

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]



Université d'Ottawa • University of Ottawa



Université d'Ottawa - University of Ottawa

FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES

FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES

SOLEIMANI, Vahab D.

AUTEUR DE LA THÈSE - AUTHOR OF THESIS

M.Sc.(Biology)

GRADE - DEGREE

Biology

FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

**GENETIC DIVERSITY ESTIMATES AND DNA FINGERPRINTING OF
CANADIAN CULTIVARS OF DURUM WHEAT**

Douglas A. Johnson

DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

J.T. Arnason

I. Lambert

J. Vierula

J.-M. De Koninck, Ph.D.

LE DOYEN DE LA FACULTÉ DES ÉTUDES
SUPÉRIEURES ET POSTDOCTORALES

SIGNATURE

DEAN OF THE FACULTY OF GRADUATE
AND POSTDOCTORAL STUDIES

Genetic Diversity Estimates and DNA Fingerprinting of Canadian Cultivars of Durum Wheat

By

Vahab D. Soleimani

Thesis submitted to

the Faculty of Graduate and Postdoctoral Studies

in partial fulfilment of the requirement for the

Master's of Science Degree in the Department of Biology

Ottawa-Carleton Institute of Biology

University of Ottawa

Ottawa, Ontario

Canada

2000



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-67864-4

Canada

Acknowledgments

I would like to thank my principal supervisors Drs. Douglas A. Johnson, and Bernard R. Baum for the completion of my Master's thesis. I am indebted to them for their constructive advice and support.

I would also like to extend my gratitude and appreciation to my research committee members Drs. John T. Arnason, and Ian Lambert for their invaluable comments and counsel.

I would also like to thank N.S.E.R.C. and Dr. D.A. Johnson for partial financial support for this project. Also I would like to acknowledge Agriculture and Agri-Food Canada Research Branch for provision of laboratory space, equipments, and supplies. I am grateful to durum wheat breeders who provided information on pedigree records, especially Dr. Ron M. DePauw, Agriculture and Agri-Food Canada, Semiarid Prairie Agricultural Research Center, Swift Current, Saskatchewan. Finally I would like to thank my wife Arezu Jahani-Asl for her support and patience.

ABSTRACT

Estimates of genetic diversity present in gene pools have important implications for breeders and germplasm curators. They constitute the raw material for plant improvement and can provide protection against genetic vulnerability to biotic and abiotic stresses. The purpose of this thesis was to derive Pedigree and Amplified Restricted Fragment Polymorphism (AFLP)-based Genetic Diversity Estimates (GDE_{PED} and GDE_{AFLP}, respectively) among all currently registered 13 Canadian durum wheat cultivars in order to test the hypothesis that the actual level of genetic variation at the DNA level is lower than what is measured from pedigree data. Two objectives were set in this thesis.

The first objective was to estimate the genetic diversity level among cultivars using two independent methods namely AFLP and pedigree. This objective was successfully met by carrying out AFLP analysis on 130 individual plants representing 13 cultivars (ten individuals per cultivar). Using ten AFLP primer pairs a total of 89 polymorphic markers were generated which were used to construct a genetic distance matrix. Concurrently a pedigree was constructed for each cultivar and a genetic distance matrix based on kinship coefficient was generated. The mean of 78 GDE_{PED} pair wise comparisons was 0.76. In contrast mean GDE_{AFLP} comparisons were 0.40. AFLP analysis not only revealed a lower average genetic diversity among cultivars compared to the pedigree, but also a significant amount of genetic similarity between cultivars that were unrelated by their pedigrees. Despite different means and distributions of the two diversity measures, a moderate rank correlation ($r= 0.45$, $P<0.002$) was detected using Mantel's test.

The second objective of this thesis was to use cultivar-specific AFLP markers for identification of cultivars and transform these markers into sequence tagged site (STS) markers for routine identification of durum wheat cultivars. The second objective was successfully met and cultivar-specific AFLP markers were used to design an identification scheme for each cultivar. Conversion of AFLP into STS markers was achieved by sequencing the cultivar-specific AFLPs and using these sequences as template for STS primers design. A cultivar diagnostic PCR assay was possible with these markers by direct amplification or after digestion of the PCR product with selected restriction endonucleases.

DNA-based markers such as AFLPs and STSs are important tools for inferring genetic variations and fingerprinting and elucidating genetic relationships among plant cultivars. AFLPs have proven to be efficient in revealing diversity at below the species level and provide a better genome coverage in a single assay compared to other DNA markers. This marker system is particularly useful for the study of crop species with a narrow genetic base such as durum wheat.

RÉSUMÉ

Les estimations de la diversité génétique actuelles dans les stocks génétiques ont des implications importantes pour les sélectionneurs et les curateurs du matériel génétique. Elles constituent la matière nécessaire pour l'amélioration des plantes et peuvent assurer la protection contre la vulnérabilité génétique aux stress biotiques et abiotiques. Le but de cette thèse était de dériver des estimations de diversité génétiques basées sur le pedigree et des "Amplified Restriction Fragment Polymorphism" (AFLP) parmi les 13 cultivars de blé durum enregistrés au Canada pour tester l'hypothèse que le niveau de variation génétique au niveau de l'ADN est plus bas que ce qui est mesuré à partir des données du pedigree. Deux objectifs ont été définis dans cette thèse.

Le premier objectif était d'estimer le niveau de variation génétique parmi les cultivars en utilisant deux méthodes, notamment AFLP et pedigree. Cet objectif a été accompli par l'analyse AFLP de 130 plantes individuelles représentant 13 cultivars (dix individus par cultivar). Une matrice de distance génétique a été construite avec 89 marqueurs polymorphiques obtenus en utilisant dix paires d'amorce. Un pedigree a été construit pour chaque cultivar et une matrice de distance génétique basée sur le coefficient de parenté a été produite. La moyenne de 78 GDE_{PED} comparaisons de paires était 0.76. La moyenne des comparaisons GDE_{AFLP} était 0.40. L'analyse de AFLP a non seulement indiquée une diversité génétique moyenne plus basse parmi les cultivars, mais également une quantité significative de similitude entre les cultivars qui étaient indépendants selon leurs pedigrees. En dépit de différentes moyennes et distributions des deux mesures de diversité, une corrélation de rang modérée ($r = 0.45$, $P < 0.002$) a été détectée en utilisant le test de Mantel.

Le deuxième objectif de cette thèse était d'utiliser les marqueurs AFLP spécifiques aux cultivars pour l'identification des cultivars et de transformer ceux-ci en marqueurs de "sequence tagged site" (STS) pour les utiliser pour l'identification des cultivars de blé durum à partir de graines ou de plantes. Ce deuxième objectif a été accompli et des marqueurs AFLP spécifiques aux cultivars ont été employés pour faire un schéma d'identification pour chaque cultivar. Les AFLPs spécifiques aux cultivars ont été séquencés et étaient utilisés comme empreinte pour le design des amorces STS. Une analyse diagnostique pour les cultivars était possible avec ces marqueurs par l'amplification directe ou après une digestion du produit de PCR avec des endonucléases de restriction.

Les marqueurs basés sur l'ADN tels que les AFLPs et les STSs sont des outils importants pour l'estimation de la variation génétique, pour faire des empreintes digitales, et pour élucider les rapports génétiques parmi les cultivars en question. Les AFLPs sont avérés efficaces pour indiquer la diversité au niveau au-dessous de l'espèce et pour fournir une meilleure assurance du génome dans une seule analyse en comparaison avec des autres marqueurs d'ADN. Ce système de marqueurs est particulièrement un bon outil pour l'étude des espèces de plantes cultivées qui ont une base génétique étroite ou pauvre comme le blé durum.

TABLE OF CONTENT

ACKNOWLEDGMENTS	2
ABSTRACT / RESUME	3
TABLE OF CONTENTS	7
LIST OF TABLES.	9
LIST OF FIGURES.	10
LIST OF ABBREVIATIONS.	11
CHAPTER ONE.	12
INTRODUCTION.	12
1.1 Durum wheat breeding and genetics.	12
1.2 Genetic diversity estimation.	15
1.3 Pedigree analysis.	19
1.4 AFLP analysis.	21
1.5 Comparison of molecular (AFLP) and pedigree-based diversity estimates.	22
1.6 DNA fingerprinting and cultivar identification.	25
1.7 Transformation of AFLPs into STSs	26
1.8 Cloning, sequencing and sequence analysis of AFLP markers.	26
1.9 Thesis objectives.	27
CHAPTER TWO.	31
2.1 Summary	33
2.2 Introduction	34
2.3 Materials and Methods	36
2.3.1 Plant material	36
2.3.2 DNA extraction and AFLP analysis	36
2.3.3 Data acquisition and diversity analysis	37
2.3.4 Pedigree analysis.	38
2.3.5 Matrix comparison	39
2.3.6 Analysis of molecular variance (AMOVA)	39
2.4 Results.	39
2.4.1 AFLP analysis.	39
2.4.2 Pedigree analysis.	41
2.4.3 Matrix comparison and Mantel test.	42

2.5 Discussion	42
2.6 Acknowledgments	46
CHAPTER THREE	57
3.1 Abstract.	59
3.2 Introduction	60
3.3 Materials and Methods	62
3.3.1 Plant materials.	62
3.3.2 DNA extraction and AFLP reactions	62
3.3.3 Isolation and cloning of AFLP markers	63
3.3.4 Sequencing, sequence analysis and restriction map construction	64
3.3.5 STS primer design and optimization.	66
3.3.6 Generation of identification keys	66
3.4 Results	67
3.4.1 Cloning and sequencing of AFLP markers and sequence analysis.	69
3.4.2 Primer design, PCR conditions and restriction digest of amplified loci.	69
3.4.3 Generation of an identification key based on STSs.	70
3.5 Discussion	72
3.5.1 AFLP-based fingerprinting.	72
3.5.2 Conversion of AFLPs into STS markers	72
3.5.3 Primer design, optimization, and generation of an identification key	75
CHAPTER FOUR	84
4.1 Conclusion.	84
4.2 Future research.	89
REFERENCES	92
APPENDIX.	102
Pedigree files on CD	

LIST OF TABLES

Table 1	p. 29	Ancestors used in the pedigree of Canadian durum wheat cultivars
Table 2	p. 47	The 13 durum wheat cultivars used in the genetic diversity study
Table 3	p. 48	The number of scored polymorphic markers detected by AFLP analysis within and among Canadian durum wheat cultivars
Table 4	p. 49	Matrix of pair-wise genetic distances between Canadian durum wheat cultivars based on kinship coefficient (above diagonal) and Nei's formula (below diagonal)
Table 5	p. 51	Analysis of molecular variance (AMOVA) for within and between cultivars
Table 6	p. 76	The DNA sequences of AFLP markers used in the design of STS primers. The primer annealing sites are shown in bold and underlined. The 5' and 3' end of each sequences contain <i>EcoRI</i> and <i>MseI</i> recognition sites respectively
Table 7	p. 79	STS markers and the AFLPs from which they were derived for cultivar fingerprinting of Canadian durum wheat.
Table 8	p. 80	STS primers used in the identification key of Canadian durum wheat cultivars.

LIST OF FIGURES

- Figure 1** **P. 52** **Autoradiograph of a portion of an AFLP fingerprint using EAAG/MCAG selective primer pair demonstrating polymorphism. Lanes 1-10 'Pelissier', 11-20 'Medora'. "A" indicates cultivar-specific markers and "B" within-cultivar polymorphic markers. The EAAG primer was labelled with γ -³³P for signal detection by autoradiography.**
- Figure 2** **p. 53** **UPGMA dendogram showing genetic relationships among the 13 durum wheat cultivars used in this study. The dendogram was constructed using 41 cultivar-specific AFLP markers and was based on the genetic distance calculated according to Nei's genetic coefficient.**
- Figure 3** **p. 54** **UPGMA dendogram depicting genetic relationships among 130 durum wheat plants representing 13 cultivars. The dendogram was constructed using 89 AFLP markers based on the genetic distance calculated according to Nei's genetic coefficient. The numbers in each cluster refer to individual plants. All 130 plants fell into their respective cultivar group.**
- Figure 4** **p. 55** **Comparison of the genetic diversity per locus among the individual plants within each cultivar based on interhaplotypic distances. The distances were measured based on the number of differences in band sharing data from pair wise comparison of cultivars.**
- Figure 5** **p. 56** **UPGMA dendogram showing genetic relationships among the 13 durum wheat cultivars used in this study. The dendogram was constructed using the matrix of kinship coefficients.**
- Figure 6** **p. 81** **Silver stained gel showing AFLP profile of two durum wheat cultivars using EAAC/MCTA selective primer pair. Arrows indicates cultivar-specific polymorphic markers that were selected for cloning and subsequent transformation into STS markers.**
- Figure 7** **p. 82** **Agarose gel showing a polymorphic profile of durum wheat cultivars using G03-255 STS primer pair. Lane 1-13 are the 13 durum cultivars, lane 14 is DNA size marker.**
- Figure 8** **p. 83** **Agarose gel showing a restriction fragment polymorphism of A05-290 digested with *Hae*III restriction enzyme. Cultivars that have one *Hae*III recognition site within the amplicon have two bands and those with two such sites have three bands.**
- Figure 9** **p. 91** **Comparison of the mean genetic diversity estimates based on AFLP and pedigree for Canadian durum wheat cultivars.**

LIST OF ABBREVIATIONS

AFLP	Amplified Restriction Fragment Polymorphism
AMOVA	Analysis of Molecular Variance
BLAST	Basic local Alignment Search Tools
COP	Coefficient of parentage
GDE	Genetic Diversity Estimate
IPTG	Isopropyl-Thio-β-D-Galactopyranoside
KIN	Kinship coefficient
NCBI	National Centre for Biotechnology Information
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SCAR	Sequence Characterized Amplified Regions
SSR	Simple Sequence Repeat
STS	Sequence Tagged Site
UPGMA	Unweighted Pair-Group Method with Arithmetic Averaging
X-Gal	5-Bromo-4-Chloro-3-Indolyl-β-D-Galactopyranoside

CHAPTER ONE

INTRODUCTION

1.1 Durum wheat breeding and genetics

Wheat was probably domesticated around 10,000 - 15,000 B.C. in the near east area known as the Fertile Crescent. Domestication of wheat, barley, and other grains resulted in the settling of local people, thus determining the evolution of human culture from the “hunter-gatherer” phase to the “agricultural” phase (Fabriani and Lintas 1988). The availability of special types of grasses that were able to grow in large grassy patches with spikes characterized by large kernels led to the growth and expansion of western Asian civilization, resulting in the Egyptian and later the European and Western civilizations.

All wheat species belong to the genus *Triticum*, a member of the grass (*Gramineae* [*Poaceae*]) family and the *Triticeae* tribe in which 1-3 flowered spikelets are sessile and alternate on opposite sides of a rachis.

Linneus proposed the first classification of *Triticum* in 1753 based on morphological differences. After Linneus, several botanists proposed various classifications of the genus. In 1918 Sakamura proposed a key for classification of wheat based on cytogenetic studies that is still in use today (Fabriani and Lintas 1988). Cytologic and cytogenetic work showed that wheat falls into three basic groups each having 14 chromosomes (seven pairs) or a multiple of 14 chromosomes in each somatic cell. At the diploid level two species that contain the A genome are recognized: *T. monococcum* L., with both wild and cultivated types and *T. urartu* L., a wild species that can be differentiated from *T. monococcum* on genetic, morphological and

biochemical bases. At the tetraploid level two main species have also been recognized: *T. timopheevi* and *T. turgidum* L. These two species have the A genome from *T. monococcum* in common, but the second genome in *T. timopheevi* is called the G genome (or Bt by some authors) and in *T. turgidum* it is the B genome. Although the origin of both B and G genomes is still debated, they most likely were derived from an ancestral species belonging to a section of the genus *Agilops* or from a modified genome A. The third group is represented by hexaploid wheat which is also characterized as having two main species: *T. zhukovskyi*, in which genome A is represented twice and genome G once, and *T. aestivum* (bread wheat) which is derived from *T. turgidum* (AB genome) and *A tauschii* (genome D)

Among all cultivated tetraploid wheats *Triticum durum* (AABB genome) types are most important even though they are cultivated in only 10% of all the wheat cultivated areas. The remaining 90% of wheat cultivated area is used up by bread wheat. Canada produces 5 per cent of the world's total wheat (20 billion bushels annually) (Fowler, D.B. 2000). However, due to its relatively small population to feed, Canada exports about 75 per cent of its annual wheat production and accounts for 20 per cent of the world's wheat export . Durum wheat is the hardiest kind of wheat in terms of growth habits and is grown predominantly in the Near East and the Mediterranean regions. In North America durum production is concentrated mainly around North Dakota and Canadian Prairies. Durum's density combined with its high protein content and gluten strength makes it the wheat of choice for producing premium pasta products. When durum is milled, the endosperm is ground into a granular product called semolina. A mixture of semolina and water produces a stiff dough which is forced through metal discs with holes to create many differently shaped pasta products.

In Canada, wheat cultivation began with the first French colonists who settled in eastern Canada in early 17th century (MacGibbon, 1932). The type of wheat grown was most likely winter wheat. Durum wheat was first introduced into western Canada after the severe rust epidemic of 1916 which destroyed the hard red spring (a bread wheat class) wheat crop. The durum cultivars which were imported from Minnesota and North Dakota were resistant to the prevalent strain of rust and as a result western Canadian farmers began to cultivate more durum wheat. Durum acreage rose steadily and reached a peak in 1928, when 56% of the total wheat acreage in Manitoba (1.5 million acres) was planted with durum wheat (Clark and Bayles, 1942). In the period 1960-1969 Canadian durum production averaged about 1 million metric tonnes (7 million bushels) per year. During the following decade (1970-1979), the average production doubled to 2 million metric tonnes. Both the production and export of durum wheat have steadily increased over the past few decades. This increase in production and also export to some degree reflects the significant improvement in overall quality of the durum wheat cultivars developed by Canadian breeders.

