

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

POLYPLOID GENOME EVOLUTION

A thesis presented in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

**Institute of Fundamental Sciences
Massey University
Palmerston North
New Zealand**

Tina Sehrish

2014

ABSTRACT

Genome duplication is a major force influencing plant genome evolution. Many plant species have shown multiple rounds of whole genome duplications in the past. Duplicated genes show variable rate of retention, silencing, subfunctionalization and neofunctionalization which are pronounced outcomes of genome duplication. This thesis addresses polyploid genome evolution focusing on the genetic and epigenetic consequences of genome duplication. *Tragopogon dubius*, *T. pratensis* and *T. porrifolius* (diploid progenitors) and their polyploids *T. miscellus* and *T. mirus* were employed as an ideal system to examine the outcomes of polyploidy. An investigation of cytonuclear coordination in *T. miscellus* polyploids showed a maternal influence which was evident from the biased retention and expression of the maternally inherited homeolog of *rbcS* possibly to facilitate its interaction with the maternally derived *rbcL* in independently formed *T. miscellus* natural polyploids. The second study involved the genetic characterization of synthetic *T. miscellus* and *T. mirus* polyploids in the context of their relationship with each other. Results showed the presence of the same multilocus genotypes reported previously in natural *T. miscellus* and *T. mirus* and also suggested that there are certain genetic rules to the formation of polyploids; that is, only some progenitor genotypes are successful in producing polyploids. In the third study, a comparative transcriptome analysis of the reciprocally formed synthetic and natural *T. miscellus* polyploids was conducted. This study demonstrated additivity in the expression of progenitor orthologs of floral identity genes in reciprocally formed *T. miscellus* polyploids, suggesting other genetic factors are responsible for the differing inflorescence and flora morphologies in *T. miscellus*. The fourth study explored the epigenetic consequences of polyploidy. The DNA methylation status of homeologous loci previously reported to be silenced in *T. miscellus* natural polyploids was investigated. This study revealed silencing of two out of five homeologous loci by DNA methylation, suggesting

other mechanisms may be responsible for silencing of the remaining three homeologous loci. In short, collectively these studies significantly contribute to our knowledge of polyploid genome evolution in *Tragopogon* in particular and in plants in general.

ACKNOWLEDGEMENTS

This thesis marks the end of my journey of perusing a Doctoral degree. Though only my name appears on the cover of this dissertation, many people have contributed to its production and completion. I owe my deepest gratitude to Almighty Allah Who has complete knowledge about everything, for blessing me a smart brain and enabling me achieving my big goal of life i.e. PhD. I couldn't see Him but I've felt him everywhere and in every difficult moment of my life, standing to guide and help me.

I would never have been able to finish my dissertation without the excellent guidance, patience, kind help and a great support of my supervisor Dr. Jennifer A. Tate, an always welcoming, smiling and my favourite personality. I always found her there whenever I came across a research question, need technical assistance in planning an experiment and/or to discuss results or even share personal experiences of expecting and raising a baby along with research. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own and at the same time the guidance to keep me on the right track. I feel proud in saying that I was no less than a lucky person for working with such a motivated, encouraging and a great scientist.

Dr. Vaughan Symonds, my co-adviser, has also been a great help throughout my degree and I am fan of his critical but insightful comments and constructive criticisms at different stages of my research. It was a great opportunity to work with him who made understanding of population genetics and all the involved experimentations and their results a trivial thing for me.

I take this opportunity to sincerely acknowledge the Higher Education Commission (HEC) Pakistan, for supporting my stay and studies at Massey University, New Zealand. Massey University exclusively Institute of Fundamental Sciences (IFS) deserves my sincere expression of thanks for providing highly equipped lab facilities, qualified staff and friendly & research conducive environment.

I am indebted to all my lab fellows especially Cindy Skema for taking me through high quality RNA extractions needed for MiSeq & HiSeq RNA sequencing; Prashant Joshi for technical and lab help and Rebecca Bloomer for her moral and academic support throughout my PhD candidature.

I find no words to express my gratitude to my parents for their love, care, time, patience, moral & financial support and encouragement to do more and more. Whatever I am today is the result of continuous support, dedication and prayers of my parents. I can't pay you back for the care which you have taken for my daughter "Ayesha" up to first nine months after her birth.

