NUCLEAR-CYTOPLASMIC INTERACTION ANALYSIS: A CRITICAL STEP FOR WHEAT GERMPLASM ENHANCEMENT

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
Of Agriculture and Applied Science

By

Ali Soltani

In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY

Major Department:

Plant Sciences

May 2014

Fargo, North Dakota

North Dakota State University
Graduate School

Title

NUCLEAR-CYTOPLASMIC INTERACTION ANALYSIS: A CRITICAL STEP FOR WHEAT GERMPLASM ENHANCEMENT

By

Ali Soltani

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Mohamed Mergoum
Chair

Shahryar F. Kianian

Michael J. Christoffers

Glenn Dorsam

Approved:

27 May 2014	Richard D. Horsley
Date	Department Chair

ABSTRACT

Wheat (*Triticum aestivum*) possesses the largest set of available alloplasmic lines in either the plant or animal kingdom. These alloplasmic lines consist of novel combinations of wheat cultivar nuclei and cytoplasm, derived from the *Triticum* and *Aegilops* genus. This set of alloplasmic lines provides a unique genetic resource for the enhancement of existing germplasm. The current study was devised to address several aspects of cytoplasm genetics, including cytoplasm evolution and its interaction with the nuclear genome. Initially, the *ATP6* gene from 46 different alloplasmic lines was isolated and sequenced. *ATP6* resides on the mitochondrial genome, and its function involves the electron transfer chain in the mitochondrial inner membrane. Our results revealed the existence of vast diversity among the cytoplasms. Two novel orthologs of *ATP6-1* genes were detected in our analysis. Subsequently, the expression activity of these orthologs was confirmed by RT-PCR. In the second part of this research, an integrative approach was employed to more elucidate the mechanism of nuclear-cytoplasmic interaction. For this study, wheat alloplasmic lines possessing *Ae. mutica* cytoplasm were compared with their corresponding euplasmic lines. Comparative QTL analysis for dry matter weight, height and spike number per plant was conducted to identify QTL which interact with each cytoplasm type. As a result, a QTL hotspot was detected on chromosome 5A. Furthermore, this analysis revealed that the genetic network, controlling traits, is dynamic. The genetic network is affected by the cytoplasm type. The methylation pattern of nuclear genes was shown to be influenced by cytoplasm. Sequencing revealed that the majority of the identified epialleles located within genic regions. Finally, this alteration of methylation was shown to change the expression pattern of corresponding genes between allo- and euplasmic lines.

ACKNOWLEDGMENTS

First and foremost, I wish to thank Dr. Shahryar Kianian for providing me the opportunity to conduct this research as part of the Wheat Germplasm Enhancement Project in the Department of Plant Sciences at North Dakota State University. I would also like to offer special thanks to Dr. Mohamed Mergoum for his gracious guidance of me in the completion of my PhD program. It was my pleasure and privilege to have Dr. Glenn Dorsam and Dr. Michael Christoffers serve my PhD committee. Thank you, both. I also am grateful for the help I received from all the members of the Wheat Germplam Enhancement Project and the Hard Red Spring Wheat Breeding Project; particularly, Justin Hegstad, for the collegiality he has established within the group. Conducting this research would not have been possible without the generous assistance of Plant Sciences department staff and faculty; in particular, Dr. Richard Horsley, Department Chair, who's stewardship of the Department has created a highly professional and collaborative atmosphere. I am also thankful to Monsanto for supporting my PhD through the Beachell-Borlaug International Scholars Program. Finally, and most importantly, I wanted to thank my wife Samira; without her love and support I would not have been able to complete this work.

DEDICATION

I would like to dedicate this work to:

My parents, Feridoun Soltani and Simin Nayyery, who dedicated their lives to me

My wife, Samira, who has blessed me with her endless, beautiful love

My sister, Anooshe. I still feel her warm hugs and support

And to all the people who strive to bring peace, food, and love to the world.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

GENERAL INTRODUCTION

Genetic diversity is a crucial prerequisite for any breeding program. For many years, genetic variation in the nuclear genome was considered the sole source of genetic diversity for breeding purposes. However, genetic variation existing in the cytoplasmic organelles, including chloroplasts and mitochondria, is a source of variation, which cannot be ignored. There are several factors that impede the wider utilization of cytoplasmic diversity in breeding programs; *i.*) narrow genetic diversity exists in chloroplast and mitochondrial genomes of domesticated crops. This narrow diversity may indicate the existence of strong selection against unfavorable organellar alleles involved in domestication. *ii.*) In some cases, increasing organellar genome diversity results in the occurrence of severely undesirable characteristics such as cytoplasmic male sterility (CMS), increased stress susceptibility, and reduced viability. These unfavorable traits may discourage researchers from investigating the potential to identify advantageous factors, and, *iii.*) hybridization barriers between cultivated crops and their wild progenitors can be a major obstacle for introgression of cytoplasmic diversity from the wild relatives.

In spite of these factors, lines have been developed in which the cytoplasmic genomes of several *Aegilops* species replace the original cytoplasm in hexaploid and tetraploid wheat by recurrent backcrossing methods. In many cases, changing the cytoplasm resulted in drastic phenotypic alterations due to new nuclear-cytoplasmic interactions caused by the alloplasmic condition. These alloplasmic lines constitute a valuable genetic resource for wheat germplasm enhancement. To improve our understanding of this genetic resource in wheat, the current research was designed around two general objectives.

The first chapter of this dissertation addresses our first objective, which was to characterize cytoplasmic diversity in alloplasmic lines. *ATP6,* a mitochondrial gene, was isolated and sequenced from several alloplasmic wheat lines. The phylogenetic relationships

among cytoplasms were elucidated based on detected SNPs. This information provided a basis from which to establish several previously unknown phylogenetic relations among the *Triticum-Aegilops* group of species.

The second chapter addresses several questions regarding the effect of cytoplasm replacement in terms of the genetic basis of traits and nuclear gene networks. In this chapter, an integrative study including QTL (Quantitative Trait Locus), epigenetic and expression analysis was employed to better elucidate the interaction between cytoplasmic and nuclear genomes.

LITERATURE REVIEW

Using Cytoplasmic Genetic Diversity in Breeding Programs

To better exploit the existing diversity in the cytoplasmic genomes, alloplasmic lines were produced in several plant species. The word 'alloplasmic' is derived from two words; *allo-* (different) and *plasmon* (cytoplasm), which indicates a line with an alien cytoplasm, relative to the nucleus. The routine method for production of an alloplasmic line is recurrent backcrossing after the initial hybridization. The progeny of this initial hybridization utilized as the female, and the conventional line as the recurrent male. Several generations of backcrossing are necessary for the near-complete restitution of the nuclear genome of one species into the cytoplasm of another species. Upon this nuclear substitution, several novel phenotypes may emerge, resultant of the new interactions between nuclear and cytoplasmic components.

The first attempts for the production of alloplasmic lines date to 1954, where Michaelis generated a few alloplasmic lines in *Epilobium* species. However, further inquiry into the characteristics of that study was obstructed by strong interspecific cross-incompatibility. Presently, the largest collection of alloplasmic lines is in wheat (*Triticum aestivum*), consisting of several plasmon types of *Triticum*, *Aegilops*, *secale* and *Agropyron* species replacing the *T. aestivum* cytoplasm (TSUNEWAKI 1980). Extensive work by several groups resulted in production and establishment of several combinations of nuclear-cytoplasm hybrids (MAAN 1975; TSUNEWAKI *et al.* 1996). This large array includes 551 alloplasmic lines, which were produced by transferring the plasmon of 46 wild relative species into twelve different hexaploid wheat nuclei (TSUNEWAKI 2009).

Several novel phenotypes emerged in wheat alloplasmic lines upon cytoplasmic substitution. One of the more interesting phenotypes is the twin seedling, which was induced

in specific plasmon types (TSUNEWAKI 1993). In this phenotype, two seedlings emerged from one seed; one of them being diploid and the other being haploid. The mechanism controlling this phenomenon is not understood. Other novel phenotypes observed in specific alloplasmic conditions include pistillody (the homeotic transformation of the stamen into a pistil), premature sprouting, and formation of seeds without germ. It is believed that the latter phenotype results from parthenogenesis of the ovum (TSUNEWAKI 1993). Upon endosperm development, the parthenogenic cells die and produce germless seeds.

It also was shown that several other phenotypes are affected by plasmon alterations (TSUNEWAKI 2009). The most pronounced included affects upon male fertility, flowering time, dry matter weight (DMW) and plant height. Indeed, Tsunewaki found 21 plant characteristics that exhibit some level of alteration in response to plasmon replacement. Based on these phenotypic effects, different plasmon types were classified into 17 distinct groups (TSUNEWAKI *et al.* 2002). These distinct groupings are confirmed by recent molecular genetic studies characterizing the molecular divergence observed among mitochondria and chloroplast genomes (Terachi and Tsunewaki 1992; Wang *et al.* 2000).

The potential utility of alloplasmic lines in hybrid wheat production led to extensive investigations into the mechanisms controlling cytoplasmic male sterility (CMS) and fertility restoring (*Rf*) genes in alloplasmic wheat lines. Early studies identified several *Rf* gene locations on different chromosomes of wheat (Tsunewaki, 1974, 1982; Tahir and Tsunewaki, 1971; Muramatsu, 1959; Mukai and Tsunewaki, 1979; Tsujimoto and Tsunewaki, 1984; Murai et al., 1990b). These analyses were mainly based on using monosomic addition and disomic substitution lines. The results showed that the majority of the *Rf* genes reside on chromosome 1B. However, genes affecting fertility also were identified on other chromosomes, including 7B, 2B and 1D. However, Maan recognized that *Rf* genes are not present in all alloplasmic systems (MAAN 1992b). In particular, compatibility between *T.*

longissimum cytoplasm and *T. turgidum* L. *var*. *durum* nuclei cannot be restored with an *Rf* gene. In this case, marginalized cytopasmic compatibility can be improved by the addition of *scs* (species cytoplasm specific) and *Vi* (vitality) gene pairs. These two genes can restore plant vigor and male fertility in alloplasmic durum lines, respectively (MAAN 1992a).

Dry matter weight (DMW) is a trait that also can be affected by cytoplasm type. This trait is adversely affected by an interaction caused by different plasmon types in most alloplasmic lines. Plasmon types C, G, M^o, Mt^2, S, S^b and S^v cause ten to twenty percent reductions in DMW compared to corresponding euplasmic lines (TSUNEWAKI 2009). Similarly, plasmon types C^u, C^{u2}, M^h, and S^l result in thirty to forty percent reduction in DMW at the maturation point. Alloplasmic wheat lines possessing M and A plasmon types exhibit the most drastic reduction in DMW, with DMW values less than sixty percent of the corresponding euplasmic line. It has been proposed that growth reduction observed in alloplasmic lines having A, C^u and C^{u2} plasmon types results from cold temperature-induced leaf variegation (TSUNEWAKI 2009). Some alloplasmic types do not exhibit statistically significant differences in DMW. These plasmon types were grouped into B and D plasmon types. Furthermore, two cytoplasm types were detected which increase DMW in combination with some specific cultivars. These plasmon types belong to two *Aegilops* species; *Ae. mutica* and *Ae. ovata*.

The affect of plasmon type on DMW can be different depending on the cultivar. For example, while the combination of *Ae. mutica* plasmon with *T. aestivum* var. Chris nucleus increases plant height and DMW drastically, the effect observed in the combination of *Ae. mutica* and *T. aestivum* var. Selkirk is insignificant. This observation implies that *T. aestivum* var. Chris has specific alleles that can increase the DMW in response to cytoplasm of *Ae. mutica*; whereas *T. aestivum* var. Selkirk lacks these alleles. Such differences can potentially

facilitate genetic studies to detect candidate genes and alleles that can improve DMW and other such desirable traits in breeding programs.

Molecular Aspects of Nuclear-Cytoplasmic Interaction

Organelles facilitate vital cellular functions including energy production, carbon assimilation, heme synthesis, apoptosis and signaling pathways in every plant species (RYAN and HOOGENRAAD 2007). The chloroplast and mitochondria are two semi-autonomous organelles with their own genomes. However, the majority of organellar proteins and components are encoded by the nuclear genome, translated in the cytoplasm, and then imported to the organelle (Woodson and Chory 2008). It is believed that more than 95% of chloroplast and mitochondrial proteins are encoded by the nuclear genome. Most of the proteins encoded by the organelle genome reside on the inner organellar membrane, and are involved in the electron transfer chain. Therefore, a strict communication is necessary between the nucleus and organelles to ensure the functionality and homeostasis of every cell component. Interruption of this communication results in alterations in physiological characteristics, which can be catastrophic. Intracellular communication can be categorized by three general criteria: *i.*) anterograde signaling, which is the communication from the nucleus to organelles. This communication is exemplified by the transfer of organellar proteins encoded by nuclear genes. These proteins can be involved in organelle structure, metabolic enzymes, or proteins that regulate the organellar gene expression. Alternatively, *ii.*) retrograde signaling is communication from organelles to the nucleus. This form of communication entails signals communicating organelle status to the nucleus in which genes in the nucleus are co-regulated. And finally, *iii.*) communication also occurs between chloroplasts and mitochondria.

Organellar gene expression can be regulated at several levels. The first level of regulation is that of transcription. For example, chloroplast genes are transcribed by two

distinct enzymes; nuclear encoded RNA polymerase (NEP) and plastid encoded RNA polymerase (PEP). It has been shown that the relative importance of each enzyme can differ during plant development. Also, PEP plays a greater role transcribing the genes involved in photosynthesis than NEP (Maliga 1998). In most of the cases, however, the abundance of organelle proteins cannot be explained solely by transcription efficiency (Giege et al. 2005). It was shown that post-transcriptional and post-translational processes are the major mechanisms by which organellar gene expression is regulated (Leon et al. 1998). The regulatory machinery is encoded by nuclear genes and then transferred to organelles. So far, two major groups of these proteins have been identified in different organisms: *i.*) first group of proteins, termed pentatricopeptide repeat (PPR) proteins, which are largely conserved among different eukaryotes, and *ii.*) a second regulatory class, termed pioneer proteins. It was proved that the latter class is less conserved among species (Saha et al. 2007). However, both groups appear to follow the same cellular mechanism, which involves attachment of protein to a particular organellar RNA and the involvement of other components for further modifications.

The expression profile of nuclear genes can be regulated by signals that originate from either mitochondria or chloroplasts. Releasing of these signals can be triggered by environmental stimuli such as biotic/abiotic stresses or developmental regulators. Reactive oxygen species (ROS) is one of the universal components of retrograde signaling pathways, which can originate in both chloroplasts and mitochondria (Woodson and Chory 2008). Under stress stimuli, non-efficient transfer of electrons in the mitochondria or chloroplast membrane results in over-production of ROS. This over-production subsequently triggers a pathway by which several nuclear stress-responsive genes such as alternative oxidase (AOX) are up-regulated (Rhoads and Subbaiah 2007). AOX is a crucial enzyme for reducing the amount of ROS and, ultimately, for organism survival under stress conditions (Rhoads et al.

2006). Another retrograde signal is Mg-protoporphyrin (Mg-proto) IX, which is specifically produced by chloroplasts (Woodson and Chory 2008). This component acts as a signal to down-regulate the nuclear genes involved in photosynthesis (Strand et al. 2003). It was shown that ABI4 is a transcription factor involved in activation of photosynthesis-related genes and down-regulated by Mg-proto IX (Koussevitzky 2007).

Another form of intracellular communication can be observed between mitochondria and chloroplasts, which is indirect and through the nucleus. For instance, a mutation in the maize mitochondrial genome (*none-chromosomal stripe 6*) will modify the expression of nuclear PSI-related genes in the chloroplast (Jiao et al. 2005).

The presence of these three-way communication modes between the nucleus, mitochondria and chloroplast requires the parallel co-evolution of all three components together. However, occasional mutations can interrupt this cross-talk and result in impairment to the organism. Some of the traits that emerge from an alteration of these interactions can be desirable for breeding proposes. One such useful phenotype is cytoplasmic male sterility (CMS), which has been utilized in hybrid breeding programs. This phenotype was first observed in flax (Bateson and Gairdner 1921). Today, it is widely accepted that CMS arises from a disruption in the nuclear-mitochondrial interaction.

There are generally three forms of CMS: i.) in this class, the stamen converts to either petals (petaloid) or pistils (pistillody). This class of CMS can be explained in a model called ABCE. In this model, accurate expression of genes that belong to each class of either A, B, C or E is necessary for correct morphogenesis of floral tissues in plants.

ii.) In the second class, anthers cannot complete development. Consequently, no pollen can be produced. Finally, under the third class of CMS, iii.) Anthers develop, but pollen is non-functional.

References

Bateson, W., and A. E. Gairdner, 1921 Male-sterility in flax, subject to two types of segregation. Journal of Genetics 11: 269-275.

Giege, P., L. J. Sweetlove, V. Cognat and C. J. Leaver, 2005 Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in Arabidopsis. Plant Cell 17: 1497-1512.

Jiao, S. X., J. M. Thornsberry, T. E. Elthon and K. J. Newton, 2005 Biochemical and molecular characterization of photosystem I deficiency in the NCS6 mitochondrial mutant of maize. Plant Molecular Biology 57: 303-313.

Koussevitzky, S., 2007 Signals from chloroplasts converge to regulate nuclear gene expression (vol 316, pg 715, 2007). Science 316: 1698-1698.

Leon, P., A. Arroyo and S. Mackenzie, 1998 Nuclear control of plastid and mitochondrial development in higher plants. Annual Review of Plant Physiology and Plant Molecular Biology 49: 453-480.

Maan, S. S., 1975 Cytoplasm variability and speciation in Triticiae, pp. 255-281 in *Prairie: A multiple view.*, edited by M. K. Wali, Univ. of North Dakota Press, Grand Forks, ND.

Maan, S. S., 1992a A Gene for Embryo-Endosperm Compatibility and Seed Viability in Alloplasmic Triticum-Turgidum. Genome 35: 772-779.

Maan, S. S., 1992b The Scs and Vi Genes Correct a Syndrome of Cytoplasmic Effects in Alloplasmic Durum-Wheat. Genome 35: 780-787.

Maliga, P., 1998 Two plastid RNA polymerases of higher plants: an evolving story. Trends in Plant Science 3: 4-6.

Rhoads, D. M., and C. C. Subbaiah, 2007 Mitochondrial retrograde regulation in plants. Mitochondrion 7: 177-194.

Rhoads, D. M., A. L. Umbach, C. C. Subbaiah and J. N. Siedow, 2006 Mitochondrial reactive oxygen species. Contribution to oxidative stress and interorganellar signaling. Plant Physiology 141: 357-366.

Ryan, M. T., and N. J. Hoogenraad, 2007 Mitochondrial-nuclear communications. Annual Review of Biochemistry 76: 701-722.

Saha, D., A. M. Prasad and R. Srinivasan, 2007 Pentatricopeptide repeat proteins and their emerging roles in plants. Plant Physiology and Biochemistry 45: 521-534.

Strand, A., T. Asami, J. Alonso, J. R. Ecker and J. Chory, 2003 Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. Nature 421: 79-83.

Terachi, T., and K. Tsunewaki, 1992 The Molecular-Basis of Genetic Diversity among Cytoplasms of *Triticum* and *Aegilops* .8. Mitochondrial Rflp Analyses Using Cloned Genes as Probes. Molecular Biology and Evolution 9: 917-931.

Tsunewaki, 1980 *Genetic diversity of the cytoplasm in Triticum and Aegilops*. Japan Society for the Promotion of Science.

Tsunewaki, K., 1993 Genome-Plasmon Interactions in Wheat. Japanese Journal of Genetics 68: 1-34.

Tsunewaki, K., 2009 Plasmon analysis in the Triticum-Aegilops complex. Breeding Science 59: 455-470.

Tsunewaki, K., G. Z. Wang and Y. Matsuoka, 1996 Plasmon analysis of Triticum (wheat) and Aegilops .1. Production of alloplasmic common wheats and their fertilities. Genes & Genetic Systems 71: 293-311.

Tsunewaki, K., G. Z. Wang and Y. Matsuoka, 2002 Plasmon analysis of Triticum (wheat) and Aegilops. 2. Characterization and classification of 47 plasmons based on their effects on common wheat phenotype. Genes & Genetic Systems 77: 409-427.

Wang, G. Z., Y. Matsuoka and K. Tsunewaki, 2000 Evolutionary features of chondriome divergence in *Triticum* (wheat) and *Aegilops* shown by RFLP analysis of mitochondrial DNAs. Theoretical and Applied Genetics 100: 221-231.

Woodson, J. D., and J. Chory, 2008 Coordination of gene expression between organellar and nuclear genomes. Nature Reviews Genetics 9: 383-395.

CHAPTER I. THE CRUCIAL ROLE OF REARRANGEMENT IN MITOCHONDRIAL GENOME EVOLUTION

Abstract

Mutation and chromosomal rearrangements are the two main forces increasing genetic diversity for natural selection to act upon, and ultimately drive the evolutionary process. Although genome evolution is a function of both forces, simultaneously, the ratio of each can be varied among different genomes and genomic regions. It is hypothesized that in plant mitochondrial genomes, rearrangements play a more important role than point mutations, but relatively few studies have directly addressed this phenomenon. To address this issue, we isolated and sequenced the ATP6-1 and ATP6-2 genes from forty-six different euplasmic and alloplasmic wheat lines. Four different ATP6-1orthologs were detected, two of them reported for the first time. RT-PCR analysis revealed that all four orthologs are expressed. Results also indicated that both point mutations and genomic rearrangement are involved in ATP6 evolution. However, rearrangement is the predominant force that triggers drastic variation. Data also indicated that speciation of domesticated wheat cultivars was simultaneous with the duplication of this gene. These results directly support the notion that rearrangement plays a significant role in driving plant mitochondrial genome evolution.

