NOTE TO USERS

This reproduction is the best copy available.

UMI®

uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Abdulvahab Soleimani
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Biology)
GRADE / DEGREE

Department of Biology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Genome Dynamics in Barley (*Hordeum vulgare* L.) Cultivars:
Molecular Diversity, Evolution, and DNA Fingerprinting

TITRE DE LA THÈSE / TITLE OF THESIS

Douglas Johnson
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Bernard Baum
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

John Arnason

Linda Bonen

Iain Lambert

Daryl Somers

Xuhua Xia

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Genome Dynamics in Barley (*Hordeum vulgare* L.) Cultivars: Molecular Diversity, Evolution, and DNA Fingerprinting

©VAHAB D. SOLEIMANI

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies

University of Ottawa

In partial fulfillment of the requirements for the Ph.D. degree

In the Ottawa-Carleton Institute of biology

Faculty of Science

Department of Biology

University of Ottawa

Ottawa, Ontario

Canada

2005

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-11027-9
Our file Notre référence
ISBN: 0-494-11027-9

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

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

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

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

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

ACKNOWLEDGMENTS

This work could not have been completed without the crucial teaching and guidance of both of my co-supervisors, Dr. Douglas A. Johnson and Dr. Bernard R. Baum. I am indebted to them for their insight, encouragement and support.

For unconditional love and support, I am grateful to my dear wife Arezu Jahani-Asl.

This study was partially funded by Automated Quality testing Inc. in collaboration with Agriculture and Agri-Food Canada, the Canadian Grain Commission (CGC) and Cereal Research Centre (CRC) Winnipeg, Manitoba. I would like to acknowledge the following individuals for their support: Dr. Tigest Demeke and Dr. Mathieu Dusabenyagasani from CGC for electronic SNP mining; Dr. Douglas Procunier and Mark Gray from CRC for the use of ABI genetic analyzer machine; Ms. Jean Mellish from CGC, Agit Singh Sahota from the Canadian Food Inspection Agency (CFIA), Dr. Thin Meiw (Alek) Choo from Agriculture and Agri-Food Canada, and Dr. Alexander Belyeyev from the Institute of Evolution, University of Haifa, Israel for providing *Hordeum spontaneum* seed material; Dr. Burton Blais, Leah Isaac, and Amelia Martinez from the CFIA for the use of LightCycler machine and Susan Floods and Judy McCarthy of Agriculture and Agri-Food Canada for graphic illustrations.

I am grateful to my advisory committee members Dr. John T. Arnason, Dr. Linda Bonen, and Dr. Ian Lambert for their constructive suggestions and critiques. For technical advice in the laboratory, I am indebted to my friends and co-workers Grant Bailey and Subbaiah Mechanda. I would also like to thank Dr. Nick Tinker, Dr. Steven Molnar, Dr. Johann Schernthaner and Dr. Brian Miki at Agriculture and Agri-Food Canada for reviewing earlier drafts of manuscripts in this thesis.

RÉSUMÉ

En l'absence du séquençage du génome complet, les marqueurs moléculaires sont des outils indispensables pour l'étude de l'évolution du génome, des mesures de la diversité génétique et de l'identification du génotype. Nous avons utilisé des marqueurs polymorphisme d'amplification séquence-spécifique (S-SAP) dérivés du BARE-1, un rétrotransposon actif de l'orge, pour mesurer la contribution de cet élément à l'évolution du génome de l'orge parmi 103 cultivars qui poussent couramment au Canada et aux États-Unis. Les résultats ont été comparés aux mesures de la diversité du génome obtenues par les polymorphismes mononucléotidiques (SNP).

La population d'orge a été divisée en groupes selon diverses propriétés agronomiques, telles que l'utilisation finale (p. ex. alimentation animale vs maltage) et la morphologie des graines (p. ex. grains sans enveloppe vs grains avec enveloppe). L'analyse de la structure génétique de la population, à l'aide de l'analyse de la variance moléculaire (AMOVA) pour les S-SAP et les SNP, a attribué la plus grande composante de covariance (90 p. 100) à la diversité génétique parmi les cultivars de ces groupes. La composante de covariance entre les groupes était d'environ 6 p. 100, ce qui ne justifiait pas la différentiation de la population dans l'ensemble d'après les propriétés agronomiques. Les relations génétiques entre les cultivars ont été évaluées par une analyse de groupement utilisant UPGMA (méthode d'analyse de groupes de paires non pondérées à partir de moyennes), ce qui a révélé une variation considérable entre les ensembles de données S-SAP et SNP.

