Polyploidization study in wild wheat
by Arunrut Vanichanon

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Plant Genetics
Montana State University

Abstract:
Polyploidization is a key component of plant evolution. The number of independent origins of polyploid species has been traditionally underestimated. The objective of this work was to study multiple origins of tetraploid wild wheat. We screened 84 primer sets to identify genome-specific primer sets for the tetraploid wild wheat [Aegilops triuncialis (UC genome)] and its diploid progenitors [Ae. umbellulata (U genome) and Ae. caudata (C genome)]. Primer sets G12 and G43 were U genome-specific and D21 was a C genome-specific primer. Restriction fragment analysis and DNA sequence comparison were used to estimate the number of polyploidization events in the formation of Ae. triuncialis. G43 data revealed at least two independent formations of Ae. triuncialis. In the chloroplast hotspot region, results suggested that at least three polyploidization origins may have occurred independently. Despite evidence for multiple origins, less genetic variation was found in Ae. triuncialis than in its diploid progenitors. Ae. triuncialis appears to be a tetraploid with multiple origins with a minimal genome change after its formation.

The polymerase chain reaction (PCR) has become a standard procedure in plant genetics. One advantage of PCR is that sequence information for primer sets can be exchanged between labs, obviating the need for exchange and maintenance of biological materials. Repeatability of primer sets, whereby the same products are amplified in different labs using the same primer set, is important for successful exchange and utilization. We have developed several hundred sequence-tagged-site (STS) primer sets, The ability of the primer sets to generate reproducible amplifications in other laboratories has been variable. We wished to empirically determine the properties of the primer sets that most influenced repeatability. In our Study, a total of 96 primer sets were tested with four genomic DNA samples on each of four thermocyclers. All major bands were repeatable across all four thermocyclers for approximately 50% of the primer sets. Characteristics most often associated with differences in repeatability included primer GC content and 3'-end stability of the primers. The propensity for primer-dimer formation was not a factor in repeatability. Our results provide empirical direction for the development of repeatable primer sets.

POLYPLOIDIZATION STUDY IN WILD WHEAT

by

Arunrut Vanichanon

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Plant Genetics

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2002

APPROVAL

Of a dissertation submitted by

Arunrut Vanichanon

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Luther E. Talbert _____ (Signature) 4/15/02 (Date)

Approved for Department of Plant Sciences and Plant Pathology

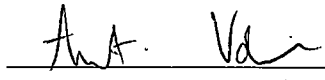
Dr. Norman F. Weeden _____ (Signature) 4/16/02 (Date)

Approved for the College of Graduate Studies.

Dr. Bruce R. McLeod _____ (Signature) 4-17-02 (Date)

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature 

Date 4/15/2002

ACKNOWLEDGMENTS

I would like to sincerely thank my major advisor, Dr. Luther Talbert, for providing the opportunity to pursue my degree and for his support, guidance, assistance and time. I would also like to thank the members of my committee, Dr Jamie Sherman, Dr. Matt Lavin, Dr. Michael Giroux and Dr. Richard Stout, whose assistance and knowledge has been appreciated.

Special thanks to members of the Spring Wheat Laboratory: Nancy Blake, Dr. Jamie Sherman, Susan Lanning for their friendship, assistance and technical expertise during five years in Montana State University. I would like to thank the members of the barley group including Dr. Tom Blake, Dr. Vladimir Kanazin and Hope Talbert for the use of the DNA sequencing machine and their assistance and technical expertise.

Lastly, I would like to thank my sister, Dr. Thanya Sripo, my husband, Panist Vanichanon, my son and daughter, Chanagun and Chanamon Vanichanon for understanding and supporting me while pursuing my education.

TABLE OF CONTENTS

TABLE OF CONTENTS –CONTINUED

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Polyploidization is a key component of plant evolution. The number of independent origins of polyploid species has been traditionally underestimated. The objective of this work was to study multiple origins of tetraploid wild wheat. We screened 84 primer sets to identify genome-specific primer sets for the tetraploid wild wheat [*Aegilops triuncialis* (UC genome)] and its diploid progenitors [*Ae. umbellulata* (U genome) and *Ae. caudata* (C genome)]. Primer sets G12 and G43 were U genome-specific and D21 was a C genome-specific primer. Restriction fragment analysis and DNA sequence comparison were used to estimate the number of polyploidization events in the formation of *Ae. triuncialis*. G43 data revealed at least two independent formations of *Ae. triuncialis*. In the chloroplast hotspot region, results suggested that at least three polyploidization origins may have occurred independently. Despite evidence for multiple origins, less genetic variation was found in *Ae. triuncialis* than in its diploid progenitors. *Ae. triuncialis* appears to be a tetraploid with multiple origins with a minimal genome change after its formation.

The polymerase chain reaction (PCR) has become a standard procedure in plant genetics. One advantage of PCR is that sequence information for primer sets can be exchanged between labs, obviating the need for exchange and maintenance of biological materials. Repeatability of primer sets, whereby the same products are amplified in different labs using the same primer set, is important for successful exchange and utilization. We have developed several hundred sequence-tagged-site (STS) primer sets. The ability of the primer sets to generate reproducible amplifications in other laboratories has been variable. We wished to empirically determine the properties of the primer sets that most influenced repeatability. In our study, a total of 96 primer sets were tested with four genomic DNA samples on each of four thermocyclers. All major bands were repeatable across all four thermocyclers for approximately 50% of the primer sets. Characteristics most often associated with differences in repeatability included primer GC content and 3'-end stability of the primers. The propensity for primer-dimer formation was not a factor in repeatability. Our results provide empirical direction for the development of repeatable primer sets.