Introduction

Mitochondria and chloroplasts are semiautonomous organelles with their own genomes. They are thought to have originated from an α-proteobacterium and a cyanobacteria, respectively (Knoop 2004). However, after their acquisition by Archaebacteria, their evolution was greatly affected by this new symbiotic condition (Timmis *et al.* 2004). Although originating from the same ancestor, mitochondrial genomes faced completely different evolutionary pathways in *Animalia* and *Plantae* kingdoms. Mitochondrial genomes in plants are 10-100 folds larger than their counterparts in animals,

and underwent extensive rearrangements (Xiong *et al.* 2008). This variation can be detected not only among different genera in the *Plantae* kingdom, but also within species of the same genus (Sloan *et al.* 2012).

Several factors are believed to influence the evolution and diversity of mitochondrial genomes in plants including point mutation, intra-molecular homologous recombination and stoichiometric shifts. The point mutation rate in plant mitochondrial genomes is slower than the nucleus and its animal counterpart (Palmer and Herbon 1988). In contrast, frequent intra-molecular recombination, which causes rearrangements and gene shuffling, is believed to be the main evolutionary force in plant mitochondrial genomes (Galtier 2011; Palmer and Herbon 1988). Homologous recombination can also modify the frequency of each mitotype within a cell or tissue in a phenomenon known as stoichiometric shift (Woloszynska 2010). These recombination events are governed by nuclear-encoded proteins, and occur in repetitive regions of mitochondrial genome (Woloszynska 2010). The presence of extensive repetitive DNA is a unique characteristic of angiosperm mitochondrial genomes and the main cause of recombination and possible genome size alterations (Kitazaki and Kubo 2010).

The wheat mitochondrial genome has several long and short repeats involved in recombination and rearrangement (Ogihara *et al.* 2005). One of these repetitive sequences resides in the ATP synthase subunit 6 (ATP6) gene region and triggers changes within this crucial gene (Bonen and Bird 1988; Kawaura *et al.* 2011; Liu *et al.* 2011; Ogihara *et al.* 2005). ATP6 provides a vital inner membrane channel that couples the proton gradient to ATP production in the mitochondria of all eukaryotes (Devenish *et al.* 2000). Vitality of function suggests an essential conservation of this gene not only among closely related species but also in diverse groups (Bonen and Bird 1988; Kawaura *et al.* 2011; Liu *et al.* 2011; Ogihara *et al.* 2005). Two distinct domains were identified in the ATP6 gene: a conserved core, and a diverse pre-sequence region (Bonen and Bird 1988; Kawaura *et al.*

2011; Liu *et al.* 2011; Ogihara *et al.* 2005). The core ATP6 sequence of wheat and maize is conserved at 96% identity (Bonen and Bird 1988; Kawaura *et al.* 2011; Liu *et al.* 2011; Ogihara *et al.* 2005). In contrast, rearrangement leads to a highly polymorphic region in the 5' pre-sequence in the *Triticea-Aegilops* complex and other plant species (Bonen and Bird 1988; Kawaura *et al.* 2011; Liu *et al.* 2011; Ogihara *et al.* 2005). Two ATP6 orthologs have been identified among *Triticea-Aegilops* species that exhibit elevated pre-sequence diversity. One ortholog was identified in *T. aestivum* (Ogihara *et al.* 2005) and another in *Ae. kotschyi* (Liu *et al.* 2011) and *Ae. crassa* (Kawaura *et al.* 2011). Both orthologs are transcriptionally active and functional (Kawaura *et al.* 2011). These studies provided evidence as to the level of variation among a few *Aegilops* species but not a complete picture of possible changes, functional alterations, and insights into the evolution of ATP6 orthologs in the entire *Triticea-Aegilops* complex.

The *Triticum-Aegilops* complex is an appropriate model for studying mitochondrial and chloroplast evolution in plants due to several unique aspects: First, a high amount of plasmon diversity, especially in the *Aegilops* genus, has been identified through a detailed alloplasmic characterization (Maan 1975; Maan 1991; Maan and Endo 1991; Tsunewaki 1980). Second, a significant amount of information on nuclear genome evolution based on detailed cytological and molecular analysis is available that can be related to cytoplasmic evolution (Badaeva *et al.* 2002; Badaeva *et al.* 2004). Lastly, a nearly complete set of alloplasmic lines facilitates studies on nuclear-cytoplasmic interaction (Tsunewaki 2009).

Efforts to elucidate the *Triticum-Aegilops* plasmon evolution began in 1978 using restriction fragment length polymorphism analysis (Vedel *et al.* 1978). Chloroplast genome analysis (Goryunova 2010; Hirosawa *et al.* 2004; Ishii *et al.* 2001; Provan *et al.* 2004), mitochondrial genome fingerprinting (Breiman 1987; Terachi and Tsunewaki 1992; Wang *et al.* 2000), and phenotypic effect of each plasmon in the alloplasmic condition were further

used for plasmon classification and evolutionary studies among these species (Tsunewaki 2009). In the present study, we isolated and sequenced ATP6-1 and ATP6-2 from several alloplasmic and euplasmic lines in this group of species to discern the evolutionary patterns of this region. This region was chosen for its active involvement in rearrangements and can be considered a hotspot of variation within the mitochondrial genome.

Material and Methods

Plant material

Forty-one different alloplasmic lines along with five euplasmic lines (Table 1.1) were included in this study. The selection of alloplasmic lines was based on two major considerations: *i.*) sampling highly diverse plasmon types based on previous classifications and *ii.*) choosing alloplasmic lines with the same cytoplasm but different nuclei to investigate potential changes in mitochondrial type due to nuclear genome substitution.

Table 1.1. List of alloplasmic and euplasmic lines used in this study

Plasmon donor species[a]	Plasmon donor chrom. No.	Genome (haploid)[b]	Plasmon type[c]	Abbr. code[d]
Alloplasmic lines with *T. aestivum* cv. Chris nucleus				
Ae. heldreichii	7	M	M^h	*hld*
Ae. columnaris	14	UM	U'	*clm*
Ae. caudata	7	C	C	*cdt*
Ae. mutica	7	T	T	*mtc*
Ae. triuncialis	14	UC	U	*trn*
Ae. ventricosa	14	DN	D	*vnt*
Ae. vavilovii	21	DMS	D^2	*vvl*
Ae. bicornis	7	S^b	S^b	*bcr-1*
Ae. kotschyi	14	US	S^v	*kts*
Ae. longissima	7	S^1	S^1	*lng*
Ae. tauschii	7	D	D	*sqr-1*
Alloplasmic lines with *T. aestivum* cv. Selkirk nucleus				
Ae. uniaristata G633	7	N	N	*unr*
T. turgidum subsp. *dicoccoides G1392*	14	AB	B	*dcd-1*
T. turgidum subsp *dicoccoides G1460*	14	AB	B	*dcd-2*
T. turgidum subsp *dicoccoides G671*	14	AB	B	*dcd-4*
T. turgidum subsp *dicoccoides #3*	14	AB	B	*dcd-3*
T. turgidum subsp *dicoccoides G1453*	14	AB	B	*dcd-5*

Table 1.1. List of alloplasmic and euplasmic lines used in this study (continued)

Plasmon donor species[a]	Plasmon donor chrom. No.	Genome (haploid)[b]	Plasmon type[c]	Abbr. code[d]
T. turgidum subsp *dicoccoides G803*	14	AB	B	*dcd-6*
T. turgidum subsp *dicoccoides okla* 11140	14	AB	B	*dcd-7*
T. turgidum subsp *dicoccoides* G1395	14	AB	B	*dcd-8*
T. dicoccoides okla 11186	14	AB	B	*dcd-9*
T. dicoccoides RL5207	14	AB	B	*dcd-10*
Haynaldia	14	VV	V	*hyn*
Ae. cylindrica	14	CD	D'	*cyl*
Ae. tauschii	7	D	D	*sqr*
Ae. crassa 6N	21	DDM	D^2	*crs*
Ae. juvenalis	21	DMU	D^2	*jvn*
Ae. crassa 4N	14	DM	D^2	*crs*
Ae. bicornis	7	S^b	S^b	*bcr-2*
Ae. searsii	7	S^s	S^v	*srs*
T. aestivum subsp. *macha 140191*	21	ABD	B	*mch-1*
T. aestivum subsp. *macha 190923*	21	ABD	B	*mch-2*
T. turgidum subsp. *durum*	14	AB	B	*trg-2*
Alloplasmic lines with *T. turgidum* CV. 56-1 nucleus				
Ae. variabilis	14	US	S^v	*vrb*
Ae. comosa	7	M	M	*cms*
Ae. longissima	7	S^1	S^1	*lng-2*
Ae. sharonensis	7	S^1	S^1	*shr*
Alloplasmic lines with *T. timopheevi* nucleus				
T. zhukovskyi	21	A^mAG	G	*zhk*
Ae. triaristata	14	UM	U	*trr*
Ae. umbellulata	7	U	U	*umb*
Euplasmic lines				
T. aestivum subsp. chris	21	ABD	B	*ast-1*
T. aestivum subsp. Selkirk	21	ABD	B	*ast-2*
T. turgidum subsp. durum 56-1	14	AB	B	*trg-1*
T. timopheevi subsp. *timopheevi*	14	AG	G	*tmp*
Ae. tauschii	7	D	D	*sqr*
Ae. longissima	7	S^1	S^1	*lng*

[a]After Miller (1987)
[b] After Kihara and Tanaka (1970) and Kimber and Tsunewaki (1988)
[c] After Tsunewaki (2009)
[d] Abbr. code is the abbreviation code given to each species and referred to in Figure 1.8

Cytoplasmic genome extraction

Cytoplasmic genomes were isolated from two-week-old seedlings grown in controlled greenhouse conditions, using the Triboush protocol (Triboush *et al.* 1998) with minor modifications. Briefly, for each sample: three to five grams of fresh tissue was ground in 20

ml cooled STE buffer [400 mM sucrose, 50 mM Tris pH 7.8, 3 mM EDTA, 0.2% (w/v) BSA and 0.1% (v/v) mercaptoethanol] and then passed through four layers of Miracloth (EMD Millipore). The homogenate was centrifuged at 2000 X g for 20 min. and the resulting supernatant centrifuged separately at 18000 X g for an additional 20 min. The pellet was then re-suspended in 20 ml ST buffer [400 mM sucrose, 50 mM Tris pH 7.8 and 0.1% (w/v) BSA] with a soft paint brush and centrifuged again at 18000 X g for 20 min. The supernatant was discarded and the pellet re-suspended in 300 µl ST buffer and 200 µl of DNaseI solution (2.5 µg/µl DNaseI in ST buffer and 10 µl of 1M $MgCl_2$). DNaseI treatment was conducted for three hours at 4°C. The reaction was stopped by adding 500 µl of 500 mM EDTA. The solution was transferred to a tube containing 25 ml NETF buffer (1.25 M NaCl, 50 mM EDTA, 50 mM Tris pH 8.0 and 50 mM NaF), mixed, and then centrifuged at 18000 X g for 20 min. The supernatant was discarded and pellets kept at -20°C for downstream genome extraction. Organellar genomic DNA was isolated using a PureLink® Genomic DNA Mini Kit (Invitrogen) per vendor protocol. Isolated organellar DNA was checked for nuclear genome contamination, using a nuclear-specific primer designed on the actin gene (TA411_4571).

Amplification and sequencing

Two sets of primers were used for amplification of ATP6-1 and ATP6-2 genes (Table 1.2). These primers were designed within the conserved sequences flanking the gene in the *Aegilops* and *Triticum* species based on sequences deposited at NCBI (NC_007579 and GU985444). PCR primers ATP6-1F and ATP6-1R were utilized for ATP6-1 amplification and ATP6-2 primers were used for ATP6-2 amplification (Table 1.2). PCR amplifications were performed using GoTaq® Flexi DNA polymerase (Promega) in an Applied Biosystems® 2720 Thermal Cycler. Amplification conditions were as follows: pre-incubation for 5 min. at 94 °C, followed by thirty-five cycles of 1 min. at 94°C, 1 min. at 55°C and 1 min. at 72°C,

and a final extension of 7 minutes at 72°C. Amplified PCR products were purified using a GenElute PCR Clean-Up kit (Sigma-Aldrich) and sequenced based on the Sanger method (GENEWIZ Inc.) using the same primers. PCR products with multiple bands were first cloned in pGem®-T Easy Vector System (Promega) and each insert was sequenced using universal SP6 and T7 primers.

RNA extraction and cDNA synthesis

Total RNA of selected lines was extracted from 100 mg seedling leaf tissue using an RNeasy® Plant Mini Kit (Qiagen). RNA was treated with RNase-Free DNase Set (Qiagen) during isolation to remove possible DNA contamination. The quantity and quality of RNA was assessed using a Nanodrop™ spectrophotometer (Thermo Scientific) and 1% agarose gel electrophoresis, respectively. M-MLV Reverse Transcriptase (Promega) and random decamers were utilized for cDNA synthesis from 1 µg of isolated RNA.

Table 1.2. Primers utilized in this study (refer to Figure 1.1 for the location of these primers on the ATP6 loci)

Primer	Forward sequence	symbol	Reverse sequence	symbol
ATP6-1	CGCAATATCTTGAGTACCCGGGAACCA	1>	TTTCCTTTCAAAAGTGAGCGAGCAG	<1
ATP6-2	CAAGTCCGTTATTATCTCAATCCG	2>	CGTTACAAGGCAACTAGCAT	<2
ATP6-1a	ATTGCCGGCGTCACAATA	a>	CTGCCAAGTTTCGCCATTTC	<a
ATP6-1b	TTGGATCCGCAGCTTGT	b>	GGGCGCAATTCCATACATTT	<b
ATP6-1c/d	GCAGGGATTACGACGAAGTTA	c/d>	GTACTGGAAGGTGTCGCTTT	<c/d
ATP6	ACCTAATTCCAGACCGGTTAATG	6>	AGTGGGTTTGCTTGGACTATG	<6
Actin	GGCAACATTGTTCTCAGTGGTGGT	-	TCCTTTCAGGAGGAGCAACAACCT	-

Data analysis

Forward and reverse sequenced fragments of each gene were assembled using CAP3 (Huang and Madan 1999) to obtain the complete sequence. Genes were aligned first using ClustalX2 (Larkin *et al.* 2007) and then analyzed for SNPs using dnaSP5 (Librado and Rozas 2009). The synonymous and non-synonymous mutation rates were calculated based on the method of Nei and Gojobori (1986) using dnaSP5 (Librado and Rozas 2009). The BioEdit package (Hall 1999) was used for sequence editing and protein prediction. Hydrophobicity profiles of predicted proteins were constructed using the Eisenberg Scale (Eisenberg *et al.* 1982) in the BioEdit package. Functional protein domains were characterized using SMART (Schultz *et al.* 1998) and SPLIT 4.0 (Juretic *et al.* 1999).

Phylogenetic analysis

To construct a comprehensive phylogenetic tree among the alloplasmic lines, three distance matrices were calculated based on 1) ATP6-1 sequence polymorphism, 2) presence-absence of ATP6-2, and 3) a combined matrix. Distance matrices of ATP6-1 and ATP6-2 were merged considering the same weight for each data point to generate a combined matrix. To construct the ATP6-1 distance matrix, sequences of all lines were first aligned using ClustalX2. Aligned sequences were then imported into R software (R Development Core Team 2010) and the distance matrix was calculated using *dist.dna*, a function of the *ape* package (Paradis *et al.* 2004). A phylogenetic tree for this gene was constructed with 100 bootstrap using *upgma, bootstrap.pml* and *plotBS* functions; each of the last three functions is available in the *phangorn* package (Schliep 2011). Another distance matrix was calculated based on presence or absence of ATP6-2 using the PowerMarker v.3.25 program (Liu and Muse 2005). A combined matrix was imported into R and used for subsequent phylogenetic analysis. The same procedure was followed to construct the phylogenetic tree among conserved and diverse regions of gene orthologs. A heatmap of plasmon similarity-

dissimilarity was generated based on a combined distance matrix in R software using the heatmap.2 function in the gplots CRAN library.

Results

ATP6-1 diversity in the *Triticum-Aegilops* complex

In this study four major orthologs of ATP6-1 were detected among *Triticum-Aegilops* species (Figure 1.1). *In silico* translation revealed that ATP6-1a encodes a 386 amino acid (aa) polypeptide with a high degree of homology to the ATP6-1 protein present in *T. aestivum* (Ogihara *et al.* 2005) and *T. turgidum* mitochondria (Noyszewski*et al.* 2014). The ATP6-1b encodes a 415 aa protein, which is identical to a previously reported gene in *Ae. crassa* (Kawaura *et al.* 2011) and *Ae. kotschyi* (Liu *et al.* 2011). This ortholog was detected in 14 different plasmon types analyzed in this study mostly from the S-, D-, U- and T-type plasmons (Table 1.1). The ATP6-1c and ATP6-1d are novel orthologs and encode 458 aa and 444 aa proteins, respectively. These two orthologs show a high degree of homology to each other, except for a 14aa insertion/deletion in their pre-sequence region.

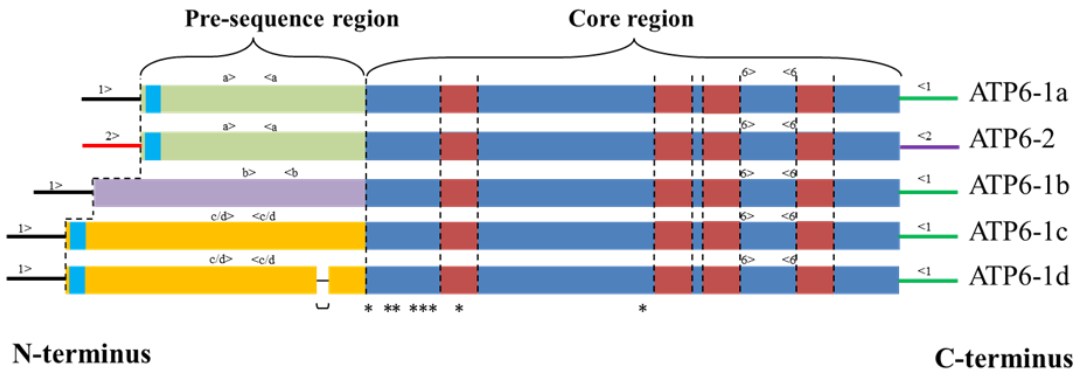

Figure 1.1. Different orthologs of *ATP6-1* and *ATP6-2*. Coding and non-coding regions are represented by boxes and weighted lines, respectively. Dark blue boxes represent similar sequence in the core region. Dark red boxes indicate the predicted alpha helix domains in the core region. Light blue boxes represent the shared regions in the pre-sequence domains of *ATP6-2*, *ATP6-1a*, *ATP6-1c* and *ATP6-1d*. SNP positions among ortholog core regions indicated by * sign and indel segment in pre-sequence of *ATP6-1c* and *ATP6-1d* is represented by a bracket. Primer positions utilized in this analysis are defined by their corresponding symbol. For information about each primer, refer to Table 1.3.

The ATP6-1c locus was detected in three alloplasmic lines: (*Ae. comosa*)*T. turgidum* line *56-1*, (*Ae. umbellulata*) *T. aestivum cv. '*Chris' and (*Ae. columnaris*)Chris. The ATP6-1d allele was present only in the (*Ae. caudata*)Chris line. Transcription activity of each ortholog was confirmed in selected alloplasmic lines (Figure 1.2).

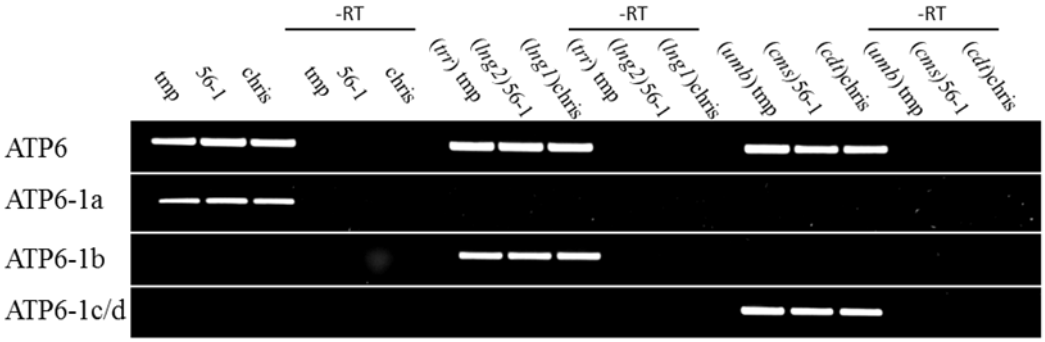

Figure 1.2. Reverse Transcript (RT)-PCR of each ATP6-1 ortholog. Results showing transcriptional activity of each *ATP6-1* ortholog. *tmp*: *T. timopheevi*, 56-1: *T. turgidum* var. 56-1, Chris: *T. aestivum* cv. Chris, *trr*: *Ae. triaristata*, *lng1*: *Ae. longissima* accession m, *lng2*: *Ae. longissima* accession c, *umb*: *Ae. umbellulata*, *cms*: *Ae.comosa* and *cdt*: *Ae. caudate*. –RT represents the minus reverse transcriptase control

Alignment of ATP6-1orthologs revealed the existence of two distinguishing regions (Figure 1.1); the C-terminal 271 aa, which are conserved, and the preceding N-termini, which are divergent in terms of both nucleotide and amino acid composition. The flanking non-coding regions also were conserved in the ATP6-1 orthologs. Comparative hydrophobicity analysis of different orthologs (Figure 1.3) indicates the presence of at least four hydrophobic regions in the conserved domain, which represent four transmembrane α-helicase motifs. Presence of these motifs also was confirmed using other predictive software. The divergent N-terminus domains mainly contain hydrophilic aa residues, which indicate their solubility in the mitochondrial inner-membrane space or matrix (Figure 1.3). Previous studies demonstrated that N-terminal residues of ATP6 protein in yeast are post-translationally removed by specific machinery that resides in the inner-membrane space (Michon *et al.* 1988).

