897	36.2
45	56.8	133	35.8	207	43.8	356	29	567	32.8	749	59.8	901	50
50	45.4	134	43.8	212	53.8	368	41	572	28.6	760	45.8	903	44.4
52	29.6	137	27.6	214	35.2	369	50.2	574	25.8	767	40	909	29
53	36.6	140	23.2	215	40.8	371	26	575	38.8	768	33.2	912	43.8
57	52	141	32.4	217	36.6	372	31	580	41.4	776	42	918	52.8
59	29.2	142	48.4	221	22.8	385	43.8	585	20.6	778	13.82	974	24.6
61	30.4	145	36.4	226	27.6	386	66.8	589	37	785	38.6	976	39.8
71	65.6	146	32.4	227	22.4	391	10.28	608	29.2	786	23.8	977	38.2
72	68.6	147	17.46	230	50.2	398	41.4	611	36.2	792	31	979	28.8
74	41.4	151	61	232	9.06	399	27.2	619	55.2	800	36.8	980	46.8
75	47.4	159	32.8	233	30.2	400	33.6	622	39	805	47.8		
77	41.2	160	23.8	235	29.2	401	35.4	628	22.2	815	37.6		
82	49.4	163	44.6	241	29.4	406	28.2	629	22.2	819	20.8		
86	55.4	164	44.2	245	40	415	35.2	639	27.6	823	33		
88	36.4	167	30.8	248	63	431	31.4	649	28.8	824	24		
95	44	171	47.6	249	37.8	435	31.6	657	48	828	28.6		
97	38.4	173	35.2	250	35.8	450	31.2	671	28.8	831	36.4		

Table C-2 Sample number and the amount of DNA extracted (in ng/µl) from 2n F2 samples

Sample	Amt (ng/µl)	Sample	Amt (ng/µl
25 (ColD)	29.6	703 (F1T)	25.6
76 (ColT)	33.6	845 (LerT)	16.35
437 (Ler D)	16.78	927 (F1D)	14.18
501 (F1T)	26	930 (F1D)	11.98

Table C-3 Sample number and the amount of DNA extracted (in ng/µl) from parental samples.

ColD and ColT represent diploid and tetraploid Columbia samples, LerD and LerT represent diploid and tetraploid Landsberg samples, F1D and F1T represent diploid and tetraploid F1 samples.