Durum cultivars that were initially grown in Western Canada were those imported from the United States mainly North Dakota, including 'Kubanka', 'Arnautka', 'Buford', 'Marouani', and 'Acme'. 'Mindum' was first introduced into Manitoba in 1921 and soon became the predominant variety. The first Canadian variety ('Stewart 63') was licensed in 1963 and became the major Canadian durum variety in 1960s. Currently, there are 13 varieties licensed in Canada (Table 2 p. 46). The variety licensing system in Canada is the key to maintaining a high average quality in grades of grains. Only the varieties that are judged to be as good as or better than the statutory standards are licensed. The statutory standards are given in the grade schedule of the

Canada Grain Act (Canada, 1970) for hard red spring wheat, amber durum wheat, as well as six row and two row malting barley.

1.2 Genetic diversity estimation

Canadian durum breeders have traditionally relied on pedigree methods to develop new durum cultivars. The new cultivars have been selected mainly for yield and disease resistance with little emphasis on maintaining a broad genetic base. Increasing crop yield through practices aimed at homogenization and achieving cultivar uniformity may inevitably lead to the susceptibility of the crops to pests and substantial losses to producers due to reduction in crop genetic base. Therefore it is important to maintain a certain level of genetic variation in the gene pool of crop species.

In applied genetics, genetic diversity is a statistical concept referring to the variance at an individual gene locus, among several loci or gene combinations, between individuals within a population, or between populations. Unless we can measure genetic diversity, we cannot predict its occurrence or assign an economic value to it. Since a plant's genotype is distinct from its phenotype, the relationships between genetic variation and phenotypic traits observed in the farmer's field are not always clear. The environment in which a plant grows can influence the expression of its genes. Also many agronomically important traits are controlled by more than one gene. Examples of such traits include grain quality, yield, and polygenic forms of pest resistance (Smale and McBride 2000; Rogers et al. 1983). It is obvious from these studies that variation at the DNA sequence level is not always manifested in the phenotype.

Measurements of the levels of genetic diversity present in germplasm have important

implications for plant breeders, germplasm curators, researchers, and policy makers. In germplasm collections, such measurements may help designate core collections to enhance the efficiency of collection management and utilization (Brown et al. 1987; Jiang et al. 1975). Diversity estimates may also be utilized for heterotic grouping (Cox and Murphy 1990) and performance which were originally developed by U.S. corn (*Zea mays* L.) breeders. Heterosis is an increase in the size or vigor of a hybrid relative to its parents. More recently, an interest in quantification of genetic diversity of pure line cultivar crops has developed. Transgressive segregation may be more likely to occur when parents in a cross are less similar allowing more favorable alleles to be combined in the offspring the combination of which is absent in either parents (Cowen and Frey, 1987; van Beuningen and Busch 1997). The viewpoint that the production of hybrid cultivars combined with the detection of heterotic patterns may become important for wheat adds to the importance of quantification of genetic distance and relatedness (Cregan and Busch, 1978; Cox and Murphy, 1990; van Beuningen and Busch, 1997). Maintaining a high level of genetic variation could prevent pest epidemics and prevent financial losses. Also a knowledge of the amount, the extent and the distribution of genetic variation is central to the development of effective conservation strategies. Just as there are significant differences in the amount of diversity present in different species, so there may be substantial differences between different groups within species (Saghai-Marouf et al. 1994). It has been commonly observed that modern cultivars of crop species have significantly lower levels of genetic variation than their wild relatives (Allard, 1992). Differences between populations within a species in the total amount of diversity appear to be greater in self-pollinated species than in cross-pollinated species (Schoen and Brown, 1991).

Distance estimation and quantification were traditionally based entirely on morphological markers and quantitative traits (Goodman, 1972). Obtaining accurate measures of genetic variation with these markers was complicated due to a lack of correspondence between genotypic content and plant phenotype as mentioned earlier and a limited number of observable markers which can be used for diversity estimation. Recently, genetic markers in the form of gene products such as isozymes (Cox et al. 1988), storage proteins (Cox et al. 1985), and direct DNA markers such as restriction fragment length polymorphism (RFLP) (Lubers et al. 1991), random amplified polymorphic DNA (RAPD) (Thorman and Osborn, 1992; Tinker et al. 1993), and amplified restriction fragment polymorphism (AFLP) (Barrett et al. 1998) have been used for genetic diversity estimation and measurements. Quantitative traits are least suited as genetic markers because of the environmental influence in the forms of epistatic and pleiotropic effects. In addition, these traits are coded by an unknown number of genes. DNA-based markers have the advantage of detecting cryptic variation at the DNA sequence level which would otherwise be undetected when phenotypic markers are used. In the case of wheat most of the alternative sources of molecular markers also have technique-specific problems such as lack of polymorphism for isozymes and RFLP markers (Cox et al. 1988), and low levels of polymorphism with RAPD markers (Devos and Gale 1993). AFLP is a recently developed DNA analysis marker system based on a novel, powerful combination of polymerase chain reaction (PCR) and restriction enzyme digestion (Vos et al. 1995). AFLPs are highly efficient compared with other DNA marker systems, are reproducible, some exhibit interspecific homology (Powell et al. 1996; Barrett and Kidwell, 1998), and can be used to analyze a large number of loci simultaneously. AFLP has been used for genetic diversity estimation in hexaploid wheat (Barrett

et al. 1998), soybean (VanToal et al. 1997), barley (Qi and Lindhout, 1997), and potato (Roupe van der Voort et al. 1997). Due to a high multiplex ratio (average number of loci detected per assay) it can overcome many of the limitations inherent to other molecular marker systems. However, like many other DNA markers AFLPs are unmapped and the extent of genome coverage with this marker system may not be always obvious. Previous studies have shown that AFLPs tend to be clustered and therefore may not represent independent variables (Karp et al. 1997). This problem can be resolved to some extent by using a large number of AFLP markers in which case there will still be some redundancy due to genetic linkage however this source of error may be small in comparison to the large number of polymorphic loci that are analyzed.

AFLP markers can also be used for monitoring allelic diversity of expressed genes by using methylation sensitive restriction enzymes such as *PstI*. Low levels of methylation tend to be associated with high levels of gene expression (Cedar, 1988), which suggests hypomethylated regions may have more effect on the phenotype than the hypermethylated DNA associated with heterochromatic regions of the genome. AFLP-based measurements of genetic variation in bread wheat were found to be lower when hypomethylated regions of genome were targeted compared to hypermethylated regions (Barrett et al. 1998). Genetic diversity measures in the same study were more correlated to the pedigree estimates when *PstI* (methylation sensitive enzyme) was used instead of *EcoRI* (methylation insensitive enzyme). The reasons for larger differences in mean diversity level and greater correlation between *PstI*-based genetic diversity measure and pedigree are largely unknown. However, a methylation insensitive enzyme is expected to produce a more random population of fragments throughout the genome than when its methylation sensitive counterpart is used. Methylation sensitive enzymes may lead to clustering of AFLP

markers in the genome and a deviation from random distribution of AFLPs may in turn effect the estimates of genetic relationships between individuals or populations. Both *EcoRI* and *Tru9I* restriction enzymes that were used in this study are methylation insensitive and therefore they are expected to cleave DNA at both hypermethylated as well as hypomethylated regions resulting in a random distribution of AFLPs in the durum wheat genome.

1.3 Pedigree analysis

The 13 currently registered Canadian durum wheat cultivars used in this study are derived from 132 cultivars, breeding lines, selection, and landraces (Table 1 p. 28). For each cultivar a pedigree was constructed based on its description, published pedigree records (Zeven and Zeven-Hessinik 1975; Zeven, and van Hintum, 1992; Kohli, 1986), and personal communication with breeders (CD appendix). Pedigree based diversity estimates were measured using kinship coefficient (KIN) (Tinker and Mather, 1993). KIN can be used as an inexpensive means of estimating genetic diversity in self pollinating crop species with a known pedigree. It is the probability that a random allele taken from a random locus in one individual is identical by descent to another random allele at the same locus in another individual based on the following assumptions: (1) all ancestors, cultivars, and breeding lines are homozygous and homogenous; (2) cultivars without a known pedigrees are unrelated; (3) parents make equal contribution of alleles to each progeny. Given these assumptions, KIN statistics may lead to overestimation of genetic distance between cultivars especially when comprehensive pedigree information is not available. The assumption that genotypes not related by pedigree do not carry homologous DNA can lead to an overestimation of genetic distance. An RFLP survey of wheat land races from Iran

and Afghanistan revealed a high level of genetic similarity (mean = 0.91) (Kim and Ward, 1997). Similar results were reported by Autrique et al. (1996) on durum wheat in which only two out of 134 alleles analyzed by RFLP were different between two ancestral lines of durum wheat.

As plant breeders continue to select new lines with desirable traits and cross them with existing materials, their pedigrees become longer. In most cases the breeders do not know the precise genealogy of the new material they use. These new materials that are brought into wheat breeding programs are typically advanced lines with long pedigrees that may contain numerous landraces. The more international the breeding program the more likely it is that the material will include landraces or lines with unknown genealogies. Since the precision of genetic diversity estimates based on pedigree data depends on the availability of accurate pedigree records, the inclusion of landraces and breeding lines with unknown genealogies can result in an overestimation of the actual genetic variation (for the reasons described earlier). Selection pressure and genetic drift can also limit allelic variation at loci controlling traits that are common objectives of selection programs (Barrett et al. 1998). Selection pressure may also limit the level of allelic diversity among linked loci in which only specific alleles conferring essential characteristic are advanced. It is obvious that selection pressure, genetic drift, and genetic linkage can play a major role in breeding programs for the reasons described earlier. When genetic diversity is estimated from the pedigree data all these factors are ignored and the result is most likely an overestimation of the actual level of diversity present in the population. On the other hand, molecular markers such as AFLPs can detect variation (as a result of selection, drift, and linkage) at the DNA level and therefore can serve as a better estimator of genetic variation.

1.4 AFLP analysis

AFLP is a novel DNA fingerprinting technique that involves PCR amplification of a selected subset of restriction fragments and their separation by polyacrylamide gel electrophoresis (Vos et al. 1995). It has been used for genetic diversity measurements, identification, and mapping in a wide variety of organisms. DNA is first digested with two restriction enzymes typically a six-base cutter such as *EcoRI* and a four-base cutter such as *MseI*. The choice of enzyme and their methylation status may depend on the objectives of the study. Typically, a combination of *EcoRI* and *MseI* is used for the study of genetic diversity in a wide variety of organisms. Adapters of known sequence are ligated to the restricted fragments and the adaptor sequence is then used as a template for specific PCR amplification of restricted fragments. PCR amplification is carried out in two steps: a pre-amplification PCR reaction with primers that are complementary to the sequence of adapter and contain a single additional selective nucleotide at their 3' end. The presence of an additional nucleotide ensures that only those fragments containing complementary bases to the selective nucleotide will be amplified resulting in a 16 fold reduction in the number of amplified restriction fragments. The product of pre-amplification PCR can be subsequently used as template for a second round of PCR amplification with primers containing 2 to 3 more selective nucleotides for subsequent reductions in the number of amplified restricted fragments. Depending on the number of selective nucleotides at the 3' end of PCR primer sequences, a subset of 1/16, 1/1056, and 1/4096 of the total restricted fragments could be amplified for 1, 2, and 3 additional nucleotides at the end of each primer pair respectively assuming that the distribution of nucleotides is random throughout the genome.

Like other PCR-based DNA markers such as RAPDs, AFLPs are mainly dominant (Powell et al. 1998). Moreover, the identity of AFLP bands on a gel are not always known and co-migrating bands may not correspond to homologous loci. The polymorphism could arise as a result of mutations and base alterations in the restriction enzyme recognition site, mutation within the sequence of selective nucleotides, or insertions/deletions in various regions in the genome. These markers are particularly important in detecting variations at below the species level as well as in species with a narrow genetic base.

1.5 *Comparison of Molecular (AFLP) and Pedigree-based diversity estimates*

Although the estimates of genetic diversity based on pedigree and DNA markers have been found to be strongly correlated in maize (*Zea mays* L.) (Gerdes and Tracy 1994; Messmer et al. 1993; Melchinger et al. 1993; Smith et al. 1990), in general a low level correlation has been reported in durum wheat (Autrique et al. 1996), barley (Tinker et al. 1993), and bread wheat (Barret et al. 1998). The discrepancy between pedigree and DNA marker-based genetic diversity estimates may be the result of the assumptions that are made in the calculation of genetic relatedness based on pedigree (Tinker et al. 1994; Cox and Murphy, 1990; Souza and Sorrells, 1989; Martin et al. 1991). Pedigree-based genetic diversity estimates may be biased by selection, genetic drift, and unequal parentage contribution (Hintum and Haalman, 1994; Cox and Murphy, 1990). The development of new varieties of plants for specific end product and agronomic characteristics requires repeated rounds of selection under specific environmental and physiological conditions. In this process, genes that are linked to the loci conferring the desired trait will also be selected. Therefore, under selection pressure the parentage contribution of

linked loci is expected to be unequal resulting in erroneous measures of pedigree-based diversity levels. In addition, the relatedness of ancestors without a known pedigree may be another major factor contributing to the discrepancy between the two methods. An RFLP analysis by Autrique et al. (1996) of two ancestral lines of durum wheat ('Khapli' and 'Vernal Emmer') which are also present as ancestral lines in most Canadian durum wheat cultivars showed that these two lines differed at only two out of the 134 alleles that were assayed. Similarly, an RFLP analysis of 55 landraces from Iran, Afghanistan, and Turkey which were unrelated by pedigree detected a high level of genetic similarity (mean=0.91) among accessions (Kim and Ward, 1997). Results of other studies show that by excluding diverse pedigree comparisons from correlation analysis the association between pedigree and DNA marker-based genetic diversity estimates could be increased. Tinker et al. (1993) compared RAPD and pedigree-based diversity estimates among 27 spring barley lines (*Hordeum vulgare* L.). When diverse pedigree values ($GDE > 0.75$) were excluded from the analysis, the rank correlation (r_s) between the two diversity estimates increased from $r_s = 0.51$ ($p < 0.01$) to $r_s = 0.73$ ($p < 0.01$). Kim and Ward (1997) showed that an increase in the correlation between RFLP and pedigree-based diversity estimates corresponded to a decrease in diversity estimates among soft winter wheat cultivars from the Eastern USA. This increased correlation may indicate an increase in the proportion of relatedness that can be accounted for by pedigree analysis as the germplasm pool moves away from its original configuration through cycles of selection. In bread wheat Barrett et al. (1998) obtained a mean genetic diversity of 0.54 with AFLP markers compared to the value of 0.96 obtained from pedigree study. Moreover, in the same study AFLP-based GDE were normally distributed but pedigree-based GDE were skewed to the right (i.e. 89% of all pedigree GDE were > 0.90). The

correlation coefficient between the two diversity measures was a low 0.28. However, when methylation sensitive enzymes were used in their AFLP analysis, the correlation coefficient was increased to 0.44. Although the reasons for this finding are mainly unknown, the increase in correlation coefficient using methylation sensitive enzymes maybe due to the distribution of their recognition sites in the genome.

In general lower genetic variation is detected with DNA markers compared to pedigree analysis (Kim and Ward, 1997; Tinker et al. 1993; Barrett et al. 1998). This is likely due to genetic drift and selection pressure in breeding programs as well as the relatedness of ancestors without a known pedigree which are not accounted for in pedigree-based genetic diversity estimation. There seems to be general agreement that DNA markers are more suited for genetic diversity estimates (Davila et al. 1998; Donoghue et al. 1994; Kim and Ward, 1997; Beer et al. 1993). This is mainly because the latter approach overcomes many of the limitations and assumptions inherent to the pedigree method. However, the extent of a DNA markers' utility may depend on the nature of the marker, their numbers, genome coverage, and the population under investigation. Despite their apparent advantage over pedigree method, DNA markers may underestimate the actual levels of genetic diversity. This may happen with co-migrating bands which are derived from different loci but are considered to be homologous. The extent of this type of error can vary depending on the marker system and the level of resolution of gel electrophoresis apparatus.

1.6 DNA fingerprinting and cultivar identification

Unequivocal identification of plant varieties is important for breeding purposes and plant proprietary right protection, and to ensure that varietal purity is maintained throughout the grain handling system. In order to maintain the varietal purity, there must be a system in place that can identify varieties quickly and at all points in the grain handling system (Baum et al. 2000). Currently, the system relies on visual identification based on morphological data (e.g. seed shape and morphology). Since these characters are limited both in number and state, their use can pose problems for identification of closely related cultivars. The development of DNA-based kits can provide a fast and accurate testing alternative for varietal identification throughout the grain handling system and can be effectively used to test varieties at any point in their development even at the seed and embryo stage.

A variety of DNA markers have been used for varietal identification in plants including RAPDs (Koller et al. 1993; Baum et al. 2000), minisatellites and oligonucleotide probes (Tzuri et al. 1991; Kaemmer et al. 1992), RFLPs (Gepts et al. 1993), microsattelites (Blair and McCouch, 1997), Sequence Characteristic Amplified Regions (SCARs) (Jiang and Sink, 1997; Deng et al. 1997), Sequence Tagged Sites (STSs) (Blair and McCouch, 1997; Baum et al. 2000), and AFLPs (Reamon-Buttner et al. 1998).

1.7 Transformation of AFLPs into sequence tagged site (STS) markers

Specific STS primers can be designed from the sequences of previously known molecular markers such as RAPDs (Talbert et al. 1996; Deng et al. 1997; Kawchuk et al. 1998), RFLPs (Blake et al. 1996; Erpelding et al. 1996; Williams et al. 1996), and AFLPs (Qu et al. 1998; de Jong et al. 1997; Bradeen and Simon, 1998). STS and SCARs were initially found useful in transforming RAPDs into reliable PCR-based markers.(Olsen et al. 1989; Paran and Michelmore 1993). They offer a cheap, easy, and fast alternative to the previously mentioned marker systems which have specific problems associated with them such as low levels of reproducibility (RAPDs), and a high cost and technical difficulties with AFLPs. STS markers are similar to SCARs (Michelmore et al. 1993) and may serve as an anchoring point between physical and genetic maps. Their reproducible amplification of defined genomic regions allows the comparative mapping (as has been done with RFLPs and AFLPs) or homology studies between related species. Once developed these markers are prone to automation for diagnostic and identification purposes.