CHAPTER 1

INTRODUCTION

Wild Wheat

The genus *Aegilops* L. comprises more than 20 species including many diploid, allotetraploid and allohexaploid species (Kimber and Feldman, 1987; Slageren, 1994). Cultivated wheat, an allohexaploid species, belongs to the genus *Triticum* L. that is closely related to *Aegilops* L. Wheat is a broadly cultivated crop and an important staple food in the world (Vasil and Vasil, 1999). Studying the wild relatives of wheat may reveal important facts that can lead to a better understanding and improvement of cultivated wheat.

In the *Triticeae*, there was presumably a common ancestor which gave rise to all diploid *Aegilops* and *Triticum* species. Based on interspecific crosses, almost every species had a distinct genome because homeologous chromosomes of different species did not completely pair. Most natural interspecific hybrids were completely or almost completely sterile. All diploid *Aegilops* and *Triticum* species were genetically isolated from each other during evolutionary time (Kimber and Feldman, 1987).

Approximately 70 % of all angiosperms are of polyploid origin (Soltis and Soltis, 1999), including wheat. Since polyploidization is a major force in plant evolution, it is a crucial component for understanding plant evolution. There are several reviews of

polyploid evolution, including the mode and rate of formation of polyploids, ecological and evolutionary attributes and genetic consequences of polyploidy (Leitch and Bennet, 1997; Soltis and Soltis, 1999; Otto and Whitton, 2000; Wendel, 2000). Recent studies indicate that the number of polyploidization events involved in the origin of polyploid species has traditionally been underestimated (Soltis and Soltis, 2000; Ben-Ari, 1998). Several studies have indicated that the evolution of wheat underwent a bottleneck situation based on observed low levels of variation using RFLPs (Harcout and Gale, 1991; Kam-Morgan et al., 1989; Autrique et al., 1996) and isozyme (Asins and Carbonell, 1989; Nevo and Beiles, 1989). However, results of Talbert et al. (1998) revealed that wheat originated at least twice based on DNA sequence analysis. Multiple origins in other *Triticum* and *Aegilops* species still need to be examined.

This study focuses on *Ae. triuncialis* because of its remarkable adaptation. *Ae. triuncialis*, an allotetraploid species, is of particular interest because it has a wider geographical and ecological distribution than its diploid progenitors. *Ae. caudata*, a diploid parent, was described as a Mediterranean species which grows abundantly mainly in western Turkey. It is a typical lowland species. *Ae. umbellulata*, another diploid parent, is of Mediterranean and Western Asiatic origin and is uncommon throughout its range. On the other hand, *Ae. triuncialis ssp. triuncialis* has widespread distribution in the Mediterranean, Western Asiatic, southern Europe and central Asia. It is common throughout its range. Conversely, *Ae. triuncialis ssp. persica* is rare in west and central Asia.

Molecular genetic studies have partially elucidated the evolution of *Ae. triuncialis*. Murai and Tsunewaki (1986) used restriction fragment analysis to hypothesize that *Ae. triuncialis* resulted from reciprocal crosses of *Ae. umbellulata* and *Ae. caudata*. The genome relationships between *Aegilops* allopolyploids and their diploid ancestors was investigated using RFLP of chloroplast genome (Ogihara and Tsunewaki, 1988) and mitochondria genome (Ogihara et al., 1993). Waines and Barnhart (1992) proposed that the female genome should be listed first in the genome formulae. Therefore, the formulae for *Ae. triuncialis ssp. triuncialis* is proposed to be UUCC and CCUU for *Ae. triuncialis ssp. persica* (Slageren, 1994). In our study, *Aegilops triuncialis* was used as a plant model to illustrate multiple origins and genetic consequences of polyploidization using DNA sequence comparison

The Polymerase Chain Reaction (PCR)

Plant molecular biology currently relies on the polymerase chain reaction (PCR) to provide a nearly unlimited source of DNA molecules for further manipulation or subsequent experimental analyses. The powerful nature of PCR has led to significant insights into biological questions including those regarding plant evolution.

PCR is a relatively simple technique conceived in 1983 by Kary Mullis (Mullis and Faloona, 1987), involving a cycle process which leads to an exponential amplification of the target DNA. The goal of PCR is to replicate a particular DNA region, not the entire genome. This is achieved by using primers, which target a specific DNA

region. A primer is a short sequence of DNA, usually 18-25 base pairs (bp) long, that is complementary to one end of the target DNA. Both strands of the DNA target are copied by using two primers; one for the 5' end of the one strand and one for the 5' end of the complementary strand. The strands of the target DNA duplex are separated by heating and then cooled to allow primer annealing. Once the primers are annealed to the template, *Taq* DNA polymerase adds nucleotides complementary to the template. Extension is accomplished by increasing the temperature to the optimum for *Taq* DNA polymerase. Each heating and cooling cycle results in the doubling of the amount of template; therefore, after 20 cycles the yield of PCR product is approximately one million copies (2^{20}) of the single target DNA molecule. Annealing temperature, length of annealing and extension steps within each cycle, and cycle numbers vary for different target DNA regions and primers (Aert et al., 1998; Bruke, 1996).

PCR technology has several advantages for plant molecular biology research. Gene content and genetic map colinearity in grasses is well accepted (Bennetzen, 1999). Therefore, cross-species primers are widely used to accelerate research such as chloroplast primers among all land plants (Petit et al., 1998), and between barley and wheat (Erpelding et al., 1996). Furthermore, primers can be widely disseminated among researchers because they are easily synthesized from their sequences. One disadvantage of PCR is that *Taq* polymerase incorporates occasional errors at a rate of approximately 10^{-4} per base per doubling in newly synthesized DNA (Andre et al., 1997). In practical terms, this may not be a major problem because the errors are apparently not biased in

favor of any specific nucleotide. One way to cope with the potential error in DNA sequence comparison is to sequence both strands of molecule (Judd et al., 1999).