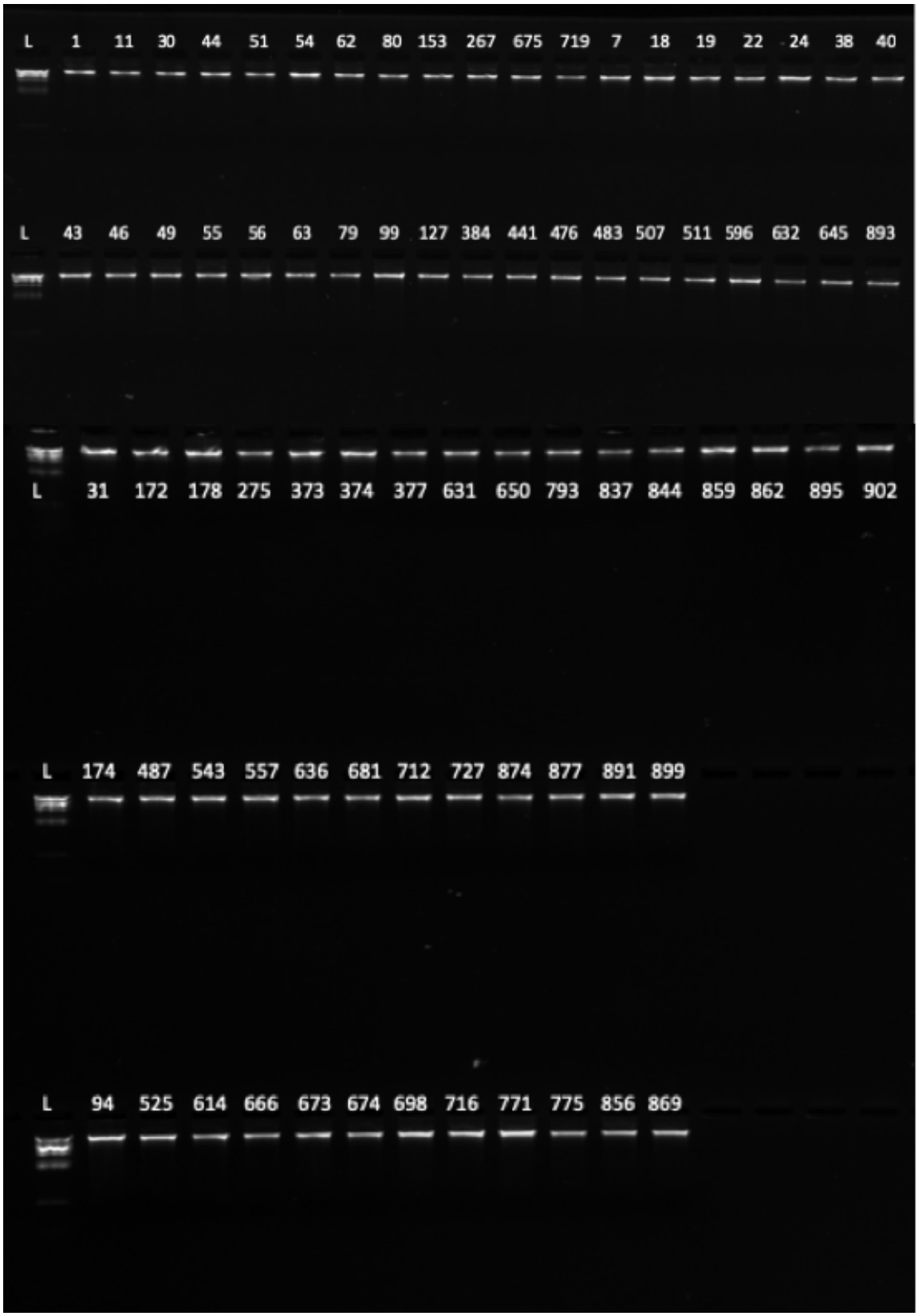


Figure C-1 1% agarose gel to check the quality of DNA extracted from 4n F2 samples.