Considering the conservation of this machinery in different species including wheat, it is possible that the N-terminal domain modification likely resides in the inner-membrane space, where it can be cleaved upon translation.

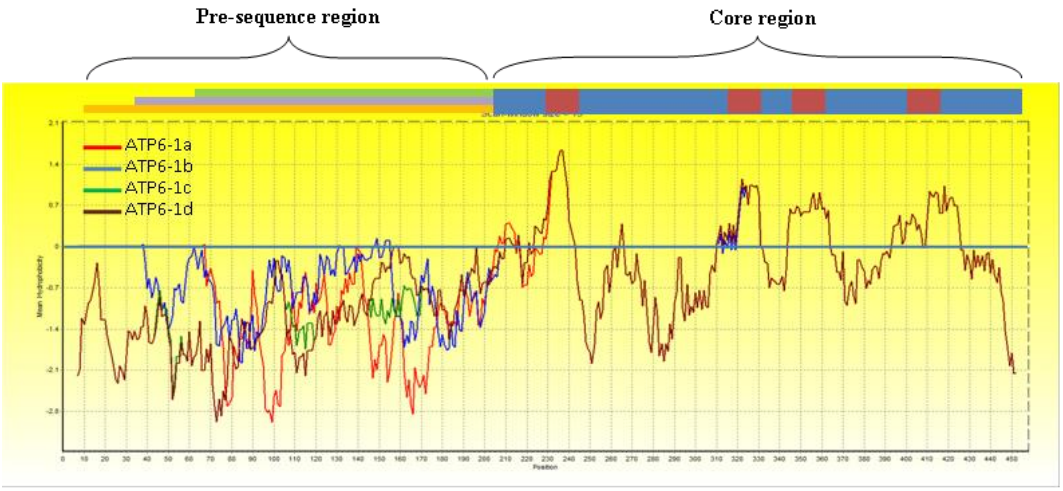

Figure 1.3. Hydrophobicity analysis of *ATP6-1* orthologs. Two distinct domains of *ATP6-1* orthologs can be observed. The conserved core domain that possesses transmembrane motifs and pre-sequences that mostly contain hydrophilic amino acids.

Polymorphism among the different ATP6-1 orthologs

Pairwise comparison of four ATP6-1 orthologs based on sequence of a relatively conserved core region indicates that ATP6-1a, with eight SNPs, is the most divergent ortholog, and that ATP6-1c/d are more similar to ATP6-1b (Table 1.3). This relationship is also observed in a phylogenetic tree based on just the core sequence (Figure 1.4a). However, when only the variable pre-sequence region was considered, ATP6-1a clustered with ATP6-1c/d (Figure 1.4b). The same analysis based on complete gene sequences revealed that the topology of the phylogram is similar to pre-sequence results indicating the importance of changes in this segment (Figure 1.4c).

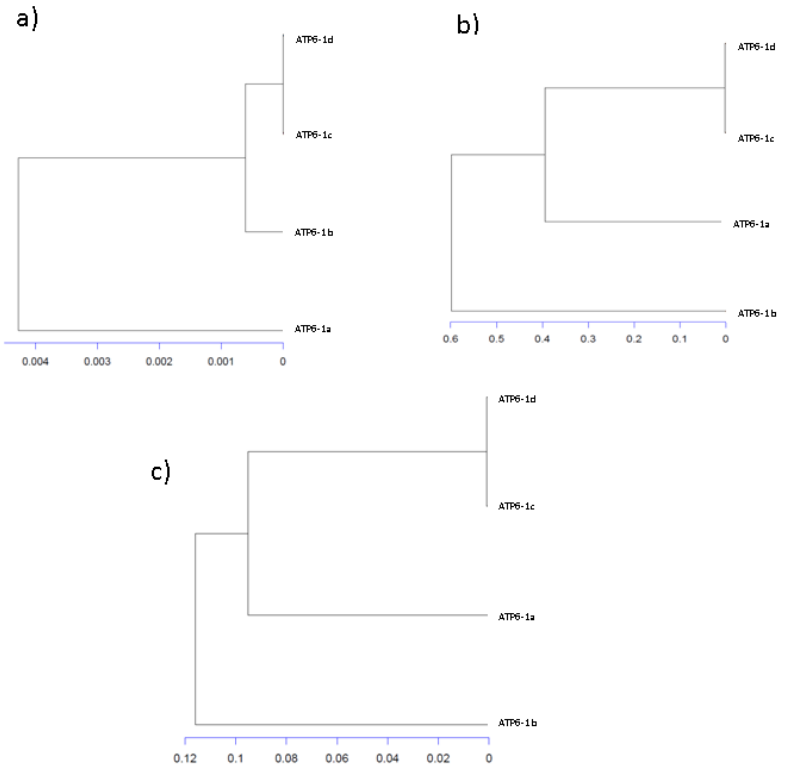

Figure 1.4. Phylogenic analysis of the *ATP6-1* orthologs. a) Phylogenetic analysis based on the core region, b) phylogenetic analysis based on the pre-sequence region, and c) phylogenetic analysis based on the entire *ATP6-1* sequence.

To further analyze the region, the N-terminus pre-sequence regions of four alleles aligned, ignoring the C- terminus. This analysis revealed a shared 40-nucleotide block present within the pre-sequence region of ATP6-1a and ATP6-1c/d (Figure 1.5). Further analysis indicated that this sequence is also present in the *T. aestivum* ATPase subunit 4 gene, as well as other species including *Bambusa oldhamii*, *Sorghum bicolor* and even in the distant species *Cycas taitungensis*. The presence of this shared block in the pre-sequence region of ATP6-1a and ATP61c/d (or its absence in ATP6-1b) likely resulted from a rearrangement event, not the gradual accumulation of point mutations. It seems that rearrangement events are common in the pre-sequence region of ATP6, since a deleted/inserted block was observed in ATP6-1c and ATP6-1d (Figure 1.1).

Figure 1.5. Alignment of the *ATP6* consensus pre-sequence. The red box indicates a 40-nucleotide block that is located at the beginning of the *ATP6-1a* and *ATP6-1c/d* orthologs but not in *ATP6-1b*.

Polymorphism of ATP6-2

The major ATP6-2 polymorphism was limited to its presence or absence among different plasmon types. ATP6-2 was detected in the mitochondrial genome of 18 alloplasmic lines and euplasmic *T. turgidum* and the two *T. aestivum* sources tested. All lines that contained *ATP6-2* belong to the B-type cytoplasm as well as *Ae. uniaristata* (N type), *Ae. variabilis* (S^v type) and *Ae. heldrachii* (M^h type). The coding region of this gene perfectly aligned to ATP6-1a sequence (Figure 1.1). Among the ATP6-2 genes sequenced, just one nucleotide deletion was detected in 3' UTR of (*T. turgidum*)*T. aestivum* cv. 'Selkirk', (*Ae. heldrachii*)Chris and (*Ae. variabilis*)56-1.

Table 1.3. Pairwise comparison of four ATP6-1 orthologs based on their core regions

	ATP6-1b				ATP6-1c				ATP6-1d			
	SNP[a]	AAS[b]	ds[c]	dn[d]	SNPs	AAS	ds	dn	SNPs	AAS	ds	dn
ATP6-1a	8	6	0.006	0.01	7	6	0.006	0.01	7	6	0.006	0.01
ATP6-1b	-	-	-	-	1	0	0	0	1	1	0	0
ATP6-1c	-	-	-	-	-	-	-	-	0	0	0	0

[a]Single Nucleotide Polymorphism
[b]Amino Acid Substitution
[c]Average Number of synonymous substitutions per synonymous site
[d]Average Number of non-synonymous substitutions per non-synonymous site

Higher non-synonymous point mutations in the ATP6-1 gene

Polymorphism analysis within each major ortholog showed that there are no nucleotide differences among lines possessing the ATP6-1a (Table 1.4). Absence of variation within plasmons possessing ATP6-1a may indicate that this ortholog is either the most recent ATP6 form or most critical to the function of mitochondria in comparison to other ATP-1 orthologs. On the other hand, alignment of ATP6-1b alleles from twenty lines detected 32 SNPs, of which 31 changed an aa residue. The average number of non-synonymous substitutions per non-synonymous site (dn) was about three-fold higher than the average number of synonymous substitutions per synonymous site (ds; Table 1.4). Interestingly, all of these non-synonymous mutation changes occurred in the pre-sequence region of ATP6-1b. The only mutation in the core region of ATP6-1b was a synonymous mutation. The synonymous mutation was a C→ T transition that occurred at the third nucleotide of the codon for leucine amino acid. Four nucleotide variations were detected among the three lines possessing ATP6-1c alleles, and they occurred in the diverse domain, resulting in an amino acid substitution. A similar higher dn number was also detected in this ortholog (Table 1.4).

The mitochondrial genome of only one line contained the ATP6-1d locus, therefore further polymorphic analysis for this gene was not possible.

A similar trend for a higher non-synonymous mutation rate was also detected in the conserved region of different ATP6 orthologs (Table 1.3). Interestingly in all cases, the non-synonymous mutations accumulated on or in close proximity to the pre-sequence region.

Table 1.4. Polymorphisms detected within lines possessing each major ortholog

ATP6-1 orthologs	Lines[a] tested	Unique[b] cytoplasm	SNP[c]	ds[d]	dn[e]	Transition	Transversion
ATP6-1a	22	9	0	0	0	0	0
ATP6-1b	21	14	32	0.0072	0.0168	16	16
ATP6-1c	3	3	4	0	0.002	1	3
ATP6-1d	1	1	-	-	-	-	-

[a]All alloplasmic lines with different cytoplasm or nucleus
[b]Alloplasmic lines that are different based on cytoplasm
[c] Single Nucleotide Polymorphism
[d]Average Number of synonymous substitutions per synonymous site
[e]Average Number of non-synonymous substitutions per non-synonymous site

Phylogenic analysis among plasmons

Two phylogenetic trees were constructed based on the data from this work; one based on polymorphisms among ATP6-1 orthologs (Figure 1.6), and the other a combined tree, considering ATP6-1 polymorphism and the presence/absence of the ATP6-2 locus (Figure 1.7). Based only on ATP6-1 sequence analysis (Figure 1.6), the mitochondria of different species were classified into three distinct clades, each representing one major ortholog. Clade I was distinguished by existence of ATP6-1a, clade II by ATP6-1b and clade III by presence of ATP6-1c and ATP6-1d. Further classification within each clade was based on nucleotide polymorphism within each ortholog.

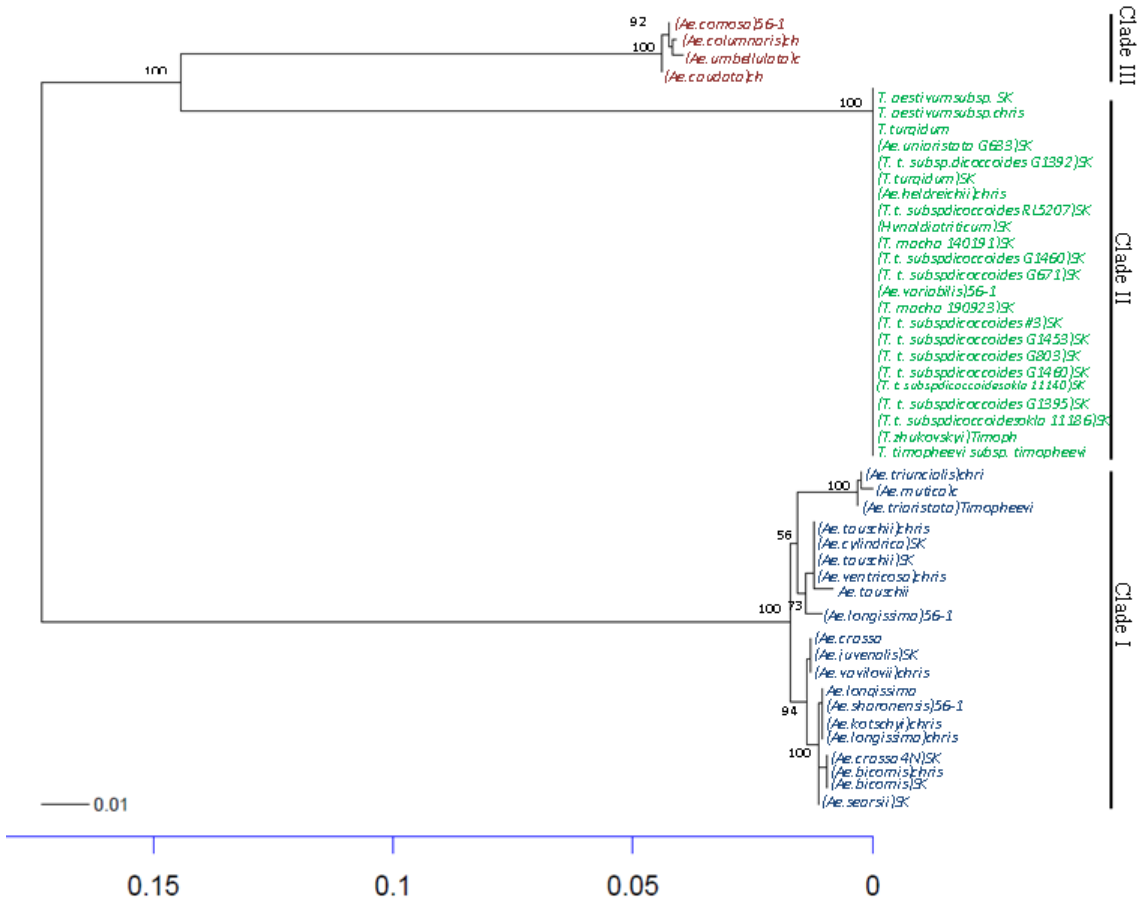

Figure 1.6. Phylogenetic tree of *ATP6-1* in different plasmons. Lines having the *ATP6-1a* gene are represented in green, lines with the *ATP6-1b* gene in blue, and lines possessing the *ATP6-1c/d* gene in red color. Bootstrap support values of each node are also indicated.

In the combined analysis, the phylogenetic tree includes two major clades representing the presence or absence of ATP6-2 in the mitochondrial genomes (Figure 1.7). The topology in each clade of both phylogenetic trees does not change except for the G plasmon classification. G-type cytoplasm lines contain ATP6-1a in their mitochondrial genomes, while ATP6-2 is missing. This plasmon type was identified in *T. timopheevi* and *T. zhukovskyi*. Considering only ATP6-1 (Figure 1.6), G-type cytoplasms grouped in the first clade, near to B plasmons. The B-type group is represented by the presence of both ATP6-1a and ATP6-2 genes. *Triticum aestivum*, *T. turgidum* and emmer lines have B-type cytoplasm. Thus, taking ATP6-2 into account, B and G mitochondria were differentiated (Figure 1.7). In the combined analysis, G-type plasmon clustered with cytoplams having ATP6-1c/d due to *i*.) absence of ATP6-2, and *ii*.) presence of a shared sequence in the pre-sequence of ATP6-1a

and ATP6-1c/d. A dissimilarity heatmap, which was generated based on a combined distance matrix (Figure 1.8), revealed that G-type cytoplasm is closer to B-plasmon. On the other hand S and D cytoplasms are more distinct compared to B-type.

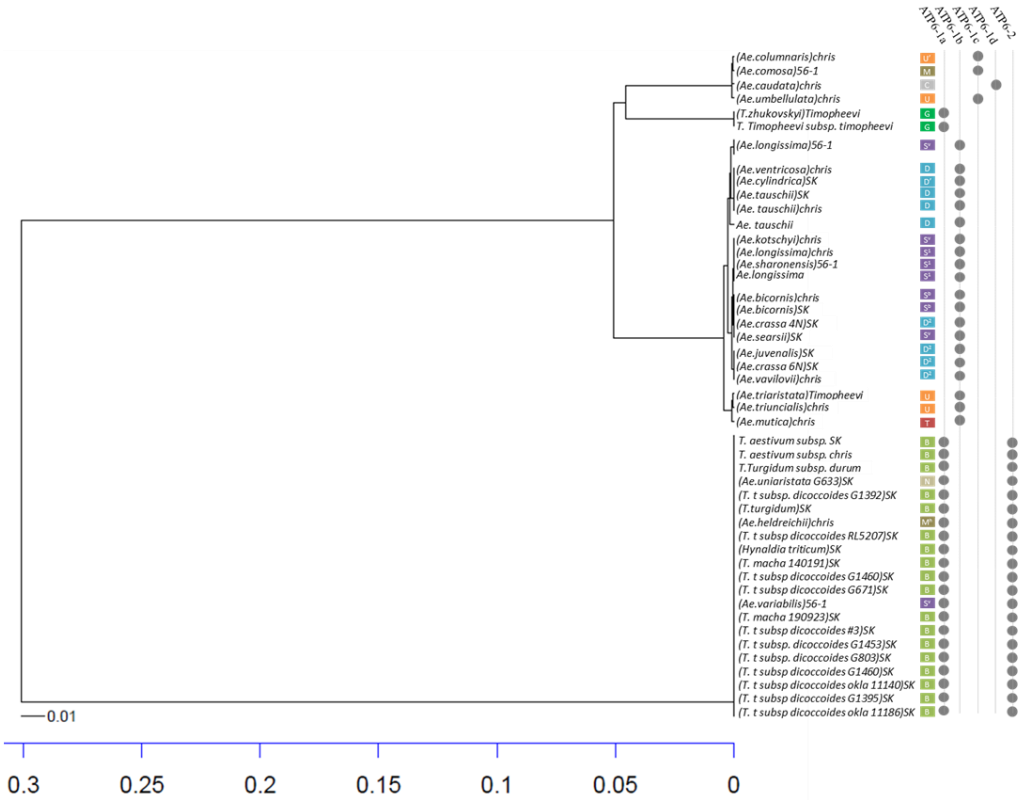

Figure 1.7. Plasmon classification of *Triticum-Aegilops* species based on the *ATP6* genes. Plasmon type of each line identified by previous research is represented in colored boxes (e.g. S in purple) in front of each line. Presence or absence of each *ATP6-1* ortholog or *ATP6-2* gene is identified by the filled circles for each plasmon type.

While alloplasmic lines with the same cytoplasm type clustered together, some exceptional cases were observed. For instance, *Ae. variabilis*, which is believed to belong to the S-type cytoplasm, clustered in the first clade with the B-type plasmons. Further, two types of *Ae. longissima* plasmons were detected. One *Ae. longissima* plasmon type was S-type, detected in euplasmic *Ae. longissima* and the alloplasmic (*Ae. longissima)*Chris. Another *Ae. longissima* plasmon type was identified in the (*Ae. longissima*)56-1 durum line that clustered with D-type plasmons. The last discrepancy involved classifying the U-type plasmons. Although ATP6-1b was identified in *Ae. triuncialis* and *Ae. triaristata* plasmons, the ATP6-

1d ortholog was identified in other U-type *Ae. columnaris* and *Ae. umbellulata*, leading to clustering of these plasmons in different clades.

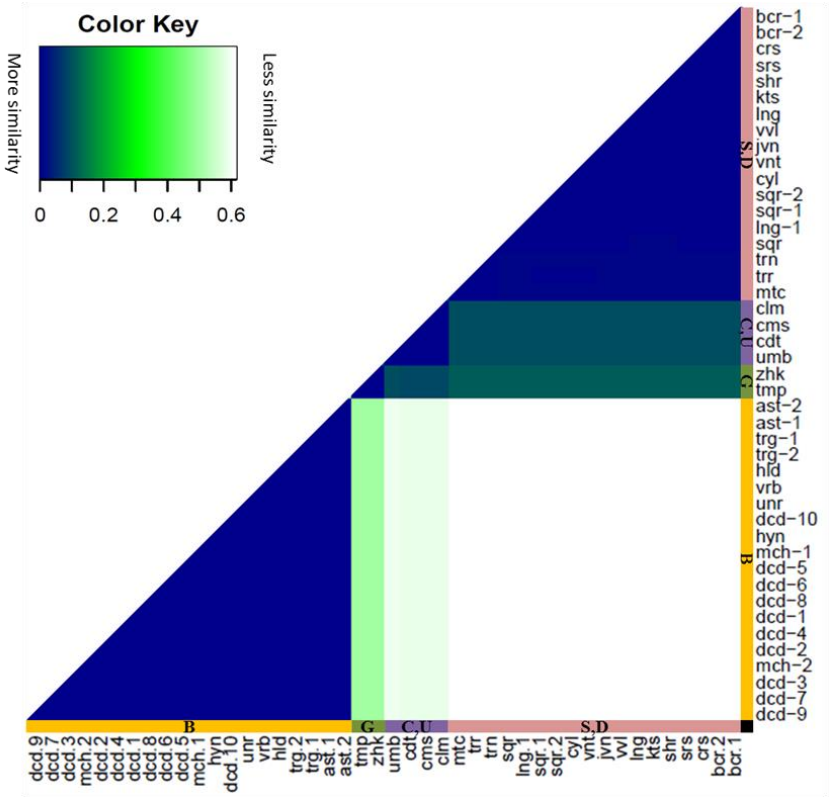

Figure 1.8. Combined distance matrix heatmap considering both the *ATP6-1* and the *ATP6-2* loci. Abbreviated codes in the outer side of each diagonal represents different lines in Table 1.1. Plasmon types for the majority of lines are identified next to each code. Lines having *ATP6-1d* ortholog considered as the C and U type plasmon. However, *Ae. tiuncialis* and *Ae. triaristata* which were previously classified as the U, clustered with the S and D type plasmon in this study.