Molecular markers have become increasingly important in plant molecular biology in its relationship to plant breeding, plant systematics and plant evolution. Most molecular marker systems are presently based on PCR technology. The objective of our study is to discern the important characteristics of primers which influence reproducible results among laboratories.

Molecular Genetic Approaches in Plant Evolution

Molecular markers of various types have helped to increase understanding and elucidate new aspects of plant evolution. Isozymes have been widely used as genetic markers in plant diversity studies since 1959. Several advantages of isozymes are that they usually produce codominant markers, they are easy to assay in large populations and they monitor the genetic variation at specific gene loci. Some limitations are evident as well: the inability to detect water insoluble or cell structure-bound enzymes and the failure to distinguish isozyme bands with identical mobility that are in reality two different alleles (Muller-Starck, 1998). Additionally, only a small set of isozyme markers are available.

Molecular markers using restriction site analysis based on DNA sequence differences are common for studying variation. Cytoplasmic DNA is extracted and then digested with restriction enzymes followed by electrophoretic gel analysis. An advantage

of this technique is that analysis covers a large DNA region. However, a disadvantage is that different mutations may result in the same sized bands. Therefore, some results may lead to incorrect interpretation.

An adapted technique, restriction fragment length polymorphism (RFLP) analysis, requires digestion of nuclear and organellar DNA using restriction enzymes, blotting and hybridization with a radioactively labeled probe (Botstein et al., 1980). RFLP has been widely used for plant breeding, plant systematic and plant evolution for over two decades.

RFLP is a laborious process whereas other currently used molecular markers based on PCR are not. Furthermore, PCR fragments can be subjected to restriction site analysis. PCR technology has generated a variety of molecular markers to study plant evolution and diversity including PCR-single-strand conformational polymorphism (PCR-SSCP) (Wang et al., 1997), randomly amplified polymorphic DNAs (RAPDs) (Williams et al., 1990), amplified fragment length polymorphism (AFLP) (Daly, 1998; Vos et al., 1995), sequence-tagged-site PCR (STS-PCR) (Talbert et al., 1996) and simple sequence repeats PCR (SSRs-PCR) (Pestsova et al., 2000). However, DNA sequencing is the ultimate method to detect genetic variation with the greatest detail. The current advances in automated sequencing continue to increase DNA sequence data. Thus, we selected DNA sequence analysis to study evolution questions in wild wheat.

CHAPTER 2

MULTIPLE ORIGINS AND GENETIC VARIABILITY OF ALLOPOLYPLOID *Aegilops triuncialis* USING NUCLEAR AND CHLOROPLAST MOLECULAR DATA

Introduction

Within the plant kingdom, polyploidization is a powerful force leading to speciation as well as an important source of genetic variation. In general, seventy percent of angiosperms have undergone polyploidization at least once (Soltis and Soltis, 1999). Polyploid species tend to be more widely distributed and found in more extreme habitats than their diploid ancestors (Soltis and Soltis, 2000). Understanding the number of polyploidization events which have occurred in the formation of a given species, and the consequences of such events, has been a major challenge. Although polyploidization has been studied for nearly a century, there are still many unknowns. The traditional point of view is that polyploidization events were rare because there is less variability within polyploid species than the diploid relatives such as wheat (Talbert et al., 1998). Putative examples of monophyletic origins include peanut *Arachis hypogaea* (Kochert et al., 1996) and salt marsh grass *Spartina anglica* (Raybould et al., 1991). However, more recently recurrent formation of polyploids has been demonstrated. Over 30 examples of allotetraploid species have been shown to have multiple origins to date (Soltis and Soltis, 1999). *Draba norvegica* (Brassicaceae) formed at least 13 times in a small area of

Scandinavia (Brochmann and Elven, 1992). *Tragopogon miscellus* and *T. mirus* have formed as many as 21 and 9 times, respectively, in eastern Washington and western Idaho, in the past 60 to 70 years (Soltis et al., 1995). Hexaploid wheat formed at least twice from *Triticum tauschii* (D genome) as a diploid progenitor (Talbert et al., 1998). The genetic and evolutionary consequence of multiple polyploidization within species still needs investigation.

Allopolyploids receive their chromosome sets from different species, unlike autopolyploids that receive multiple sets of chromosomes from one species. The prevalence of allopolyploid formation and the degree of genetic separation between allopolyploids and their progenitors have important consequences for the accumulation of genetic variability within an allopolyploid. Wheat and its wild relatives provide opportunity for studying allopolyploidization. For example, wheat, as an allohexaploid, has three genomes: A, B and D genomes. *Tragopogon miscellus* is the best known example of an allotetraploid species that occurred recently by reciprocal crosses of the diploids, *T. dubius* and *T. pratensis*, according to molecular data (Soltis et al., 1995). Like *Tragopogon miscellus*, *Aegilops triuncialis* (UC genome) is an allotetraploid that resulted from reciprocal crosses of diploids, *Ae. umbellulata* (U genome) and *Ae. caudata* (C genome) (Wang et al., 1997).