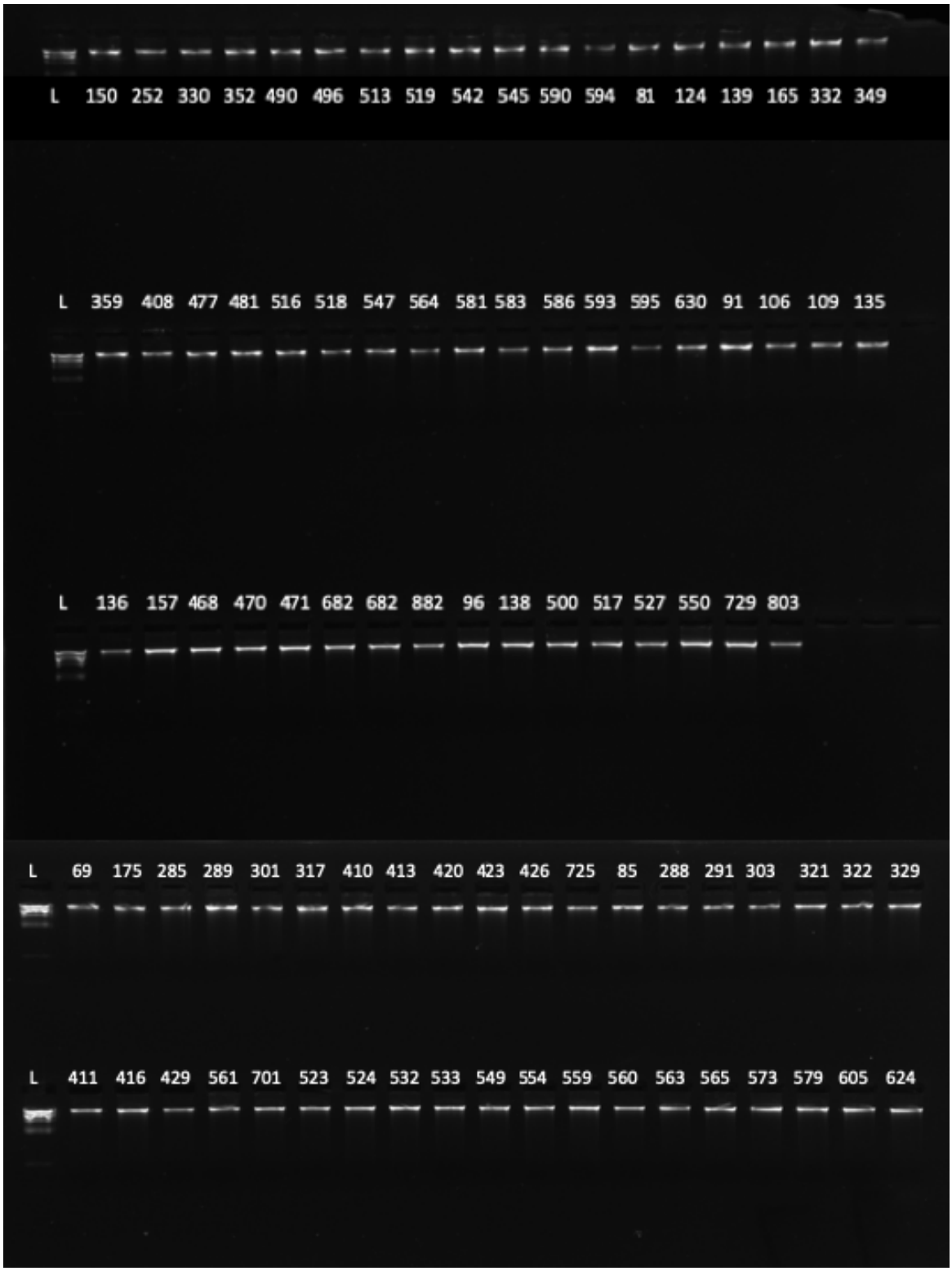


Figure C-2 1% agarose gel to check the quality of DNA extracted from 4n F2 samples.