No stoichiometric shift was detected for the ATP6-1 gene

To investigate the possible effect of nuclear alleles on stoichiometric shift of the ATP6-1 in alloplasmic condition, four cases were analyzed in which a different nucleus was substituted into the same plasmon. In one case, *Ae. tauschii* cytoplasms were clustered together whether the nucleus was Chris, Selkirk or euplasmic *Ae. tauschii* (Figures 1.7 and 1.8). The same results were observed in the other three cases (*T. turgidum*, *Ae. longissima* and *Ae. bicornis*). The evidence presented leads to the conclusion that nuclear alleles do not affect the stoichiometric shifting of the ATP6-1 locus.

Discussion

This research revealed the existence of two novel orthologs of ATP6-1, in addition to those previously reported in the *Triticum-Aegilops* group of species. Although all shared a conserved sequence in the core region, their pre-sequences were different. The conserved region is predicted to provide the essential domains that embed the ATP6 protein in the inner-mitochondrial membrane and facilitate proton transfer (Devenish *et al.* 2000). The role of the diverse N-terminal domain is not well understood. In yeast, ATP6 undergoes a proteolytic cleavage between the Thr-10 and Ser-11 residues of the precursor (Michon *et al.* 1988) removing the N-terminal domain. Interestingly, this cleavage site is co-localized at the border of the core and pre-sequence domains of ATP6 orthologs in the *Triticum-Aegilops* group among other species (Bonen and Bird 1988). Furthermore, this site is similar to a post-translational cleavage site of the yeast cytochrome c oxidase precursor (Maccecchini *et al.* 1979).

At least two nuclear-encoded proteins are involved in efficient post-translational modification and assembly of ATP6 in yeast mitochondria (Zeng *et al.* 2007). One is ATP23p, which is a metalloprotease that cleaves 10 amino acid residues from the ATP6 N-terminus in the inner-membrane space. Two potential homologs of ATP23 were detected on chromosomes 4 and 12 of rice. Data mining in the TIGR transcript assemblies (TAs) database (Zeng *et al.* 2007) identified two corresponding TAs with significant homology confidence in wheat (CJ64383 at 1.8 e-64 and BJ269864 at 3.4 e-74) indicating the conservation of the ATP23 gene among species. However, precise analysis of the transcriptional activity between detected orthologs would require further confirmation. The second protein involved in ATP6 processing is ATP10 (Tzagoloff *et al.* 2004), which together assemble into the ATP synthase complex in close interaction with ATP9 subunits. ATP10 resides on chromosome 5 of rice. Its homolog in wheat (CK171397 at 9.3 e-73) is also conserved and transcriptionally active.

Additionally, It has been demonstrated that unprocessed ATP6 can be assembled in the F_0 complex and be functional (Zeng *et al.* 2007).

These evidences strongly suggest that N-terminal residues of wheat ATP6 may be removed upon translation by a homologous mechanism similar to that characterized in yeast. Conservation of the processing machinery in rice and wheat strengthens this hypothesis. Since the N-terminus cleavage occurs in the mitochondrial inner-membrane space, and the amino acid composition of this region contains mainly hydrophilic residues (Figure 1.3), it can be speculated that the soluble N-terminus domain is essential for functionality of ATP6 and its downstream processing. However, it seems that this mechanistic pathway is different in animals, as evidenced in *Bovine* (Fearnley and Walker 1986).

Two different forces are involved in ATP6-1 evolution

The ATP6-1 orthologs can be distinguished mainly by their diverse pre-sequence. The exact origin of this region is not clear, but it was hypothesized that DNA rearrangements played a more important role than the gradual accumulation of point mutations in their evolution (Bonen and Bird 1988; Kawaura *et al.* 2011; Liu *et al.* 2011; Ogihara *et al.* 2005).

These rearrangements more likely originated from break-induced recombination (BIR) which is a subcategory of homologous recombination (LLORENTE *et al.* 2008). It was recently proposed that BIR plays an important role in plant mitochondrial genome evolution (CHRISTENSEN 2013). BIR promotes high mutability and structural rearrangements especially at the BIR initiation site (MALKOVA and IRA 2013).This is consistent with our results which show higher variability at the ATP6 pre-sequence. This segment is probably closer to the BIR initiation site. Microhomology-mediated BIR (MMBIR) was detected in eukaryotes and can be responsible for copy number variation (HASTINGS *et al.* 2009). Existence of two copies of ATP6 (ATP6-1 and ATP6-2) in wheat and some other species highlights the potential role of

MMBIR in the evolution of ATP6. It was shown that in *E. coli* (PONDER *et al.* 2005) and cancerous cells (COQUELLE *et al.* 1998) the classical BIR switched to MMBIR more frequently in stress conditions. This switch can be more frequent in mitochondria due to greater influence of exposure to stress-induced elements. On the other hand, gradual accumulation of point mutations, observed in the core sequence region of the ATP6 gene, can be a reliable representative of the evolutionary time clock among these species.

In this study two regions were detected within the ATP6 gene that mark different evolutionary pathways: *i.*) the core region (Figure 1.1) in which point mutations are the main evolutionary force and *ii.*) the pre-sequence region (Figure 1.1), where rearrangements are the main cause of differentiation among orthologs. Based on this data, the ATP6-1b and the ATP6-1c/d core sequences shared a more recent common ancestor as compared with the ATP6-1a. This relationship can be interpreted based on the core region point mutation data that follows a slow stepwise mechanism of change. However, a similar pattern of rearrangement that occurred in the pre-sequence of ATP6-1a and ATP6-1c/d resulted in greater homology among these orthologs as compared with ATP6-1b. Point mutation in the pre-sequence further increased variation leading to differentiation within an ortholog.

The nature of point mutations and greater dn than ds in the pre-sequence region indicates that this region is under strong positive selection (YANG and NIELSEN 2000). Most point mutations occurred in the N-terminal region of each ATP6 ortholog, resulting in the replacement of an aa residue (Table 1.4). The non-silent nature of such mutations was previously identified in a comparison between wheat and maize ATP6 sequences (Bonen and Bird 1988). Considering the amount of variation in ATP6 pre-sequence among different species, it can be concluded that this region is under strong positive selection pressure compared with the other regions of the mitochondrial genome (Boer *et al.* 1985; Bonen and

Bird 1988; Bonen *et al.* 1984; Bonen *et al.* 1987). Indeed, the existence of only hydrophilic residues seems sufficient for functionality of this domain.

Classification of the *Triticum-Aegilops* plasmons

To date several criteria have been utilized for classification of the *Triticum-Aegilops* plasmons. However, due to the complex nature of mitochondrial evolution, these criteria should be considered simultaneously in a comprehensive study. Phenotypic effects of each plasmon in a set of different alloplasmic lines was the primary criteria considered for plasmon classification by Tsunewaki and colleagues (Tsunewaki 2009). Forty-seven plasmons have been classified into 15 different groups considering phenotypic effects (Tsunewaki 2009). Chloroplast and mitochondrial genome fingerprinting data clustered the same set of plasmons into 16 and 20 groups, respectively (Tsunewaki 2009). Defining the critical threshold for classification breakpoint is an optional parameter and can be adjusted separately for each criterion. However, more distinct classification of species based on mitochondrial genome sequence may indicate that more mutations can accumulate in this genome due to presence of more repetitive and non-coding DNA. Accumulation of mutations and an increased amount of intra-molecular recombination in this genome facilitates more diversification among species based on this organelle compared with the chloroplast genome. Indeed, four cases were identified in which chloroplast genomic RFLP data did not differentiate plasmons where mitochondrial genome fingerprinting did (Tsunewaki 2009).

This study revealed that ATP6 has several features that make it ideal for evolutionary studies. First, the pre-sequence region of this gene resides on a recombinationally active site, which provides information regarding mitochondrial genome rearrangement patterns. Although these rearrangements cannot be considered good estimators of the evolutionary clock, they are valuable for plasmon classification. Second, this region seems to be unaffected by nuclear mediated stoichiometric shifting facilitating the use of alloplasmic lines

in plasmon classification. The only possible exception was the plasmon of the alloplasmic *Ae. longissima* with a 56-1 nucleu,s which did not cluster with that of euplasmic *Ae. longissima*. This exception may not be the result of stoichiometric shifting but rather because a different *Ae. longissima* accession was used for alloplasmic construction (see below). Third, the point mutation rate in the pre-sequence region is higher than the rest of the gene, which may help elucidate evolutionary relationship among plasmons possessing the same ATP6 ortholog. Finally, the existence of conserved flanking regions facilitates isolation and sequencing of this gene from different species. Based on these observations, the ATP6 gene provides an excellent basis for evolutionary studies in the *Triticum-Aegilops* complex.

Duplication of ATP6 and the B and G plasmon divergence were simultaneous

Both the B and G type plasmons possess identical ATP6-1a genes, indicating they diverged from a common ancestor. Although these plasmons are identical in terms of their ATP6-1a, they can be clearly distinguished by chloroplast and mitochondrial DNA fingerprinting (Mori *et al.* 2009; Wang *et al.* 1997). ATP6-2 in the mitochondria of the B type plasmons co-exists with ATP6-1a; however, in the G-type plasmons it is missing (Figure 1.7). The absence of ATP6-2 in the G plasmon also was supported by the data from Mohr and colleagues (Mohr *et al.* 1993). They showed that *T. timopheevi* mitochondria have only one version of ATP6, which is similar to that of *T. aestivum* ATP6-1. *Triticum zhukovskyi* has the G type plasmon and a hexaploid genome. The absence of ATP6-2 in *T. zhukovskyi* was also shown in this study. Tsunewaki (2009) proposed that the *T. zhukovskyi* plasmon is derived from *T. timopheevi*. The absence of ATP6-2 in both plasmons, the presence of identical ATP6-1a sequence, and their clustering into one group in this study strengthens this hypothesis. Based on this evidence and the constant presence of at least one ATP6-1ortholog in all plasmons, it can be speculated that the emergence of ATP6-1a in the G-type plasmons occurred earlier than the B-type plasmons. The ATP6-2 upstream region (start codon to

position -200) has a higher homology than the upstream region of ATP6-1 to the COXII upstream region (NCBI, NC_007579). This homology indicates that ATP6-2 possibly utilized the COXII regulatory elements. In several species, both the ATP6 and the COX genes reside on recombinationally active sites of the mitochondrial genome (Hanson and Bentolila 2004). A conserved sequence between the upstream regions of ATP6-1a and COXII might provide the necessary homology for BIR and the emergence of a chimeric ATP6-2 gene. Due to the multipartite nature of mitochondrial genomes (Butow 1986) both genes could co-exist and be passed down to the progeny. It is likely that this recombination-mediated duplication was simultaneous with B-type plasmon speciation. Interestingly, no aa change was detected between the coding regions of ATP6-1a from *T. aestivum* and *T. timopheevi* and ATP6-2. The absence of variation among these genes may indicate recent duplication of ATP6 in B-type cytoplasm.

In this study, three plasmons derived from *Ae. variabilis*, *Ae. heldereichii* and *Ae. uniaristata* were clustered with the B type cytoplasm (Figure 1.7). Previously distinct plasmon types were assigned to these species (Tsunewaki 1993). This discrepancy may stem from insufficient coverage of RFLP analysis employed in the previous study (Wang *et al.* 2000). This limitation can hinder detection of all possible polymorphism across the mitochondrial genome, including the ATP6 region (Wang *et al.* 2000). It is also possible that different research groups utilized different accessions of these species for alloplasmic line construction. Precise evolutionary analysis of these plasmons requires a more comprehensive analysis utilizing many more accessions from diverse origins to account for possible heterogeneity and classification discrepancies.

The ATP6-1b gene is associated with the S and D plasmons

The ATP6-1b ortholog was detected in most of the D and S type plasmons. Results of this study indicate that these cytoplasmic genomes are similar based on the absence of ATP6-2. Previous studies (Tsunewaki 1993; Wang *et al.* 2000) classified the D plasmon into three sub-classes; D, D^2 and D'. Based on this classification, D plasmon exists in *Ae. tauschii* and *Ae. ventricosa*, and the D^2 plasmon in *Ae. crassa*, *Ae. juvenalis* and *Ae. vavilovii*. The only member of D' sub-class identified so far is *Ae. cylindrica*. The analysis here shows that the D and D' sub-classes have identical ATP6-1b sequence. This clearly indicates that diploid *Ae. tauschii* was the maternal parent of tetraploid *Ae. ventricosa* and *Ae. cylindrica*. This phylogenic relationship is further substantiated by other studies utilizing mitochondrial- and chloroplast-based markers (Goryunova 2010; Tsunewaki 2009).

The D^2 plasmon members clustered in a group closer to the S-type plasmons (Figure 1.7). Results of this study suggest that tetraploid *Ae. crassa* is the potential parent of the hexaploid D^2 type species. This evolutionary relationship also was proposed by Tsunewaki (2009) utilizing plasmon genome fingerprinting. However, the diploid origin of tetraploid *Ae. crassa* plasmon has not been recognized. Based on nuclear C-banding and FISH analysis (Badaeva *et al.* 2002), one genome (X^{cr}) of tetraploid *Ae. crassa* was believed to be derived from a species belonging to the *Sitopsis* section and another (D^{cr1}) from *Ae. tauschii*. These results indicate that tetraploid *Ae. crassa* was derived from a hybridization between *Ae. tauschii* and a S genome donor. Although mitochondrial and chloroplast genome fingerprinting implicated *Ae. tauschii* as the maternal parent in this hybridization, results presented here indicate to the contrary that the S donor genome was indeed the maternal parent (Figure 1.7). The close relationship of tetraploid *Ae. crassa* to the S^b type species implicates diploid *Ae. bicornis* as the potential S genome and maternal donor (Figure 1.7). Chloroplast and nuclear analysis revealed vast intra-specific diversity within *Ae. tauschii*

(Badaeva *et al.* 2002; Goryunova 2010). Thus it is possible that an extinct accession of *Ae. tauschii* differentiated to contain the D^2 plasmon. Furthermore, it also can be postulated that although *Ae. bicornis* was the maternal parent of tetraploid *Ae. crassa* plasmon, the introgression of the D nuclear genome from the paternal donor resulted in drastic structural modifications in the S mitochondrial genome, altering its fingerprinting pattern to be more similar to the D counterpart. Similar changes in the nuclear S genome resulted in the formation of B and G genomes of polyploid wheat. It can then be argued that both the nuclear and cytoplasmic genomes of the ancestral S genome were dramatically modified during the evolution of the *Triticum-Aegilops* group of species.

Five different species that contained the S type cytoplasm were utilized in this study; *Ae. bicornis*, *Ae. searsii*, *Ae. kotschyi*, *Ae. sharonensis* and *Ae. longissima*. Previous studies (Tsunewaki 1993) demonstrated high similarity among these S subgroups for both mitochondria and chloroplast fingerprinting. Diploid *Ae. sharonensis* and *Ae. longissima* were clustered together in this study, confirming previous reports regarding their close relationship. Tetraploid *Ae. kotschyi* plasmon was clustered with these diploid species, indicating that one of these species was the likely donor of the cytoplasm. Based on chromosome C-banding and FISH analysis, Badaeva and colleagues (Badaeva *et al.* 2004) proposed that *Ae. kotschyi* is derived from a hybridization of *Ae. umbellulata* and *Ae. sharonensis*. The results presented here indicate that *Ae. sharonensis* served as the maternal parent in that hybridization.

A unique form of nuclear-cytoplasmic interaction was characterized in alloplasmic lines with *Ae. longissima* (S type) and *Ae. tauschii* (D type) cytoplasm (Asakura *et al.* 1997; Maan 1992b). In these lines, nuclear-cytoplasmic compatibility can be established by the presence of a specific nuclear gene(s) derived from the cytoplasm donor or other species. Two accessions of *Ae. longissima* were identified previously which behaved differently in

regard to their interaction with the durum nucleus in the alloplasmic condition (Maan 1996). One accession, which was designated as *b*, produced a fertile (*Ae. longissima)* durum line while the other, designated *c*, did not produce viable alloplasmic lines (Maan 1996). In the *c* accession, the presence of two genes, *scs* (*Species Cytoplasm Specific*) and *Vi* (*Vitality*), are necessary for complete recovery of vigor and fertility, respectively (Maan 1992a). Here two types of *Ae. longissima* plasmon also were identified according to their ATP6-1b sequence (Figure 1.7). One type grouped with the S plasmons and the other with the D plasmons. The later *Ae. longissima* type is indeed the *c* accession that was incompatible with the durum nucleus. Interestingly, the nuclear-cytoplasmic compatibility of both *Ae. tauschii* and *Ae. longissima* (*c* accession) cytoplasm with the durum nucleus can be ameliorated by the same gene system (Asakura *et al.* 1997; Maan 1992b). This phenomenon may be due to the fact that these plasmons share the same mitochondrial genome sequence and/or structure.

Classification of the U, C and T cytoplasm

Three different plasmon types have been identified in *Ae. triuncialis* species (Endo 1975). The first type is similar to *Ae. umbellulata*, the second is closer to *Ae. caudata*, and the last shows higher similarity to *Ae. mutica* plasmon (Endo 1975). The *Ae. triuncialis* accession used in this study possesses the ATP6-1b, and classified with *Ae. mutica* and *Ae. triaristata* cytoplasm, indicative of their close relationship. Homology between *Ae. triuncialis* and *Ae. triaristata* cytoplasm also was indicated by previous studies (Tsunewaki 1993). Previous studies revealed that, although the chloroplast genome of *Ae. mutica* and *Ae. triuncialis* plasmon are similar to each other (Provan *et al.* 2004; Tsunewaki 1993; Wang *et al.* 1997), their mitochondrial genomes are quite different (Tsunewaki 1993; Wang *et al.* 2000). In our study, based only on mitochondrial gene sequence and not the structure, similar classification as that based on chloroplast genome RFLP is derived. There has been much debate over the rate of evolutionary change in the mitochondria and chloroplast genomes - both in terms of

point mutation and recombination. Although Tsunewaki (1993) concluded evolutionary changes occur at a faster rate in the mitochondria than in the chloroplast, the opposite conclusion was reached by other groups (Wang *et al.* 1997). Results presented here indicate that point mutations accumulate at the same rate in the mitochondrial ATP6-1 region and the chloroplast genome. However, structural alterations in the mitochondrial genome, which occur more frequently, can lead to higher diversification and observed discrepancy between mitochondria and chloroplast fingerprinting data of *Ae. mutica*- and *Ae. triuncialis* -type cytoplasm.

Aegilops umbellulata also carries the U type cytoplasm, possessing ATP6-1c in its mitochondrial genome. The ATP6-1c ortholog also has been identified in *Ae. comosa* and *Ae. columnaris* plasmons. Results presented here indicate that tetraploid *Ae. columnaris* ATP6-1c has higher homology to diploid *Ae. columnaris* ATP6-1c. This evidence is consistent with previous results regarding the likelihood of *Ae. umbellulata* as the cytoplasm donor of *Ae. columnaris* (Tsunewaki 2009). *Aegilops caudata* was the only plasmon with ATP6-1d orthologs.

Conclusions

Cytoplasmic genetics is intricate, elegant, and complex. However, close study reveals insights into mitochondrial genome alteration and evolution. Detailed analysis of genes and gene classes provides a useful and informative model for evolutionary studies within species. In the current study, we investigated ATP6 variability in the *Triticum-Aegilops* species in an evolutionary context, characterizing gene gain/loss, rearrangement patterns and stoichiometric shifts. Four major ATP6-1 orthologs were detected, point mutations were characterized within each ortholog, and presence/absence of ATP6-2 was established as a possible recent evolutionary event. These findings illuminate the critical nature of ATP6 functionality and diversification, and contribute to a more complete understanding of the

evolutionary process in the *Triticum-Aegilops* species. Although ATP6 can provide a suitable model for evolutionary studies within species, whole genome sequencing data of both chloroplast and mitochondria are needed for a comprehensive explanation of plasmon genetics and its evolution.

Acknowledgements

We would like to thank Professor SS Maan and Professor K Tsunewaki for the development of the alloplasmic wheat lines and providing us the material used in this study. We are also grateful for all the assistance received from members of the Wheat Germplasm Enhancement and the Hard Red Spring Wheat programs at NDSU. Particular thanks to Drs. Philip McClean, Justin Faris and Steven Xu for their valuable suggestions that improved the manuscript. This research was supported by the Monsanto Beachell-Borlaug International Scholars program to SFK.