Aegilops triuncialis is the most widespread *Aegilops* species in the world. This grass is distributed between altitudes of 300 m and 1000 m and has become a troublesome weed on U.S. rangelands (Watanabe and Kawahara, 1999). Additionally, *Ae. triuncialis* is of agronomic interest as the source of resistance for leaf rust (*Puccinia*

triticina), Karnal bunt (*Ustilago tritici*), powdery mildew (*Erysiphe graminis*) and cereal cyst nematode (*Heterodera avenae*) resistance in cultivated wheat (Singh et al, 2000). Its tremendously successful adaptation results from genetic attributes of polyploids such as high genetic variability and the evolution of new gene functions (Soltis and Soltis, 2000).

Morphological, cytological and electrophoretic approaches are used to determine multiple origins of allopolyploid species. Molecular approaches including restriction fragment analysis, comparative sequencing and various PCR-based techniques are now available to uncover recurrent origins of allopolyploid. It is important to carefully select DNA regions when using these approaches because various genomes and DNA regions are suitable for different taxonomic levels (Soltis et al., 1998). Within a plant cell, three different types of DNA are found: nuclear, chloroplast and mitochondria DNA. Most current molecular data have come from the chloroplast, the highly repetitive sequences of ribosomal RNA (rRNA) and low copy genes (Soltis etal., 1998). The most frequently chosen nuclear DNA studied has been rRNA including 5S, 18S-26S rRNA, and internal transcribed spacer (ITS). Based on rRNA, Soltis and Soltis (1991) suggested multiple origins of the allopolyploid, *Tragopogon mirus*.

Because the nuclear genome is inherited biparentally, genome specific primers are preferred. Genome specific primers lessen ambiguous results by demonstrating direct inheritance of an allele from a specific ancestral genome. For example, using the D genome specific primer A1 sequence comparison between wheat and the D genome diploid progenitor indicated that hexaploid wheat was formed at least twice (Talbert et al., 1998).

Badaeva et al. (1996a) chose 5S and 18S-5.8S-26S(18S-26S) rRNA gene families to study the relationship of the U, C and D genomes among *Aegilops* diploid species. All three genomes have the same 5S rRNA banding pattern. The D genome had a unique pattern in 18S-5.8S-26S(18S-26S) rRNA. However, U and C genomes showed identical patterns. Highly repetitive nuclear DNA clones from rye and the D genome were also selected. The results indicated that U and C genomes were closely related; however, the D genome was more closely related to the M and M^h genomes than other *Aegilops* diploid species (Badaeva et al., 1996b). Another study used 46 probes of nuclear repetitive nucleotide sequences to resolve the relationship among *Aegilops* diploid species. The results showed that the U and C genomes were closely related and in the same clade (Dvorak and Zhang, 1992).

Because of the size and slow mutation rate of chloroplast genome, it has several advantages for taxonomic and evolutionary study. The genome is quite small, approximately 120-200 kilobases (kb). In Chinese Spring wheat, the chloroplast genome is 134,540 base pairs (bp) long (Ogihara and Tsunewaki, 2000). Chloroplast DNA is usually maternally inherited in plants. Most chloroplast genes are a single copy (Palmer et al., 1985). The genome is relatively conserved during its evolution throughout plant species. The chloroplast genome evolves four to five times slower than the nuclear genome and three times faster than mitochondria (Page and Holmes, 1998). Because of the slow mutation rate of plant mitochondria DNA, its genome was not selected in this study. Because different regions of DNA evolved at different rates, some parts of the chloroplast DNA might be appropriate to resolve relationships at different taxonomic

levels (Soltis and Soltis, 1998). For instance, some noncoding regions and microsatellites may be applicable at an intraspecific level due to their greater variation (Karp et al., 1998; Gielly and Taberlet, 1994).

All of these advantages allow the chloroplast genome to be a useful tool for evolutionary studies. The chloroplast gene encoding the large subunit of ribulose-1, 5-bisphosphate carboxylase/ oxygenase (*rbc*L) was used extensively in molecular systematic studies among angiosperms (Chase et al., 1993). Other chloroplast genes such as *atp*B, *ndh*F and *mat*K, were also applied (Soltis and Soltis, 1998; Judd et al., 1999). RFLP analysis of both chloroplast and ribosomal DNA sequences were also used to identify the parental origin of polyploid genomes with multiple origins (Soltis et al., 1992). Multiple origins of *Tragopogon miscellus* and *T. mirus* were studied using chloroplast and rRNA data (Soltis and Soltis, 1989; Soltis and Soltis, 1991).

All previous studies confirmed a maternal lineage of chloroplast inheritance among *Triticum* and *Aegilops* species (Ogihara and Tsunewaki, 1982; Murai and Tsunewaki, 1986; Tsunewaki, 1993; Wang et al., 1997). Genetic diversity in *Ae. triuncialis* was studied using the chloroplast genome. Results indicated its multiple origins from reciprocal crosses between *Ae. umbellulata* and *Ae. caudata* (Murai and Tsunewaki,1986). Ogihara and Tsunewaki (1988) studied chloroplast genomes of alloplasmic lines of 35 species of *Triticum* and *Aegilops* with 13 restriction enzymes. Their results showed 33 nucleotide substitutions distributed equally throughout the genome, in contrast to 14 insertion/deletion mutations (indels). Six of fourteen indels were located between the genes *rbc*L and *pet*A. The region was designated as a hotspot