L'analyse quantitative du rétrotransposon BARE-1, avec réaction en chaîne de la polymérase (PCR) en temps réel, dans un petit groupe de cultivars a démontré des

différences importantes dans le nombre de copies de cet élément parmi les cultivars. La plupart des éléments du BARE-1 étaient sous forme de longues répétitions terminales (solo LTR), révélant un taux élevé de recombinaisons homologues entre les copies de rétrotransposons du génome. Des différences pouvant atteindre 3 000 copies du BARE-1 par génome haploïde ont été repérées parmi les cultivars développés au cours des trois dernières décennies. Les SNP informatifs tels ceux ayant un PIC (taux d'information sur le polymorphisme) élevé ont servi à produire les clés d'identification pour différencier les cultivars d'orge autrement indifférenciables sur les plans de la morphologie et de la biochimie.

ABSTRACT

In the absence of whole genome sequencing, molecular markers are indispensable tools for the study of genome evolution, genetic diversity measurements, and genotype identification. We have used sequence-specific amplified polymorphism (S-SAP) markers that were derived from the BARE-1, an active retrotransposon of barley, to measure the contribution of this element to the evolution of barley genome among 103 cultivars that are commonly grown in Canada and the United States. The results were compared to the genome diversity measures obtained by the single nucleotide polymorphisms (SNP).

The barley populations were divided into groups based on various agronomic traits such as end use, i.e., feed versus malting, and seed morphology, i.e., naked versus covered kernel. Analysis of the genetic structure in the population using analysis of molecular variance (AMOVA) for both S-SAP and SNP attributed the largest co-variance component (90%) to the genetic diversity among cultivars within groups. Co-variance component between groups was about 6% which indicated that there was no justification for population differentiation along the set based upon agronomic traits. Genetic relationships among cultivars was assessed by cluster analysis with UPGMA and found to vary substantially between S-SAP and SNP datasets.

Quantitative analysis of BARE-1 retrotransposon with real-time PCR in a small group of cultivars showed significant differences in the copy number of the element among cultivars. Most of the BARE-1 elements were in the form of solo LTRs, indicating a high rate of homologous recombination between retrotransposon copies in the genome.

Differences of up to 3000 BARE-1copies per haploid genome were found among cultivars that have been developed and registered within the past three decades. Informative SNPs such as those with high polymorphic information content (PIC) values were used to generate identification keys to distinguish barley cultivars which were otherwise indistinguishable at the morphological and biochemical levels.

TABLE OF CONTENTS

Genome Dynamics in Barley (*Hordeum vulgare* L.) Cultivars: Molecular Diversity, Evolution, and DNA Fingerprinting I
ACKNOWLEDGMENTS II
RÉSUMÉ III
ABSTRACT V
LIST OF TABLES XI
LIST OF FIGURES XIII
LIST OF ABBREVIATIONS XV
CHAPTER ONE 1
INTRODUCTION 1
1.1 Taxonomy and origin of barley 1
1.2 Barley genome 2
1.3 Plant breeding 5
1.4 Identification of plant cultivars 7
1.5 Genomes and evolution 8
1.6 Transposable genetic elements 10
1.7 Retrotransposons 11
1.8 Retrotransposons and Host Genome evolution 20
1.9 Regulation of Transposition 22
1.10 Interaction between the host and retrotransposons 24
1.11 Maintenance of retrotransposons in genomes 26
1.12 Retrotransposons as DNA markers for genetic studies 28
1.13 Single nucleotide polymorphism (SNP) 31
1.14 Measuring genetic diversity 32
1.15 Rational for the present study 34
1.16 HYPOTHESES AND OBJECTIVES 37
CHAPTER TWO 42
Genetic diversity among barley cultivars assessed by sequence-specific amplification polymorphism 42
2.1 Abstract 44
2.2 Introduction 46
2.3 Materials and Methods 54
2.3.1 Plant materials and S-SAP reactions 54
2.3.2 BARE-1 primers 55
2.3.3 Data analysis 58
2.3.3.1 Clustering 58
2.3.3.2 Gene diversity estimates 58
2.3.3.3 Analysis of molecular variance 59
2.3.3.4 Regression analysis 59
2.3.3.5 Discriminant Analyses 59
2.4 Results 61
2.4.1 Genetic diversity/similarity estimates 61
2.4.2 Clustering and DA 63