1.8 Cloning, sequencing, and sequence analysis of AFLP markers

The nature and genomic location of the AFLP markers used in the development of STS loci in this study were unknown. However, all these markers exhibited cultivar-specific behavior (i.e. monomorphic pattern among individual plants within a cultivar and polymorphic pattern between cultivars). In order to overcome the cost and technical constraints associated with AFLP technology, we have transformed these markers into a PCR-based marker system which would be easier to use widely in breeding programs and cultivar purity testing. The transformation of

AFLPs into STS markers has been described by Reamon-Buttner and Jung (2000). In this method the AFLP marker of interest is recovered from polyacrylamide gel, cloned and sequenced as described. The sequence can then be used for designing STS primers. This method has been used to obtain markers linked to sex in *Asparagus officinalis* L. (Reamon-Buttner and Jung 2000), and for tagging of the dwarf BREIZH (*Bzh*) gene in *Brassica napus* L. (Barrett et al. 1998).

1.9 Thesis objectives

Our hypothesis was that the AFLP-based diversity estimates would be lower than the pedigree-based estimates and furthermore AFLP would be a better estimator of the actual diversity level than the pedigree method. To this end two main objectives were set. The first objective was to estimate the level of genetic diversity/relatedness present in the Canadian cultivars of durum wheat using two independent methods namely molecular (AFLPs) and pedigree methods. To this end 89 AFLP polymorphic markers were obtained and a genetic distance matrix was constructed based on the presence/absence of bands at polymorphic loci. Dendograms representing genetic relationships among various durum wheat cultivars were constructed based on Nei's genetic distance obtained from the AFLP data. Part of the first objective was to construct a pedigree for each durum cultivar and use this information to estimate genetic diversity based on the kinship coefficient

The second objective of the thesis was to fingerprint durum wheat cultivars using AFLPs and to transform these markers into a simple PCR-based STS marker system for rapid identification of cultivars at any stage of plant development.

In this thesis the relevance of DNA markers as better estimators of genetic diversity in

durum wheat is reported. A substantial level of genetic variation within each cultivar and in all cultivars was detected with the AFLP method. Finally, two identification keys, one based on AFLP polymorphism and the other based on their STSs, are presented.

This thesis is presented in the form of two papers of submitted status. Chapter 2 and 3 represent the reproduction of each paper. The references have been removed from the ends of each chapters and have been combined into one general section at the end of the thesis together with references from the introduction and discussion (chapter 4). Chapter 4 provides an overview of the results and a synthesis of discussion section of each chapter with the view to future work. A CD containing detailed pedigree of all cultivars excluding 'Pelissier' which has no pedigree information has been presented in the appendix under JPE folder.

Table 1 Cultivars name and their codes used in the ancestry of Canadian durum wheat cultivars

code	cultivar	code	cultivar
AD21	8267-AD2A	LD32	LD357
AVN1	AC AVONLEA	LD33	LD308
BR11	BR180	LDN1	LANGDON
BRV1	BREVOR	LDS1	LEEDS
BRV2	BREVON	LYD1	LLOYD
CAR1	CARLETON	MCN1	MACOUN
CLR1	COULTER	MED1	MEDORA
CND1	CANDO	MEL1	AC MELITA
D211	D21563	MEX1	MEXICALI 75
D651	D6530	MRS1	AC MORSE
D652	D65114	NAV1	AC NAVIGATOR
D741	D74112	ND71	ND74112
DT15	DT182	NDK1	NOTH DAKOTA 61-130
DT18	DT188	NUG1	NUGGET
DT21	DT224	PAT1	AC PATHFINDER
DT31	DT354	PLN1	PLENTY
DT33	DT322	PLS	PELISSIER
DT34	DT320	QLF1	QUILAFEN
DT36	DT367	RL31	RL3097
DT37	DT379	RL32	RL3304
DT41	DT430	RL33	RL3380
DT43	DT437	RL34	RL3607
DT61	DT612	RMS1	RAMSEY
EDM1	EDMORE	RUG1	RUGBY
EDN1	EDEN	SEP1	SCEPTRE
FED1	FEDERATION	SNT1	SENTRY
FLR1	FLORENCE	STR1	STEWART
HRB1	HORNBLLENDE	TEH1	TEHUACAN
HRC1	HERCULES	UK1	UNKNOWN1
HSR1	HUSSAR	UK10	UNKNOWN10
IND1	INDIAN	UK11	UNKNOWN11
KPL1	KHAPLI	UK12	UNKNOWN12

KYL1	KYLE	UK13	UNKNOWN13
LAK1	LAKOTA	UK14	UNKNOWN14
LD31	LD379	UK15	UNKNOWN15
UK16	UNKNOWN16	UK52	UNKNOWN52
UK17	UNKNOWN17	UK54	UNKNOWN54
UK19	UNKNOWN19	UK56	UNKNOWN56
UK2	UNKNOWN2	UK57	UNKNOWN57
UK20	UNKNOWN20	UK58	UNKNOWN58
UK21	UNKNOWN21	UK59	UNKNOWN59
UK22	UNKNOWN22	UK6	UNKNOWN6
UK23	UNKNOWN23	UK60	UNKNOWN60
UK24	UNKNOWN24	UK62	UNKNOWN62
UK25	UNKNOWN25	UK63	UNKNOWN63
UK26	UNKNOWN26	UK64	UNKNOWN64
UK27	UNKNOWN27	UK65	UNKNOWN65
UK28	UNKNOWN28	UK66	UNKNOWN66
UK29	UNKNOWN29	UK67	UNKNOWN67
UK3	UNKNOWN3	UK68	UNKNOWN68
UK30	UNKNOWN30	UK69	UNKNOWN69
UK31	UNKNOWN31	UK7	UNKNOWN7
UK32	UNKNOWN32	UK70	UNKNOWN70
UK33	UNKNOWN33	UK8	UNKNOWN8
UK34	UNKNOWN34	UK9	UNKNOWN9
UK35	UNKNOWN35	VRN1	VERNUM
UK36	UNKNOWN36	WAK1	WAKOOMA
UK37	UNKNOWN37	WBD2	WESTBRED881
UK38	UNKNOWN38	WEL1	WELLS
UK39	UNKNOWN39	WRD1	WARD
UK4	UNKNOWN4	WSC1	WASCANA
UK40	UNKNOWN40	YAN1	YANDILLA
UK41	UNKNOWN41		
UK42	UNKNOWN42		
UK43	UNKNOWN43		
UK44	UNKNOWN44		
UK45	UNKNOWN45		
UK46	UNKNOWN46		
UK47	UNKNOWN47		
UK48	UNKNOWN48		
UK49	UNKNOWN49		
UK50	UNKNOWN50		
UK51	UNKNOWN51		

CHAPTER TWO

AFLP and Pedigree-based genetic diversity estimates in modern cultivars of durum wheat (*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.)

The following chapter is a reproduction of a paper submitted to the Theoretical and Applied Genetics journal by Soleimani D. Vahab, Baum, B.R., and Johnson, D.A. The chapter fulfils the first objective of the thesis: Estimating genetic diversity among currently registered Canadian durum wheat cultivars based on molecular (AFLP) and pedigree approaches. AFLP technology was used to measure the levels of genetic variation between individual plants within a cultivar as well as between cultivars. This was accomplished by analyzing polymorphic markers at various loci and subsequent transformation of these markers into a genetic distance matrix which was used for estimation of genetic diversity/relatedness. In addition, a comprehensive pedigree was constructed for each durum wheat cultivar based on cultivar description, breeding record, personal communication with breeder, and published pedigree data. The data obtained from this study was transformed into a genetic distance matrix using Kinship coefficient.

The results from the two independent approaches mentioned above were compared using various statistical methods in order to assess their utility for future durum wheat breeding.

AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat**(*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.)****Soleimani VD, (1,2), *Baum BR, (1), Johnson DA (2)**

1) Eastern Cereal and Oilseed Research Centre, Agriculture & Agri-Food Canada, Central
Experimental Farm, Ottawa, Canada, ON K1A 0C6

2) Ottawa-Carleton Institute of Biology, University of Ottawa, 30 Marie Curie, Ottawa, Canada,
ON K1N 6N5

***Corresponding author**

Tel: 613-759-1821

Fax: 613-759-1924

e-mail: baumbr@em.agr.ca

2.1 Summary

A substantial amount of between and within cultivar genetic variation was detected in modern Canadian durum wheat (*Triticum turgidum* L. ssp. *durum* (Desf.) Husn.) cultivars based upon Amplified Restriction Fragment Polymorphism (AFLP). The ancestry of Canadian durum wheat cultivars was traced back to 132 cultivars, selections, and breeding lines including 17 landraces. Mean pair-wise genetic distance based on KIN was 0.76. In contrast AFLP-based mean pair-wise genetic distance was 0.40. Despite a large difference between the means of the two diversity measures a moderate positive correlation ($r = 0.457$, $p < 0.002$) was detected between the two distance matrices. Cluster analysis with the AFLP data divided all cultivars into three major groups reflecting their breeding origins. One group contained 'Pelissier' alone, which was a selection from a landrace introduced into US from Algeria. On the other hand such groupings among cultivars were not evident when KIN was used for genetic diversity measures instead. The level of genetic variation among individuals within a cultivar at the breeders' seed level was estimated based on an interhaplotypic distance matrix derived from the AFLP data. We found that the level of genetic variation within the most developed cultivars is fairly substantial despite rigorous selection pressure aimed at cultivar purity in breeding programs. Comparison of AFLP and pedigree-based genetic diversity estimates in crop species such as durum wheat can provide important information for plant improvement.

Key words: Amplified Restriction Fragment Polymorphism (AFLP), kinship coefficient (KIN), genetic diversity estimates, durum wheat

2.2 Introduction

Measurements of genetic diversity in cultivated crops have important implications for breeding programs and for the conservation of genetic resources. Both pedigree (Cox et al. 1985; Martin et al. 1991) and DNA markers (Karp et al. 1996; Barrette et al. 1998; Davila et al. 1998) have been used to measure genetic diversity levels among genotypes. In self pollinating crop species with a known pedigree, the kinship coefficient (KIN) (Tinker and Mather, 1993) can be used as an inexpensive tool for genetic diversity estimates. KIN is the probability of identity by descent for a particular allele between two genotypes. It gives a theoretical measure of genetic diversity/relatedness among cultivars based on the assumption of equal parental contribution. Pedigree-based diversity measurements can result in an overestimation of the actual level of genetic diversity present in the gene pool as a result of the assumption that are made regarding genetic drift, selection pressure, and the relatedness of ancestors without a known pedigree (Cox et al 1985; Kim and Ward, 1997; Barrette et al. 1998; Graner et al 1994).

DNA markers have the advantage of directly detecting sequence variation among cultivars and therefore the ability to bypass the assumptions that are inherent to pedigree analysis. Autrique et al. (1996) identified Restriction Fragment Length Polymorphisms (RFLPs) in 113 improved cultivars and landraces of durum wheat and reported a mean genetic distance of 0.21 and 0.31 within the improved lines and landraces respectively. In bread wheat, a mean genetic distance of 0.54 was reported by Barrette et al. (1998) with AFLP markers.

Correlation between DNA markers and pedigree-based diversity estimates has varied from 0.21 between RFLP and Pedigree methods in barley (Graner et al. 1994) and durum wheat (Autrique et al. 1996) to 0.81 between RFLP and pedigree data among 37 maize (*Zea mays* L.)

inbred lines (Smith et al. 1990). A moderate correlation coefficient value of 0.42 was found between AFLP and pedigree data in bread wheat cultivars from the Pacific Northwest (Barrette et al. 1998).

DNA markers are most suitable for genetic diversity estimates (Davila et al. 1998; Donoughue et al. 1994; Kim and Ward, 1997; Ahnert et al. 1996; Plaschke et al. 1995), however, the extent of their utility may depend on the nature of the marker, their number, genome coverage, and the population under investigation. DNA-based markers are particularly useful for wheat or other crop species with a narrow genetic basis. Markers that can detect higher levels of polymorphism between wheat cultivars can be utilized to derive more accurate genetic diversity estimates. The objective of this study was to investigate the utility of AFLP fingerprinting for estimating the levels of genetic diversity in a modern crop with an apparent narrow genetic variation, i.e. Canadian durum wheat cultivars, and compare AFLP and pedigree-based genetic diversity estimates. As with RFLP analysis, AFLP results from mutations that alter restriction endonuclease recognition sites. AFLP technology is highly reproducible and, in addition, its multiplex ratio can be changed simply by altering the number of selective nucleotides at the 3' end of PCR primers. We have exploited these two properties in this study of genetic diversity in a species with a narrow genetic base, namely durum wheat.

2.3 **Materials and Methods**

2.3.1 **Plant Materials**

Breeders seeds from 13 cultivars (Table 2 p. 46) were obtained from the Canadian Food Inspection Agency, Ottawa. Ten seeds from each cultivar were grown in a phytotron at 22° C for three weeks and 1 - 2 grams of fresh leaf material was harvested for DNA extraction.

2.3.2 **DNA extraction and AFLP analysis**

DNA from 130 plants representing 13 cultivars (10 plants per cultivar) was isolated by a modified CTAB procedure (Doyle and Doyle 1990). Working solutions of 50 ng/ μ l were prepared by dilution of the original stock solution in TE (10 mM Tris-HCl, pH 7.5, 0.1mM EDTA, pH 8.0). 500 ng of DNA from each plant was digested with *Tru9I* (an isoschizomer of *MseI*) and *EcoRI* restriction enzymes. Adapters of known sequence were ligated to the ends of restricted fragments with T4 DNA ligase (Promega). The AFLP method has been extensively described by Vos et al. (1995). Amplification of the restriction fragments was carried out using two sets of primers. The *EcoRI* primer sets (E) all contain the core sequence 5'GACTGCGTACCAATTC while all the *MseI* primers (M) have the sequence 5'GATGAGTCCTGAGTAA in common. AFLP-PCR amplifications were carried out in two steps: a pre-amplification PCR reaction with E+A and M+C primers (where the symbol following the E or M represents an additional selective nucleotide at the 3' end of each primer and N represents one of the four possible nucleotides) was performed to reduce the number of

amplified restriction fragments by a factor of 16. Subsequently, the product of the pre-amplification reaction was used as template for selective amplification using ³³P labelled E+ANN in combination with M+CNN primers resulting in a calculated 4096 fold overall reduction in the number of generated PCR fragments. The PCR products were mixed with an equal volume of denaturing dye (98% deionized formamide, 0.025% bromo-phenol-blue, 0.025% xylene cyanol) and denatured at 94°C for 3 minutes. Amplification products were resolved in a 5% denaturing polyacrylamide gel prepared with 1X TBE buffer (100 mM Tris, 100 mM boric acid, 2 mM EDTA, pH 8.0,) as the running buffer and electrophoresed at 80 W (constant power) for 2.5 hours. The gels were dried and exposed to Kodak X-Omat film for 2-3 days. DNA fingerprints were evaluated by visual inspection of autoradiographs.

2.3.3 Data acquisition and diversity analysis

Autoradiographs were scored based on the presence or absence of bands at polymorphic loci generating a matrix of 1s and 0s representing the presence and absence of a band respectively. Only those bands falling within the range of 50-500 bp were considered. These bands were used to generate a genetic similarity matrix with the SIMQUAL routine based on DICE (1945) formula from the NTSYS-pc statistical package (Rohlf, 1990). The genetic similarity matrix was transformed into a distance matrix using the TRANSF subroutine of NTSYS-pc. Clustering of genotypes was performed using the SAHN in NTSYS-pc based on the genetic distance matrix with the UPGMA method. Within cultivar genetic diversity levels were estimated based on the number of pair-wise differences between pairs of haplotypes as measured by the average gene diversity per locus using the Arlequin program (Schneider et al. 1995).

2.3.4 Pedigree analysis

Variety descriptions of Canadian durum wheat cultivars were obtained from the Canadian Food Inspection Agency. Information regarding their known ancestors was obtained from cultivar descriptions, breeding records, personal communication with breeders, published pedigree databases (Zeven and Zeven-Hissink, 1976), and the Germplasm Resource Information Network (available at URL <http://www.ars-grin.gov>). The ancestry of each cultivar was traced as far back as possible. For each cultivar a pedigree was constructed using the Cultivar Registry System software package (Baum et al. 1990). Kinship Coefficients (r) were calculated using the Bigkin87 program (Tinker, N.A. 1993) based on the following assumptions: (1) all ancestors, cultivars, and breeding lines are homozygous and homogeneous; (2) cultivars without a known pedigree are unrelated; (3) parents make equal contribution of alleles to each progeny; and (4) the relationship between an ancestor and a line derived from that ancestor was $r = 0.75$. Pair-wise r values were used to generate a kinship coefficient matrix. The TRANS subroutine of NTSYS-pc was used to transform this matrix into a distance matrix based on $(1-r)$. Clustering of cultivars was performed with subroutine SAHN of NTSYS-pc from the genetic distance matrix with the UPGMA clustering method

2.3.5 Matrix Comparison

The AFLP-based Nei's genetic distance matrix and the pedigree-based distance matrix based on the kinship coefficients were compared by the MAXCOMP routine of NTSYS-pc. The normalized Mantel statistic Z (Mantel, 1967) was used to determine the level of association between the two matrices. The COPH subroutine of NTSYS-pc was used to determine the cophenetic correlation coefficient between similarity matrix and the similarity cluster for both AFLP and pedigree data sets. The statistical considerations for these analysis were described by Beer et al. (1993).

2.3.6 Analysis of molecular variance (AMOVA).

The matrix of the 130 genotypes was summarized into haplotypes. The reduced matrix of haplotypes was then submitted to Arlequin (Excoffier 1995) to compute the genetic structure, i.e. variance components.