and not only contained many direct and inverted repeats near the indel region but also was AT rich (Ogihara et al., 1992). They suggested that these two characteristics might be responsible for the mutation. In previous studies, *Ae. caudata*, *Ae. triuncialis* and synthetic *Ae. triuncialis* had chloroplast type 2 which included a 300 base pair deletion within the hotspot region. Unlike *Ae. caudata*, *Ae. umbellulata* which had chloroplast type 3 did not have this deletion (Ogihara and Tsunewaki, 1988). This deletion was one of many structural changes distinguishing between *Ae. caudata* and *Ae. umbellulata.* This region seemed promising for distinguishing these genomes; although, only one alloplasmic line per species was analyzed. The region was studied further using nucleotide sequence comparison among two alloplasmic lines of *Triticum. aestivum* cv. Chinese Spring carrying cytoplasm of *Ae. crassa* and *Ae. squarrosa* and one euplasmic line, *T. aestivum* cv. Chinese Spring. Ogihara et al. (1991) found that nucleotides at the intergenic regions diverged ten times faster than those of coding regions. Wang et al. (1997) concluded that the genetic relationship of the chloroplast genomes of the U genome of *Ae. umbellulata* and the C or D genomes of *Ae. caudata* and *Ae. squarrosa* are only moderately close.

Due to superior colonizing abilities of *Aegilops triuncialis* compared to its diploid ancestors, many questions still need to be answered. What is the level of genetic differentiation between the diploid parents? How much genetic variability is within *Aegilops triuncialis* relative to diploid progenitors? Was this genetic variability due to the genetic consequences of polyploidization? How frequently did this species form? Will

DNA sequence data from chloroplast and nuclear genomes give similar results? Soltis and Soltis (1993) stated that different types of DNA might lead to different conclusions.

Investigation of these questions was carried out to gain a better understanding of the evolutionary dynamics of *Ae. triuncialis*. Molecular genetic data were the approach selected as they provide a wealth of new insights into polyploid evolution. Moreover, molecular genetic approaches have provided critical data regarding the genetic consequences of polyploid evolution. From previous molecular results (Badaeva et al., 1996; Dvorak and Zhang, 1992), the U and C genomes appear to be closely related, such that coding regions are too similar to provide enough discrimination between *Ae. triuncialis* and its diploid progenitors. Therefore, noncoding regions, both in chloroplast and nuclear genomes, were selected to assess genetic diversity between *Ae. triuncialis* (UC genome) and its ancestors *Ae. umbellulata* (U genome) and *Ae. caudata* (C genome). First, we screened primers to identify those which amplify all three genomes to observe genetic variability among them. Second, genome specific primers for U, C and chloroplast genomes were selected to enhance data interpretation. Polyploid organisms receive nuclear genome from both maternal and paternal parents. It is difficult to identify the origin of specific alleles if more than one genome are amplified. Genome specific primers allow unambiguous genome assessment. Since DNA sequence comparison provides the greatest discrimination of evolutionary relationships, sequence data were obtained from DNA segments amplified using genome specific primers. Phylogenetic analysis based on DNA sequence data was performed to determine the relationships of

the diploid ancestors and the allotetraploid species and to address the possibility that the allopolyploid *Ae. triuncialis* formed multiple times.

Materials and Methods

Plant Materials

Three species of wild wheats were chosen. Thirty one accessions of *Aegilops caudata*, 33 accessions of *Aegilops umbellulata* and 212 accessions of *Aegilops triuncialis* were obtained from Harold E. Bockelman USDA National Small Grain Collection, Aberdeen, Idaho, USA. *Aegilops caudata* and *Aegilops umbellulata* are diploid progenitors of *Aegilops triuncialis*. Three accessions of *Aegilops caudata*, two accessions of *Aegilops umbellulata* and six accessions of *Aegilops triuncialis* were requested from Wheat Genetics Resource, Kansas State University. Four accessions of *Aegilops triuncialis* were obtained from Plant Germplasm Institute, Kyoto University Japan (http://www.shigen.nig.ac.jp/wheat/wheat.html?228,31) (Murai and Tsunewaki 1986). *Aegilops caudata* came from only Turkey and Greece. *Aegilops umbellulata* predominantly originated from Turkey. *Aegilops triuncialis* is widely distributed throughout the Mediterranean, Black Sea region, the Middle East of Asia and western Africa. *Triticum tauschii* accession KU2050 from Afghanistan was used as an outgroup species.

Genomic DNA Isolation

All plants were grown in the Plant Growth Center at Montana State University and young leaves were collected for total genomic DNA extraction (Riede and Anderson, 1996). A single plant was used per each accession. Young leaves, weighing between 20 and 300 mg, were collected and ground in liquid nitrogen using an electric drill and Kontes pestle in a 2 ml microcentrifuge tube. After grinding, 750 µl of prewarmed extraction buffer [0.5M NaCl, 0.1M Tris-HCl pH 8.0, 0.05 M ethylenediaminetetra-acetic acid (EDTA), 8.4 gm/L sodium dodecyl sulfate (SDS), 3.8 gm/L sodium bisulfite] was added and vortexed until suspended. Samples were incubated at 65 °C in a waterbath for 45 minutes and mixed by gentle inversion every ten minutes. After the incubation period, samples were cooled on ice and 750 µl of chloroform was added. Samples were mixed to homogeneity and centrifuged at 14,000 G for ten minutes. Approximately 600 µl of the upper phase was drawn off and transferred to a new tube containing 1 ml cold 95 % ethanol. DNA pellets were precipitated by gentle inversion followed by centrifugation at 14,000 G for ten minutes. Genomic DNA was washed with 1 ml 70 % ethanol, shaking at least 1 hour and followed by centrifugation at 14,000 G for four minutes. The supernatant was decanted. The DNA pellet was air dried and resuspended in 100 µl TE buffer. Genomic DNA sample were quantified by running 1 µl on a 1 % agarose gel with 1 X TBE buffer and comparing with a precision molecular mass standard (Bio-Rad®). The working concentrations of genomic DNA were then adjusted to approximately 100 ng/µl for use as template DNA in PCR reaction.