2.4.3 AMOVA analysis: Partitioning of genetic variance 63
2.4.4 Genetic heterogeneity within cultivars 66
2.5 Discussion 68
2.5.1 Measurements of genetic diversity, cluster and DA analysis 69
Efficient validation of Single Nucleotide Polymorphisms in plants by allele-specific PCR, with an example from barley 74
3.1 Abstract 76
3.2 Introduction 77
3.3 Materials and Methods 78
3.3.1 Plant Material and DNA extraction 78
3.3.2 Detection and identification of SNPs 80
3.3.3 SNP validation 81
3.3.3.1 AS-PCR 81
3.3.3.2 Primer extension 81
3.4 Results 82
3.5 Discussion 84
3.6 Acknowledgments 87
4.1 Abstract 91
4.2 Introduction 92
4.3 Materials and Methods 98
4.3.1 Plant material and DNA extraction 98
4.3.2 PCR primers 98
4.3.3 S-SAP analysis 99
4.3.4 SNP analysis 102
4.3.4.1 Detection 102
4.3.4.2 Validation 103
4.3.5 Data analysis 103
4.3.5.1 Data acquisition and scoring 103
4.3.5.2 GS estimation 104
4.3.5.3 AMOVA 104
4.3.5.4 Matrix comparison and Mantel test 105
4.3.5.5 Pedigree analysis 105
4.4 Results 106
4.4.1 Genetic diversity as revealed by S-SAP 106
4.4.2 Pedigree analysis 106
4.4.3 SNP analysis 110
4.4.4 Low rate of SNP validation with PCR analysis 117
4.4.5 AMOVA 117
4.4.6 Matrix comparison 118
4.5 Discussion 120
4.6 Conclusion 126
4.7 Acknowledgments 126
CHAPTER FIVE 128
Quantification of the retrotransposon BARE-1 reveals the dynamic nature of the barley genome 128

5.1 Abstract ... 130
5.2 Introduction ... 132
5.3 Materials and Methods ... 134
5.3.1 Plant material ... 134
5.3.2 DNA isolation ... 134
5.3.3 PCR primer design and amplification reactions ... 134
5.3.4 Cloning and sequencing of RT, and LTR domains for use as external standards ... 136
5.3.5 Quantitative real-time PCR ... 137
5.3.6 Standard curves and copy number estimation ... 138
5.3.7 Differences between means ... 138
5.4 Results ... 139
5.4.1 PCR amplification and cloning of RT and LTR sequences ... 139
5.4.2 Standard and melting curves ... 139
5.4.3 RT and LTR copy number estimates ... 144
5.5 Discussion ... 147
5.6 Acknowledgements ... 150
CHAPTER SIX ... 151
Identification of six-rowed spring malting barley cultivars by using allele-specific PCR to detect Single Nucleotide Polymorphisms ... 151
6.1 Abstract ... 153
6.2 Introduction ... 154
6.3 Materials and Methods ... 157
6.3.1 Plant Material ... 157
6.3.2 DNA Extraction ... 157
6.3.3 DNA sequencing and identification of SNPs ... 160
6.3.4 SNP Validation ... 161
6.3.5 Data analysis and cultivar identification ... 161
6.3.6 Cluster analysis ... 162
6.4 Results ... 163
6.4.1 SNP discovery and confirmation ... 163
6.4.2 SNP validation with allele-specific PCR ... 164
6.4.3 Cultivar identification ... 167
6.5 Discussion ... 167
6.6 Acknowledgments ... 171
CHAPTER SEVEN ... 195
7.1 General discussion ... 195
7.1.2 Survey of barley genome with S-SAP and SNP markers ... 197
7.1.3 Comparative analysis of genetic diversity using S-SAP, SNP and KIN 200
7.1.4 Analysis of genetic variation in the population ... 201
7.1.5 Incongruence between S-SAP and SNP diversity estimates ... 202
7.1.6 Retrotransposition and the evolution of barley genome ... 204
7.1.7 Intra cultivar Genetic variation ... 205
7.2 Conclusion ... 207