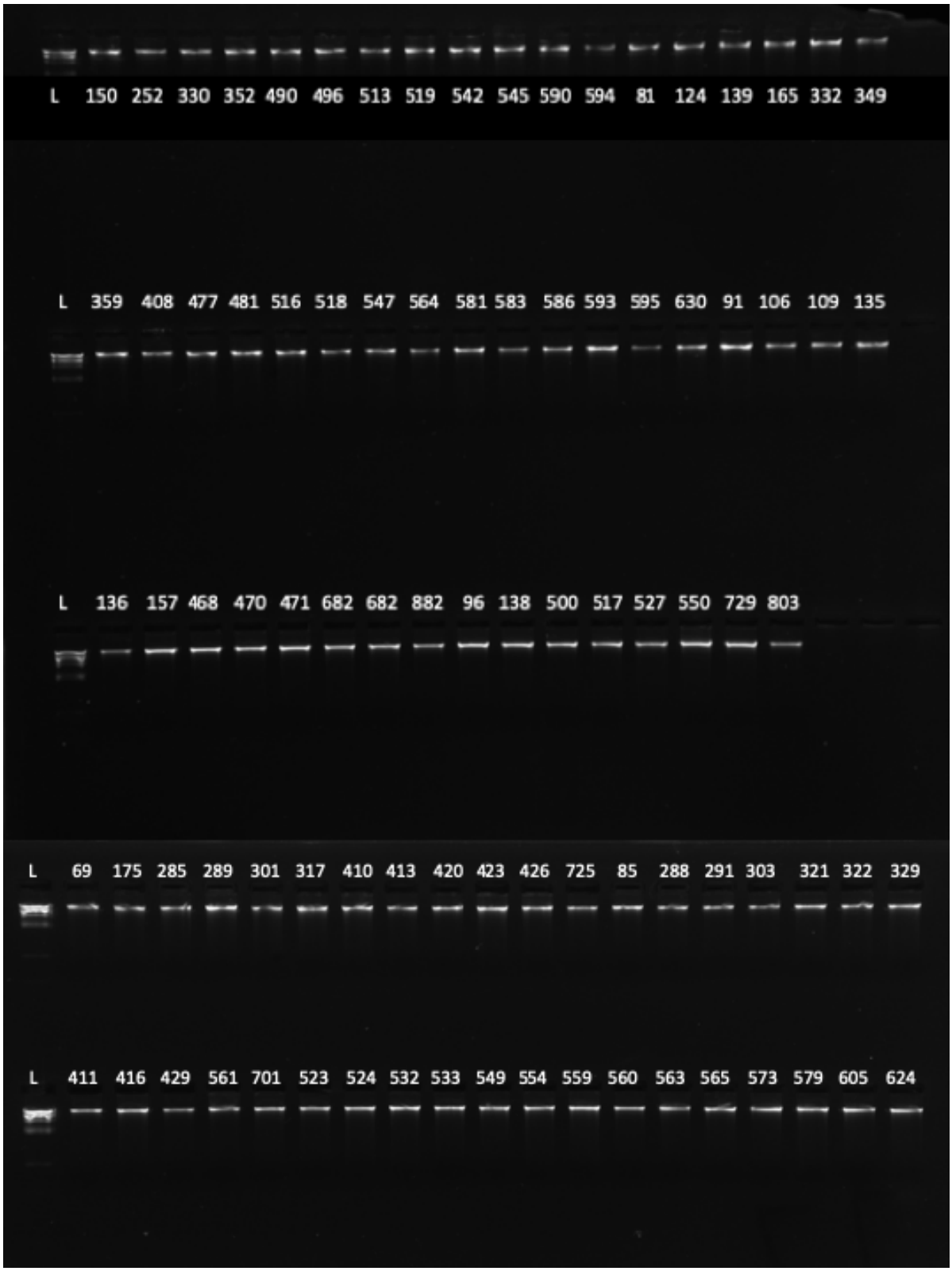


Figure C-3 1% agarose gel to check the quality of DNA extracted from 4n F2 samples.