2.4 Results

2.4.1 AFLP Analysis

Ten AFLP primer pairs were sufficient to detect 89 polymorphic markers among 130 genotypes resulting in an average of 8.9 polymorphic loci per primer pair (Table 3 p. 47). Forty one markers showed cultivar-specific polymorphism i.e. either present or absent within all 10 plants in each cultivar. The pair-wise genetic distance matrix for the 41 cultivar-specific markers is presented in Table 4 p. 48 (Lower triangle). The remaining 48 markers showed polymorphism

both within and among cultivars. An example of an AFLP autoradiograph, showing both polymorphism within and among cultivars, is shown in Figure 1 p. 51. There was no linear relationship between the number of amplified loci with the number of polymorphic markers for any given primer pair. (Results not shown). Average genetic distance between pairs of cultivars was estimated at 0.40 when only the cultivar-specific markers were used. This value dropped to 0.38 when all the markers were pooled in the analysis.

Clustering of genotypes based on the cultivar-specific AFLP markers resulted in two distinct clusters at the phenon line of 0.48 (Fig. 2 p. 52). The first cluster contains 'AC Avonlea', 'AC Melita', 'Medora', 'Sceptre', 'Hercules', 'Plenty', and 'AC Morse'. The second cluster contains 'AC Navigator', 'AC Pathfinder', 'Kyle', 'Wakooma', 'Wascana' and the selection from a landrace 'Pelissier'.

The AFLP data set was first split into two subsets, a subset of 41 cultivar-specific markers and a subset consisting of the 48 markers which are polymorphic within cultivar fingerprints. Cluster analysis was performed on each data set alone as well as the entire AFLP data to see if there was any difference in the clustering and in the genetic relationships among cultivars in each data subset. Clustering of genotypes with all three data subsets separately recovered cultivars developed at the Winnipeg Research Station and University of Saskatchewan in one group, and cultivars developed at the Swift Current research Station in the second group. When the entire data set (the two subsets together) was input into the cluster analysis all individual plants from each cultivar remained grouped together under each cultivar (Figure 3 p. 53). However, when the subset of the 48 within cultivar polymorphic AFLP markers was used all clustered together into their cultivar grouping except for two individual plants, one from 'AC Morse' and one from

'Pelissier'.

The level of genetic diversity/homogeneity within each cultivar was estimated from the 48 AFLP markers that showed polymorphism within cultivars. The 130 single plants in our study fell into 104 unique haplotypes. 'Pelissier' and 'Wascana' showed the highest and the lowest levels of genetic heterogeneity respectively (Fig. 4 p. 54). A clear trend of increasing level of within cultivar genetic heterogeneity is shown in Figure 4 p. 54.

2.4.2 Pedigree analysis

The ancestry of Canadian durum wheat cultivars was traced back to 132 cultivars and breeding lines including 17 landraces. We found that eleven ancestral lines and cultivars were present in at least 85% of the pedigrees and that three of them, 'Vernal Emmer', 'Mindum', and 'Heiti' were present in all pedigrees. The pair-wise genetic distances based on KIN is presented in Table 4 p. 48 (Upper triangle). The average pair-wise genetic distance between pairs of cultivars based on KIN was 0.76. Clustering of genotypes based on KIN did not result in any major groupings. At the 0.75 phenon line five clusters were evident. 'AC Avonlea' and 'AC Pathfinder' in the first cluster, 'AC Melita', 'Medora', 'AC Morse', 'Hercules', and 'Plenty' in the second cluster 'AC Navigator', 'Kyle', 'Wakooma' and 'Wascana' in the third cluster. 'Sceptre' and 'Pelissier' formed the fourth and fifth cluster respectively (Fig. 5 p.55)

2.4.3 Matrix comparison and Mantel test.

The degree of relationship between the distance matrices based on AFLP and KIN was measured by comparing the distance matrices with the normalized Mantel statistics. A moderately positive correlation ($r=0.457$, $p<0.002$) between the two matrices was observed.

2.4.4 Analysis of molecular variance (AMOVA).

The matrix of genotypes was summarized to a matrix of 104 haplotypes. The genetic structure is summarized in Table 5 p. 50. The variance component between cultivars is about 8.5-fold larger than the within cultivars variance component.

2.5 Discussion

Clustering of Canadian durum wheat cultivars based on 89 polymorphic AFLP markers revealed a low level of genetic variation compared to the estimates obtained by pedigree analysis. The AFLP and pedigree-based mean pair wise genetic distances were 0.40 and 0.76 respectively. The moderate correlation coefficient ($r = 0.45$) found between matrices based on AFLP genetic diversity estimates and KIN is similar to that reported for coefficient of parentage (COP) and AFLP in hexaploid wheat (Barrette et al. 1998), but higher than those obtained for durum wheat using RFLP and pedigree data (Autrique et al. 1996). On the other hand this value is lower than those obtained for maize (Smith et al. 1990).

The discrepancy between pedigree and AFLP-based genetic diversity estimates maybe the result of assumptions that were made in the calculation of genetic relatedness based on pedigrees. (Tinker et al. 1993, Cox and Murphy, 1990, Souza and Sorrells, 1989, Martin et al. 1991;

Murphy et al. 1986). These estimates may be biased due to selection pressure, unequal parental contribution, and relatedness of ancestors without a known pedigree. The assumption of no genetic relationship between ancestors without a known pedigree may be a major factor contributing to the disparity between pedigree and molecular marker based diversity estimates especially when a comprehensive pedigree information is not available. An RFLP analysis of two ancestral durum cultivars ('Khapli' and 'Vernal Emmer'), which are also present in the parentage of most Canadian durum cultivars, by Autrique et al (1996) showed that they were different for only two out of 134 alleles that were assayed. This result demonstrates that pedigree-based diversity measurements may result in overestimation of the actual levels of genetic diversity.

Lower genetic variation measured at the DNA level among modern cultivars is likely the result of selection pressure and genetic drift in breeding programs. Such practices which are aimed at genetic homogenization and purity of cultivars, may result in an improvement in yield and other agronomically important traits at the expense of reduction in the genetic base of crops.

UPGMA cluster analysis using 41 AFLP markers (polymorphic between cultivars but monomorphic between the individual plants within cultivars) produced two clusters (Fig. 2 p. 52). And the addition of 48 AFLP markers that were polymorphic between the individual plants within cultivars to the data set did not produce any major structural rearrangements in the dendrogram, except that 'Pelissier', an early selection from a landrace fell into a distinct cluster (Fig. 3 p. 53). When the subset of 48 within cultivar polymorphic AFLP markers was used 2 out of 130 plants were placed in wrong cultivar groups (result not shown). This result was expected given the relatively higher level of heterogeneity that was detected among individual plants

within these two cultivars compared to the corresponding levels in other cultivars (Fig. 4 p. 54). Another cause for this result may be impurities in the 'breeders seed' material although at this level the seed material is expected to be pure. The relative consistency in the results of using different subsets of AFLP marker data suggests that an adequate number of markers were used over the whole range of the genome although the nature and genomic locations or distribution of these markers are unknown.

Pairs of cultivars such as such as 'Kyle' and 'AC Navigator' that were closely related in the pedigree-based dendrogram (Fig. 5 p. 55) were more closely related in the AFLP-based dendrogram (Fig. 2 p. 52). On the other hand 'Wakooma' and 'Wascana' clustered closely in the pedigree dendrogram (Fig. 5 p. 55) but much less so in the AFLP-based dendrogram (Fig. 2 p. 52). 'Sceptre' and 'AC Melita' clustered closely in the AFLP-based but not in the pedigree-based dendrogram. These differences may be due to the repeated cycle of recombination and selection during cultivar development which are unaccounted for by pedigree analysis but revealed at the DNA level.

The new or modern cultivars have one feature in common - a high degree of genetic homogeneity (Frankel and Soule 1981). Although we have witnessed and demonstrated such a progression among the 13 cultivars in this investigation (Fig. 4 p. 54), a fairly substantial level of genetic heterogeneity is still maintained in spite of the rigorous selection pressure aimed at variety purity, uniformity, and associated breeding practices directed towards agronomic homogeneity.

Variance components within and between cultivars were measured as 1.72 and 14.67 (Table 5 p. 50) respectively when all 89 polymorphic markers were used indicating a much broader range

of variation between cultivars compared to the corresponding values within a cultivar. This indicates clearly the results of rigorous selection applied during the breeding process.

The reproducibility of AFLP data coupled with the high multiplex ratio of this marker system makes this approach ideal for the studies involving genetic diversity in crop species with a narrow genetic base such as durum wheat. Furthermore it also provides a means to fingerprint closely related cultivars for identification, variety purity maintenance and for intellectual property protection of newly bred cultivars. Although the pedigree approach remains as an inexpensive way of inferring estimates of genetic relationship in crop plants, molecular markers such as AFLPs are more likely to reflect a true measure of genetic relationships. Unlike pedigree analysis, the AFLP approach can detect the effects of genetic drift, and selection as well as giving a realistic measure of relatedness among ancestors.

Criticism of the use of some DNA markers such as AFLPs for genetic diversity estimates stems from their nature as unmapped markers whose genomic distribution is unclear. Some studies have shown that AFLPs tend to be genetically clustered (Karp et al. 1997) and as a result would not represent independent variables (Reeves et al. 1999). But with the relatively large number of markers used in this study, some redundancy resulting from genetic linkage may have been expected, however, this source of error is likely to be insignificant in comparison to the large number of polymorphisms detected and analysed.

In conclusion, there still exists a substantial level of genetic variation within modern cultivars of durum wheat as detected by AFLP, despite rigorous selection pressure aimed at variety purity and associated breeding practices. Measurements of genetic diversity can be used in breeding programs in order to maximize the level of variation present in segregating populations by

crossing cultivars with greater genetic distance.

2.6

Acknowledgments

The authors would like to thank Ajit Sahota of the Canadian Food Inspection Agency, Ottawa, for providing plant material and the plant breeders who provided information on breeding records, especially Dr. Ron M. DePauw, Agriculture & Agri-Food Canada, Semiarid Prairie Agricultural Research Centre, Swift Current, Saskatchewan. We appreciate Dr. DePauw's comments on a draft of the manuscript. We also thank Grant Bailey, Agriculture & Agri-Food Canada, Research Branch, Ottawa for technical advice. This work was supported, in part, by a grant from the Natural Sciences and Engineering Research Council of Canada to Douglas A. Johnson.

The experiments comply with the laws of Canada

Table 2. The 13 cultivars used in this genetic diversity study of durum wheat.

Cultivar Name	Code	Origin
'AC Avonlea'	AVN1	SCRS*
'AC Melita'	MEL1	WRS†
'AC Morse'	MRS1	WRS
'AC Navigator'	NAV1	SCRS
'AC Pathfinder'	PAT1	SCRS
'Hercules'	HRC1	WRS
'Kyle'	KYL1	SCRS
'Medora'	MED1	WRS
'Pelissier'	PLS1	Selection from a landrace (Algeria)
'Plenty'	PLN1	University of Saskatchewan
'Sceptre'	SCP1	University of saskatchewan
'Wakooma'	WAK1	SCRS
'Wascana'	WAS1	SCRS

*** Semiarid Prairie Agricultural Research Centre, Swift Current, Saskatchewan, Canada**

† Cereal Research Centre, Winnipeg, Manitoba, Canada

Table 3. The number of scored polymorphic markers detected by AFLP analysis within and among the 13 Canadian durum wheat cultivars.

Primer pair	Cultivar-specific	Polymorphic within cultivar	Total
EAAG/MCGC	7	5	12
EAAG/MCAG	8	7	15
EAAC/MCTA	6	6	12
EAGC/MCAG	6	8	14
EAGC/MCTT	3	7	10
EATG/MCTC	7	4	11
EAAG/MCCG	0	5	5
EAAC/MCTC	0	3	3
EAGG/MCTG	3	2	5
EAGG/MCTC	1	1	2
Total	41	48	89

Table 4. Matrix of pair-wise genetic distances between Canadian durum wheat cultivars based on kinship coefficient (above diagonal) and on Nei's formula (below diagonal).

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 AC Avonlea	0.0000	0.7830	0.6970	0.8680	0.6860	0.8480	0.8490	0.7010	0.9770	0.8770	0.9130	0.8460	0.8140
2 AC Melita	0.3617	0.0000	0.5830	0.7550	0.8420	0.6660	0.7740	0.3280	1.0000	0.7390	0.8040	0.7870	0.7870
3 AC Morse	0.3636	0.3191	0.0000	0.7930	0.7510	0.6700	0.7960	0.4380	0.9880	0.7840	0.7770	0.8010	0.7860
4 AC Navigator	0.4500	0.4419	0.5500	0.0000	0.6310	0.8120	0.3820	0.7580	0.9410	0.7820	0.9020	0.4740	0.5210
5 AC Pathfinder	0.4500	0.4419	0.5000	0.2778	0.0000	0.8890	0.7950	0.8410	0.9650	0.8730	0.9390	0.7890	0.7420
6 Hercules	0.4667	0.3333	0.4222	0.5122	0.5610	0.0000	0.7990	0.5690	1.0000	0.6420	0.7680	0.8110	0.8110
7 Kyle	0.4737	0.5122	0.6842	0.1765	0.4118	0.5385	0.0000	0.7820	0.9060	0.7340	0.9030	0.1840	0.3090
8 Medora	0.2917	0.1765	0.2500	0.4545	0.5000	0.2245	0.3810	0.0000	1.0000	0.7290	0.7410	0.7950	0.7950
9 Pelissier	0.4500	0.5814	0.6000	0.2778	0.3333	0.6585	0.3529	0.6364	0.0000	0.9530	1.0000	0.8120	0.8120
10 Plenty	0.3778	0.2917	0.3778	0.4634	0.4634	0.2174	0.4872	0.2653	0.5610	0.0000	0.8840	0.7240	0.6620
11 Sceptre	0.3636	0.2766	0.3182	0.4000	0.5000	0.2444	0.3684	0.1667	0.6000	0.2000	0.0000	0.9090	0.9090
12 Wakooma	0.3333	0.4222	0.4762	0.3158	0.5263	0.4419	0.2222	0.3478	0.4211	0.3488	0.2381	0.0000	0.2500

13 Wascana 0.4419 0.4348 0.4419 0.3333 0.2308 0.5000 0.4054 0.4043 0.3333 0.4091 0.4419 0.4634 0.0000

Table 5. Analysis of molecular variance (AMOVA).

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	12	1781.446	14.67239 Va*	89.45
Within populations	117	202.400	1.72991 Vb*	10.55
Total	129	1983.846	16.40231	100.00

***Va: among populations (cultivars) variance; Vb: within population variance**

Figure 1 Autoradiograph of a portion of an AFLP fingerprint using EAAG/MCAG selective primer pair demonstrating polymorphism. Lane 1-10 'Pelissier', 11-20 'Medora'. "A" indicate cultivar-specific markers and "B" within cultivar polymorphic markers. The EAAG primer was labeled with γ -³³P for signal detection by autoradiography.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