7.3 Future work **208**
LITERATURE CITED **212**
APPENDIX **242**

LIST OF TABLES

Table 1.1. Examples of transposable genetic elements from various hosts........................39

Table 2.1 Material used in S-SAP analysis of Canadian barley cultivars 49

Table 2.2. Sequences of PCR primers used in S-SAP study 56

Table 2.3. S-SAP primer pair combinations, number of amplified bands and percentage polymorphism detected among barley cultivars 57

Table 2.4. Summaries of AMOVA analyses among various agronomic groups of barley cultivars...64

Table 3.1. A comparison between the AS-PCR and the primer extension methods 83

Table 4.1 Spring six-rowed barley cultivars used for comparative study.........................94

Table 4.2. DNA sequence of synthetic adapters and PCR primers that were used in S-SAP analysis 100

Table 4.3. AMOVA analysis with S-SAP and SNP data for partitioning the variance...107

Table 4.4. List of SNP alleles found within their respective loci and indices of their gene diversity within cultivars...115

Table 4.5. Pair-wise matrix correlation analysis with Mantel test for kinship, S-SAP and SNP matrices...119

Table 4.6. Maximum, minimum and mean GS estimates from pedigree (kinship), SNP, and S-SAP data for six-rowed spring barley..121

Table 5.1. Barley material used for BARE-1 quantification experiment with real-time PCR...135

Table 5.2. Copy number of BARE-1 LTR and RT genes among barley samples...........145

Table 5.3. Pair-wise squared Mahalanobis distances estimated from BARE-1 copy number between genomes...146

Table 6.1. Six-rowed malting barley cultivars used for cultivar identification study......158

Table 6.2. Consensus sequence of barley genomic loci derived from sequencing five to eight cultivars...172

Table 6.3. Identification key of six-rowed spring malting barley cultivars using SNPs..191

LIST OF FIGURES

Figure 1.1. Area planted with major crops in Canada (2003 statistics) 3

Figure 1.2a. Haploid DNA content of seven cereal species 4

Figure 1.2b. Phylogenetic relationships among cereals depicted in figure 2a 4

Figure 1.3. Progressive reduction in the level of genetic diversity 12

Figure 1.4. Structural organization of retroelements and retroviruses 15

Figure 1.5. Structural organization of BARE-1 retrotransposon 23

Figure 1.6. Mechanisms of genome contraction as a result of retrotransposon-mediated recombination 21

Figure 1.7. A schematic outline of S-SAP procedure 30

Figure 2.1. Portion of S-SAP profile of three barley cultivars 62

Figure 2.2. Discriminant analysis (DA) of barley cultivars 65

Figure 2.3. The effect of inbreeding on the intra cultivar genetic variation 67

Figure. 3.1. Schematic diagram showing the AS-PCR approach used to validate SNPs 79

Figure. 3.2. AS-PCR profile of two barley cultivars 85

Figure. 3.3. Comparison of genotyping with primer extension (left) and AS-PCR (right) 88

Figure. 4.1. Autoradiogram of a portion of an S-SAP fingerprint 108

Figure. 4.2. UPGMA dendrogram showing genetic relationship among barley cultivars 109

Figure. 4.3. UPGMA dendrogram showing genetic relationship among 35 barley cultivars 111

Figure. 4.4. Allele-specific PCR to detect a G/T transversion 112

Figure. 4.5. Multiplex detection of SNPs with the primer extension method 113

Figure. 4.6. UPGMA dendrogram showing genetic relationship among barley cultivars....114

Figure. 5.1. Agarose gel electrophoresis showing PCR amplified product of RT and LTR....140

Figure. 5.2. PCR product detection in real time through five orders of magnitude....141