I am highly grateful to Muhammad Faisal, my husband and also a Ph.D scholar was the greatest motivation for my higher studies. His intellectual thoughts, love for science and encouragement always helped me to achieve my dream of life. I appreciate his patience and permission to leave him for the first year of my studies in New Zealand.

I find no one deserving for dedication this dissertation more than my cuddly, fairy and the loveliest daughter "Ayesha Bint e Sehrish Faisal" Who couldn't enjoy the warmth of her mother for first nine months for being away from me and missing my company, love, care, and proper time that she deserved.

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES.....	viii
LIST OF TABLES	x
LIST OF ABBREVIATIONS.....	xi
CHAPTER ONE.....	1
1. Introduction.....	2
1.1 Polyploidy.....	2
1.2 Types of polyploids.....	2
1.3 Chromosomal pairing behavior of polyploids.....	3
1.4 Formation of polyploids.....	4
1.5 Phenotypic consequences of polyploidy	5
1.6 Genetic and epigenetic consequences of polyploidy	6
1.6.1 Change in genome size.....	7
1.6.2 Gene loss	8
1.6.3 Changes in the transcriptome and proteome	9
1.6.4 DNA methylation.....	11
1.6.5 Transposon activation, small RNAs and RNAi.....	12
1.7 Synthetic allopolyploid lines as a useful resource.....	13
1.8 <i>Tragopogon</i> as a study system	15
1.9 Thesis chapters	19
1.9.1 Cytonuclear coordination in <i>T. miscellus</i> polyploids (Chapter 2).....	19
1.9.2 Exploring genetic structure of <i>T. mirus</i> and <i>T. miscellus</i> synthetic polyploids (Chapter 3).....	19
1.9.3 Comparative analysis of floral transcriptomes (Chapter 4)	20
1.9.4 DNA methylation: A gene silencing mechanism post-polyploidization (Chapter 5)	20
CHAPTER TWO	21
2. Biased paternal genomic loss and maternal expression of <i>rbcS-1</i> homeologs in <i>Tragopogon miscellus</i> (Asteraceae) allopolyploids: insight into cytonuclear compatibility.....	22
2.1 Abstract.....	22
2.2 Introduction.....	23
2.3 Materials and Methods.....	26
2.3.1 Plant material.....	26

2.3.2 DNA and RNA extraction.....	26
2.3.3 Primer design, PCR and sequencing of <i>rbcL</i> and <i>rbcS-1</i>	27
2.3.4 Genomic and cDNA CAPS analysis	29
2.3.5 5' Genome walking and 3' RACE of <i>rbcS-1</i>	30
2.3.6 Prediction of <i>rbcS-1</i> gene structure.....	32
2.3.7 Homeolog-specific RT-PCR	32
2.4 Results	34
2.4.1 <i>rbcS</i> gene family	34
2.4.2 Characterization of <i>rbcS-1</i> in <i>Tragopogon</i>	34
2.4.3 Divergence between <i>rbcS-1</i> and <i>rbcL</i> homeologs in the diploids and their pattern of retention in <i>T. miscellus</i>	35
2.4.4 Expression of <i>rbcS-1</i> homeologs in <i>T. miscellus</i> polyploids.....	38
2.5 Discussion.....	41
2.5.1 Characterization of <i>rbcS-1</i> in <i>Tragopogon</i> diploid species.....	41
2.5.2 Genomic loss and expression of <i>rbcS-1</i> homeologs biased towards the maternal parent in <i>T. miscellus</i> polyploids	42
2.6 Acknowledgements	46
2.7 Supplementary material	47
CHAPTER THREE.....	55
3. Genetic characterization of synthetic <i>Tragopogon</i> polyploids using microsatellite markers.....	56
3.1 Abstract.....	56
3.2 Introduction.....	57
3.3 Materials and Methods	62
3.3.1 Plant material.....	62
3.3.2 Microsatellite PCR.....	63
3.3.3 Comparison of F ₁ , S ₀ and S ₁ synthetics with natural <i>Tragopogon</i> polyploids	64
3.3.4 Microsatellite data analysis	66
3.3.5 Exploration of multilocus genotypes in the synthetic polyploids	67
3.4 Results	68
3.4.1 Amplification efficiency and diversity of microsatellite markers.....	68
3.4.2 Occurrence of multilocus genotypes in the synthetic polyploids	69
3.4.3 Genetic structure of the synthetic polyploids.....	70
3.4.3.1 <i>T. miscellus</i>	70
3.4.3.2 <i>T. mirus</i>	72
3.5 Discussion.....	77
3.5.1 Genetic structure of synthetic polyploids at microsatellite level.....	77