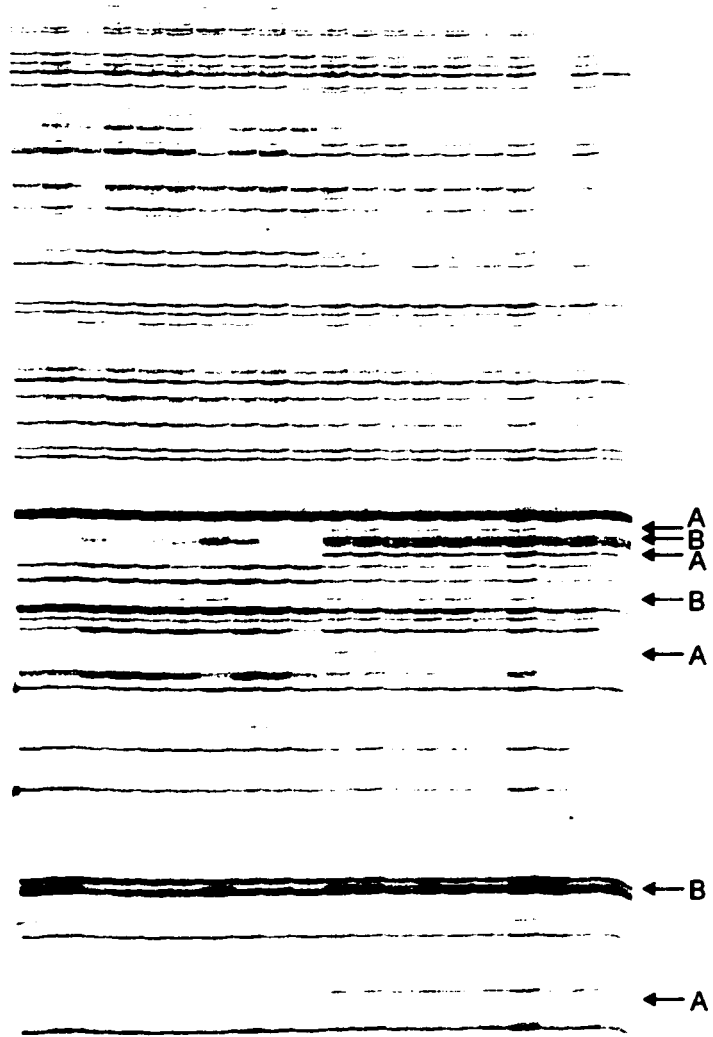
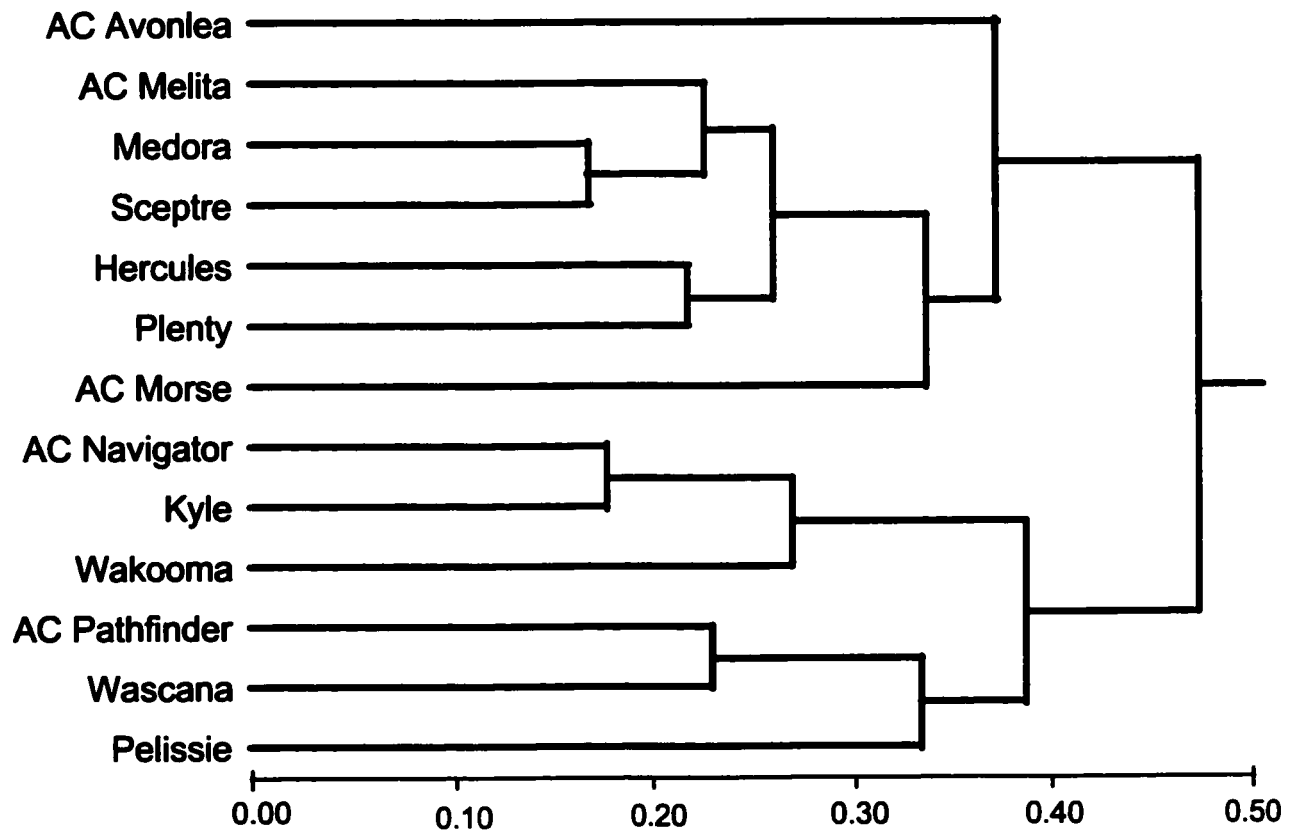


Figure 2 UPGMA dendrogram showing genetic relationships among the 13 durum wheat cultivars used in this study. The dendrogram was constructed using 41 cultivar-specific AFLP markers and was based on the genetic distance calculated according to Nei's genetic coefficient.



Nei's genetic distance

Figure 3 UPGMA dendrogram depicting genetic relationships among 130 durum wheat plants representing 13 cultivars. The dendrogram was constructed using 89 AFLP markers based on the genetic distance calculated according to Nei's genetic coefficient. The numbers in each cluster refer to individual plants. All 130 plants fell into their respective cultivar groups.

Winnipeg Research Station

Swift Current Research Station

- AC Avonlea
- AC Melita
- Medora
- Sceptre
- Hercules
- Plenty
- AC Morse
- AC Navigator
- Kyle
- Wakooma
- AC Pathfinder
- Wascana
- Pelissier

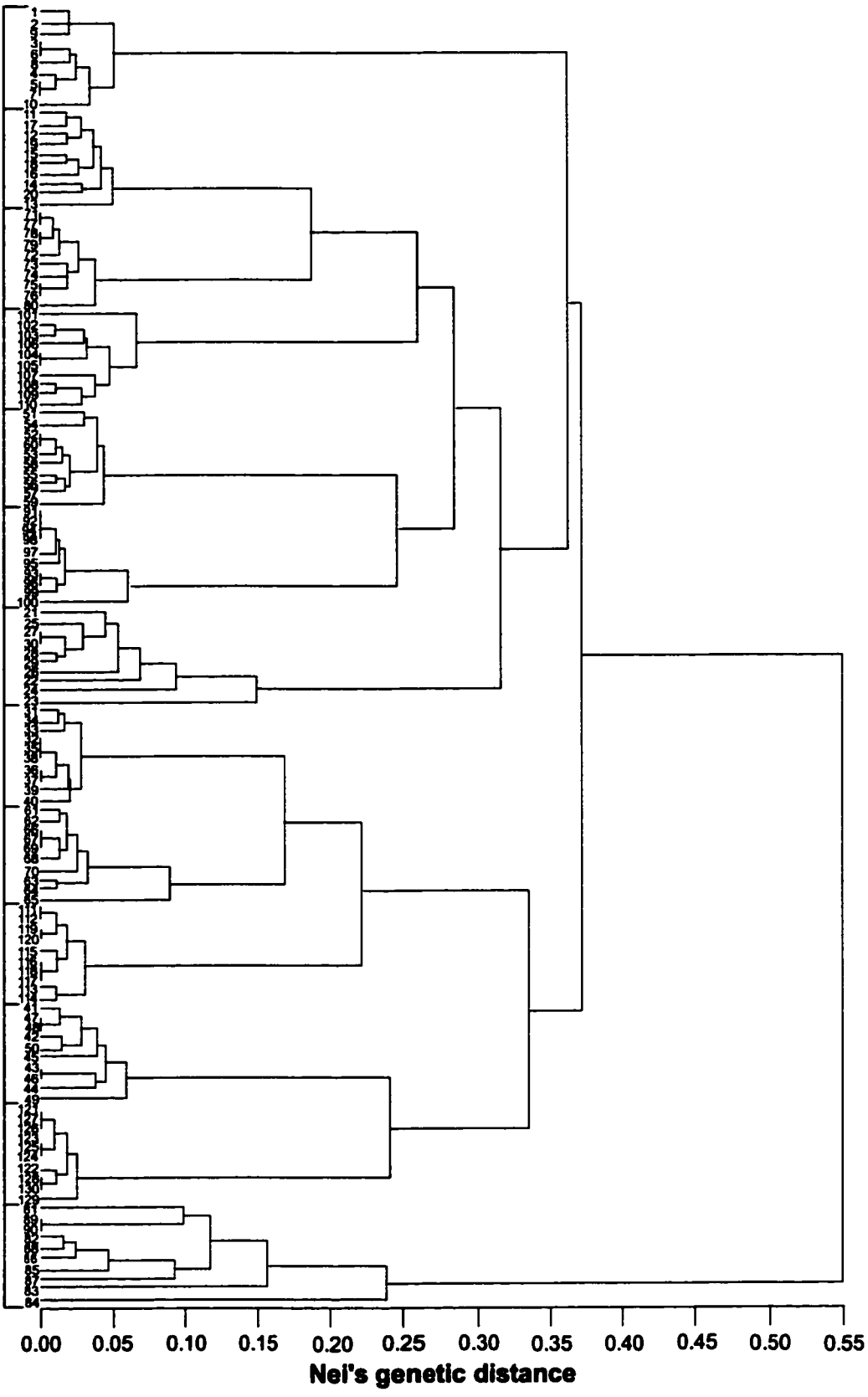


Figure 4 Comparison of the genetic diversity per locus among the individual plants within each cultivar based on interhaplotypic distances. The distances were measured based on the number of differences in band sharing data from pair wise comparison of cultivars.

Average gene diversity per loci in Canadian durum wheat cultivars

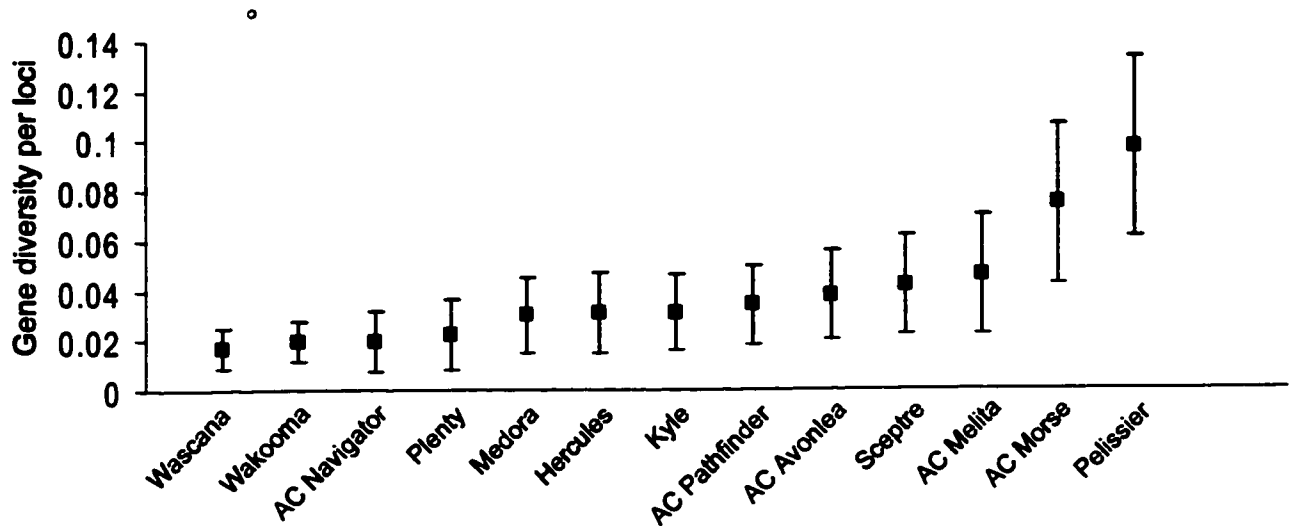
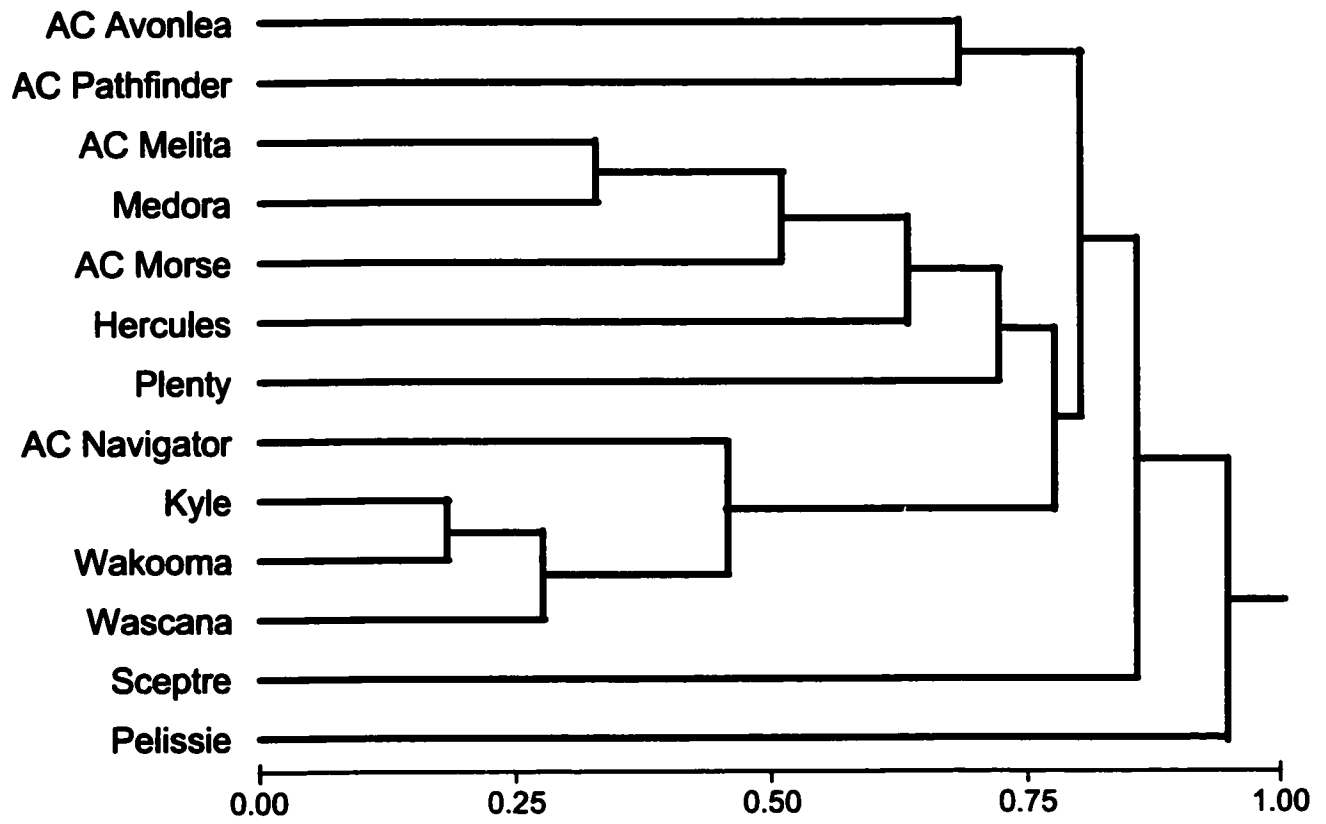


Figure 5 UPGMA dendrogram showing genetic relationships among the 13 durum wheat cultivars used in this study. The dendrogram was constructed using the matrix of kinship coefficients.



Genetic distance based on Kinship coefficient

CHAPTER THREE

Identification of Canadian durum wheat (*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.) Cultivars using AFLP_s and their STS markers.

The following chapter is a reproduction of a paper submitted to the Canadian Journal of Plant Science by Soleimani, D. Vahab, Baum, B.R., and Johnson, D.A. The chapter details the fulfilment of the second objective of this thesis: DNA fingerprinting of 13 currently registered Canadian durum wheat cultivar based on AFLP polymorphic markers and transformation of these markers into STSs for rapid cultivar identification. A subset of cultivar-specific AFLP markers were selected for cloning and sequencing and specific PCR primers were designed for each sequence. Reaction conditions were optimized for each primer pair. For those primers that did not yield any polymorphic profile, the PCR products were digested with restriction enzymes resulting in identifiable restriction fragment polymorphism. Two separate identification keys were generated one of them based on AFLP markers, the other based on their STS polymorphism.

Identification of Canadian durum wheat (*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.)**Cultivars using AFLP and their STS markers.**

Soleimani, D.V. (1,2), Baum, B.R. (1), and Johnson, D.A. (2)

1) Eastern Cereal and Oilseed Research Centre, Agriculture & Agri-Food Canada, Central Experimental Farm, Ottawa, Canada, ON K1A 0C6

2) Ottawa-Carleton Institute of Biology, University of Ottawa, 30 Marie Curie, Ottawa, Canada, ON K1N 6N5

***Corresponding author**

Tel: 613-759-1821

Fax: 613-759-1924

e-mail: baumbr@em.agr.ca

3.1**Abstract**

We have developed two identification schemes for currently registered Canadian cultivars of durum wheat (*Triticum turgidum* L. subsp. *durum* (Desf) Husn) based on cultivar-specific amplified restriction fragment polymorphism (AFLP) and their sequence tagged sites (STS) markers. Each identification key required 7 markers. Transformation of AFLPs into STS markers was done in order to develop a PCR-based identification assay which was cost effective and required minimum technical expertise. Selected AFLPs were cloned, sequenced, and specific PCR primers were designed from their sequences. A cultivar diagnostic PCR assay was carried out for each STS primer pair. Five STS primers showed polymorphism between cultivars. But sixty per cent of STS primers (7 out of 12) did not produce any polymorphism indicating the presence of the sequence in all cultivars. The PCR products of the latter primers were digested with selected restriction enzymes resulting in restriction fragment polymorphism for two loci. An STS-based identification key was generated for cultivar identification based on either the presence/absence of a DNA band or presence/absence of a restriction enzyme recognition site after digestion of the PCR products with a restriction enzyme. DNA based markers can be used as an efficient alternative to morphological traits for cultivar identification and finger printing at any stage of plant development. Moreover, an STS-based assay can be used with a minute amount of plant tissue such as fraction of a seed.

Key Words: Amplified restriction fragment polymorphism (AFLP), sequence tagged sites (STS), cloning, Identification, durum wheat

3.2

Introduction

DNA markers such AFLPs (Vos et al. 1995) and their derived STSs (Olsen et al. 1989) are important tools for genome analysis, plant breeding programs and identification of plant cultivars. STS primers can be designed from a previously known sequence such as AFLPs (Reamon-Butter and Jung 2000; de Jung et al 1997), random amplified polymorphic DNA (RAPDs) (Deng et al. 1997, Kawchuk et al. 1998), and restriction fragment length polymorphism (RFLPs) (William et al. 1996; Erpelding et al. 1996). The resulting STS-based polymorphism is either detected by gel electrophoresis or following digestion of PCR products with restriction endonucleases.

AFLP is based on the amplification by PCR of selected restriction fragments in a total genomic digest. It has proven to be a powerful tool in elucidating genetic variation in a wide range of species (Sharma et al. 1996; Maughan et al. 1996; Hill et al. 1996; de Jung et al. 1997). Polymorphism can result from the creation or loss of a restriction enzyme recognition site, mutation within the priming sites for selective bases at the 3' end of AFLP primers or insertions/deletions within a locus in the genome. These mutations fall into two categories: those that are inherited by all the individuals within a cultivar and can be used for cultivar identification, and those that occur among individual plants within a cultivar. Although AFLP markers are reproducible and can be used for cultivar identification purposes the cost and the levels of technical expertise associated with this marker system limits their applications for routine screening and testing. To overcome these limitations AFLPs can be transformed into STSs by designing specific PCR primers from their sequences for rapid diagnostic applications.