STS-PCR Primers for PCR Amplification

A total of 84 primer sets were used to screen accessions of *Ae. caudata*, *Ae. umbellulata* and *Ae. triuncialis*. The nucleotide primers were synthesized by Sigma Genosys, USA. Twenty-nine primer sets were developed from mapped RFLP clones of D genome diploid *Triticum tauschii* (Talbert et al., 1994). Thirty-two primer sets were designed in the chloroplast hotspot region. Nine and six primer sets were developed from wheat and barley genomic DNA, respectively. Six and two primer sets were developed from barley and oat cDNA(Tragoonrung et al., 1992). Some primers are shown in Table 1. The Cp6 (U6/R6) primer set was designed in this study to amplify the noncoding region between *ycf*4 and *cem*A genes within the chloroplast genome of *Triticum* and *Aegilops* species. The last four primers in the Table 1 were used for sequencing cloned inserts (see below).

PCR Amplification and Analysis

PCR amplifications were conducted in 50 µl reactions consisting of 1 X Promega reaction buffer, 200 µM of dNTPs, 1.5 mM $MgCl_2$, 400 nM of left and right primers, 0.75 U of Taq polymerase, and 100 ng of genomic DNA. The PCR cycle is shown in Table 2 (Talbert et al. 1994). Since the primers were derived from non target species, we chose lower annealing temperatures. The annealing temperature, a controlling factor for PCR amplification, will enhance or suppress artifact formation when it is decreased or increased (Watson, 1989). The PCR was performed in a PTC-100 programmable thermocycler (MJ Research, Inc.).

Table 1. Primers and sequence of primers for PCR amplification

Name of primer	Sequence	Location[a]	Genome specific
G43 Forward	5'-GGCGC ATGCA CCAAA ATGTT-3'	6D	U
G43 Reverse	5'-ACCTT GTCGT GCATA GGAAC-3'	6D	U
D21 Forward	5'-TCTTC CAGTT AGAGA TCTCC-3'	4D	C
D21 Reverse	5'-TCGTT CGTAC TAGTA GTACC-3'	4D	C
Cp6 Forward	5'-GCTGC CGAAT TGGCC TATTT-3'	Cp[b]	cp
Cp6 Reverse	5'-GCCTG GTATT CCACC AATTC-3'	cp	cp
D2 Forward	5'-CGAAT GTTTC TACTG CGCTG T-3'	7ABD	NA
D2 Reverse	5'-CTCCC TGTTT GTGGA AAGCT-3'	7ABD	NA
D15 Forward	5'-GTCTT CACGG AGATC TGTAT-3'	1D,5BD, 7BD	NA
D15 Reverse	5'-GCTGC CTGTT TTGTT TCGCA-3'	1D,5BD, 7BD	NA
F8 Forward	5'-GCATT ATCAT CAGCT GAAAG-3'	4ABD	NA
F8 Reverse	5'-GTTCA AGGCA GACCT TGACT-3'	4ABD	NA
G12 Forward	5'-CCAGT GTTGT AGTTC TCTAT-3'	2B,2D	U
G12 Reverse	5'-TATAC TTCTG AGCTG CCGAG-3'	2B,2D	U
WG232 Forward	5'-CCTCA GTGTT TCAGG GTAAA-3'	7A,4AD	NA
WG232 Reverse	5'-TGGAC TCGTG TTCAA TAATG-3'	7A,4AD	NA
M13 Forward	5'-GTAAA ACGAC GGCCA G -3'		
M13 Reverse	5'-CAGGA AACAG CTATG AC -3'		
SP6	5'-TATTT AGGTG ACACT ATAG-3'		
T7	5'-TAATA CGACT CACTA TAGGG-3'		

[a] Map location in wheat or *T. tauschii.*
[b] Chloroplast genome

Table 2. PCR parameters for DNA amplification

Segment	Step	Temperature (°C)	Time	Number of cycles
1	Denaturing	94	4 minutes	1
2	Denaturing Annealing Extension	94 45 or 50 72	1 minute 1 minute 1.2 minutes	30
3	Extension Holding Holding	72 4 10	7 minutes 4 minutes ∞	1 1 1

PCR Product Evaluation

The PCR products were analyzed on either 1 % agarose gel with 1 X Tris-borate EDTA running buffer or on 7 % polyacrylamide gel with 0.5 X Tris-borate EDTA running buffer. The gels were stained with ethidium bromide, visualized with UV light, and photographed.

Restriction Endonuclease Digestion

Some of the PCR products were digested with restriction endonucleases to observe differences. Restriction endonucleases are groups of enzymes that bind and cleave double-stranded DNA at specific sequences. Restriction digest contained approximately 1 U of restriction enzyme, 1 X reaction buffer and sterile distilled water to

give a total volume of 20 µl. The restriction enzymes used in this study including their restriction sequences and optimal temperatures are shown in Table 3. After one hour digestion had been completed, and the digested products are analyzed by 7 % polyacrylamide gel electrophoresis.

Table 3. Restriction endonucleases with their recognition sequences and optimal temperatures

Restriction enzyme	Recognition sequences	Optimal temperature (°C)
*Dde*I	C↓TNAG	37
*Eco*RI	G↓AATTC	37
*Hha*I	GCG↓C	37
*Hin*dIII	A↓AGCTT	37
*Hinf*I	G↓ANTC	37
*Mse*I	T↓TAA	37
*Rsa*I	GT↓AC	37
*Xba*I	T↓CTAGA	37

Diversity Analysis

The allelic diversity of nuclear and chloroplast genomes was calculated as indicated by the polymorphism information content (PIC) value described by Botstein et al. (1980) and modified by Anderson et al. (1993) for self-pollinated species. PIC value is the best indication for genetic diversity because it indicated the relative polymorphism value of each primer. The formulae is:

$$PIC_i = 1 - \sum_{j}^{n} p_{ij}^2$$

where p_{ij} is the frequency of the jth pattern of primer i and summation extends over n patterns.