3.5.2 Genetic contribution of parental diploids into synthetic polyploid lineages.....	78
3.5.3 Implication of genetic variation present in synthetic polyploids on the genetic changes observed in the synthetics.....	80
CHAPTER FOUR.....	83
4. Comparative analysis of floral transcriptomes.....	84
4.1 Abstract.....	84
4.2 Introduction.....	84
4.3 Materials and methods.....	89
4.3.1 Plant Material.....	89
4.3.2 RNA extraction.....	91
4.3.3 RNA quantification and quality control	91
4.3.4 RNA Sequencing.....	92
4.3.5 Data analysis.....	92
4.4 Results	93
4.4.1 Divergence between parental species.....	93
4.4.2 Expression of floral development genes in <i>Tragopogon</i> diploids and polyploids.....	95
4.4.2.1 Expression of A-class genes	95
4.4.2.2 Expression of B-class genes.....	96
4.4.2.3 Expression of C-class genes.....	98
4.4.2.4 Expression of E-class genes.....	99
4.4.2.5 Floral symmetry genes.....	99
4.5 Discussion.....	100
4.5.1 Divergence among parental species.....	101
4.5.2 Transcript abundance or expression of floral identity genes	101
4.5.2.1 Transcript abundance for A-class genes	101
4.5.2.2 Transcript abundance for B-class genes	102
4.5.2.3 Transcript abundance for C- and E-class genes.....	103
4.5.2.4 Floral symmetry genes.....	104
CHAPTER FIVE.....	116
5. Gene silencing via DNA methylation in naturally occurring <i>Tragopogon miscellus</i> (Asteraceae) allopolyploids	117
5.1 Abstract.....	117
5.2 Introduction.....	117
5.3 Materials and methods.....	119
5.3.1 Plant material.....	119
5.3.2 Bisulfite conversion	119

5.3.3 Amplification and sequencing of genomic and bisulfite-converted DNA	120
5.3.4 Genome walking.....	120
5.4 Results and Discussion	121
5.5 Supplementary material and methods.....	126
5.5.1 Principle of bisulfite conversion	126
5.5.2 Cloning of BS-converted sequences.....	126
5.5.3 Genome walking protocols	126
CHAPTER SIX.....	133
6. General discussion	134
6.1 Conclusions and future perspectives.....	136
Bibliography.....	138