In durum wheat, application of molecular markers such as AFLPs and their STSs is

useful for varietal identification and for ensuring that varietal purity is maintained throughout the grain handling system and possibly for variety registration. These markers could also be used in breeding to enhance the efficiency of breeding programs and marker assisted selection (MAS). A simple PCR-based assay can be used for cultivar identification at any stage of plant development including a fraction of seed material. This approach can provide an efficient alternative which can simplify and accelerate cultivar identification and eliminate the need for morphological markers. Currently, there are 13 registered varieties of durum wheat in Canada (Table 2 p. 46). The objective of this study was to develop DNA-based cultivar identification schemes using AFLPs and their STS markers. These markers can be used in an automated cultivar identification apparatus for fingerprinting, quality, and purity testing in grain handling system.

Abbreviations: AFLP, amplified restriction fragment polymorphism; STS, sequence tagged site

3.3

Material and Method

3.3.1 Plant Material

Breeders seeds from the 13 registered Canadian cultivars (Table 2 p. 46) were obtained from the Canadian Food Inspection Agency, Ottawa. Ten seeds from each cultivar were grown for three weeks and 1 - 2 grams of fresh leaf material were harvested for DNA extraction

3.3.2 DNA Extraction and AFLP Reactions

DNA from 130 plants representing the 13 cultivars (10 plants per cultivar) was isolated according to a modified CTAB procedure (Doyle and Doyle,1991). Working solutions of 50 ng/µl were prepared by dilution of the original stock solution in TE (10 mM Tris-HCl, pH 7.5, 0.1mM EDTA, pH 8.0). 500 ng of DNA from each plant was digested with 2.5 units each of *Tru9I* (an isoschizomer of *MseI*) and *EcoRI* restriction enzymes. Double stranded DNA adapter molecules of known sequences were ligated to the ends of restricted fragments with T4 DNA ligase (Promega). The AFLP method which was followed is extensively described in Vos et al. (1995). Amplification of the restriction fragments was carried out using two sets of primers: The *EcoRI* primer set (E) all have the sequence 5' GACTGCGTACCATTC3', the *MseI* primers (M) have the sequence 5' GATGAGTCCTGAGTA3' A in common. The letter following the E or M represent additional selective nucleotide at the 3' end of each primer. AFLP-PCR amplifications were carried out in two steps: a pre-amplification PCR reaction with E+A and M+C primers was performed to reduce the number of amplified restriction fragments by a factor of 16. The products of pre-amplification PCR were diluted 50 fold in TE buffer and used as template for

selective amplification. PCR products were mixed with an equal volume of denaturing solution (98% deionized formamide, 0.025% bromophenol blue, 0.025% xylene cyanol) and denatured at 94 °C for 3 minutes. Amplification products were loaded on 5% denaturing polyacrylamide gels using 1X TBE buffer (100 mM Tris-HCl pH 7.5, 100 mM boric acid, 2 mM EDTA, pH 8.0) and electrophoresed at 80 W (constant power) for 2.5 hours. The gels were dried and exposed to X-ray film X-Omat (Kodak) for 2-3 days. A duplicate sample was run on a denaturing polyacrylamide gel and silver stained (Bassam et al. 1991) for the isolation and cloning of selected AFLP fragments. DNA fingerprints were evaluated by visual inspection of autoradiographs. The size of amplification products were determined by comparison with a 330-30 bp AFLP DNA ladder (Promega).

3.3.3 Isolation and cloning of AFLP markers

AFLP markers of interest were detected on a silver stained denaturing polyacrylamide gel. The ssDNA fragments were isolated from the gel based on the crush and soak method of Sambrook et al. (1989) with minor modifications. Briefly, the band of interest was excised from the gel using a clean razor blade, placed in a 1.5 ml Eppendorff tube and rinsed with sterile water. Two volumes of elution buffer (0.5M Ammonium Acetate, 10mM Magnesium Acetate, 1mM EDTA, pH 8.0, 0.1% SDS) were added and the gel was crushed using a pipet tip. The samples were incubated overnight at 37 °C with constant shaking. Gel debris was separated from the DNA solution by filtering the solution through a Xcluda style D tip (aerosol barrier tip) which was placed in a collection tube (Bio-Rad Laboratories) and centrifuged at 3000 rpm (3K) for 20 seconds. 1/10 volume of 3M sodium acetate (pH 5.5) and 2 volumes of 100% ethanol

were added to the filtrate and the DNA was precipitated at -80°C for 2 hours. DNA was recovered by spinning of samples for 20 minutes at 13K rpm in a microfuge at 4°C . The pellet was washed with 70% ethanol and vacuum dried for 1 hour. The DNA was resuspended in $10\ \mu\text{l}$ of TE buffer and used as template for PCR amplification using AFLP pre-amplification primers. The PCR products were ligated into pGEM-T Easy Vector System (Promega) and used to transform "JM109 High Efficiency Competent" *E. coli* cells following the manufacturer's recommendations. The cells were plated on a liquid broth medium containing ampicillin to a final concentration of $100\ \mu\text{g/ml}$. The plates were supplemented with $100\ \mu\text{l}$ of IPTG (0.5mM) and $20\ \mu\text{l}$ of X-Gal ($50\ \text{mg/ml}$). Selection of transformants was carried out based on blue/white colour. Eight white colonies per plate were selected for colony PCR analysis to confirm the presence of an insert and size variability among the clones generated from the same diagnostic DNA band. Plasmid DNA was extracted from colonies containing the insert DNA based on alkaline lysis method (Sambrook et al. 1989). The plasmid DNA was digested with *EcoRI* restriction enzyme (New England Biolabs) and run on 1.5% agarose gel to determine the size of insert DNA.

3.3.4 DNA sequencing, sequence analysis, and restriction map construction

DNA was sequenced from two clones for each amplicon using the Sanger's dideoxy-chain terminator method (Sanger et al. 1977) on an ABI PRISM 377 DNA sequencer (Perkin Elmer, Applied Biosystems) using ABI Prism Dye terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase and following the manufacturer's recommendation. The sequences that were obtained for each amplicon were aligned using CLUSTAL X (Thompson et

al. 1994) in order to assess their similarity. A sequence similarity search was done using Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990) available at the National Centre for Biotechnology Information (NCBI) to identify similarities with known sequences in the Genebank. For each amplicon a restriction map was constructed using the Web Cutter computer program available at URL: <http://www.medkem.gu.se/cutter>

3.3.5 STS primer design and optimization

Primers were designed using the Oligo Selection Program- OSP (Hillier and Green 1991). Selection criteria for each oligomer included GC content, T_m , maximum length of amplification products and other criteria to obtain an optimal specificity for each primer pair. STS primers were commercially synthesized and their respective PCR conditions were optimized for annealing temperature and magnesium concentration. For STS primers that produced monomorphic profiles across all cultivars, a touchdown PCR reaction was performed. This strategy was important to prevent non-specific primer/template annealing during the first few amplification cycles and therefore increasing the numbers of specific target sequences for subsequent cycles of amplification. In the cases where touchdown PCR did not result in a polymorphic banding profile the product of amplification was digested with restriction enzymes in order to detect possible restriction fragment polymorphism.

3.3.6 Generation of identification keys

A matrix of 1s and 0s was generated for the presence and absence of a DNA band for both AFLP and their STS markers respectively. When polymorphism was the result of restriction enzyme digest, the presence of an enzyme recognition site was scored as 1 and absence of such site scored as 0. The keys were generated using the DELTA program (Dallwitz et al. 1993) and consisted of subordinated couplets consecutively dividing cultivars into two groups until all cultivars were distinguished.

3.4**Results**

Cultivar specific polymorphic markers were scored as those that were monomorphic between individual plants within a cultivar and polymorphic among cultivars. A total of 41 such markers were obtained using 10 AFLP primer pairs (Chapter 2). Seven out of 41 such markers that showed polymorphism between cultivars were sufficient to distinguish all the 13 cultivars as shown below via an identification key. The key consists of subordinated couplets, the first dividing all the cultivars into two groups. Each couplet gives a further division into two groups, and this process continues until all cultivars are distinguished. The numbers on the right indicate the following step to be taken, the number on the left refers to the number of couplets, and the number in bracket refers to the preceding couplet.

1	(0)	B03-330	ABSENT	2
		B03-330	PRESENT	8
2	(1)	B12-155	ABSENT	3
		B12-155	PRESENT	5
3	(2)	A05-420	ABSENT	4
		A05-420	PRESENT	WASCANA
4	(3)	A05-415	ABSENT	AC AVONLEA
		A05-415	PRESENT	AC PATHFINDER
5	(2)	A05-415	ABSENT	6
		A05-415	PRESENT	7
6	(5)	B12-320	ABSENT	KYLE
		B12-320	PRESENT	WAKOOMA
7	(5)	B12-300	ABSENT	PELISSIER
		B12-300	PRESENT	AC NAVIGATOR
8	(1)	A05-420	ABSENT	9
		A05-420	PRESENT	11
9	(8)	A05-415	ABSENT	10
		A05-415	PRESENT	AC MORSE
10	(9)	B12-530	ABSENT	SCEPTRE
		B12-530	PRESENT	PLENTY
11	(8)	A05-415	ABSENT	MEDORA
		A05-415	PRESENT	12
12	(11)	B12-300	ABSENT	AC MELITA
		B12-300	PRESENT	HERCULES

In the construction of the cultivar identification key based on AFLP data all 41 markers were given equal weight in terms of reliability and reproducibility of each marker. Since only a subset of these markers were required to distinguish all durum wheat cultivars many possible taxonomic keys were generated. An example of an identification key is shown above.

3.4.1 Cloning and sequencing of the AFLP markers and sequence analysis

DNA fragments from 12 AFLP diagnostic markers were cloned and sequenced according to the methods described. The size of the cloned fragments ranged from 156 bp for B12-185 to 418 bp for G03-420 (Table 6 p. 75). The AFLP markers that were selected for sequencing were those between 100 to 500 bp range and did not contain any other amplified locus in their immediate vicinity as visualized on an polyacrylamide gel (Figure 6 p. 80). This was important in order to avoid introducing possible contaminant molecules. Colony PCR analysis using universal AFLP primers showed only one clone (A06-425) with variable size which was subsequently eliminate form further analysis. The DNA sequence data were subjected to BLAST search. Two of these sequences were known in wheat as indicated. Most of the remaining sequences showed no significant similarity to other known sequences available in GeneBank.

3.4.2 Primer design PCR condition and restriction digest of amplified loci

For each selected sequence of AFLP marker a specific STS primer pair was designed (Table 8 p. 79). The primers were assayed on all 13 cultivars. An example of such an assay is shown in Figure 7 p. 81. In order to increase the specificity of the amplification assay a touch down PCR was performed as described. This was done by performing PCR at a high annealing temperature and gradually reducing the annealing temperature during subsequent cycles until reaching the optimal PCR condition for that particular primer pair. Two out of the 7 primer pairs (A05-290 and B03-340) did not produce any polymorphic profile under any PCR condition. A restriction map for these two sequences was constructed and amplification products were digested with selected restriction endonucleases. Both A05-290 and B03-340 yielded

polymorphism between cultivars when cut with *Hae*III and *Mnl* I enzymes respectively (Figure 8 p. 82).

3.4.3 Generation of an identification key based on STSs

An identification key was generated as described by Baum et al. (2000) for STS data sets. The STS-based key follows the same format as the AFLP-based key. The key was generated from a matrix of 1s and 0s for presence and absence of a DNA band respectively. When polymorphism was the result of restriction fragment polymorphism, the presence of a restriction enzyme site was scored as one and its absence scored as zero. The key required seven markers for the identification of all cultivars.

1(0).	A05-290 <i>HaeIII</i> absent.....	2
	A05-290 <i>HaeIII</i> present.....	7
2(1).	G03-255 absent.....	3
	G03-255 present.....	5
3(2).	H07-320 absent.....	4
	H07-320 present.....	11. SCEPTRE
4(3).	A05-220 absent.....	7. KYLE
	A05-220 present.....	12. WAKOOMA
5(2).	H07-320 absent.....	13. WASCANA
	H07-320 present.....	6
6(5).	A05-220 absent.....	9. PELISSIER
	A05-220 present.....	4. AC NAVIGATOR
7(1).	B03-340 <i>MnlI</i> absent.....	8
	B03-340 <i>MnlI</i> present.....	9
8(7).	H07-320 absent.....	6. HERCULES
	H07-320 present.....	10. PLENTY
9(7).	G03-255 absent.....	1. AC AVONLEA
	G03-255 present.....	10
10(9).	H07-320 absent.....	11
	H07-320 present.....	8. MEDORA
11(10).	B12-185 absent.....	3. AC MORSE
	B12-185 present.....	12
12(11).	G03-420 absent.....	5. AC PATHFINDER
	G03-420 present.....	2. AC MELITA

3.5

Discussion

3.5.1 AFLP-based fingerprinting

We have found AFLP to be an efficient marker system for fingerprinting of durum wheat. In contrast to other molecular approaches used in wheat such as RFLPs and RAPDs (Devos and Gale 1992; Vierling and Nguyen 1992) a relatively high level of between-cultivar variation was observed. This finding is mainly due a high multiplex ratio of the AFLP marker system. In other words, a large number of loci were targeted with each AFLP assay which in turn increased the number of detected polymorphisms. Since a single nucleotide mutation could alter a restriction enzyme recognition site it can potentially give rise to AFLP polymorphisms. In addition, mutations within the selective nucleotide in AFLP primer pairs and/or insertion/deletions in a particular region of genome could also result in polymorphism.

3.5.2 Conversion of AFLPs into STS markers

STS and SCAR markers were originally developed to overcome the problems associated with lack of reproducibility of RAPD markers (Michelmore et al.1993; Olson et al. 1989). Although the cultivar-specific AFLPs that were generated in this study were reproducible as tested twice on DNA samples extracted independently (results not shown), their use for routine identification purposes is made complicated by the costs and the level of technical expertise required for this marker system. These problems could be mostly overcome by converting AFLPs into STS markers. As has previously been done for *Asparagus officinalis* (Reamon-Buttner and Jung 2000), *Triticum aestivum* (Qu et al.1998), and *Solanum tuberosum* (de Jong et

al. 1997), *Oryza sativa* (Cho et al 1996). In this study, we cloned and sequenced 12 AFLP markers and designed specific PCR primers from each amplicon. The most useful clones were identified by amplifying the insert with the AFLP primer pair which gave rise to the polymorphic marker. Only one in 12 clones (A06-425) gave a heterogeneous population of fragments. In 11 out of 12 clones, all the assayed amplicons exhibited the same mobility. In contrast Reamon-Buttner and Jung (2000) found that most of their fragments were of heterogeneous nature. In our study, care was taken to isolate only the well-spaced AFLP fragments to prevent the possibility of contaminating DNA from nearby sequences.

We aimed to design specific PCR primers for each amplicon based on the entire sequence of each amplicon and narrowed selection criteria for GC content, annealing temperature, and the length of the amplified products. Although under touch down PCR conditions each STS primer pairs amplified only one locus (Figure 7 p. 81) when the annealing temperature was reduced additional bands were detected (results not shown). This finding may indicate the repetitive nature of some STS loci under investigation.

Previous workers have designed primers from AFLP markers such that the STS primers contained the complete *EcoRI* or *MseI* restriction site sequence plus the selective nucleotides as a common domain and extended this domain by adding additional nucleotides from the cloned AFLP markers (Qu et al. 1998; Reamon-Buttner and Jung 2000). A similar approach has been taken with SCARs developed from RAPD markers (Deng et al. 1997; Fang et al. 1997). In the conversion of AFLPs into STS markers Reamon-Buttner and Jung (2000) reported that as the primer sequences moved further away from the restriction sites, different PCR products were observed which did not correspond to the sequence of the AFLP marker from which they had

been derived. Similar findings were reported by Qu et al. (1998). All concluded that in order to amplify a sequence which is identical to its corresponding AFLP marker, the STS primer must contain the restriction site and selective nucleotides.

In the present study we have chosen the entire sequence of the AFLP marker for STS primer design without restricting the primer selection to the ends of the sequences. In this case, polymorphism can result if there is a mutation in the STS primer annealing sites. Theoretically, these types of mutation are expected to result only from a portion of cloned AFLP markers and in the absence of such mutations a monomorphic profile is expected with STS-PCR assay as was observed for 60% or 7 out of 12 STS primer pairs that were originally designed. Other types of mutations that could lead to AFLP polymorphism may include: mutation in the restriction enzyme recognition sites and/or mutations within selective nucleotides. Since most of our STS primers were designed from the internal sequences of AFLP markers (Table 6 p. 75) it is unlikely that these mutations were detected. For one AFLP marker "A05-220" the flanking regions were cloned using inverse PCR (results not shown). This was done to effectively transform an AFLP marker into an STS by replacing the AFLP adapter sequences with the genomic sequences to gain insight into the sequence of flanking region. A BLAST search of A05-220 flanking region showed the sequence to be similar to BARE-1-like retrotransposon long terminal repeat (LTR) in barley, a close relative of wheat which is likely to be present in multiple copies in wheat genome. Another sequence (G03-420) was also found to be similar to BARE-1-like retrotransposon in barley. Others were not known to have any similarity to the known sequences in the Genebank. Pervious studies have shown that almost 80% of the hexaploid wheat genome (AABBDD) is composed of repetitive DNA sequences (Devos and Gale 1993). Durum wheat has a tetraploid

genome (AABB) sharing A and B genome with the hexaploid bread wheat and therefore it too is expected to contain large amounts of repetitive and non-coding DNA.

3.5.3 Primer design, optimization and generation of an identification key

PCR primers (Table 7 p. 78) were designed to achieve optimal diagnostic differentiation as described. Most primers did not yield sufficient differentiation of cultivars at theoretical T_m . Therefore we used a touch down PCR technique in order to minimize non-specific PCR amplification.