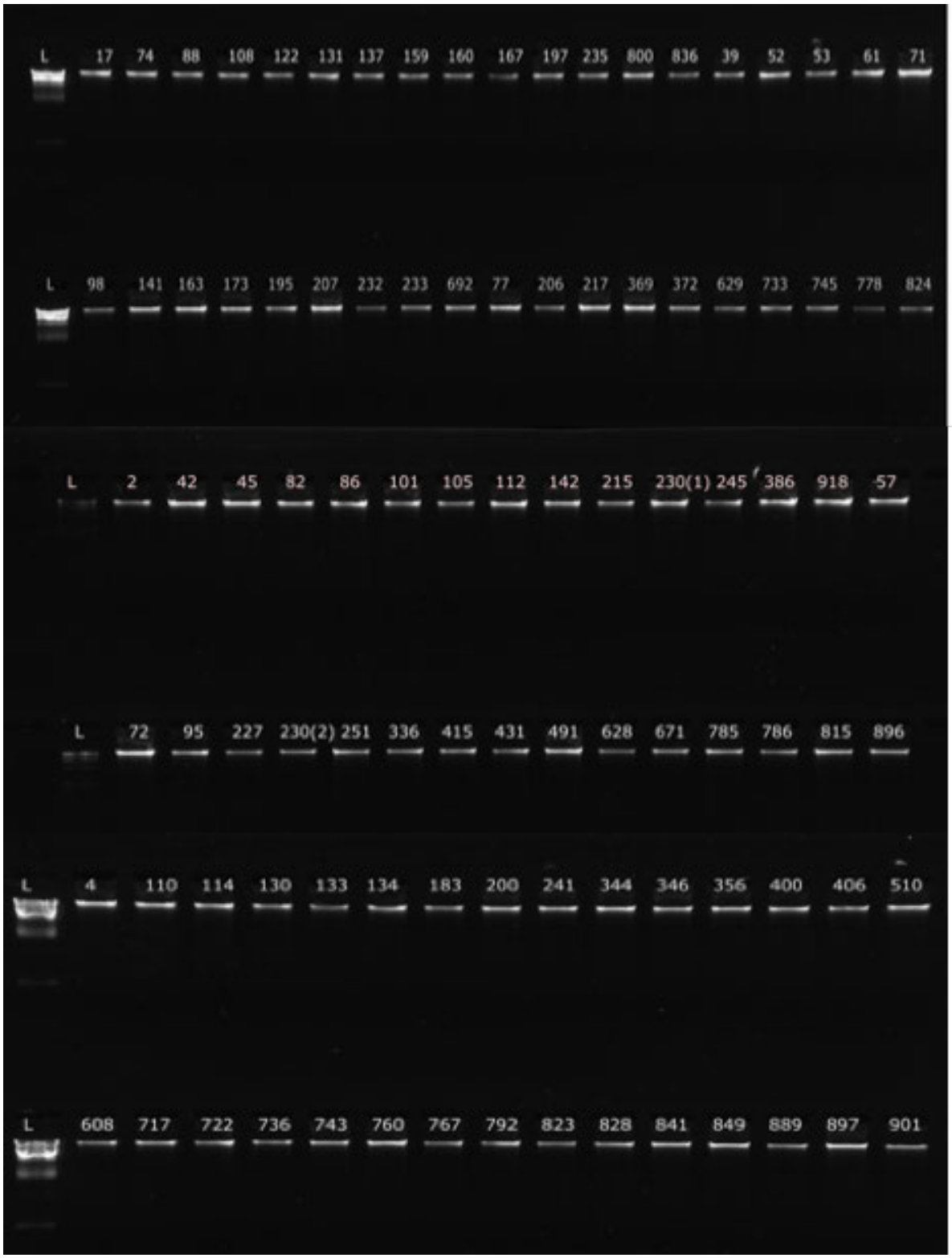


Figure C-4 1% agarose gel to check the quality of DNA extracted from 2n F2 samples.