Statistical Analysis

A χ^2 goodness-of-fit test was performed to test the nuclear and chloroplast correspondence for G43 and U6/R6 loci within *Ae. triuncialis*. A 2 x 2 two-way table was used for the χ^2 goodness-of-fit test. Significance was tested at the a = 0.05 level.

Cloning

PCR products were cloned to improve the quality of DNA sequencing. Ten µl of each PCR product was separated on 1 % agarose gel with Tris-borate EDTA running buffer. Band size was confirmed. Cloning for PCR product was accomplished using the pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA) or pGEM®-T vector (Promega, Madison, WI). Five to ten white colonies were selected from each plate and each single colony was cultured overnight in LB broth containing 50 µg/ml amplicillin. One µl of each culture was amplified by the same PCR condition as described previously. Twenty-five µl of each PCR product was analyzed on 1 % agarose gels with Tris-borate EDTA running buffer to ensure the insert was the correct size.

Plasmid DNA Preparation

Ten µl of each selected colony from previous procedure was cultured overnight in 3 ml of 2 X YT broth containing 50 µg/ml amplicillin at 37 °C and rotating 125 rpm. Three ml of the culture was collected and placed in two ml microcentrifuge tubes. One point five ml of cell culture was centrifuged at 12,000 G for 20 seconds twice at room temperature in the same microcentrifuge tube. The supernatant was discarded and the

bacterial pellet was resuspended in 200 µl cell suspension buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA) then mixed by vortexing. Two hundred µl of freshly prepared alkaline lysis solution (1 % SDS in 0.2 M NaOH) was added into each tube and mixed by inverting the tube several times. After that, 200 µl of 2.55 M potassium acetate, pH 4.8, was added to neutralize the suspension and gently mixed by inverting. The precipitate was visible at this point. The samples were centrifuged 12,000 G for 5 minutes. The supernatant was transferred to a new clean tube that contained 300 µl of Prep-A-Gene plasmid binding buffer (Bio-Rad). Sixty µl of resuspended Prep-A-Gene matrix was added to each tube with ten minutes shaking incubation. The well-mixed solution was transferred into a spin column, placed into a 1.5 ml microcentrifuge tube, and centrifuged 12,000 G for 30 seconds to remove the liquid phase. Five hundred µl of Prep-A-Gene washing buffer was added into each column and centrifuged 30 seconds. This wash step was performed twice. A dry centrifugation was performed at 12,000 G for 4 minutes to ensure that all wash buffer was eluted from the spin column. One hundred µl sterile deionized water was added to the spin column to elute the plasmid DNA into a new catch tube. Alcohol precipitation was performed after this step to concentrate the sample. Eleven point one µl of 3M sodium acetate pH 5.2 and 227.78 µl of 95 % ethanol were added, mixed and incubated one hour at –80 °C. The samples were centrifuged at 12,000 G for 10 minutes and then the supernatant was removed. The pellet was washed with 100 µl 80 % ethanol and centrifuged again. Finally, the samples were air dried and resuspended in 11 µl sterile deionized water. One µl of the sample was quantified by

running on a 1 % agarose gel and compared to the precision molecular mass standard (Bio-rad). The purified plasmid was then ready for sequencing.

DNA Sequencing

Standard PCR was performed in 50 µl reactions using the same condition as in PCR amplification except 0.2 µl of Taq polymerase was used per reaction and segment 2 of the PCR cycle was repeated 35 times. PCR products were cleaned by Qiagen columns to remove excess primers and salts followed by precipitation with 3 M sodium acetate pH 4.8 and 95 % ethanol. Sequencing was done on an ABI377 automated DNA sequencer and the Perkin Elmer BigDye™ sequencing reaction kit (PE Biosystems). Sequences were read in both the forward and reverse directions either using the original primer sets or plasmid primer sets. Some samples were sent out for sequencing using the same protocol at Washington State University (Pullman).

Reconstructed Phylogenetic tree

The sequences from the same primer set were first aligned by ALIGN (Scientific and Educational Software, 1989) followed by manual alignment to minimize gaps. The data were analyzed using the parsimony heuristic search, neighbor joining, and bootstrap parsimony of the Phylogeny Using Parsimony Analysis (PAUP*) program version 4.0beta8 for 32 bit Microsoft window (Swofford, 1998). The stepwise addition option was used to find the most parsimonious bootstrap trees. Bootstrap was performed by using the full heuristic search option of PAUP* to calculate the robustness of each

branch. The analysis was set with the following parameters: 100 bootstrap replicates (Felsenstein, 1985) with gaps treated as missing data, tree bisection-reconstruction branch swapping and random sequence addition. All characters were weighed equally. Bootstrap values indicated the percentage of time that resampling yielded the same clade. The goodness of fit statistic was determined to estimate reliability of each phylogenetic tree. Consistency index (CI), retention index (RI), and rescaled consistency index (RC) were calculated (Kluge and Farris, 1969; Farris, 1989). Pairwise genetic distances were calculated using the kimura2-parameter option in PHYLIP version 3.572c (DNADIST program) (Felsenstein, 1997) and MEGA (Kumar et al. 1993). *Triticum tauschii* was used as an outgroup taxon.