The identification of durum wheat cultivars was possible both with AFLP as well as STS primers. Whereas AFLP yields multiple bands, STS primers are expected to produce presence/absence of a single band. The STS-based identification scheme is a simple PCR-based method which is cost effective and requires very little template DNA. Although we have demonstrated that cultivar identification can also be reliably achieved with the AFLP markers, the cost and labour of the latter method may limit its routine application for variety identification.

Table 6. The sequences of AFLP markers used in the design of STS primers. Primer annealing sites are shown in bold and underlined. The 5' and 3' ends of each sequence contain *EcoRI* and *MseI* recognition sequences respectively.

A05-570

5'GAATTCAACTGGCCGGTGC GGAGGC**ATTCCTGGAAGCTCTTCTGAAAAGACGTC**
 TTGATATTCGCAAACGACTGGAATTTGCGAGATAGCATCCAGTTCACCATTCTCATT
 GAGAGAAAACAGACGGATGGTATTATCACGAGCGGCAAAGACA**ATTACATCCTCA**
GACGAATGAGTCAATTGAATCTGCCTGGCTGCATAGTTAA3'

A05-940

5'GAATTCAACACGGAGGAGGTA**ACCATCGTGCTCAAATTAGGTTTGCTATGCACG**
 TACCCATCGCCCAACGCGCGGCCTATCATGCGGAAGGTCATGCAGTACCTTGACCAT
 GATCAATCGCCTCCTAATTTATCACCGTCCTACATTAGCTACATCATGATGGCCCAA
 CTGCAATATGAAGGGTTTGATTACACAACATGCCATGTTCTCAACCCATAGCGAGT
 GTTGCCACTGTGTCAGGAGAGTCCTCGGCGACTATTT**TGCGAGAGGGGAGAAGAA**
GTGTGTAGTTAA3'

B03-1100

5'GAATTCAAGAAGGGTACAAGGAAAACAAGCAAGAAGCTGGCCTAATTAGGTAG
 GTGTAAGGAGAGGGACAAATAAACTTGGACTACCAAACCAACCTTTGAGATGGAT
 GGATGTCCTGCTCATTACAGCTCCGGAAAAGCAACTCGATCCTGGTGAAGAGGAACTC

ACATGGAGAGGAAATCCTTGCTGGATTTGTGAGTCTCAGTCGAACCAAATCGCCTAC
 CGTGAGACGATCGACGGCAGAGAGAAGACGTGGTGGTCAGCGCGAAGGCAACCT
 CACCTCTGTGAGCAGCAGCTCTGCTCCTCCTCTGGGGTGTCCCTTCCTGTAA3'

G03-1600

5'GAATTCAGGCTTTTTGGAGCTCGGTTACACCCCGTAGCTAGTTC AAGAACTCAGA
ATACAGCCACCAAAGTAGGACTAGGGTTTTACGCATCCTCGCGGCCTGAACCTGGA
 CAAACAACCACATGCTCTCTGTTAGACCCACTCTTCTCTCAACCCCGCGCCCCGCA
 ACCGCAGTAGGGATTGTTGTGATCCATAAGTGTCTGTTCCACCGACACCTTTATA
 ATCACCCAGTTACGTTGTGATGTTTGGTAGCACACAAAGTGTTCCCTCCAGTAAACGG
 GAGTTGCATAATCTCATAGTCATAGGAACATGTATAAGTCATGGAGAAAGCAATAG
 CAACATACTAAACGATCGGGTGCTAAGCTAACGGAATGGGTCAAGTCAATCACAT
CATTCTCCTAATGGTTTGATCCTGTAA3'

B12-400

5'GAATTC AAGGGCAATTTTTGGCTTTTTTATACAGGGAGTCATAGGTGGGCATA
 GTTCGCTCCCAGGCTGGTCCAAAAATTCAGTACGTCAGAACTTTATGGCCGCGACA
 AGGTTATCTCAAAGGGGGCAAATCCTATTTGTCGCTTATCGCGTTAA3'

G03-700

5'GAATTCAGCCCTGAAATCTGACTTCTTGGTGTAATGAGCACCCAGGTTTTCTTTA
 GCCTTCTTCAGACGTGCCATTTGCACCAGCGATGTGTCGAATCGGGATACGCCTCCT

TGATCGCCTTTTCCATTGCTCTTGCTTGATCCGCCAAGGAAACATAAAAAGGTCATGA
AACTAGTTTGCGGGACTTCAAACGATATAAACAGATTTTACGTTTGGTAGATGATC
CCATTACCTGTTAA3'

H07-1000

5GAATTCAGGCATATAGCATAAGGAACTATACGGTTTTTAGTTTGTGTTACACTGGT
CCACCCATAAAAAGGACAGACCCAGTGCA TAGAAGCTCCCACACAAGGTGGGGTC
TGGGGAGGGATTATAGGAACCTAGTCTTACCCCTGCAAAGTGCAATGCAGAGAGGC
CGGTTCGAACCAGGACCTCTTGGCACAAGTGGGGAGGACTTCACCACTGCGCCAG
GCCTGCCCTCACACTGGTCCACCCATATGGTCCAAAATTCTTACTCTCCTGGCACTT
CATCTCATGCATTTGAGCAGTTAA3'

Table 7. STS markers and the AFLPs from which they were derived for cultivar fingerprinting of durum wheat.

AFLP primer pair	AFLP marker name	Size in base pair	Identity in the GeneBank	Derived STS marker
1 EAGC/MCAG	G03-700	237	nsm†	G03-255
2 EAAC/MCTA	A05-570	204	nsm	A05-220
3 EAAC/MCTA	A05-940	306	nsm	A05-290
4 EAAG/MCAG	B03-1100	328	nsm	B03-340
5 EAGC/MCAG	G03-1600	418	BARE-1 LTR*	G03-420
6 EAGG/MCTG	H07-1000	301	nsm	H07-320
7 EAAG/MCGC	B12-400	156	nsm	B12 185

† no significant match to other sequences available in the GeneBank found.

* BARE-1-like retrotransposon long terminal repeat (LTR) in barley (*Hordeum vulgare* L.)

Table 8. STS Primers used in the identification key of Canadian durum wheat cultivars

Primer name	DNA ng	MgCl ₂ mM	Annealing temperature °C	Touch down PCR*	Cycles number	T _m °C	Primer sequences
1 G03-255	50	3.0	59	no	35	60	5'AGCCCTGAAATCTGACTTCTTGGTGTA3' F 5'CAGGTAATGGGATCATCTACCAAACGTAA3' R
2 A05-220	50	3.0	68	yes	31	58	5'ATTCCTGGAAGCTCTTCTGAAAAGACG3' F 5'AATTGACTCATTCGTCCTGAGGATGTA3' R
3 A05-290	50	3.0	69	yes	31	62	5'TCAACACGGAGGAGTAACCATCGTGC3' F 5'CTACACACTTCTTCTCCCTCTCGCAA3' R
4 B03-340	50	2.5	62	yes	31	61	5'TTCAAGAAGGGTACAAGGAAACAAGCAAG3' F 5'GCTGACCACCACCGTCTTCTCTCTGCCCCG3' R
5 G03-420	50	2.5	65	yes	31	60	5'GTTCAAGAACTCAGAAATACAGCCACCAA3' F 5'TTAGGAGAATGATGTGATTGACTTGACCCCAT3'
6 H07-320	50	2.5	55	no	35	62	5'TAGTTTGTGTTACACTGGTCCACCCATA3' F 5'ATGCATGAGATGAAGTGCCAGAGAAGTAA3' R
7 B12-185	50	3.0	58	no	31	58	5'TTTTATACAGGGAGTCATAGGTGGCA3' F 5'CAAATAGGATTTGCCCTTTGAGATA3' R

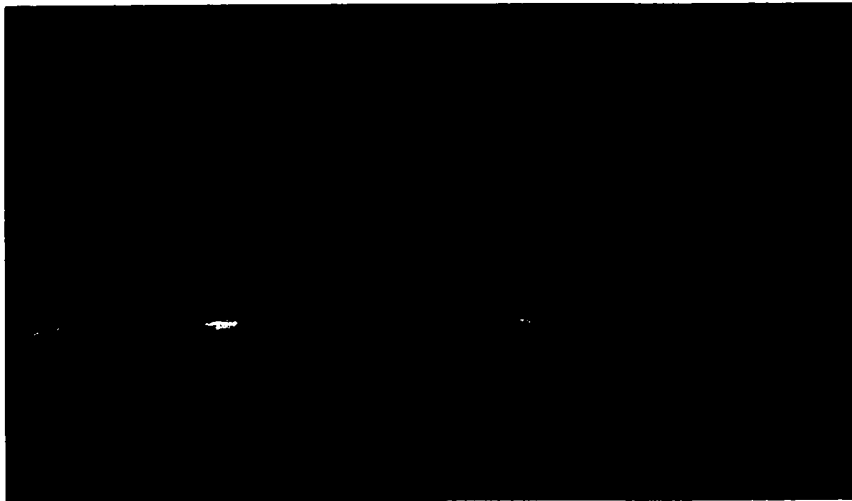
Figure 6 Silver stained polyacrylamide gel showing AFLP profile of two durum wheat cultivars using EAAC/MCTA selective primer pair. Arrows indicate cultivar-specific polymorphic markers that were selected for cloning and subsequent transformation into STS markers.

1 2 3 4 5 6



Figure 7 Agarose gel showing a polymorphic profile of durum wheat cultivars using G03-255 primer pair. Lanes 1-13 are the 13 durum cultivars. Lane 14 is the DNA size marker.

1 2 3 4 5 6 7 8 9 10 11 12 13 14



←
500 bp

←
200 bp

Figure 8 Agarose gel showing a restriction fragment polymorphism of A05-290 digested with *HaeIII* restriction enzyme. Cultivars that have one *HaeIII* recognition site within the amplicon have two DNA bands and those with three such sites have three bands.

1 2 3 4 5 6 7 8 9 10 11 12 13 14



CHAPTER FOUR

4.1

Conclusion

The hypothesis in the present thesis was that AFLP-based diversity estimates would be lower than pedigree-based estimates and furthermore AFLP would be a better estimator of the actual diversity level than the pedigree method. In this thesis two independent methods (AFLP technology and pedigree data) were used in order to estimate the level of genetic variation that is present in Canadian durum wheat cultivars. In addition, two DNA based identification schemes were developed for cultivar identification.. To this end, two objectives were set. The first objective was to screen a population of 130 genotypes representing all currently registered durum wheat cultivars to generate sufficient polymorphic markers for diversity analysis. A comprehensive pedigree search was conducted to include all ancestors, selections and breeding lines for which the breeding records were available. The second objective was to use cultivar specific AFLP markers to identify each cultivar and transform these markers into STS loci for rapid PCR-based identification of durum wheat cultivars.

The first objective was successfully met. A total of 89 polymorphic markers were generated using 10 AFLP primer pairs (Chapter 2) and average genetic diversity levels were measured. These markers exhibited two patterns of polymorphic profile. 41 of them were polymorphic between but monomorphic within cultivars and were hence termed cultivar-specific. The remaining 48 markers showed polymorphism both within and between cultivars. Three separate cluster analyses were performed using the UPGMA clustering method. The first clustering was done using 41 cultivar specific markers, the second with the remaining 48

markers, and for the third clustering analysis both classes of markers were pooled. By subdividing the data set into two subsets and subsequent cluster analysis it was concluded that genetic relationships among cultivars remained fairly unchanged when the two different subsets were compared. Furthermore, the combination of the two subsets into a larger data set of 89 markers did not result in any major structural change in the genetic relationships as represented by UPGMA dendograms. These results show that an adequate number of markers over the whole range of genome was obtained.

Measurements of the levels of genetic variation within each cultivars were determined as average gene diversity over loci. We found that there is still a substantial amount of genetic variation present within each cultivar at the breeders' seed level despite rigorous selection pressure aimed at variety purity. In addition to the possibility of contamination in the seed samples, activation of transpositions maybe a more probable reason that could potentially contribute to this variation. The effects of micro-climatic changes on the rate of transposition in barley were studied by Kalandar et al. (1999). It was found that selection pressure as a result of climate change had a significant impact on the activity of BARE-1-like retrotransposons. In this thesis, the level of genetic variation was highest within 'Pelissier' which is a selection from a landrace and lowest in 'Wascana' (a modern cultivar). This result was expected since landraces and wild populations contain more diversity than cultivated cultivars (Autrique et al. 1996; Kim and Ward 1997 Saghai-Marroof et al. 1994). Variance components for within and between cultivars were measured using AMOVA. The between cultivar variance component was 8.5 fold higher than within cultivar component.

Pedigree-based diversity measurements were obtained from the pedigree of each cultivar

using kinship coefficient. Average genetic diversity among cultivars was measured at 0.76. In contrast AFLP-based average genetic diversity estimates were measured as 0.40. Despite a large difference between the two diversity measures, a moderate positive correlation ($r = 0.457$, $p < 0.002$) was detected between AFLP and pedigree matrices using Mantel's test.

Our hypothesis was supported in that the level of genetic diversity was much higher when the pedigree approach was used compared to the AFLP method. Similar studies involving comparison of molecular markers and pedigree data in other cereal species such as bread wheat (Barrett et al. 1998), barley (Tinker et al. 1993), and durum wheat (Autrique et al. 1996) have revealed similar trends.

The results of this thesis indicate that durum wheat cultivars which are unrelated by pedigree share a substantial amount of genetic similarity which is undetected by the pedigree analysis but can be studied using the AFLP method. In other words, the actual level of genetic variation can be overestimated when pedigree analysis is used (Figure 9 p. 90). In this thesis, the AFLP analysis of 'Pelissier' and 'Hercules' revealed a genetic similarity of 0.45. Since these two cultivars were unrelated by pedigree the KIN value between them was zero. The results also indicate that there is still a significant level of genetic variation present within modern cultivars of durum wheat which could be used in breeding programs to maximize the level of allelic variation. As well, genetic variation was detected within each cultivar contrary to the assumption that modern cultivars are genetically uniform.

The second objective of this thesis was to use the subset of cultivar-specific AFLP markers for identification of cultivars. These markers were called cultivar-specific because they exhibited a monomorphic profile for all the individuals within a cultivar but a polymorphic

profile among cultivars. Since each marker has two states (either the presence or absence of a DNA fragment), a minimum number of five such markers is needed to identify all cultivars. However, in this thesis the shortest identification key required a minimum number of seven markers. Although it was possible to identify each Canadian durum cultivar with these AFLP markers, the application of labourious AFLP technology for routine identification purposes had serious limitations. Therefore, the other part of the second objective was to transform these AFLPs into sequence tagged sites (STSs) which are simple PCR-based amplicons. We have developed specific PCR primers (Table 7 p. 78) based on the entire sequence of AFLP markers. Primer selection were based on annealing temperature, GC content and maximum length of the amplified products. Previous workers have developed STS primers from AFLPs by incorporating the restriction enzyme recognition site plus the selective nucleotides and extended 3' to incorporate few additional nucleotides in order to recover the original AFLP markers (Qu et al. 1998; Reamon-Buttner and Jung 2000). In this study, we chose a different approach. Most of the STS primers in our study were designed in such a way that they did not include the restriction enzyme recognition sites on either ends of the sequence, but were rather designed from the internal sequences. 60% (7 out of 12) of the STS primers didn't produce any polymorphic profile when screened against all Canadian durum wheat cultivars while 40% (5 out of 12) showed polymorphism between cultivars. Two of the former loci were polymorphic when digested with *HaeIII* and *MnII* restriction enzymes. There may be several reasons for these result. STS polymorphism can occur when there are mutations in the primer annealing sites. On the other hand AFLP polymorphism could be the result of mutations in the restriction enzyme recognition site, mutations in the selective nucleotides, or insertions/deletions in a locus in the genome.

Because the majority of STS loci were monomorphic the most probable reason seems to be the presence of mutations within the restriction enzyme recognition sites. In this case the STS primer annealing site could be present in all genomes assayed.

In summary, this thesis deals with genetic diversity estimates and cultivar fingerprinting of Canadian durum wheat. Average genetic diversity was measured using AFLP and pedigree methods. It has been demonstrated that pedigree-based diversity estimates tend to be higher compared with AFLP-based estimates. Furthermore, a significant level of genetic similarity between cultivars with unrelated pedigrees has been shown using AFLPs. As well, there is still a substantial amount of genetic variation present within each modern cultivar despite selection pressure in the breeding programs and the assumption of genetic uniformity of modern cultivars. In addition, this thesis had led to the development of two cultivar identification schemes based on AFLP and STS polymorphism.

4.2 Future Research

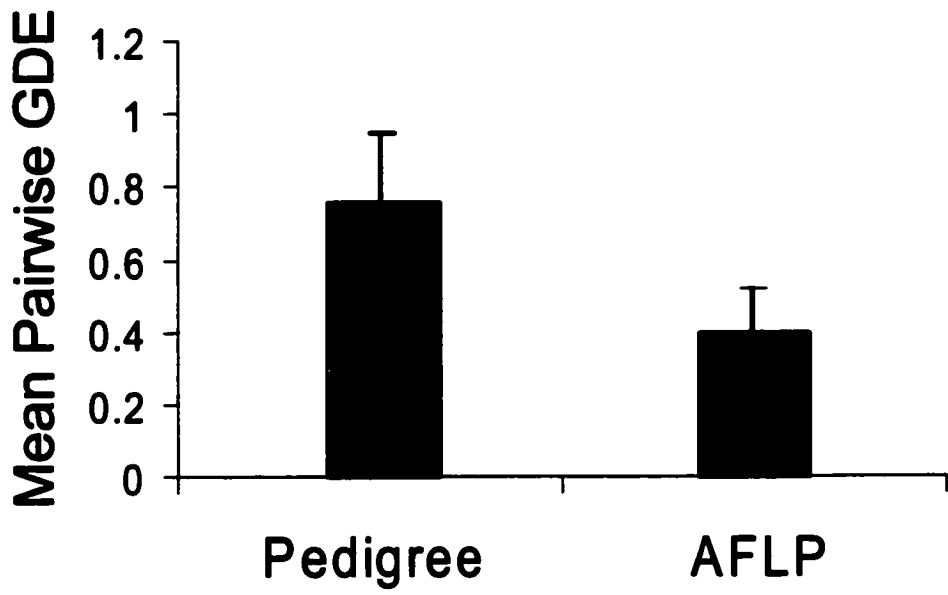
Although this thesis has answered questions pertaining to the levels of genetic variation among Canadian durum wheat cultivars and their genetic relationships as well as providing DNA-based identification assays for cultivar identification, many questions remain unanswered.