References

Asakura, N., C. Nakamura and I. Ohtsuka, 1997 A nuclear compatibility gene, Ncc-tmp, of *Triticum timopheevi* for the cytoplasm of *Aegilops tauschii*. Genes & Genetic Systems 72: 71-78.

Badaeva, E. D., A. V. Amosova, O. V. Muravenko, T. E. Samatadze, N. N. Chikid et al., 2002 Genome differentiation in *Aegilops*. 3. Evolution of the D-genome cluster. Plant Systematics and Evolution 231: 163-190.

Badaeva, E. D., A. V. Amosova, T. E. Samatadze, S. A. Zoshchuk, N. G. Shostak et al., 2004 Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. Plant Systematics and Evolution 246: 45-76.

Boer, P. H., J. E. McIntosh, M. W. Gray and L. Bonen, 1985 The Wheat Mitochondrial Gene for Apocytochrome-B - Absence of a Prokaryotic Ribosome Binding-Site. Nucleic Acids Research 13: 2281-2292.

Bonen, L., and S. Bird, 1988 Sequence-Analysis of the Wheat Mitochondrial Atp6 Gene Reveals a Fused Upstream Reading Frame and Markedly Divergent N Termini among Plant Atp6 Proteins. Gene 73: 47-56.

Bonen, L., P. H. Boer and M. W. Gray, 1984 The Wheat Cytochrome-Oxidase Subunit-Ii Gene Has an Intron Insert and 3 Radical Amino-Acid Changes Relative to Maize. Embo Journal 3: 2531-2536.

Bonen, L., P. H. Boer, J. E. McIntosh and M. W. Gray, 1987 Nucleotide-Sequence of the Wheat Mitochondrial Gene for Subunit-I of Cytochrome-Oxidase. Nucleic Acids Research 15: 6734-6734.

Breiman, A., 1987 Mitochondrial-DNA Diversity in the Genera of *Triticum* and *Aegilops* Revealed by Southern Blot Hybridization. Theoretical and Applied Genetics 73: 563-570.

Butow, R. A., 1986 Mitochondrial-DNA - Rearranging the Plant Genome. Nature 324: 620-620.

Christensen, A. C., 2013 Plant Mitochondrial Genome Evolution Can Be Explained by DNA Repair Mechanisms. Genome Biology and Evolution **5:** 1079-1086.

Coquelle, A., F. Toledo, S. Stern, A. Bieth And M. Debatisse, 1998 A new role for hypoxia in tumor progression: Induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs. Molecular Cell **2:** 259-265.

Devenish, R. J., M. Prescott, X. Roucou and P. Nagley, 2000 Insights into ATP synthase assembly and function through the molecular genetic manipulation of subunits of the yeast mitochondrial enzyme complex. Biochimica Et Biophysica Acta-Bioenergetics 1458: 428-442.

Eisenberg, D., R. M. Weiss, T. C. Terwilliger and W. Wilcox, 1982 Hydrophobic Moments and Protein-Structure. Faraday Symposia of the Chemical Society: 109-120.

Endo, T. R., Tsunewaki K., 1975 Genetic diversity of the cytoplasm of *Triticum* and *Aegilops*. I. On the origin of the cytoplasm of *Ae. triuncialis* L. Seiken Ziho 25-26: 55-66.

Fearnley, I. M., and J. E. Walker, 1986 2 Overlapping Genes in Bovine Mitochondrial-DNA Encode Membrane-Components of Atp Synthase. Embo Journal 5: 2003-2008.

Galtier, N., 2011 The intriguing evolutionary dynamics of plant mitochondrial DNA. BMC Biology 9.

Goryunova, S. V., Chikida, N. N., Badaeva, E. D., & Pukhal'skii, V. A., 2010 Polymorphism and phylogeny of D-genome species of *Aegilops* L. Based on the results of analysis of the microsatellite loci of the chloroplast genome. Moscow University Biological Sciences Bulletin 65: 148-151.

Hall, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95-98.

Hanson, M. R., and S. Bentolila, 2004 Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16: S154-S169.

Hastings, P. J., G. Ira and J. R. Lupski, 2009 A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation. Plos Genetics **5**.

Hirosawa, S., S. Takumi, T. Ishii, T. Kawahara, C. Nakamura et al., 2004 Chloroplast and nuclear DNA variation in common wheat: insight into the origin and evolution of common wheat. Genes & Genetic Systems 79: 271-282.

Ishii, T., N. Mori and Y. Ogihara, 2001 Evaluation of allelic diversity at chloroplast microsatellite loci among common wheat and its ancestral species. Theoretical and Applied Genetics 103: 896-904.

Juretic, D., A. Jeroncic and D. Zucic, 1999 Sequence analysis of membrane proteins with the Web server SPLIT. Croatica Chemica Acta 72: 975-997.

Kawaura, K., A. Saeki, T. Masumura, S. Morita and Y. Ogihara, 2011 Heteroplasmy and expression of mitochondrial genes in alloplasmic and euplasmic wheat. Genes & Genetic Systems 86: 249-255.

Kitazaki, K., and T. Kubo, 2010 Cost of Having the Largest Mitochondrial Genome: Evolutionary Mechanism of Plant Mitochondrial Genome. Journal of Botany 2010

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan et al., 2007 Clustal W and clustal X version 2.0. Bioinformatics 23: 2947-2948.

Libardo, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.

Liu, H. T., P. Cui, K. H. Zhan, Q. Lin, G. Y. Zhuo et al., 2011 Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. BMC Genomics 12.

Liu, K. J., and S. V. Muse, 2005 PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128-2129.

Llorente, B., C. E. Smith and L. S. Symington, 2008 Break-induced replication - What is it and what is it for? Cell Cycle **7:** 859-864.

Maan, S. S., 1975 Cytoplasm variability and speciation in *Triticeae*, pp. 255-281 in Prairie: A multiple view., edited by M. K. WALI, Univ. of North Dakota Press, Grand Forks, ND.

Maan, S. S., 1991 Nucleo-cytoplasmic genetics of wheat., pp. 175-194 in Proc. Int. Symp. Nuclear and Organeller Genetics of Wheat., edited by T. Saskuma and Kinoshita, Hokkaido University, Sapporo, Japan.

Maan, S. S., 1992a The *Scs* and *Vi* Genes Correct a Syndrome of Cytoplasmic Effects in Alloplasmic Durum-Wheat. Genome 35: 780-787.

Maan, S. S., 1992b Transfer of a *Species Cytoplasm Specific* (*Scs*) Gene from *Triticum-timopheevii* to *Triticum turgidum*. Genome 35: 238-243.

Maan, S. S., 1996 Natural or induced nucleocytoplasmic heterogeneity in *Triticum longissimum*. Genome 39: 71-76.

Maan, S. S., and T. R. Endo, 1991 Nucleo-cytoplasmic interactions stabilize ploidy level in wheat interspecific hybrids. Genome 34: 983-987.

Maccecchini, M. L., Y. Rudin and G. Schatz, 1979 Transport of Proteins across the Mitochondrial Outer-Membrane - Precursor Form of the Cytoplasmically Made Intermembrane Enzyme Cytochrome C Peroxidase. Journal of Biological Chemistry 254: 7468-7471.

Malkova, A., and G. Ira, 2013 Break-induced replication: functions and molecular mechanism. Current Opinion in Genetics & Development **23:** 271-279.

Michon, T., M. Galante and J. Velours, 1988 Nh2-Terminal Sequence of the Isolated Yeast Atp Synthase Subunit-6 Reveals Post-Translational Cleavage. European Journal of Biochemistry 172: 621-625.

Mohr, S., E. Schultekappert, W. Odenbach, G. Oettler and U. Kuck, 1993 Mitochondrial-DNA of Cytoplasmic Male-Sterile *Triticum timopheevi* - Rearrangement of Upstream Sequences of the Atp6 and Orf25 Genes. Theoretical and Applied Genetics 86: 259-268.

Mori, N., Y. Kondo, T. Ishii, T. Kawahara, J. Valkoun et al., 2009 Genetic diversity and origin of timopheevi wheat inferred by chloroplast DNA fingerprinting. Breeding Science 59: 571-578.

Nei, M., and T. Gojobori, 1986 Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. Molecular Biology and Evolution **3:** 418-426.

Noyszewski, A., F. Ghavami, L. Al-Nimer, A. Soltani, Y. Gu *et al.*, 2014 Accelerated evolution of the mitochondrial genome in an alloplasmic line of durum wheat. BMC Genomics **15**.

Ogihara, Y., Y. Yamazaki, K. Murai, A. Kanno, T. Terachi et al., 2005 Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Research 33: 6235-6250.

Palmer, J. D., and L. A. Herbon, 1988 Plant Mitochondrial-DNA Evolves Rapidly in Structure, but Slowly in Sequence. Journal of Molecular Evolution 28: 87-97.

Paradis, E., J. Claude and K. Strimmer, 2004 APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20: 289-290.

Provan, J., P. Wolters, K. H. Caldwell and W. Powell, 2004 High-resolution organellar genome analysis of *Triticum* and *Aegilops* sheds new light on cytoplasm evolution in wheat. Theoretical and Applied Genetics 108: 1182-1190.

Ponder, R. G., N. C. Fonville and S. M. Rosenberg, 2005 A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. Molecular Cell **19:** 791-804.

R Development Core Team, 2010 R: A Language and Environment for Statistical Computing, pp. R Foundation for Statistical Computing, Vienna, Austria.

Schliep, K. P., 2011 phangorn: phylogenetic analysis in R. Bioinformatics 27: 592-593.

Schultz, J., F. Milpetz, P. Bork and C. P. Ponting, 1998 SMART, a simple modular architecture research tool: Identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America 95: 5857-5864.

Sloan, D. B., A. J. Alverson, J. P. Chuckalovack, M. Wu, D. E. Mccauley et al., 2012 Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. Plos Biology 10.

Terachi, T., and K. Tsunewaki, 1992 The Molecular-Basis of Genetic Diversity among Cytoplasms of *Triticum* and *Aegilops* .8. Mitochondrial Rflp Analyses Using Cloned Genes as Probes. Molecular Biology and Evolution 9: 917-931.

Timmis, J. N., M. A. Ayliffe, C. Y. Huang and W. Martin, 2004 Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. Nature Reviews Genetics 5: 123-U116.

Triboush, S. O., N. G. Danilenko and O. G. Davydenko, 1998 A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. Plant Molecular Biology Reporter 16: 183-189.

Tsunewaki, K., 1980 Genetic diversity of the cytoplasm in *Triticum* and *Aegilops*. Japan Society for the Promotion of Science.

Tsunewaki, K., 1993 Genome-Plasmon Interactions in Wheat. Japanese Journal of Genetics 68: 1-34.

Tsunewaki, K., 2009 Plasmon analysis in the *Triticum-Aegilops* complex. Breeding Science 59: 455-470.

Tzagoloff, A., A. Barrientos, W. Neupert and J. M. Herrmann, 2004 Atp10p assists assembly of atp6p into the F-0 unit of the yeast mitochondrial ATPase. Journal of Biological Chemistry 279: 19775-19780.

Vedel, F., F. Quetier, F. Dosba and G. Doussinault, 1978 Study of Wheat Phylogeny by *Eco*RI Analysis of Chloroplastic and Mitochondrial DNAs. Plant Science Letters 13: 97-102.

Wang, G. Z., Y. Matsuoka and K. Tsunewaki, 2000 Evolutionary features of chondriome divergence in *Triticum* (wheat) and *Aegilops* shown by RFLP analysis of mitochondrial DNAs. Theoretical and Applied Genetics 100: 221-231.

Wang, G. Z., N. T. Miyashita and K. Tsunewaki, 1997 Plasmon analyses of *Triticum* (wheat) and *Aegilops*: PCR single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs. Proceedings of the National Academy of Sciences of the United States of America 94: 14570-14577.

Xiong, A. S., R. H. Peng, J. Zhuang, F. Gao, B. Zhu et al., 2008 Gene duplication and transfer events in plant mitochondrial genome. Biochemical and Biophysical Research Communications 376: 1-4.

Yang, Z. H., and R. Nielsen, 2000 Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution **17:** 32-43.

Zeng, X. M., W. Neupert and A. Tzagoloff, 2007 The metalloprotease encoded by ATP23 has a dual function in processing and assembly of subunit 6 of mitochondrial ATPase. Molecular Biology of the Cell 18: 617-626.

CHAPTER II. AN INTEGRATIVE APPROACH TO STUDY NUCLEAR-CYTOPLASMIC INTERACTIONS (NCI) IN WHEAT (*TRITICUM AESTIVUM*) ALLOPLASMIC LINES

Abstract

Three different genome components, *vis.* mitochondria, plastids, and the nucleus co-exist in every plant and algae cell. A compatible interaction between these genomes is required for the evolution and adaptation of organisms. This interaction, however, can be dissociated and re-constituted by frequent interspecific hybridizations. Several studies have addressed the effect of interspecific hybridization on nuclear intra-genomic interactions. However, the magnitude and essence of inter-genomic interactions between nuclear and cytoplasmic genomes remains unknown. To study these aspects of inter-genomic interactions, we utilized alloplasmic wheat lines *Triticum aestivum* cv. Chris and *T. aestivum* cv. Selkirk possessing alien cytoplasm derived from *Ae. mutica*. An integrated approach was employed including comparative QTL (Quantitative Trait Locus), epigenetic and expression analyses to elucidate the effect of cytoplasmic genomes on nuclear gene interaction and expression. QTL analysis was conducted for three physiological traits: dry matter weight, height and spike number per plant. QTL controlling each trait were detected separately for two F_2 populations, both segregating for the same nuclear alleles (Selkirk and Chris) but one possessing the original B type cytoplasm and another with *Ae. mutica* T-type cytoplasm. The methylation profile of alloplasmic lines was compared with their corresponding euplasmic line. Polymorphic methylated fragments were sequenced and annotated. The expression patterns of polymorphic methylated genes were investigated using q-RT-PCR. Our results indicate that different QTL control the same trait in alloplasmic and euplasmic populations. A QTL on chromosome 5A was detected that controls several traits in an interactive way with cytoplasm. Furthermore, eight genic regions were detected that exhibited methylation

patterns affected by cytoplasm type. The genes in those regions belong mostly to plant defense and carbon metabolism pathways. Expression analysis revealed that the transcription level of two genes significantly differ between alloplasmic and euplasmic lines. In general, the present study showed that cytoplasmic genomes drastically affect the nuclear gene network controlling physiological traits. This effect can be at least partially due to epigenetic regulation that can consequently influence gene expression

Introduction

More than ninety-five percent of organellar-targeted proteins including structural, enzymatic and regulatory components are encoded by the nucleus and then transferred to organelles (anterograde signaling). Alternately, organelles regulate nuclear genes through retrograde signaling. These retrograde signals mostly regulate nuclear genes involved in photosynthetic and metabolic pathways. Maintenance of regulatory cross-talk between nuclear and organellar genomes is crucial for the functionality and survival of organisms under evolutionary pressure. However, regulatory communication can be dissociated and re-constituted by interspecific hybridization.

Interspecific hybridization is widely recognized as a common evolutionary phenomenon that imparts new species with superior characteristics (ADAMS and WENDEL 2005; SOLTIS and SOLTIS 2009). This superiority is believed to stem from the emergence of novel intra-genomic interactions between nuclear genomes, as well as interactions between nuclear and cytoplasmic genomes. These novel interactions can ultimately facilitate the evolution of organisms and speciation.

In the first step following hybridization, early polyploids experienced “genetic shock”, a term first proposed by McClintock (1984). Genetic shock enables neoallopolyploids to overcome two major barriers: correct chromosomal pairing, and

balanced gene expression. Vast alterations in genetic and epigenetic content seem necessary to facilitate the co-existence of divergent genomes within the same cell (CHEN 2007). In this coordinated process, retro-element activation is enhanced, which results in genetic-level modifications (MADLUNG *et al.* 2005). This trimming step, along with other related modifications, is associated with elimination of genome-specific sequences and chromosomal rearrangements. This process ensures efficient pairing of homologous chromosomes, which is crucial for allopolyploid reproduction (CHEN 2007).

Along with genetic modifications, drastic regulation in gene expression - particularly among homeologous genes - is necessary to maintain the functionality and homeostasis of metabolic pathways in new allopolyploids. Several interrelated mechanisms have been identified that regulate the expression network including cis- and trans- mechanisms. These mechanisms establish the novel intra- and inter-genic interactions in allopolyploids. The majority of homeologous gene interactions are additive, which indicates the divergence of cis-elements between genomes. However, non-additive interactions have been detected among regulatory components or/and epigenetic regulation (JACKSON and CHEN 2010).

Although there are species [for example, *Brassica* (NAGAHARU 1935) and *Tragopogon* (SOLTIS and SOLTIS 1999) genera] where the reciprocal crosses produced natural allopolyploid lines, most historic interspecific crosses are recognized as only maternal-specific hybridizations. In *Triticum* (TSUNEWAKI 2009), *Aegilops* (TSUNEWAKI 2009), *Gossypium* (WENDEL and CRONN 2003) and *Arabidopsis* (SALL et al. 2003) genera, allopolyploid lines were produced and established in a maternal-specific manner. Furthermore, specific cytoplasm types have been identified that contribute adaptability to specific habitats (ACOSTA and PREMOLI 2010; ALLAINGUILLAUME *et al.* 2009). This evidence suggests the importance of nuclear-cytoplasmic interaction (NCI) upon hybridization.

There are generally two interrelated mechanisms by which cytoplasmic organelles, including mitochondria and chloroplasts, can control the nuclear gene expression: *i.*) retrograde signaling and *ii.*) imprinting. The retrograde mechanism is a signaling pathway, originating from mitochondria and/or chloroplasts, which coordinates the corresponding nuclear genes (WOODSON and CHORY 2008). Several candidate gene systems have been proposed in plants that function as retrograde signals, including redox (reduction/oxidation) signals (FEY *et al.* 2005), Mg-proto (STRAND *et al.* 2003) and reactive oxygen species (ROS) (LEE *et al.* 2007; RHOADS and SUBBAIAH 2007). These particular signals co-regulate a subset of nuclear genes involved in photosynthesis, stress-response or organelle structure. As such, these signals are an indirect representation of mitochondria/chloroplast states by which nuclear genes can co-regulate their expression to maintain balance between different pathways.

Parental regulation of gene expression is a well-documented mechanism known as imprinting (GEHRING 2013). In this process, the allele transmitted by either the maternal or paternal parent is expressed in the progeny. Epigenetic regulation seems to be the predominant mechanism controlling imprinting (GEHRING 2013). Although imprinting regulation was detected in different developmental stages in mammals, the expression of detected imprinted genes in plants has been limited to endosperm development (BELMONTE *et al.* 2013; HE *et al.* 2010; KOHLER *et al.* 2003).

Development of a near-complete set of alloplasmic lines (lines with alien cytoplasm) provides a valuable resource to study nuclear-cytoplasmic interactions (NCI) in wheat (TSUNEWAKI 2009). Several alien cytoplasms from *Aegilops* species have replaced the B-type cytoplasm commonly detected in euplasmic *Triticum aestivum* and *T. turgidum* cultivars (MAAN 1991; TSUNEWAKI *et al.* 1996). In these alloplasmic lines, a vast range of phenotypic modifications have been observed due to new NCI.

To elucidate novel insights into NCI, two alloplasmic wheat lines possessing *Ae. mutica* cytoplasm were compared with their corresponding euplasmic lines (see below). These lines were used in the present study to fulfill three critical objectives: *i.*) to identify major loci that control the same physiological traits in alloplasmic and euplasmic lines. This comparative analysis will help us to ultimately identify the gene(s) that interact with each type of cytoplasm. Such gene(s) can be considered as potential candidates that channelize the cytoplasm-specific direction of polyploid evolution. *ii.*) to assess the possible effect of alien cytoplasm on the methylation profile of nuclear genes; and *iii.*) to analyze the effect of methylation modification on the expression patterns of alloplasmic and euplasmic lines for selected genes.

Material and Methods

Plant material

The nucleus of two wheat cultivars including *T. aestivum* cv. 'Selkirk' and *T. aestivum* cv. 'Chris' were transferred to the *Ae. mutica* cytoplasm through recurrent backcrossing (MAAN 1977). In the alloplasmic condition, the cytoplasm derived from *Ae. mutica*, which is categorized as a T-type plasmon, induces hybrid vigor when it replaces the wheat cultivar Chris cytoplasm (Figure 2.1). However, this drastic alteration is not detectable in alloplasmic lines with the Selkirk nucleus. This differential effect facilitates the identification of causal loci via a comparative QTL analysis approach.