LIST OF FIGURES

Fig. 2.1 <i>rbcS-1</i> gene structure and locations of SNPs. The structure of the <i>rbcS-1</i> gene is shown at the top with both coding regions and non-coding regulatory elements indicated. Locations for SNPs between <i>T. dubius</i> and <i>T. pratensis</i> homeologs have been scaled along the length of the <i>rbcS-1</i> gene.....	33
Fig. 2.2 CAPS analysis showing additivity and expression of <i>rbcS-1</i> homeologs. (a) Genomic DNA CAPS results for six representative samples and (b) cDNA CAPS results for four representative samples of naturally occurring <i>T. miscellus</i> polyploids along with representative diploid parents, <i>T. dubius</i> and <i>T. pratensis</i> . Chromatograms belong to the same polyploid samples (from top to bottom) as in the gel photo below (from left to right). Chromatograms show sequence polymorphisms at the third SNP in exon 1 for both genomic DNA and cDNA sequences of the natural polyploids. “L” following the polyploid name denotes the long-liguled form, and “S” denotes the short-liguled form.	37
Fig. 2.3 Graphical illustration of retention and expression of parental <i>rbcS-1</i> homeologs in <i>T. miscellus</i> (short and long liguled) natural polyploids. Bar charts show (a) retention and (b) expression of parental <i>rbcS-1</i> homeologs in <i>T. miscellus</i> polyploids. Black, grey and white colours of the bars correspond to retention/expression of <i>T. pratensis</i> <i>rbcS-1</i> homeolog, both progenitor <i>rbcS-1</i> homeologs and <i>T. dubius</i> <i>rbcS-1</i> homeolog respectively. Short liguled and long liguled individuals are represented with separate bars.....	38
Fig. 2.4 Homeolog-specific RT-PCR of <i>rbcS-1</i> . (a) <i>T. dubius</i> and (b) <i>T. pratensis</i> homeolog-specific RT-PCR results are shown for 11 <i>T. miscellus</i> natural polyploids with representative diploid parents, <i>T. dubius</i> and <i>T. pratensis</i> . Six individuals of the allotetraploid <i>T. miscellus</i> (indicated by an asterisk (*)) showed genomic loss of one parental fragment (<i>T. dubius</i> homeolog was lost in five short-liguled polyploids, and the <i>T. pratensis</i> homeolog was lost in one long-liguled Pullman individual 2605-28). The remaining five natural polyploids show expression biased towards one of the parents. “L” following the polyploid name denotes the long-liguled form, and “S” denotes the short-liguled form.	40
Fig. S2.1 cDNA-CAPS for TDF-85 as a control to check equal expression of parental copies in <i>T. miscellus</i> polyploids showing biased maternal expression for <i>rbcS-1</i>	52
Fig. S2.2 Alignment of <i>rbcS-1</i> and <i>rbcS-2</i> cDNA sequences of <i>Tragopogon dubius</i> along with protein translation for both genes. Arrows indicate position of exons and premature stop codons.	53
Fig. 3.1 Formation of synthetic polyploids.....	62
Fig. 3.2 Neighbor-net of <i>T. miscellus</i> synthetic polyploids. (a) Based on all microsatellite loci. (b) Based on only <i>T. dubius</i> loci and includes their <i>T. dubius</i> parents and generic types named as G1, G2 and G3 based on multilocus genotype I, II and III. Each split is corresponding to the cluster of individuals having distinct multilocus genotype found in the natural polyploid populations.	71
Fig. 3.3 Neighbor-net of <i>T. mirus</i> synthetic polyploids. (a) Based on all microsatellite loci (b) Based on only <i>T. dubius</i> loci and include their <i>T.</i>	

<p><i>dubius</i> parents and generic types named as G1, G2 and G3 based on multilocus genotype I, II and III. Each split corresponds to the cluster of individuals having distinct multilocus genotypes.</p> <p>Fig. 3.4 (a) Neighbor-net of <i>T. miscellus</i> polyploids along with their <i>T. pratensis</i> parents.....</p> <p>Fig. 4.1 Diagrammatic illustration of the ABC model.....</p> <p>Fig. 5.1 Sequence polymorphisms between the diploid parents (<i>Tragopogon dubius</i> and <i>T. pratensis</i>) were used to determine if there is homeolog-specific silencing in <i>T. miscellus</i> allopolyploids. (a) Diagrammatic illustration of the expected chromatogram peaks for genomic and bisulfite-converted sequences when un-methylated or methylated in allopolyploid <i>T. miscellus</i>. This example shows silencing of the <i>T. dubius</i> homeolog. (b) Chromatograms of TDF-44 indicating the position of a methylated CpG adjacent to a polymorphic site (red box) in <i>T. miscellus</i> compared to the diploids. (c) Chromatograms from S18 showing an un-methylated CpG site in <i>T. miscellus</i> (black box) and the location of a polymorphic site between parental copies (red box). Red, blue, green and yellow colors of the chromatogram correspond to A, C, T and G, respectively. BS-converted=bisulfite-converted.....</p>	<p>73</p> <p>74</p> <p>86</p> <p>124</p>
---	--