Figure. 5.3. Standard curve generated from plasmid DNA containing RT domain of BARE-1 retrotransposon....142

Figure. 5.4. Melting curve analysis of RT domain....143

Figure 6.1. Multiple alignment of DNA sequences from the MWG 2218 locus....165

Figure 6.2. SNP identification by agarose gel electerophoresis....166

Figure 7.1 Distribution of genetic similarities around the mean for KIN, S-SAP and SNP data....211

LIST OF ABBREVIATIONS

AFLP	Amplified Restriction Fragment Length Polymorphism
AMOVA	Analysis of Molecular Variance
AP	Aspartic Protease
AS-PCR	Allele-Specific Polymerase Chain Reaction
BLAST	Basic Local Alignment Search Tool
CAP	Contig Assembly Program
cDNA	Complementary DNA
CFIA	Canadian Food Inspection Agency
CGC	Canadian Grain Commission
Contig	Contiguous DNA sequence assembled using overlapping DNA sequences
DA	Discriminant Analysis
DELTA	Descriptive Language for Taxonomy
DNA	Deoxyribonucleic Acid
DUS	Distinct Uniform Stable
EST	Expressed Sequence Tag
GS	Genetic Similarity
IN	Integrase
INDEL	Insertion/deletion
IR	Inverted Repeat
IRAP	Inter-Retrotransposon Amplified Polymorphism
KVD	Kernel Visual Distinguishability
LARD	Large Retrotransposon Derivative

LTR	Long Terminal Repeat
MAS	Marker Assisted Selection
MITE	Miniature Inverted-repeat Transposable Element
MYA	Million Years Ago
NCBI	National Center for Biotechnology Information
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PBR	Plant Breeders' Rights
PBS	Primer Binding Site
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
PPT	Polypurine Tract
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RBIP	Retrotransposon Based Insertion Polymorphism
REMAP	Retrotransposon-Microsatellite Amplified Polymorphism
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RT	Reverse Transcriptase
SAHN	Sequential Agglomerative Hierarchic Non overlapping Clustering Method
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
SNuPE	Single Nucleotide Primer Extension

S-SAP	Sequence-Specific Amplification Polymorphism
SSR	Simple Sequence Repeat
STS	Sequence Tag Site
TRIM	Terminal Repeat Retrotransposon in Miniature
UPGMA	Unweighted Pair Group Method with Arithmetic Average
UPOV	International Convention for the Protection of New Varieties of Plants
VLP	Virus-Like Particle