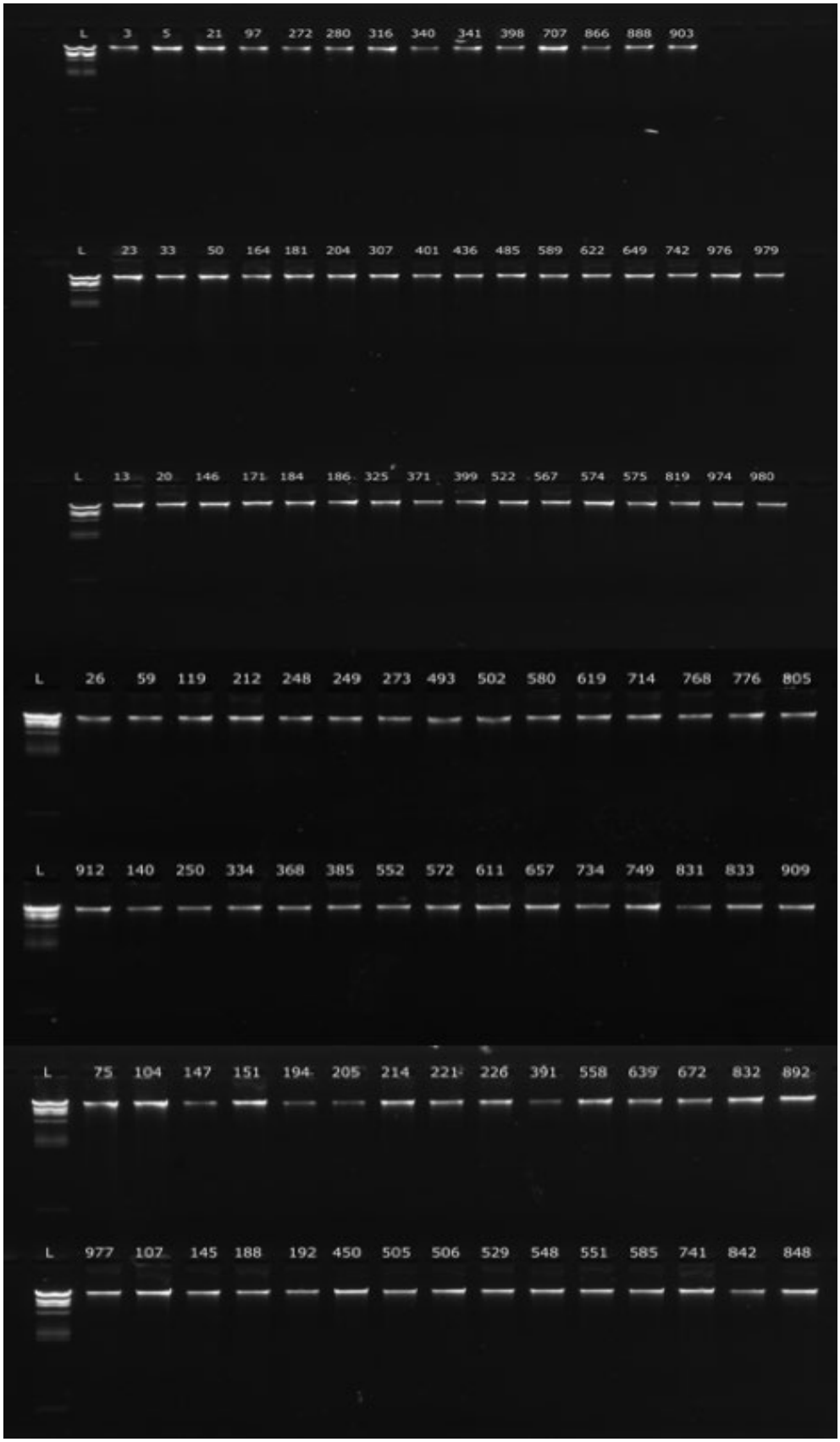


Figure C-5 1% agarose gel to check the quality of DNA extracted from 2n F2 samples.

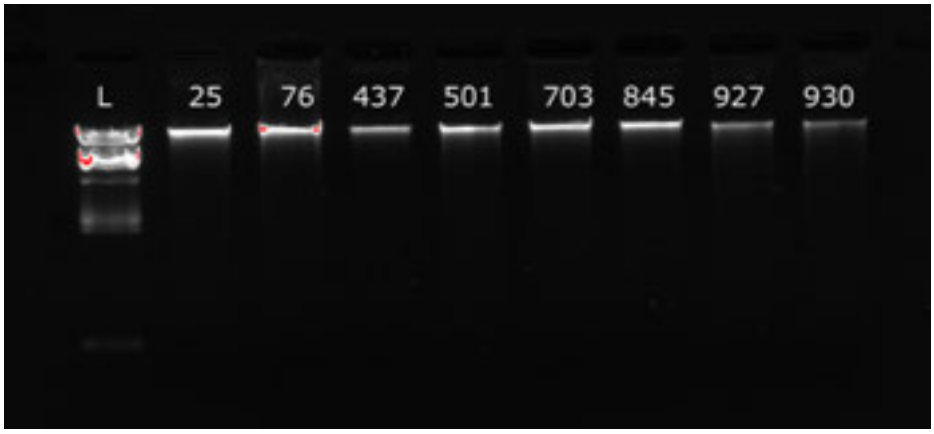


Figure C-6 1% agarose gel to check the quality of DNA extracted from parental samples.

Quality of the samples extracted was checked by running the extracted DNA on 1% agarose gel as shown in **Figure C-1**, **Figure C-2**, and **Figure C-3** for tetraploid F2s, **Figure C-4**, and **Figure C-5** for diploid F2s, and **Figure C-6** for parents. After confirming the quantity and quality of the samples extracted, they were sent to China for further processing for RAD-sequencing.