Results and Discussion

A total of 84 primer sets were screened on at least two accessions of each of *Ae. umbellulata*, *Ae. caudata* and *Ae. triuncialis* in order to identify genome-specific primer sets. PCR products were amplified using 45 °C (71.43 %) and 50 °C (68.65 %) annealing temperatures (Table 4). Forty-six primer sets amplified all three genomes and only three primer sets were genome-specific. U genome-specific primers were G12 and G43 and primer set D21 was C genome-specific. All primer sets did not amplified in all accessions since they were derived from related species (*Triticum tauschii*, wheat and barley), and not target species (see chapter 3) (Vanichanon et al, 1999). For preliminary work of cereal genomes, it is necessary to use primer sets that were created from across cereal

species. Due to grass genome colinearity study, Bennetzen (1999) proposed that the grasses have a single genetic system with some limitation.

Table 4. Genome specificity of primers based on annealing temperature

Annealing temperature	Number of Primers				% of primers that amplified at least some accessions
	U &UC specific	C & UC specific	All three genomes	Total	
45 °C	3	1	46	68	71.43%
50 °C	2	1	35	84	68.65%

Diversity in Polyploid *Aegilops triuncialis* and Diploid Progenitors

Six primer sets shown in Table 5 were selected to amplify PCR products from all 276 accessions of the three species to study genetic diversity. PIC values (Table 5) revealed that *Ae. caudata* was a more diverse species than *Ae. umbellulata* as reported by Chee et al. (1995). Average PIC values from all four nuclear primers (D2, D15, F8 and WG232) are 0.34, 0.84 and 0.62 for the U, C and UC genomes, respectively. These PIC values illustrated that *Ae. caudata* had the greatest diversity. *Ae. triuncialis* had more diversity than *Ae. umbellulata*, perhaps because of genetic diversity introduced from *Ae. caudata*. For example, the polymorphism seen in *Ae. triuncialis* using primer set D2 may

Table 5. Polymorphism restriction fragments distribution among U, C and UC genomes as a measure of genetic diversity.

Primer name/ Restrition Enzyme	Number of unique patterns (Number of accessions analyzed)[c]			Percentage of the most common pattern (%) observed in the U, C and UC genomes			Percentage of the most common pattern (%) observed in the U, C and UC genomes			PIC value		
	U	C	UC	U	C	UC	U	C	UC	U	C	UC
A. Nuclear Primers												
D2/ *Rsa*I	5 (34)	4 (29)	5 (99)	85.3	3.5	0	2.9	51.7	44.4	0.28	0.70	0.66
D15/ *Hha*I	6 (30)	6 (8)	5 (155)	67.6	9.1	71.6	67.6[a]	9.1[a]	71.6[a]	0.53	0.99	0.50
F8/ *Hinf*I	2 (30)	3 (20)	12 (70)	85.3	4.8	11.4	0	54.6	0	0.28	0.71	0.85
WG232/ *Mse*I	4 (30)	3 (12)	2 (170)	85.0	6.1	89.4	0	21.2	8.7	0.28	0.94	0.46
Average PIC value for all nuclear primers										**0.34**	**0.84**	**0.62**
Standard deviation of PIC value for all nuclear primers										0.13	0.15	0.17
B. Genome Specific Primers												
U specific G43/ *Dde*I	2 (28)	- (33)	2 (185)	52.9	-[b]	54.1	-	-	-	0.64	-	0.59
Chloroplast U6/R6/ NA	2 (33)	2 (27)	2 (203)	88.2	12.1	50.7	8.8	69.7	51.2	0.22	0.41	0.48

[a] There was the same bandind pattern in U diploids

[b] G43 did not amplify any products in the C genome accessions tested

[c] A total number of 34, 33 and 220 accessions of *Ae. umbellulata, Ae.caudata* and *Ae. triuncialis*, respectively, were analyzed

have originated from *Ae. caudata* because the most common pattern in *Ae. caudata* was also common in *Ae. triuncialis* (Table 5). Depending on the primer sets, *Ae. triuncialis* had similar variability to either *Ae. umbellulata* (D15) or *Ae. caudata* (D2). For instance, primer set D15 showed six banding patterns within *Ae. umbellulata* and 67.6 % of the 34 accessions had a common pattern. This pattern was also seen in 71.6 % of *Ae. triuncialis* (209 accessions). Only four accessions of *Ae. triuncialis* showed different banding patterns. The genetic variation within the allotetraploid may result from the genetic variation introduced from the diploid progenitors through multiple polyploidization or introgression or genetic variation that accumulated by mutation after polyploidization. It appears that polyploid *Ae. triuncialis* have less variation than diploid progenitor. PIC values from D15 were calculated 0.53, 0.50 and 0.99 for the U, UC and C genomes, respectively. From the D15 result, polyploidization event may occur only a few times resulting in less variation in the UC genomes.

Multiple Origins of *Aegilops triuncialis*
Inferred Using Nuclear DNA Sequence Analysis

Nuclear DNA analysis was chosen to assess multiple origins of polyploid wild wheat. Genome-specific primers were preferred because they provided unambiguous evidence that a specific *Ae. triuncialis* pattern came from either *Ae. umbellulata* or *Ae. caudata*. Therefore, G43 was selected for further evaluation because it was a U genome-specific primer. *Dde*I restriction digestion of the G43 amplified product yielded two banding patterns: allele A (Figure 1 lane U05, U"C10; fragment sizes: 300, 240, 220 base pairs) and allele B (Figure 1 lane U08, UC.04; fragment sizes: 240, 220, 200, 100 base