An examples of such a question relating to the results of the thesis which should be considered is where do the cultivars-specific AFLPs map? These markers were used for both diversity estimates and cultivar fingerprinting. However, they are unmapped markers with unknown genomic location and whether or not they could be used as markers linked to some agronomic traits remains to be determined. Once their genomic location is determined these markers could potentially be used for marker assisted selection (MAS). The high multiplex ratio of AFLP system coupled with equally high degree of detected polymorphism in wheat could be used for the construction of a dense genetic map (Roupe van der Voort, et al. 1997). Also the nature of AFLPs generated in this thesis with respect to dominance/recessivity could be explored in a segregating population.

In choosing the best AFLPs for the development of more robust STS markers it is important to know the nature of these markers. Ideally, these AFLPs should be unique sequences which are present in a single copy in the genome. Although the nature of AFLPs in this study is unknown, in other species such as *Asparagus officinalis* (Reamon-Buttner et al. 1999) they were of a highly repetitive nature. Designing STS markers from repetitive sequences may complicate the screening assay because the STS primer may have more than one annealing site in the genome or amplify a different locus. Some of these problems were encountered during the execution of this thesis. Finally, the results of this thesis together with future work on other

grains could lead to the development of DNA-based identification kits which can be used in the grain handling system for identification, purity, and quality testing.

Figure 9 Comparison of the mean genetic diversity estimates based on AFLP and pedigree for Canadian durum wheat cultivars.



REFERENCES

- Ahnert, D., Lee, M., Austin, D.F., Livini, C., Woodman, W.L., Openshaw, S.J., Smith, J.S.C., Porte, K., and Dalton, G. 1996. Genetic diversity among elite sorghum inbred lines assessed with DNA markers and pedigree information. *Crop Sci.* **36**: 1385-1392
- Allard, R.W. 1992. Predictive methods for germplasm identification. CAB International, Wallingford, Oxon, UK. Pp. 119-146.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology.* **215**: 403-410.
- Autrique, E., Nachit, M.M., Monneveux, P., Tanksley, S.D., and Sorrells, M.E. 1996. Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage. *Crop Sci* **36**: 735-742.
- Barrett, B.A., Kidwell, K.K., and Fox, P.N. 1998. Comparison of AFLP and pedigree-based genetic diversity assessment methods using wheat cultivars from the Pacific Northwest. *Crop Sci* **38**: 1271-1278.
- Bassam, B.J., Caetano-Anolles, Gresshoff, P.M. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* **196**: 80-83.
- Baum, B.R., Bailey, L.G., Nugent, J., Royds, W.G., and Elvidge, R. 1990. The *triticales* cultivar international register computer system. A flexible information system suitable for international registration. *Taxon* **39**: 9-15.
- Baum, B.R., Mechenda, S., and Soleimani, V. 2000. Identification of Canadian six row barley (*Hordeum vulgare* L.) Cultivars with primers derived from STSs obtained from RAPDs diagnostic bands. *Seed Sci. & Technol.* **28**: 445-466.

- Beer, S.C., Goffreda, J., Phillips, T.D., Murphy, J.P., and Sorrells, M.E. 1993. Assessment of genetic variations in *Avena sterilis* using morphological traits, isozymes, and RFLPs. *Crop Sci.* **33**: 1386-1393.
- Blair, M.W., and McCouch, S.R. 1997. Microsatellite and sequence-tagged-site markers diagnostic for the rice bacterial leaf blight resistance gene xa-5. *Theor. Appl. Genet.* **95**: 174-184.
- Brown, A.H.D. 1989. Core collections: a practical approach to genetic resources management. *Genome* **31**: 818-824. Cambridge, UK.
- Canada, 1970. The Canada Grain Act, Statutes of Canada, 19 Elizabeth II, Chapter 7, assented to 18th December.
- Cedar, H. 1988. DNA methylation and gene activity. *Cell* **53**: 3-4.
- Chao, S., Sharp, P.J., Warland, J., Warham, E.J., Koebner, R.M.D., and Gale, M.D. 1989 RFLP-based genetic maps of wheat homeologous group 7 chromosomes. *Theor Appl Genet* **78**: 495-504
- Cho, Y.G., Blair, M.W., Panaud, O., and McCouch, S.R. 1996. Cloning and mapping of variety-specific rice genome DNA sequences: amplified fragment length polymorphism (AFLP) from silver-stained polyacrylamide gels. *Genome* **39**: 373-378.
- Clark, J.A., and Bayles, B.B. 1942. Classification of wheat varieties grown in the United States in 1939. USDA Bull. 795, Washington, DC.
- Cox, T.S., Kiang, Y.T., Gorman, M.B., and Rodgers, D.M. 1985 Relationship between coefficient of parentage and genetic similarity indices in soybean. *Crop Sci* **25**: 529-532.

- Cox, T.S., Murphy, J.P., and Harrell, L.G. 1988. Isoelectric focussing patterns of kernel isozymes from 80 North American winter wheat cultivars. *Can. J. Plant Sci.* **68**: 65-72.
- Dallwitz, M.J., Paine, T.A., and Zurcher, E.J. 1993. User's guide to DELTA system: a General System for Processing Taxonomic Description. 4th edition *URL*
http://biodiversity.uno.edu/delta.
- Davila, J.A., Sanchez, de la Hoz, M.P., Loarce, Y., and Ferrer, E. 1998. The use of random amplified microsatellite polymorphic DNA and coefficient of parentage to determine genetic relationships in barley. *Genome* **41**: 477-486.
- de Jong, W., Forsyth, A., Leister, D., Gebhardt, C., and Baulcombe, D.C. 1997. A potato hypersensitive resistance gene against potato virus X maps to a resistance gene cluster on chromosome 5. *Theor. Appl. Genet.* **95**: 246-252.
- Deng, Z., Huang, S., Xiao, S., and Gmitter, F.G. Jr. 1997. Development and characterization of SCAR markers linked to the citrus triteza virus resistance gene from *Poncirus trifoliata*. *Genome* **40**: 697-704.
- Devos, K.M., and Gale, M.D. 1992. The use of random amplified polymorphic DNA markers in wheat, *Theor. Appl. Genet.* **84**: 567-572.
- Dice, L.R. 1945. Measures of the amount of ecological association between species. *Ecology* **26**: 297-302.
- Doyle, J.J., and Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15.
- Erpelding, J.E., Blake, N.K., Blake, T.K., and Talbert, L.E. 1996. Transfer of sequence tagged sites PCR markers between wheat and barley. *Genome* **39**: 802-810.

- Fabriani, Giuseppe and Lintas, Claudia. 1988. Durum wheat chemistry and technology. American Association of Cereal Chemists. Minn.
- Fowler, D.B. 2000. Winter wheat production manual. University of Saskatchewan. URL http://www.usask.ca/agriculture/plantsci/winter_wheat/contents.htm.
- Frankel, O.H., and Soulé, M.E. 1981. Conservation and evolution. Cambridge University Press.
- Gepts, P., Stockton, T., and Sonnate, G. 1993. Use of hypervariable markers in genetic diversity studies, *in application of RAPD technology to plant breeding*, Neff, M., Ed., ASHS Publishers, St. Paul, pp. 41-45.
- Gerdes, J.T., and Tracy, W.F. 1994. Diversity of historically important sweet corn inbreds estimated by RFLPs, morphology, isozymes, and pedigree. *Crop Sci.* **34**: 26-33.
- Goodman, M.M. 1972. Distance analysis in biology. *Syst. Zool.* **21**: 174-186.
- Graner, A., Ludwig, W.F., and Melchinger, A.E. 1994. Relationships among European barley germplasm: II. Comparison of RFLP and pedigree data. *Crop Sci* **34**: 1199- 1205.
- Hill, M., Witsenboer, H., Zabeau, M., Vos, P., Kesseli, R. et al., 1996. PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lectuca ssp.* *Theor. Appl. Genet.* **93**: 1202-1210.
- Hillier, L., and Green, P. 1991. OSP: A computer program for choosing PCR and DNA sequencing primers. *PCR Methods and Applications* **1**: 124-128.
- Hintum, T.J.L. van, and Haalman, D. 1994. Pedigree analysis for for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare* s.lat). *Theor. Appl. Genet.* **28**: 70-74.

- Jian, S.K., Qualset, C.O., Bhatt, G.M., and Wu, K.K. 1975. Geographical patterns of phenotypic diversity in a world collection of durum wheats. *Crop Sci.* **15**: 700-704.
- Kaemmer, D., Ramser, J., Schon, M., Weigand, F., Saxena, M.C., Driesel, A.J., Kahl, G., and Weising, K. 1992. DNA fingerprinting of fungal genomes: a case study with *Ascochyta rabiei* in DNA polymorphisms in eucaryotic genomes, Kahl, G., Appelhans, H., Komf, J., and Driesel, A.J., Eds., *Biotech-Forum 10, Adv. Mol. Genet.* **5**, Huthig Verlag, Heidelberg, pp. 255-270.
- Karp, A., Seberg, O., and Buiatti, M. 1996. Molecular techniques in the assessment of botanical diversity. *Ann. Bot. (London)* **78**: 143-149.
- Kawchuk, L.M., Hachey, J., and Lynch, D.R. 1998. Development of sequence characterized DNA markers linked to dominant verticillium wilt resistance gene in tomato. *Genome* **41**: 91-95.
- Kempthorne, O. 1969. An introduction to genetic statistics. Iowa State University Press. Ames, Iowa.
- Kim, H.S., and Ward, R.W. 1997. Genetic diversity in Eastern U.S. soft winter wheat (*Triticum aestivum* L. em. Thell.) based on RFLPs and coefficient of parentage. *Theor Appl Genet* **94**: 472-479.
- Kohli, M.M., 1986. Wheat varieties of the Southern Cone region of South America. Names; parentage; pedigrees and origins. CIMMYT, Mexico, D.F.
- Lubbers, E.L., Gills, K.S., Cox, T.S., and Gills, B.S. 1991. Variations of molecular markers among geographically diverse accessions of *Triticum tauschii*. *Genome* **34**: 354-361.

- MacGibbon, Duncan Alexander 1932. *The Canadian grain trade*. Toronto.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209-220.
- Martin, J.M., Blake, T.K., and Hockett, E.A. 1991. Diversity among North America spring barley cultivars based on coefficient of parentage. *Crop Sci* **31**: 1131-1137.
- Maugham, P.J., Saghai-Marrof, M.A., Buss, G.R., and Huestis, G.M. 1996. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near isogenic line analysis. *Theor. Appl. Genet.* **93**:392-401.
- Messmer, M.M., Melchinger, A.E., Hemann, R.E., and Boppenmaier, J. 1993. Relationships among European maize inbreds: II. Comparison of pedigree and RFLP data. *Crop Sci.* **33**: 944-950.
- Michelmore, R.W., Paran, I., and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulk segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. USA* **88**: 9828-9832.
- Murphy, J.P., Cox, T.S., and Rogers, D.M. 1986. Cluster analysis of red winter wheat cultivars based upon coefficient of parentage. *Crop Sci.* **26**: 672-676.
- Nei, M., and Li, W. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**: 5269-5273.
- O'Donoghue, L.S., Souza, E., Tanksley, S.D., and Sorrell, M.E. 1994. Relationships among North American oat cultivars based on restriction fragment length polymorphism. *Crop Sci* **34**: 1251-1258.

- Olsen, M., Hood, L., Cantor, C., and Botstein, D. 1989. A common language for physical mapping of human genome, *Science* **245**: 1434-1435.
- Paran, I., and Michelmore, R.W. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce *Theor. Appl. Genet.* **85**: 985-993.
- Powell, w., Morgante, M., and Andre, C. 1996. The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, **2**:225-238.
- Qi, X., and Lindhout, P. 1997. Development of AFLP markers in barley. *Mol. Gen. Genet.* **254**: 330-336.
- Qu, L.J., Foote, T.N., Roberts, M.A., Money, T.A., Aragon-Alcaide, L. Snape, J.W., Moore, G. 1998 . A simple PCR-based method for scoring the phlb deletion in wheat. *Theor. Appl. Genet.* **96**: 371-375.
- Reamon-Buttner, S.M., and Jung, C. 2000. AFLP-derived STS markers for the identification of sex in *Asparagus officinalis* L. *Theor. Appl. Genet.* **100**: 432-438.
- Reamon-Buttner, S.M., Schmidt, T., and Jung, C. 1999. AFLPs represent highly repetitive sequences in *Asparagus officinalis* L. *Chrom. Res.* **7**: 297-304.
- Reeves, C.J., Law, J.R., Donini, P., Koebner, M.D., and Cooke, J.R. 1999. Changes over time in the genetic diversity of UK cereal crops. National Institute of Agricultural Botany, Cambridge, UK. URL: <http://apps3.fao.org/wiews/prague/paper12.htm>
- Rogers, D.M., Murphy, J.P., and Frey, K.J. 1983. Impact of plant breeding on the grain yield and genetic diversity of spring oats. *Crop Sci.* **23**: 737-740.
- Rohlf, F.J. 1990. NTSYS-pc. Numerical taxonomy and multivariate analysis system. ver. 2.02. Exeter software, Setauket, NY.

- Roupe van der Voort, J.N.A.M., van Zandvoort, P., van Eck, H.J., Folkertsma, R.T., Hutten, R.C.B., Draaistra, J., Gommers, F.J., Jacobson, E., Helder, J., and Bakker, J. 1997. Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. *Theor. Appl. Genet.* **255**: 438-447
- Saghai-Marouf, M.A., Zhang, Q., and Biyashev, R. 1994. Comparison of restriction length polymorphisms in wild and cultivated barley. *Genome*, **38**: 298-306.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning. A laboratory manual.* Cold Spring Harbor Laboratory press. 3 vols.
- Sanger, F., Nicklen, S., and Coulson A.R. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.
- Schneider, S., Kueffer, J.M., Roessli, D., and Excoffier, L. 1995. Arlequin: a software for population genetic data analysis. Genetics and Biochemistry Laboratory, University of Geneva, Switzerland. URL: <http://anthropologie.unige.ch/arlequin>
- Schoen, D.J., and Brown, A.H.D. 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system. *Proc. Natl. Acad. Sci. USA* **88**:4494-4497.
- Sharma, S.K., Knox, M.R., and Ellis, T.H. 1996. AFLP analysis of the diversity and phylogeny of lens and its comparison with RAPDs analysis. *Theor. Appl. Genet.* **93**:751-758.
- Smith, O.S., Smith, J.S.C., Bowen, S.L., Tenborg, R.A., and Wall, S.J. 1990. Similarity among a group of elite maize inbreds as measured by pedigree, F1 heterosis, and RFLPs. *Theor Appl Genet* **80**: 833-840.
- Sneath, P.H.A., and Sokal, R.R. 1973. *Numerical taxonomy. The principles and practice of*

- numerical classification. W.H. Freeman and Co., San Francisco CA.
- Souza, E., and Sorrells, M.E. 1989. Pedigree analysis of North America oat cultivars released from 1951 to 1985. *Crop Sci* **29**: 595-601.
- Talbert, L.E., Bruckner, P.L., Smith, L.Y., Sears, R., and Martin, T.J. 1996. Development of PCR markers linked to resistance to wheat streak mosaic virus in wheat. *Theor. Appl. Genet.* **93**: 463-467.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**: 4673-4680.
- Tinker, N.A., and Mather, D.E. 1993. KIN: Software for computing kinship coefficients. *J. Hered.* **84**: 238.
- Tinker, N.A., Fortin, M.G., and Mather, D.E. 1993. Random amplified polymorphic DNA and pedigree relationships in spring barley. *Theor Appl Genet* **85**: 976-984.
- van Beuningen, L.T., and Busch, R.H. 1997. Genetic diversity among North American Spring Wheat Cultivars: I. Analysis of the coefficient of parentage matrix. *Crop Sci.* **37**: 570-579.
- Vierling, R.A., and Nugyen, H.T. 1992. Use of RAPDs markers to determine the genetic diversity of diploid wheat genotypes, *Theor. Appl. Genet.* **84**: 835-838.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van der Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**: 4407-4414.

- William, K.J., Fisher, J.M., and Langridge, P. 1996. Development of a PCR-based allele-specific assay from an RFLP probe linked to resistance to cereal cyst nematode in wheat. *Genome* 39:798-801.**
- Zabeau, M. 1993. Selective restriction fragment amplification: A general method for DNA fingerprinting. European Patent Office. Publication number 0 534 858 A1.**
- Zeven, A.C., and van Hintum, T.J.L. 1992. Classification of landraces and improved cultivars of hexaploid wheats (*Triticum aestivum*, *T. compactum*, and *T. spelta*) grown in the USA and described in 1922 *Euphytica* 59: 33-47.**
- Zeven, A.C., and Zeven-Hissink, N.C. 1976. Genealogies of 14000 wheat varieties. The Netherlands Cereals Centre, Wageningen.**

APPENDIX

The pedigree of all registered Canadian durum wheat cultivars except for 'Pelissier' (a selection from a landrace) is presented on a CD in appendix. The scanned pedigree files are saved in JPEG format and can be viewed using adobe photoshop, Microsoft photoEditor, or similar programs.

Each cross is represented by two arrows, each coming from one parent. The number on the side of some arrows indicate the number of back crosses that were made in the development of that particular cross. Each cross or cultivar is represented by a code (Table 1 p. 28). Unnamed selection, crosses, and lines are marked as unknown.