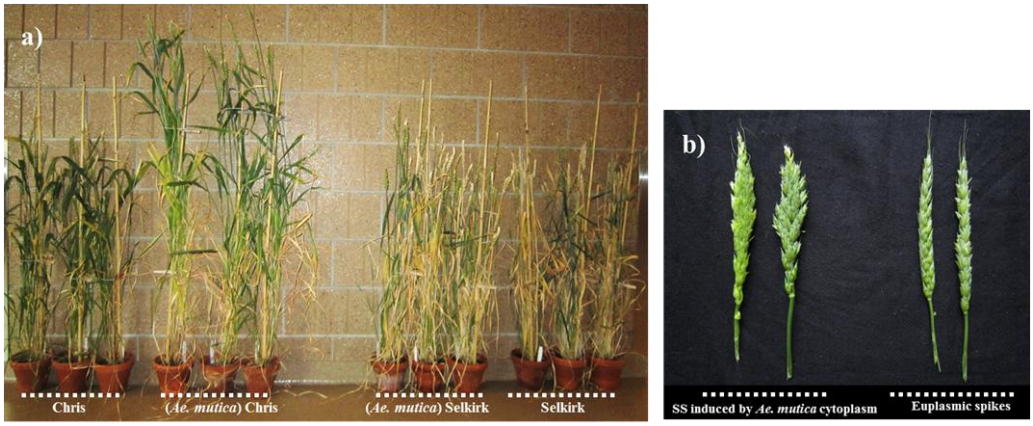

Figure 2.1. Phenotypic differences between euplasmic and alloplasmic lines. a) Drastic increase in height is observable in (*Ae. mutica*) Chris compared with the euplasmic Chris line. b) Supernumerary spikelets (SS) is another phenotype observed in alloplasmic wheat lines with *Ae. mutica* cytoplasm

Developing the mapping populations

Two F_2 populations were developed to identify the potential QTL that are interacting with cytoplasm. The first one contained 282 individuals derived from (*Ae. mutica*) *T. aestivum* cv. Selkirk × (*Ae. mutica*) *T. aestivum* cv. Chris, which is segregating for nuclear alleles in *Ae. mutica* cytoplasm. The second F_2 population contained 245 individuals developed based on a cross between euplasmic *T. aestivum* cv. Selkirk × *T. aestivum* cv. Chris. Cytoplasmic interacting loci were then identified through a comparison between population-specific QTL.

Phenotyping

Both eu- and alloplasmic F_2 populations were planted in controlled conditions in two seasons at the Agricultural Experiment Station Research Greenhouse Complex at North Dakota State University, Fargo. The first season (Spring 2013) included 102 eu- and 148 alloplasmic individuals and the following season (Fall 2013) consisted of 143 eu- and 134 alloplasmic lines. Corresponding parental lines along with the F_1 plants were also grown under the same condition in each season. Three traits including above-ground dry matter weight (DMW), head number and height were measured after plant maturation for both populations. Supernumerary spikelet (SS) was measured in alloplasmic populations.

Genotyping

Total DNA of each F_2, F_1 and parental line was isolated using a glass-filter high-throughput DNA extraction protocol. DNA quality was assessed through visual observation on a 1% agarose gel, and DNA concentrations were measured with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). Isolated DNA from the 2013 spring and fall seasons were genotyped using wheat 9K and 90K SNP Infinium assays (Illumina), respectively.

Genetic map construction

GenomeStudio software (Illumina, Inc. San Diego, CA, USA) was used to visualize, cluster and call each SNP class. Genotypic data from both 9 and 90 K platforms were imported into R software (R DEVELOPMENT CORE TEAM 2010) and two genetic maps were constructed for each platform separately. Individuals with high missing genotypic data and markers with high missing individuals were removed. Furthermore, markers with significant (P=0.05) distortion from the expected 1:2:1 ratio were removed from the analysis. Linkage groups were formed considering 0.35 and 10 for maximum recombination fraction (*max.rf*) and minimum LOD (*min.lod*), respectively. The markers of each linkage group were ordered by the *orderMarker* function. The number of total crossover events and crossovers per chromosome were calculated for each individual using the *countXO* function in the same package.

Single QTL analysis

To detect main effect QTL, both alloplasmic and euplasmic populations were analyzed separately using two different software packages. A whole genome QTL scan with a single QTL model was performed using the function *scanone*, available in R/qtl package (version 2.15.1). For this purpose, default settings were used to perform standard interval mapping (Lander and Botstein 1989). To measure the QTL effect: first, the closest marker to

each QTL was determined followed by the calculation of phenotypic mean for each genotypic class. The data was also analyzed using composite interval mapping (CIM), available in QTLCarthographer V2.5 (WANG *et al*. 2012). In this method, for QTL detection, model 6 with forward and backward step-wise regression, five markers as cofactors to control genetic background, and a 10 cM genome-wide scan window were used. In both QTL detection methods, one thousand permutations were used to determine the significant threshold LOD at P=0.05. Every locus above the threshold LOD score was called a definitive QTL; those loci with P=0.1 to 0.05 were considered putative QTL.

Two-dimensional QTL analysis

The R/qtl package (version 2.15.1) also was used to identify major two-locus interactions. For this purpose, the *scantwo* function was performed to analyze the two-dimensional QTL relationship between all possible marker pairs. Four models were considered while calculating interactions using the *scantwo* function. Briefly, *i.*) full model [$y = \mu + \beta_1 q_1 + \beta_2 q_2 + \beta_3 (q_1 \times q_2) + \varepsilon$]; *ii.*) additive model [$y = \mu + \beta_1 q_1 + \beta_2 q_2 + \varepsilon$]; *iii.*) single model [$y = \mu + \beta_1 q_1 + \varepsilon$]; and *iv*) null model, without any interactions [$y = \mu + \varepsilon$] were considered.

Full LODs were calculated comparing full model values to the null model values. Additive LODs were estimated by comparing additive model values to the null model values, and interaction LODs were measured by comparing full model values to the additive model values. Interactions with highest LODs were selected to make the QTL model using the *makeqtl* function. The effect of each interaction was estimated by the *addint* function.

Comparative methylation analysis

Three replicates of both alloplasmic lines, *i.e.*, (*Ae. mutica*) Chris and (*Ae. mutica*) Selkirk, and their corresponding euplasmic lines *T. aestivum* cv. Chris and *T. aestivum* cv.

Selkirk were grown in controlled greenhouse conditions. Total DNA was isolated from leaf tissue of each replicate at the seedling stage. The DNA quality and quantity was assessed through visual observation on a 1% agarose gel and with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA).

Each DNA sample was double-digested by *Eco*RI-HF/*Hpa*II (NEB, Ipswich, MA, USA) enzymes at 37ºC for three hours. The digestion reaction was prepared in a volume of 20 µl and contained 20 units of each enzyme, 200 ng of template DNA and 1X CutSmart buffer. After digestion, the enzymes were de-activated by incubating the reaction at 80 ºC for 20 min. Two different adaptors including *Eco*RI and *Hpa*II were ligated to the digested fragments. The ligation buffer contained 5 µM of *Eco*RI and 50 µM of *Hpa*II adaptors (Table 2.1), 100 units of T4 DNA ligase (NEB, Ipswich, MA, USA) and the entire digested reaction from the previous step. This reaction was carried out by incubating the samples overnight at 16ºC. The digest/ligation reaction was then diluted at a ratio of 1:5 with distilled water.

Pre-selective PCR amplification was performed using primers (Table 1.1) designed based on the adaptor sequence plus an additional single selective base at the 3' end. Each pre-selective PCR reaction contained 10 pmol of each pre-selective primer, 0.25 mM of each dNTP, 1X Taq DNA polymerase buffer (Promega Corporation, Madison, WI, USA), 0.75 units of Go Taq® Flexi DNA polymerase (Promega Corporation, Madison, WI, USA) and 5µl of the 1:5 diluted digest-ligation reaction. PCR amplification was conducted using the Applied Biosystems® 2720 Thermal Cycler and the following program: initial denaturation for 4 min. at 94 °C followed by twenty-five cycles of one min. each at 94°C, 56°C and 72°C and a final extension of five min. at 72 °C.

Selective PCR amplification was conducted using selective primer combinations (Table 1.1). Selective primers were similar to pre-selective primers, but with two or three

additional variable nucleotides at the 3' end to selectively reduce the amplified fragment complexity. The selective amplification was conducted using an Applied Biosystems® 2720 Thermal Cycler and the following program: initial denaturation for 4 min. at 94°C followed by 35 cycles of one min. each at 94°C, 56°C and 72°C and a final extension of five min. at 72°C.

After the selective amplification step, DNA samples were denatured by heating at 94°C for 5 min. Amplified fragments were separated on 6% denaturing polyacrylamide gels running at 70W for four hours. Subsequently amplified fragments were visualized following a silver staining protocol (BASSAM *et al.* 1991) and polymorphic bands were detected and isolated for downstream analysis.

Sequencing of the polymorphic bands

Bands that showed consistent polymorphism between the lines with the same nucleus but different cytoplasm were isolated from the denaturing polyacrylamide gel. These selected amplified PCR products were purified using a GenElute PCR Clean-Up kit (Sigma-Aldrich, St. Louis, MO, USA) and sequenced based on the Sanger method (GENEWIZ Inc., South Plainfield, NJ) using the original primers.

Annotation of polymorphic sequences

To expand the size of polymorphic sequences, homologous contigs of the sequenced fragments were identified by a blastN algorithm in the wheat 454 sequence database (WILKINSON *et al.* 2012). The best significant hit was then used as a query to search for potential homologous annotated regions in all plant species (NCBI, ncbi.nlm.nih.gov). The homologous wheat contigs also were used in a query of the wheat EST database (CHILDS *et al.* 2007) to identify possible exon-intron boundaries of differentially methylated contigs.

Table 2.1. Sequences of adaptors and primers used for methylation analysis

Adaptors			
*Eco*RI-adaptor-F	CTCGTAGACTGCGTACC	*Eco*RI-adaptor-R	AATTGGTACGCAGTC
*Hpa*II-adaptor-F	GATCATGAGTCCTGCT	*Hpa*II-adaptor-R	CGAGCAGGACTCATGA
Pre-selective primers			
Pre-*Eco*RI	GACTGCGTACCAATTCA	Pre-*Hpa*II/*Msp*I	ATCATGAGTCCTGCTCGGT
Selective primer combinations			
*Eco*RI-ACA + *Hpa* II-TCCA		*Eco*RI-AGG + *Hpa* II-TCGA	
*Eco*RI-ACA + *Hpa* II-TTC		*Eco*RI-ACC + *Hpa* II-TTC	
*Eco*RI-ACA + *Hpa* II-TGC		*Eco*RI-ACC + *Hpa* II-TGC	
*Eco*RI-ACA + *Hpa* II-TTCT		*Eco*RI-ACC + *Hpa* II-TTCT	
*Eco*RI-ACA + *Hpa* II-TTGC		*Eco*RI-ACC + *Hpa* II-TTGC	
*Eco*RI-ACA + *Hpa* II-TCA		*Eco*RI-ACC + *Hpa* II-TCA	
*Eco*RI-ACA + *Hpa* II-TCCA		*Eco*RI-ACC + *Hpa* II-TCCA	
*Eco*RI-ACA + *Hpa* II-TCGA		*Eco*RI-ACC + *Hpa* II-TCGA	

RNA extraction, cDNA synthesis and real-time quantitative PCR

To analyze the potential effect of methylation modification on the expression profile of nearby genes, we conducted a comparative expression analysis between eu- and alloplasmic lines. Total RNA was isolated from leaf tissues of two-week-old seedlings using the RNeasy® Plant Mini Kit (Qiagen, Valencia, CA, USA). To calculate the statistical significance threshold, two biological replications were used for each parental line. The quantity and quality of isolated RNA was confirmed using a NanoDrop 1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA) and 1% agarose gel electrophoresis, respectively. A QuantiTect® Reverse Transcription kit (Qiagen, Valencia, CA, USA) was used for cDNA synthesis and genomic DNA removal. Complete degradation of genomic DNA was confirmed by checking for any amplicon from the minus RT control.

Specific primers for qPCR were designed based on EST sequences near the methylated polymorphic regions (Table 2.2). Primers were designed using PrimerQuest software, available at the IDT website (www.idtdna.com), targeting sequences 100-200 bp in length with a 61ºC annealing temperature. Primers were used for qPCR using the SYBR® Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA). The primers for the wheat β-actin gene (TA411_4571) were used for an endogenous control. The qPCR data was

analyzed using the $2^{-\Delta\Delta Ct}$method (LIVAK and SCHMITTGEN 2001). In this analysis, the euplasmic lines of each wheat cultivar were used as the calibrator.

Table 2.2. Sequences of primers used for expression analysis

Forward	Sequence	Reverse	Sequence
MDH-F	TAAGGCTTCTCCTCTGACTCTC	MDH-R	CAGTGGATGAGAAGGGATTGAC
Ac-CoA-F	GGCAGTGGTGGGTAATTAGAG	Ac-CoA-R	CACCTCGACCGTCAAATCAT
A1du-F	GCTGCTACTGTTGTGTGTTTC	A1du-R	GGCTTACCGGAGTTGCTATTA
Mla-F	GTCTACTAGTGGTGGAAGAGGA	Mla-R	CTGCCTCGTCGAATCAATCA
c394-qRT-F	CGAGGAGACTGTACGCAAAT	c394-qRT-R	CTCTTCCTCTTCACGCTTTCT
Actin-F	GGCAACATTGTTCTCAGTGGTGGT	Actin-R	TCCTTTCAGGAGGAGCAACAACCT

Results

Plasmon effect on phenotypes of F_2 populations

Our results show that replacing the original B-type cytoplasm by T-type cytoplasm will increase the DMW and height in F_2 individuals (Figure 2.2). However head numbers were decreased as a result of plasmon replacement (Figure 2.2).

Plotting the values of different traits showed that a positive correlation is present between DMW and height, DMW and head number and between height and head number for both populations. However, supernumerary spikelets, which were induced in alloplasmic situations had negative correlation with other traits (Figure 2.2).

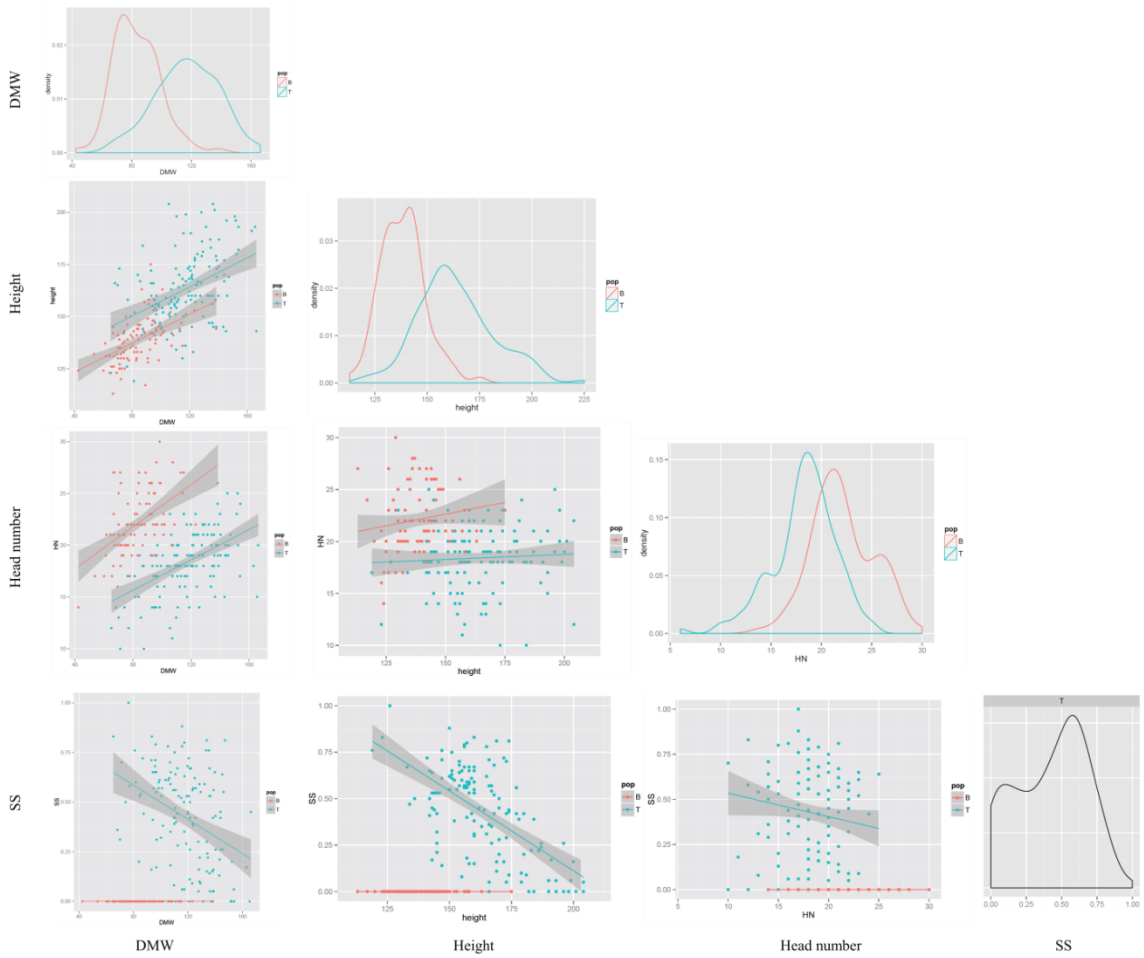

Figure 2.2. Effect of cytoplasm on phenotype of F_2 population. Phenotype of dry matter weight (DMW), height, head number (HN) and supernumerary spikelet (SS) in F_2 populations possessing different cytoplasm types. The diagonal plots show the distribution frequency of each trait in two populations. The scatterplots indicate the interaction of two traits in alloplasmic and euplasmic populations. The red points indicate the individuals possessing *Ae. mutica* cytoplasm (T type) and blue points indicate the individuals having original B-type cytoplasm. Red and blue lines in each scatterplot correspond to the regression line of each population. Confidence level is indicated by the darker gray background.

Genetic map construction

The genotypic data of alloplasmic and euplasmic populations from each growing season were combined together to construct the genetic map. For the 2013 spring season in which the 9K platform was employed, 1,808 SNPs were detected between Selkirk and Chris cultivars. Seventeen markers were discarded after filtration for distorted segregation markers. The remaining markers were mapped to 769 loci which formed 28 linkage groups with an overall length of 2060.4 cM (Table 2.3). In this map, the highest number of loci were located on the B genome, followed by A and D genomes, respectively. The A genome was the

longest genome at 974 cM in length (Table 2.3). The D genome was only 248 cM in length. The average spacing between two markers across all chromosomes was 2.8 cM, with a range from 1.1 cM to 15.1 cM (Table 2.3). The maximum gap between markers was 40.5 cM on chromosome 5A (Table 2.3). In total, 5535 SNP markers have been detected between Chris and Selkirk cultivars when the 90K platform is employed. These markers were mapped on 1750 loci and 26 linkage groups (Table 2.3). The total length of the constructed map was 3796 cM. In this map, the B genome possessed the highest number of markers and longest genome. In comparison with the constructed map from the 9k platform, using the 90 K platform improved the map coverage and resolution. The validity of each linkage group and the marker order within each group of both maps were assessed by comparing the constructed genetic maps with the previous consensus map.

Effect of cytoplasm on recombination frequency

Our experiments with two genotyping platforms revealed that a number of recombination events per chromosome followed a general reduction trend in linkage groups 4A, 1B.1, 6B, 6A.1, 7B, 7A.2, 2A and 2D.1 for the spring experiment (Figure 2.3). Among them, more drastic reduction was observed in linkage group 4A. The 90 K platform which was employed for the fall season improved the resolution of the genetic map (Table 2.3). Higher resolution revealed that the alloplasmic lines show a reduction in the number of recombinations for a majority of chromosomes (Figure 2.3).