CHAPTER ONE

INTRODUCTION

1.1 Taxonomy and origin of barley

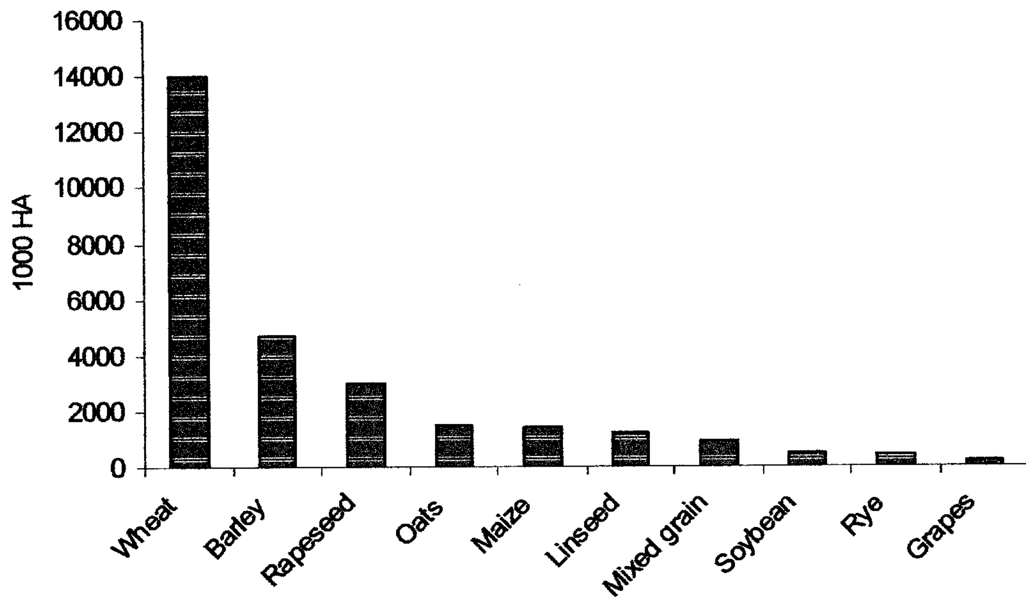
Cultivated barley (*Hordeum vulgare* L.) belongs to the tribe Triticeae in the grass family Poaceae. The *Hordeum* genus comprises about 32 species and 45 taxa (Bothmer et al. 1991). Based on spike morphology, barley can be divided into two-rowed and six-rowed types (Briggs et al. 1978). Barley is the world's fourth most important cereal crop after wheat, maize, and rice. In Canada, barley cultivation ranks number two among cereal crops (Fig. 1.1). Most of Canada's barley is produced in Alberta, Manitoba, and Saskatchewan and to a lesser extent in Ontario (230,000 acres in 2000) (Ontario Ministry of Agriculture Food and Rural Affair. URL: *http://www.gov.on.ca/OMAFRA/english/stats/crops/ctybarley00.html*). Barley is believed to be one of the earliest domesticated crop species and is grown over a broad environmental range (Zohary and Hopf 1993). Among the major cereal crops, barley is the most responsive and sensitive to the environment such that specific varieties are adapted to specific habitats. The most widely accepted hypothesis on the origin of cultivated barley defines the Fertile Crescent as its centre of origin (Harlan 1976). However, a hypothesis of multi-centric origin has also been proposed (Molina-Cano et al. 1999). In Canada, barley grain is mainly used to make malt, which in turn is used in brewing industry to produce alcoholic beverages such as beer and whisky. The grain is also used as feed for cattle and swine and in various food products for human consumption.

1.2 Barley genome

Cultivated barley is a self-pollinating diploid species (2n = 14). However, tetraploid (2n = 28) species in the genus *Hordeum* also exist (Pickering et al. 1999). The genome of cultivated barley has been estimated to contain approximately 5.5 picograms of DNA per haploid nucleus, equivalent to about 5.4×10^{9} bp (Bennet and Smith 1976). Only about 5 % of the barley genome is occupied by genes (Kankaanpaa et al. 1997; Kalendar et al. 2004). The remaining 95 % is composed of various types of non-genic DNA elements such as transposons, retrotransposons, microsatellite loci and special classes of repetitive elements found in telomeric and centromeric regions of cereals chromosomes. Approximately 10-20 % of the barley genome is composed of tandemly repeated sequences while 50-60 % of the genome is composed of repeated sequences interspersed throughout the genome (Rimpau et al. 1980; Manninen and Schulman 1993; Kalendar et al. 2004). Current estimates of gene number in plants vary from 25,000 to 60,000 (Goff et al. 2000; Kurata et al. 2002; Yu et al. 2002). The smallest plant genome, Arabidopsis contains about 25,000 genes (Arabidopsis Genome Initiative 2000). Estimates of gene number in plants are largely derived from large-scale Expressed Sequence Tag (EST) and genome sequencing projects (Arabidopsis Genome Initiative, 2000; Casacuberta and Puigdomenech 2000).

Within the grass family, the smallest genome (4.3×10^{8} bp) belongs to rice (*Oryza sativa* L.) and the largest genomes to oats (*Avena*) (3.9×10^{9} to 3.8×10^{9} bp) and wheat (*Triticum*) (4.8×10^{9} to 16.9×10^{9} bp). Barley, with a genome size of 5.4×10^{9} bp occupies an intermediate position among the grasses (Fig. 1.2). All estimates of genome