LIST OF TABLES

Table 2.1 <i>rbcS-1</i> primers designed in this study.....	27
Table 2.2 Naturally occurring individuals of <i>Tragopogon miscellus</i> that showed bias in the retention and expression of parental <i>rbcS-1</i> homeologs. A dash (-) indicates that we were not able to study a particular individual for both retention (genomic DNA) and expression (cDNA).....	39
Table S2.1 List of natural and synthetic polyploids (<i>T. miscellus</i>) and diploid parents (<i>T. dubius</i> and <i>T. pratensis</i>) examined. Data are summarized from genomic DNA and cDNA sequencing, genomic and cDNA CAPS, and homeolog-specific RT-PCR. Note: Letters “D” and “P” correspond to the diploid parents <i>T. dubius</i> and <i>T. pratensis</i> , respectively. A ‘D’ or a ‘P’ indicates that only one parental homeolog was detected in genomic DNA or expressed. P>D indicates that the <i>T. pratensis</i> homeolog showed higher relative expression than the <i>T. dubius rbcS-1</i> homeolog in the <i>T. miscellus</i> individual and vice versa for D>P.	47
Table S2.2 Transcription factor binding sites in <i>rbcS-1</i> promoter region as determined by Plant Promoter Analysis Navigator (PlantPAN).....	51
Table 3.1 Crossing information and number of F ₁ , S ₀ and S ₁ synthetic polyploid lines examined in the study.....	65
Table 3.2 Occurrence of the multilocus genotypes in the synthetic polyploids.....	67
Table 4.1 List of diploid and polyploid samples.	90
Table 4.2 List of MADS-box orthologs from other species of Asteraceae that were used as reference genes.....	94
Table 4.3 Total read count for <i>Tragopogon</i> transcripts mapping to ABC genes for each of the diploid parents and <i>T. miscellus</i> polyploid.....	106
Table 4.4 Percentages of parental transcripts in <i>Tragopogon</i> diploid and polyploids orthologous to ABC class genes from other groups of Asteraceae.	108
Table 5.1 Individual plants used in the study and their methylation status for the genes studied; silencing data from Tate <i>et al.</i> (2006) and Buggs <i>et al.</i> (2009).....	122
Table S5.1 List of primers.....	128

LIST OF ABBREVIATIONS

A	Adenosine
<i>AG</i>	<i>AGAMOUS</i>
AS	Antisense
<i>AP1</i>	<i>APETALA1</i>
bp	Base Pair
BS- Converted	Bisulfite converted
°C	Degrees celcius
CAPS	Cleaved amplified polymorphic sequence
cDNA	Complementary DNA
CTAB	Cetyl trimethylammonium bromide
Contig	Contiguous sequence
<i>CYC</i>	<i>CYCLOIDEA</i>
C	Cytosine
<i>DEF</i>	<i>DEFICIENS</i>
dNTP	Deoxynucleoside 5'-triphosphate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tags
F ₁	First hybrid generation
FUE	Far upstream elements
gDNA	Genomic DNA
<i>GLO</i>	<i>GLOBOSA</i>
g	Gram
G	Guanine
HS	Homeolog specific
ID	Identification number
Indel	Insertion or deletion
Kb	Kilobase
♀	Maternal parent
MADS	MCM1, Agamous, Deficiens, Serum response factor (family of transcription factors with conserved DNA binding site)
Mb	Megabase
m NA	Messenger RNA
µg	Microgram
µl	Microliter
µM	Micromolar
ml	Milliliter
mM	Millimolar
MYA	Million years ago
min	Minute(s)
M	Molar
ng	Nanogram
NUE	Near upstream elements
♂	Paternal parent
<i>PI</i>	<i>PISTILLATA</i>

PCR	Polymerase chain reaction
PLACE	Plant cis-acting regulatory DNA elements
PVP	Polyvinyl pyrrolidone
3' RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
RNase	Ribnuclease
rRNA	Ribosomal RNA
RuBisCO	Ribulose-1,5-bisphosphate carboxylase oxygenase
rbcL	Ribulose bisphosphate carboxylase large subunit
rbcS	Ribulose bisphosphate carboxylase small subunit
RIN	RNA integrity number
S ₀	First generation of synthetic polyploids
S ₁	Second generation of synthetic polyploids
S	sense
<i>SEP</i>	<i>SEPALLATA</i>
SNP	Single nucleotide polymorphism
Spp.	Species
T	Thymine
TDF	Transcript derived fragment
TE	Transposable elements
TSS	Transcription start site
U	Unit(s)
UTR	Un-translated region
U	Uracil
WGD	Whole genome duplication