Table 2.3. Summary of two genetic maps, constructed by 9K and 90K Illumina platforms

	Constructed genetic map using 9K platform (2013 Spring)						Constructed genetic map using 90K platform (2013 Fall)				
	Chr.	n.mar	length	ave.spacing	max.spacing			n.mar	length	ave.spacing	max.spacing
A genome	1A	85	113.4	1.3	14.8		1A	144	214	1.5	19.5
	2A	28	112.8	4.2	33.7		2A	60	211.9	3.6	43.4
	3A	39	125.7	3.3	24.5		3A	89	187.3	2.1	13.6
	4A	41	166.6	4.2	33.6		4A	73	228.5	3.2	30.1
	5A	30	152.2	5.2	40.5		5A	49	193.7	4	44.8
	6A.1	49	142.3	3	45.8		6A.1	82	199.2	2.5	25.7
	6A.2	2	10	10	10		6A.2	6	13	2.6	9.4
	7A.1	22	63.8	3	14.7		7A.1	41	90.2	2.3	9.3
	7A.2	15	87.9	6.3	39.1		7A.2	40	120.8	3.1	31.9
	Total	**311**	**974.7**	**4.5**	**45.8**		**Total**	**584**	**1458**	**2.8**	**44.8**
B genome	1B.1	61	67.9	1.1	18.2		1B.1	124	126.8	1	8.7
	1B.2	3	5.5	2.8	5.3		1B.2	10	38.1	4.2	28.3
	2B	110	155.2	1.4	29.3		2B	248	293.4	1.2	19.4
	3B.1	46	132.5	2.9	39.7		3B	194	321.5	1.7	49.2
	3B.2	9	11.1	1.4	4.2		4B	144	189.9	1.3	16.1
	4B	43	102.7	2.4	21.3		5B	107	238.2	2.2	29.7
	5B	41	124	3.1	17.3		6B	79	137.2	1.8	16.1
	6B	52	132.2	2.6	26.8		7B	103	232.2	2.3	40.3
	7B	33	96.1	3	17.8						
	Total	**398**	**827.2**	**2.3**	**39.7**		**Total**	**1009**	**1577.3**	**1.96**	**49.2**
D genome	1D	5	17	4.2	11.3		1D	18	52.8	3.1	19.8
	2D.1	10	76.2	8.5	23.8		2D	28	226.7	8.4	42.2
	2D.2	6	23.9	4.8	16.5		3D.1	16	86.5	5.8	44.4
	3D	4	12.1	4	7		3D.2	9	37.4	4.7	22.4
	4D	3	3.6	1.8	2.8		4D	4	28.6	9.5	24
	5D.1	5	60.3	15.1	38.9		5D.1	10	85.9	9.5	37
	5D.2	2	10	10	10		5D.2	4	33.9	11.3	29.8
	6D	8	13.4	1.9	9.6		6D	25	95.4	4	31.9
	7D	6	32.3	6.5	15		7D	17	113.7	7.1	40.8
	Total	**49**	**248.8**	**6.3**	**38.9**		**Total**	**131**	**761**	**7**	**44.4**
Overall		**758**	**2050.7**	**2.8**	**45.8**			**1724**	**3796.3**	**2.2**	**49.2**

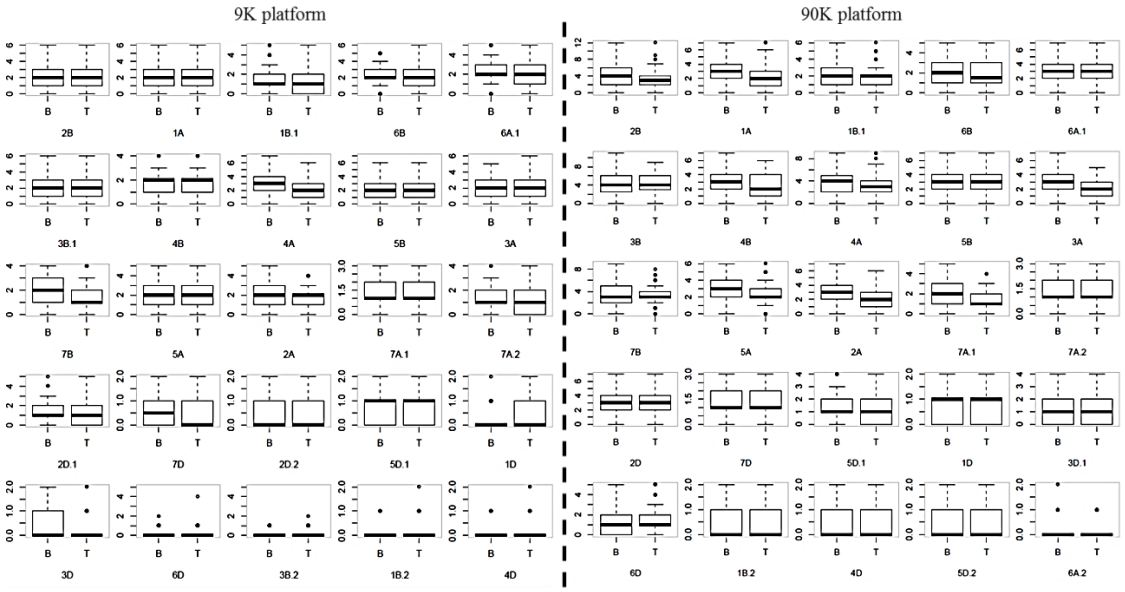

Figure 2.3. Effect of cytoplasm on recombination. T indicates the number of recombinations in the population with *Ae. mutica* cytoplasm and B indicates the number of recombinations in the normal (euplasmic) cytoplasmic condition.

In fall season, a QTL has been detected on chromosome 5A which controls the DMW in alloplasmic lines. Synteny analysis revealed that this region is located at the same end of the chromosome 5A long arm where a QTL was detected in the euplasmic lines in spring season. Another QTL was detected on chromosome 2D that controls DMW in both allo- and euplasmic lines in fall season.

Table 2.4. Major QTL controlling different traits

	2013 Spring								2013 Fall						
	Alloplasmic				Euplasmic				Alloplasmic				Euplasmic		
Trait	Chr.	Pos.	R^2		Chr.	Pos.	R^2		Chr.	Pos.	R^2		Chr.	Pos.	R^2
DMW															
	7D	19-32	10		**5A**	130-152	4		**5A**	0-12	10		**2D**	42-56	43
	7A	40-54	10						**2D**	48-60	17				
Height															
	5A	138-152	12		**5A**	140-152	12		**5A**	0-21	24		5A	0-12	
					2D.1	67-76	21						**2D**	122-150	8
Spike number															
	1B.1	25-39	19		7B	6-12	11		**6A**	29-53	17		2B	93-113	0.1
	7D	4-32	8		5A	16-29	14		2D	176-208	13		2A	142-151	0.1

Two-dimensional QTL analysis for DMW indicates that the interaction pattern between different nuclear loci depends on the cytoplasmic type and the environmental season. In spring 2013, in the alloplasmic population, the major interaction was detected between linkage groups 7B/6A.2 and was responsible for 8% of DMW variation in the alloplasmic condition (Table 2.5). However in euplasmic population, this interaction was not detected. Indeed, in euplasmic lines, four major interactions control 43% of the variation. In the fall 2013 season, however, an interaction between chromosomes 2B and 2D is the most prominent interaction that could not be detected in euplasmic population. On the other hand a locus on chromosome 7B was detected that interacts with two loci on chromosomes 2D and 6D (Table 2.5)

Plant height: QTL analysis for plant height in spring season detected a significant QTL on linkage group 5A for both alloplasmic and euplasmic lines (Table 2.4). Furthermore, in euplasmic lines, another QTL was detected on linkage group 2D.1 (Table 2.4). The location of QTL on chromosome 5A was similar in both alloplasmic and euplasmic population. The interval of this QTL was also similar to the QTL that controls the DMW in the euplasmic and alloplasmic lines in the spring and fall seasons, respectively. This similarity in interval location may indicate that there is one major gene or gene cluster on the long arm of chromosome 5A that is responsive to plasmon type.

Table 2.5. Major epistatic interactions in spring

Population	Chr. 1	Pos. 1	Chr. 2	Pos. 2	LOD	R^2(%)
Dry matter weight (DMW)						
T	7B	85	6A.2	10	5.2	8.4
B	2D.1	76	6D	1	5.6	13.5
B	3B.2	1	1A	55	4.3	11.1
B	3A	8	5D.1	0	4.6	9
B	5A	12	4B	147	4.54	9
Plant height						
T	4B	48	4A	83	5.3	13.3
T	7A.1	37	6D	13	5	12.8
T	2B	20	4A	96	4.8	10.4
B	5A	143	2A	1	5	9.8
B	2B	22	4A	30	4.4	7.9
B	5D.1	41	4D	0	4.3	8.3
Spike number						
T	7A.2	49	1B.1	40	6.5	10
T	1A	55	4A	104	8	12.7
T	2B	72	2A	2	6.8	11
B	4A	77	7B	39	10.8	23
B	4A	38	7A.2	73	7	16
B	7A.1	18	7A.2	23	7.8	18

Table 2.6. Major epistatic interactions in fall

Population	Chr. 1	Pos. 1	Chr. 2	Pos. 2	LOD	% var
Epistatic interactions controlling dry matter weight						
T	2D	59	1D	30	3.6	9.2
T	5B	208	4D	0	3.8	9.7
T	2B	231	2D	0	4.3	11
B	7B.3	188	2D	59	5.7	14
B	7B.3	143	6D	28	6	14
Epistatic interactions controlling height						
T	6A	68.9	4D	0	6.8	18
T	4D	0	5B	238	3.5	10
T	5D.2	0	1B.2	0	5.7	15
B	3B	12	2D	59	3	8
B	6A	79	2D	59	3	8
B	2D	83	5A	78	2.8	7
Epistatic interactions controlling head number						
T	3B	217	1B	52	5.2	18
T	3B	127	6B	136	5.8	16
T	3B	321	7B.3	135	5.1	16
B	4A	76	2D	116	5.6	18
B	3A	76	4D	28	5	16
B	5A	78	6D	92	4.6	14

Two-dimensional QTL analysis for alloplasmic population in spring identified a significant epistatic interaction between chromosomes 4A and 4B (Table 2.5). Along with two other putative interactions, epistatic interactions could explain about 36% of detected variation for height in alloplasmic population. For the euplasmic population in the same season, three major interactions were identified that could explain 26% of observed variation. A locus on chromosome 2B was detected that was involved in epistatic interaction in both euplasmic and alloplasmic conditions. This may indicate the possible importance of this locus in controlling height in both plasmon conditions. Furthermore an interacting locus on chromosome 5A was detected that resides on the same position where other significant QTL were detected. Two dimensional QTL analysis in fall season revealed a major gene on chromosome 4D that is interacting with two genes on 6A and 5B. On the other hand chromosome 2D is the prominent interacting locus in the euplasmic population in the same season.

Spike number: For alloplasmic condition in spring season, CIM revealed one significant QTL on chromosome 1B.1 that controls spike number (Table 2.4). Furthermore a putative QTL was detected on chromosome 7D. However in the euplasmic condition two putative QTL were detected (Table 2.4). In fall season for alloplasmic lines, a QTL was detected on chromosome 6A that control 17% of head number variation in alloplasmic lines. However two putative QTL residing on group 2 were detected in euplasmic lines for this trait.

In the spring, three major interactions could explain 34% and 57% of head number variation in alloplasmic and euplasmic lines, respectively. Our data indicates that the majority of these loci reside on group 7 and 4 of the wheat genome. In fall season, however, different loci on chromosome 3B were detected that were involved in interaction. Interestingly all of these major interactions were detected among different loci of the B genome. In euplasmic populations the major interactions were detected in 4A/2D, 3A/4D, 5A, 6D which can explain 48% of spike number variation.

Methylation profile of nuclear genes altered by plasmon type

To assess the effect of cytoplasm on epigenetic regulation of the nuclear genome, we compared the methylation pattern of alloplasmic lines with their corresponding euplasmic lines. Two classes of polymorphic bands were detected in our analysis: *i*.) Presence of an amplicon in alloplasmic lines and absence in corresponding euplasmic lines, which indicates methylated alleles in the euplasmic lines, and *ii*.) absence of an amplicon in alloplasmic lines and presence in the euplasmic line, indicating a methylated allele in the alloplasmic line, possibly the result of a cytoplasmic interaction.

In total, thirteen bands were detected, isolated and sequenced that showed polymorphism between alloplasmic and euplasmic lines. Homologous regions could not be identified for five of these fragments in any species. However, for other eight fragments, significant homology was identified (Table 2.7). Mapping the positions of *Eco*RI/*Hpa*II restriction recognition sites, polymorphic fragments and ESTs on the wheat contigs revealed that six of these potential methylated sites reside directly on or nearby expressed regions (Figure 2.4).

Table 2.7. List of the genes for which the methylation pattern was modified by the plasmon type.

Fragment	Length (bp)	Homologous sequence	E-value	Comment *
EHM-2-	509	*Triticum aestivum* cultivar Carazinho transposon	5e-11	Methylated
EHM-2-	434	*Triticum aestivum* cytosolic acetyl-CoA carboxylase (Acc-2)	5e-106	Methylated
EHM-2-	232	*Hordeum vulgare* Mla locus	3e-30	Methylated
EHM-2-	311	*Brachypodium distachyon*acyl amino-acid-releasing enzyme-like	5e-33	Methylated
EHM-2-	196	*Triticum aestivum* cultivar Chinese Spring hexose carrier, LR34,	3e-34	Methylated
EHM-3-	603	*T. aestivum* stress-induced receptor-like kinase A1du gene	0	Methylated
EHM-3-	150	*T. aestivum*, malate dehydrogenase, Partial cds	2e-08	Methylated
EHM-3-	613	*T. aestivum*AWJL175 (LRR gene family)	1e-48	Methylated
* Indicates methylation status in comparison to corresponding alloplasmic/euplasmic line. For example “Methylated in Chris” indicats methylated in euplasmicchris, but unmethylated in alloplasmic (*Ae. mutica*)Chris and vice versa				

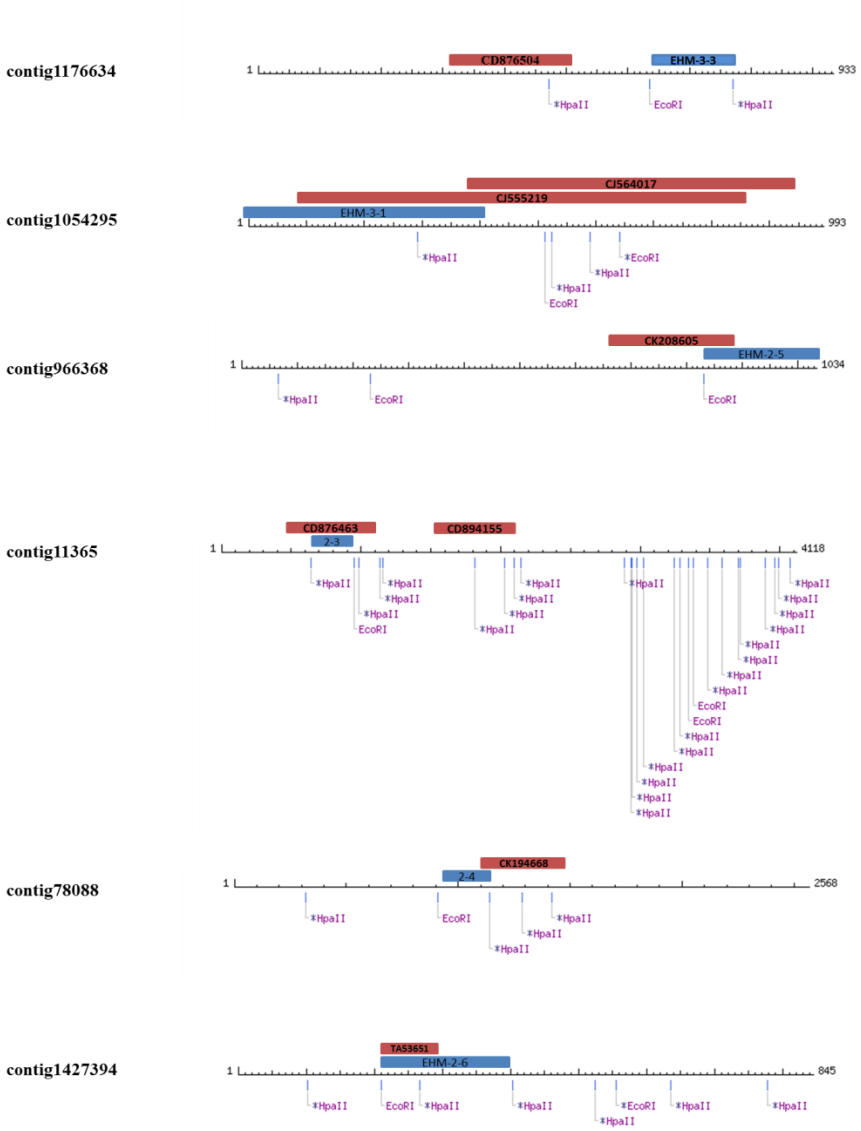

Figure 2.4. Methylated sites annotation. The blue bars indicate the polymorphic fragments detected in our experiment. Solid black lines represent the contigs showing significant homology to polymorphic fragments. Restriction recognition sites of both *Eco*RI and *Hpa*II enzymes are indicated by purple texts. The position of potential ESTs is represented by red bars.

Expression analysis of polymorphic methylated genes

To analyze the effect of methylation polymorphism on expression levels, the relative transcription abundance of five ESTs were investigated. These ESTs include CK194668 (Mla gene), CD876504 (malate dehydrogenase), CD876463 (cytosolic acetyl-CoA carboxylase), A1du gene (CJ564017) and TA53651. From these ESTs, no expression was detected for the *Mla* gene. Expression levels of malate dehydrogenase and cytosolic acetyl-CoA carboxylase

were comparable between eu- and alloplasmic lines. This suggests that different methylation profiles in these genes do not affect the transcript level; or, more likely, that expression alteration occurs in a tissue- or developmental- specific manner. However, drastic down-regulation (five-fold in SK and 16-fold in Chris) was detected in alloplasmic lines for A1du gene (Figure 2.5). Alternately, significant overexpression of TA53651 (contig 1427394) was detected in both alloplasmic lines relative to their euplasmic lines. This overexpression was about 1.5-fold compared to euplasmic lines. These alterations in A1du and TA53651 expressions may indicate the general effect of *Ae. mutica* cytoplasm on the transcription level of nuclear genes.

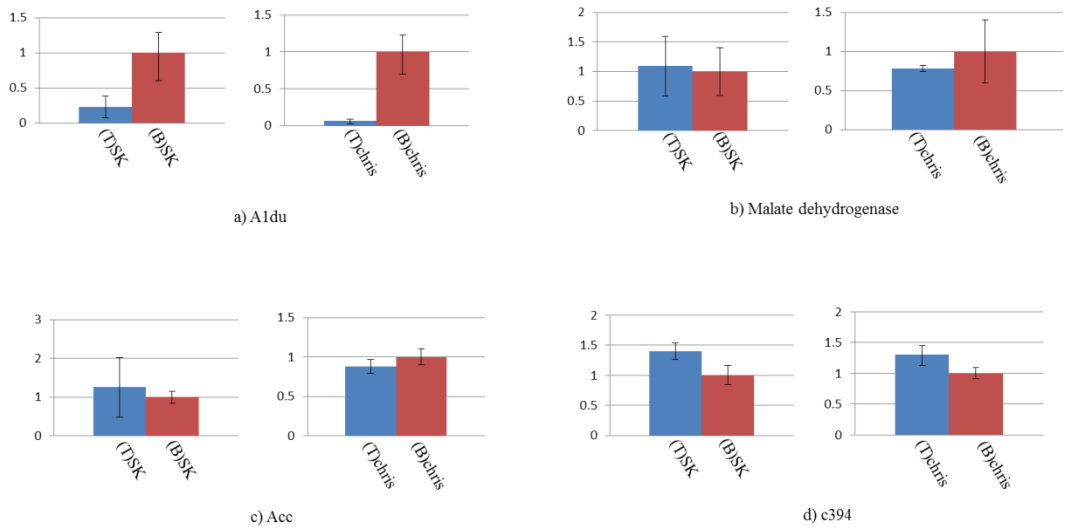

Figure 2.5. Expression analysis of detected polymorphic methylated genes. Blue bars indicate the expression level of genes in alloplasmic condition relative to euplasmic expression level (red bars).

Discussion

Interspecific hybridization has played an important role in the speciation of organisms in both *Plantae* and *Animalia* kingdoms. It is believed that more than 10% of animal and 25% of plant species experienced interspecific hybridizations during their evolution (MALLET 2005). Neo-allopolyploid species derived from these hybridizations usually show more adaptive characteristics compared with their parental species. These characteristics enable them to occupy new niches (MALLET 2007). To elucidate the role of interspecific

hybridization on speciation, attention has mostly focused on nuclear hybridization (HEGARTY and HISCOCK 2005; MALLET 2005; MALLET 2007). However, interspecific hybridizations also result in novel nuclear-cytoplasm combinations and it was recently proposed that cytoplasmic genomes in interaction with nuclear alleles can convey adaptive superiority to the previously less-adapted species (BUDAR and ROUX 2011). From a biological point of view, the co-evolution of three genomes including nuclear, mitochondria and chloroplast is critical for the superiority and survival of the organism (RAND *et al.* 2004; WOODSON and CHORY 2008). Hybrids with different degrees of NCI compatibility can emerge as a result of natural interspecific hybridization (LEVIN 2003). Upon genome stabilization, the survival or extinction of the new nucleus-cytoplasm hybrids depends upon the magnitude and the nature of selection pressure, natural or imposed. In most cases, historical hybridizations took place in a maternal-specific direction, although their reciprocal crosses were possible. This may indicate the importance of parental cytoplasm in cross-ability or post-zygotic adaptive characteristics of the progeny.

Production of alloplasmic wheat lines provides the means to study NCI. These lines possess novel hybrids that enable us to study NCI in a systematic way. A vast portion of these alloplasmic lines include unique nuclear and cytoplasmic combinations that cannot be studied in natural populations (MAAN 1992; TSUNEWAKI 2009; TSUNEWAKI *et al.* 1996).

The current study was devised to elucidate the role of NCI in controlling the major physiological traits including plant biomass, height and spike number per plant. These traits have been selected as an indicator of potential heterosis. Furthermore, the effect of plasmon on epigenetic and expression of nuclear genes was investigated.

Nuclear genes are interacting with cytoplasm and the environment

We identified that QTL controlling physiological traits are dynamic, affected by both environment and cytoplasm type. This led us to conclude that cytoplasmic genomes like environment modify the nuclear gene network controlling these traits in plants. A major QTL for DMW was detected on chromosome 5A that interacts with cytoplasm type in both seasons. In previous studies, the Q gene, which is involved in wheat domestication, has been mapped on the same arm of 5A chromosome (ZHANG *et al.* 2011). This gene encodes an AP2-like transcription factor that controls several traits in wheat, including height and head morphology (SIMONS *et al.* 2006). It is likely that this gene or other gene(s) involved in wheat domestication interact with cytoplasm and facilitate the superiority of the organism under natural or artificial selection. Another QTL that controls DMW in alloplasmic lines in the spring season was detected on chromosomes 7A and 7D. Interestingly, these QTL did not express in the euplasmic line in the same growing season. In *Arabidopsis lyrata*, QTL analysis in populations derived from reciprocal crosses identified a QTL that improved fitness in interaction with cytoplasm (LEINONEN *et al.* 2013).

In alloplasmic conditions, nuclear genes can affect plant vigor in interaction with mitochondria, chloroplasts, or both organelles. Alterations in nuclear-mitochondrial interaction can be observed, in most cases, in terms of male fertility/sterility (CHASE 2007; HANSON and BENTOLILA 2004). Alternately, nucleus-chloroplast interactions affect photosynthesis efficiency and consequently plant vigor (KUSHNIR *et al.* 1991; PETER *et al.* 1999; VANDERME.JP 1974; ZUBKO *et al.* 2001). In wheat, a QTL associated with a gene involved in photosynthesis and chlorophyll b accumulation was identified on chromosome 7A (FREEMAN *et al.* 1987). We suggest that the homeologous QTL on chromosome 7D and 7A or other cytoplasm-associated genes may affect photosynthesis performance and, consequently, biomass.

QTL analysis for plant height in the alloplasmic and euplasmic population in different environmental conditions detected a common region on chromosome 5A. Detection of the same region in both populations may indicate that the same gene(s) are involved in controlling plant height, regardless of cytoplasmic type. This QTL on chromosome 5A may also correspond to the Q gene, which has been shown to have a pleiotropic effect on several traits (SIMONS *et al.* 2006). A QTL that controls the height and interacting with cytoplasm was detected on chromosome 2D and is expressed just in euplasmic conditions in both seasons. Previous studies showed that almost all of the wheat chromosomes are involved in governing plant height under different developmental and environmental conditions (WU *et al.* 2010). Among these genetic components, a QTL on chromosome 4A was detected that had several epistatic interactions with other loci (WU *et al.* 2010). In our analysis, a QTL on chromosome 4A was detected that demonstrated an epistatic interaction in both the euplasmic and alloplasmic condition.

Epigenetic regulation affected by plasmon type

The results of two-dimensional QTL analysis revealed that the non-additive gene interaction network is dynamic and depends upon the cytoplasm type (Table 2.5). Non-additive gene interaction could result from the interaction of heterogeneous regulatory components or epigenetic regulations (JACKSON and CHEN 2010). In the present study, both the alloplasmic and euplasmic population segregated for the same sets of nuclear alleles. This suggests that the observed differences between the alloplasmic and euplasmic populations might have resulted from an epigenetic regulatory phenomenon, affected by plasmon.

To assess this hypothesis, we conducted a comparative analysis using methylation profile of parental lines. Our analysis revealed that the plasmon alteration can affect the methylation profile of the nuclear genes. Most of the differentially detected methylated genes belong to stress-responsive and energy production pathways. It has previously been shown

that stress-responsive genes are mostly expressed in a non-additive manner in polyploids (WANG *et al.* 2006). This non-additive expression pattern was suggested to result from a pronounced epigenetic regulation (JACKSON and CHEN 2010). Recently, a genome-wide epigenomic study among *Arabidopsis* accessions also revealed that stress-responsive genes are under a high level of methylation regulation (SCHMITZ *et al.* 2013). Based on our analysis, a fraction of genes involved in plant defense can be epigenetically regulated by the signals that originate from organelles (Table 2.6). It was previously shown that replacing the plasmon with an alien counterpart affects the resistance level of alloplasmic lines (DHITAPHICHIT *et al.* 1989; KEANE and JONES 1990; WASHINGTON and MAAN 1974; WORLAND *et al.* 1987). These modifications of the resistance levels may be the result of changes in the methylation profile of stress-responsive genes.

One of the genes detected as a potential target for cytoplasmic methylation regulation was *Malate dehydrogenase* (*MDH*). Several isoforms of this enzyme exist in plant cells (MILLER *et al.* 1998). One isoform is located in the mitochondria, which is involved in the citric acid cycle. Another isoform is present in the chloroplast and is important for maintaining the redox equivalent balance between stroma and cytosol. The cytosolic form assists the malate-aspartate shuttle, and the peroxisomic form is involved in β-oxidation (MILLER *et al.* 1998). Interestingly, it was shown that in transgenic tomato plants, down-regulation of the mitochondrial MDH is associated with enhancement of photosynthetic performance and growth (NUNES-NESI *et al.* 2005). The role of this gene in photosynthetic performance and growth of (*Ae.mutica*) Chris lines is a topic of investigation for further study.

In the present study, the methylation profile of cytosolic acetyl-CoA carboxylase (Acc-2) was also altered in the alloplasmic lines. This enzyme is involved in fatty acid elongation and flavonoid biosynthesis (PODKOWINSKI *et al.* 2003). It is believed that the

cytosolic and plastid isoforms of acetyl-CoA carboxylase have a critical role in carbon flux from photosynthesis to synthesis of primary and secondary metabolites (PODKOWINSKI *et al.* 2003).

Detecting these genes as potential methylation sites may indicate the presence of a strict co-regulation between retrograde and anterograde signaling pathways in the cells. In this co-regulation process, changing the plasmon type will affect the methylation profile of organelle-targeted nuclear genes via a retrograde signaling pathway.

Modification in methylation affects the expression level

Our expression analysis showed that the expression levels of two polymorphic methylated regions were significantly modified in alloplasmic lines (Figure 2.4). Interestingly, the methylated alleles (epialleles) were associated with the higher expressions of overlapped EST. Recent evidences suggest that the CG methylation in the gene body is positively correlated with the expression level (KULIS *et al.* 2012; SCHMITZ *et al.* 2013). This observation can explain the higher expression of A1du and TA53651 in euplasmic and alloplasmic lines, respectively.

Although polymorphic epialleles have been found in malate dehydrogenase and cytosolic acetyl-CoA carboxylase (Table 2.6), no expression differences were detected between them. Both of these genes are presented as a gene family (MILLER *et al.* 1998; PODKOWINSKI *et al.* 2003) and their expression has been shown to be tissue-specific (HAWKE and LEECH 1987; O'HARA *et al.* 2002; THELEN *et al.* 2001). It is possible that methylation differentiation does not have any effect on these genes, or more likely the expression difference cannot be detected in particular tissue. More specific expression analysis seems necessary to provide more resolution for these two genes

References

Acosta, M. C., and A. C. Premoli, 2010 Evidence of chloroplast capture in South American Nothofagus (subgenus Nothofagus, Nothofagaceae). Molecular Phylogenetics and Evolution **54:** 235-242.

Adams, K. L., and J. F. Wendel, 2005 Novel patterns of gene expression in polyploid plants. Trends in Genetics **21:** 539-543.

Allainguillaume, J., T. Harwood, C. S. Ford, G. Cuccato, C. Norris *et al.*, 2009 Rapeseed cytoplasm gives advantage in wild relatives and complicates genetically modified crop biocontainment. New Phytologist **183:** 1201-1211.

Bassam, B. J., G. Caetanoanolles and P. M. Gresshoff, 1991 Fast and Sensitive Silver Staining of DNA in Polyacrylamide Gels. Analytical Biochemistry **196:** 80-83.

Belmonte, M. F., R. C. Kirkbride, S. L. Stone, J. M. Pelletier, A. Q. Bui *et al.*, 2013 Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proceedings of the National Academy of Sciences of the United States of America **110:** E435-E444.

Budar, F., and F. Roux, 2011 The role of organelle genomes in plant adaptation: time to get to work! Plant Signal Behav **6:** 635-639.

Chase, C. D., 2007 Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends in Genetics **23:** 81-90.

Chen, Z. J., 2007 Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annual Review of Plant Biology **58:** 377-406.

Childs, K. L., J. P. Hamilton, W. Zhu, E. Ly, F. Cheung *et al.*, 2007 The TIGR plant transcript assemblies database. Nucleic Acids Research **35:** D846-D851.

Dhitaphichit, P., P. Jones and E. M. Keane, 1989 Nuclear and Cytoplasmic Gene-Control of Resistance to Loose Smut (Ustilago-Tritici (Pers) Rostr) in Wheat (Triticum-Aestivum L). Theoretical and Applied Genetics **78:** 897-903.

Fey, V., R. Wagner, K. Brautigam, M. Wirtz, R. Hell *et al.*, 2005 Retrograde plastid redox signals in the expression of nuclear genes for chloroplast proteins of Arabidopsis thaliana (vol 280, pg 5318, 2005). Journal of Biological Chemistry **280:** 17572-17572.

Freeman, T. P., M. E. Duysen and N. D. Williams, 1987 Effects of Gene Dosage on Light Harvesting Chlorophyll Accumulation, Chloroplast Development, and Photosynthesis in Wheat (Triticum-Aestivum). Canadian Journal of Botany-Revue Canadienne De Botanique **65:** 2118-2123.

Gehring, M., 2013 Genomic Imprinting: Insights From Plants. Annual Review of Genetics **47:** 187-208.

Hanson, M. R., and S. Bentolila, 2004 Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell **16:** S154-S169.

Hawke, J. C., and R. M. Leech, 1987 Acetyl-Coa-Carboxylase Activity in Normally Developing Wheat Leaves. Planta **171:** 489-495.

He, G. M., X. P. Zhu, A. A. Elling, L. B. Chen, X. F. Wang *et al.*, 2010 Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. Plant Cell **22:** 17-33.

Hegarty, M. J., and S. J. Hiscock, 2005 Hybrid speciation in plants: new insights from molecular studies. New Phytologist **165:** 411-423.

Jackson, S., and Z. J. Chen, 2010 Genomic and expression plasticity of polyploidy. Current Opinion in Plant Biology **13:** 153-159.

Keane, E. M., and P. W. Jones, 1990 Effects of Alien Cytoplasm Substitution on the Response of Wheat Cultivars to Septoria-Nodorum. Annals of Applied Biology **117:** 299-312.

Kohler, C., L. Hennig, C. Spillane, S. Pien, W. Gruissem *et al.*, 2003 The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes & Development **17:** 1540-1553.

Kulis, M., S. Heath, M. Bibikova, A. C. Queiros, A. Navarro *et al.*, 2012 Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nature Genetics **44:** 1236-1242.

Kushnir, S., E. Babiychuk, M. Bannikova, V. Momot, I. Komarnitsky *et al.*, 1991 Nucleocytoplasmic Incompatibility in Cybrid Plants Possessing an Atropa Genome and a Nicotiana Plastome. Molecular & General Genetics **225:** 225-230.

Lee, K. P., C. Kim, F. Landgraf and K. Apel, 2007 EXECUTER1- and EXECUTER2-dependent transfer of stress-related signals from the plastid to the nucleus of Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **104:** 10270-10275.

Leinonen, P. H., D. L. Remington, J. Leppala and O. Savolainen, 2013 Genetic basis of local adaptation and flowering time variation in Arabidopsis lyrata. Molecular Ecology **22:** 709-723.

Levin, D. A., 2003 The cytoplasmic factor in plant speciation. Systematic Botany **28:** 5-11.

Livak, K. J., and T. D. Schmittgen, 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods **25:** 402-408.

Maan, S. S., 1977 Cytoplasmic Homology between Aegilops-Mutica-Boiss and Aegilops-Ovata L. Euphytica **26:** 601-613.

Maan, S. S., 1991 Nucleo-cytoplasmic genetics of wheat., pp. 175-194 in *Proc. Int. Symp. Nuclear and Organeller Genetics of Wheat.*, edited by T. Sasakuma and Kinoshita, Hokkaido University, Sapporo, Japan.

Maan, S. S., 1992 A Gene for Embryo-Endosperm Compatibility and Seed Viability in Alloplasmic Triticum-Turgidum. Genome **35:** 772-779.

Madlung, A., A. P. Tyagi, B. Watson, H. M. Jiang, T. Kagochi *et al.*, 2005 Genomic changes in synthetic Arabidopsis polyploids. Plant Journal **41:** 221-230.

Mallet, J., 2005 Hybridization as an invasion of the genome. Trends in Ecology & Evolution **20:** 229-237.

Mallet, J., 2007 Hybrid speciation. Nature **446:** 279-283.

Mcclintock, B., 1984 The Significance of Responses of the Genome to Challenge. Science **226:** 792-801.

Miller, S. S., B. T. Driscoll, R. G. Gregerson, J. S. Gantt and C. P. Vance, 1998 Alfalfa malate dehydrogenase (MDH): molecular cloning and characterization of five different forms reveals a unique nodule-enhanced MDH. Plant Journal **15:** 173-184.

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

Nunes-Nesi, A., F. Carrari, A. Lytovchenko, A. M. O. Smith, M. E. Loureiro *et al.*, 2005 Enhanced photosynthetic performance and growth as a consequence of decreasing mitochondrial malate dehydrogenase activity in transgenic tomato plants. Plant Physiology **137:** 611-622.

O'Hara, P., A. R. Slabas and T. Fawcett, 2002 Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. Plant Physiology **129:** 310-320.

Peter, S., O. Spang, P. Medgyesy and C. Schafer, 1999 Consequences of intergeneric chloroplast transfers on photosynthesis and sensitivity to high light. Australian Journal of Plant Physiology **26:** 171-177.

Podkowinski, J., J. Jelenska, A. Sirikhachornkit, E. Zuther, R. Haselkorn *et al.*, 2003 Expression of cytosolic and plastid acetyl-coenzyme A carboxylase genes in young wheat plants. Plant Physiology **131:** 763-772.

RDevelopment Core Team, 2010 R: A Language and Environment for Statistical Computing, pp. R Foundation for Statistical Computing, Vienna, Austria.

Rand, D. M., R. A. Haney and A. J. Fry, 2004 Cytonuclear coevolution: the genomics of cooperation. Trends in Ecology & Evolution **19:** 645-653.

Rhoads, D. M., and C. C. Subbaiah, 2007 Mitochondrial retrograde regulation in plants. Mitochondrion **7:** 177-194.

Sall, T., M. Jakobsson, C. Lind-Hallden and C. Hallden, 2003 Chloroplast DNA indicates a single origin of the allotetraplold Arabidopsis suecica. Journal of Evolutionary Biology **16:** 1019-1029.

Schmitz, R. J., M. D. Schultz, M. A. Urich, J. R. Nery, M. Pelizzola *et al.*, 2013 Patterns of population epigenomic diversity. Nature **495:** 193-198.

Simons, K. J., J. P. Fellers, H. N. Trick, Z. Zhang, Y. S. Tai *et al.*, 2006 Molecular characterization of the major wheat domestication gene Q. Genetics **172:** 547-555.

Soltis, D. E., and P. S. Soltis, 1999 Polyploidy: recurrent formation and genome evolution. Trends in Ecology & Evolution **14:** 348-352.

Soltis, P. S., and D. E. Soltis, 2009 The Role of Hybridization in Plant Speciation. Annual Review of Plant Biology **60:** 561-588.

Strand, A., T. Asami, J. Alonso, J. R. Ecker and J. Chory, 2003 Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. Nature **421:** 79-83.

Thelen, J. J., S. Mekhedov and J. B. Ohlrogge, 2001 Brassicaceae express multiple isoforms of biotin carboxyl carrier protein in a tissue-specific manner. Plant Physiology **125:** 2016-2028.

Tsunewaki, K., 2009 Plasmon analysis in the Triticum-Aegilops complex. Breeding Science **59:** 455-470.

Tsunewaki, K., G. Z. Wang and Y. Matsuoka, 1996 Plasmon analysis of Triticum (wheat) and Aegilops .1. Production of alloplasmic common wheats and their fertilities. Genes & Genetic Systems **71:** 293-311.

Vanderme.Jp, 1974 Hybrid Chlorosis in Interspecific Crosses of Oenothera - Polygenic Inheritance of Nuclear Component. Canadian Journal of Genetics and Cytology **16:** 193-201.

Wang, J. L., L. Tian, H. S. Lee, N. E. Wei, H. M. Jiang *et al.*, 2006 Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics **172:** 507-517.

Wang, S., C. J. Basten and Z. B. Zeng, 2012 Windows QTL Cartographer 2.5, pp. Department of Statistics, North Carolina State University, Raleigh, NC.

Washington, W. J., and S. S. Maan, 1974 Disease Reaction of Wheat with Alien Cytoplasms. Crop Science **14:** 903-905.

Wendel, J. F., and R. C. Cronn, 2003 Polyploidy and the evolutionary history of cotton. Advances in Agronomy, Vol 78 **78:** 139-186.

Wilkinson, P. A., M. O. Winfield, G. L. A. Barker, A. M. Allen, A. Burridge *et al.*, 2012 CerealsDB 2.0: an integrated resource for plant breeders and scientists. Bmc Bioinformatics **13**.

Woodson, J. D., and J. Chory, 2008 Coordination of gene expression between organellar and nuclear genomes. Nature Reviews Genetics **9:** 383-395.

Worland, A. J., M. D. Gale and C. N. Law, 1987 Wheat Genetics, pp. 129-171 in *Wheat breeding: its scientific basis*, edited by F. G. H. Lupton. Chapman & Hall, New York.

Wu, X. S., Z. H. Wang, X. P. Chang and R. L. Jing, 2010 Genetic dissection of the developmental behaviours of plant height in wheat under diverse water regimes. Journal of Experimental Botany **61:** 2923-2937.

Zhang, Z., H. Belcram, P. Gornicki, M. Charles, J. Just *et al.*, 2011 Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. Proc Natl Acad Sci U S A **108:** 18737-18742.

Zubko, M. K., E. I. Zubko, A. V. Ruban, K. Adler, H. P. Mock *et al.*, 2001 Extensive developmental and metabolic alterations in cybrids Nicotiana tabacum (plus Hyoscyamus niger) are caused by complex nucleo-cytoplasmic incompatibility. Plant Journal **25:** 627-639.

GENERAL CONCLUSION

Increasing genotypic diversity is one of the crucial requirements for sustainable wheat breeding programs. Organelles, including chloroplast and mitochondria, have their own genomes. These genomes can be considered a source of additional genetic diversity to existing nuclear genome diversity. However, systemic analyses need to be conducted to characterize this genetic diversity in cytoplasmic components, and to characterize the differential effects those components impart.

Our first experiment was conducted to characterize the mitochondrial diversity among *Triticum-Aegilops* species. We chose the ATP6 gene sequence for our analysis because it resides on a recombination-active site of the mitochondrial genome. In this analysis, we have shown that cytoplasmic diversity is very narrow among hexaploid and emmer wheat lines. The narrow germplasm base of the mitochondria is a limiting factor for further selection and improvement of wheat cytoplasmic genome breeding efforts. Alternatively, significant diversity was detected in mitochondrial genomes of wheat wild relatives, including *Aegilops* species. This diversity can be introduced into wheat cultivars via a recurrent backcross method. In this crossing methodology, the wild species needs be the female parent, followed by recurrent backcrossing with the wheat cultivar as the male parent. The end product of this procedure is considered as an "alloplasmic" line that can be further utilized in breeding or pre-breeding programs. In this analysis, we concluded that two general forces drive mitochondrial genome diversity. These forces include genomic rearrangements that can be associated with gene duplication (such as the *ATP6* gene in different species) and point mutations, which further diversify cytoplasmic genomes. Our data showed that strong positive selection occurs in the ATP6 gene region, indicating the likely importance of this gene in the evolution of mitochondria and the consequent wheat species. Further analysis, however, is necessary to investigate the evolutionary trajectory of this crucial gene in other

species. Ancestral lines of some species also were identified based on the phylogenetic relationships of the ATP6 sequence. These evolutionary relationships help us improve our understanding of wheat cytoplasm evolution and its diversification during domestication.

Our second experiment showed that the cytoplasmic diversity existing in wheat alloplasmic lines affects phenotypic diversity observed among those lines. Several traits were identified that can be altered by the introduction of an alien cytoplasm. Plant height, head number and dry matter weight are among the traits that are greatly affected by cytoplasmic genomes. These effects are controlled via interactions between nuclear and cytoplasmic alleles. In one of the cases, alloplasmic lines with Chris nuclei and *Ae. mutica* cytoplasm demonstrated drastic effects on the observed traits. However, no such effect was observed in the Selkirk cultivars having the same cytoplasm.

To further elucidate the genetic mechanism of this interaction, we employed an integrative approach including genomic scan (QTL analysis), epigenetic (methylation polymorphism) and transcription (qPCR) analysis. In the QTL analysis, we identified a QTL on the long arm of chromosome 5A that affects several traits via interaction with *Ae. mutica* cytoplasm. We postulate this QTL is related to the Q gene, which is involved in wheat domestication. The Q gene was characterized previously as a transcription factor that controls several traits, including height and spike morphology. This gene could facilitate the domestication of wheat in coordination with cytoplasmic genomes. However, this hypothesis needs to be addressed in future experiments.

We also showed that cytoplasm substitution is associated with changes in genome methylation profiles. Several genes were detected as targets for cytoplasm-originated methylation polymorphism. Interestingly, the majority of these genes are categorized as organelle-targeted genes. This finding indicates the existence of a highly controlled

communication between organelles and nuclear genes. In this cross-talk, organelles regulate the methylation pattern of nuclear-encoded organelle-targeted genes.

Furthermore, several stress-responsive genes were detected based upon a changing methylation profile in response to cytoplasm replacement. Mitochondria and chloroplasts are the primary sources for environmental stimuli perception. These organelles can trigger a set of nuclear genes by a retrograde signaling pathway. Our findings revealed that methylation modification of stress-responsive genes can be a part of this retrograde signaling.

Our investigation in transcriptome level proved that modification in methylation affects the expression pattern of target genes. In a most extreme case, the expression of the A1du gene was shown to be drastically down-regulated in alloplasmic lines with *Ae. mutica* cytoplasm.

In general, our research revealed that a wide range of cytoplasmic diversity exists in wild relatives of wheat. Introgression of this diversity into present wheat cultivars may be beneficial toward improving wheat germplasm and for the development of superior cultivars.