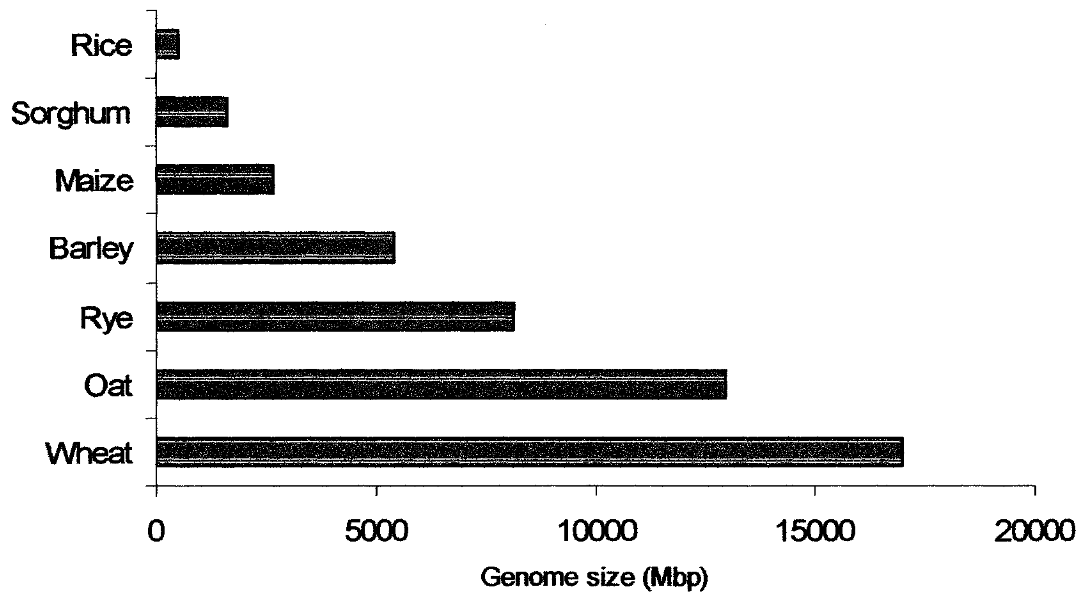
Figure 1.1: Area planted with major crops in Canada (2003 statistics). About 42 million hectares were planted with barley. Most of Canada barley is produced in Alberta (approximately 50 %). Most of the rest is produced in Saskatchewan and Manitoba.

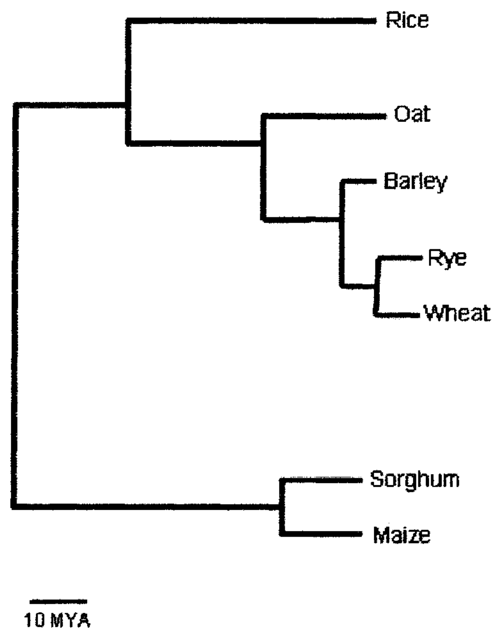


16000
14000
12000
10000
8000
6000
4000
2000
0
1000 HA
Wheat
Barley
Rapeseed
Oats
Maize
Linseed
Mixed grain
Soybean
Rye
Grapes

Figure 1.2a: Haploid DNA content of seven cereal species. The largest and the smallest genomes belong to wheat and rice respectively (35fold difference). Data was taken form the Royal Botanical Garden, Kew web site (http://www.rbgkew.org.uk/cval/database1.html).

Figure 1.2b: Phylogenetic relationships among cereal species depicted in figure 1.2a. The tree was drawn based on the information from Bennetzen and Kellogg (1997) and Grass Phylogeny Working Group (GPWG) (http://www.virtualherbarium.org/grass/gpwg).


Rice
Sorghum
Maize
Barley
Rye
Oat
Wheat
0
5000
10000
15000
20000
Genome size (Mbp)


Rice
Oat
Barley
Rye
Wheat
Sorghum
Maize
10 MYA

size in plants are taken from the Royal Botanical Garden, Kew web site (http://www.rbgkew.org.uk/cval/database1.html) unless stated otherwise.

Among cereals, analysis of the first draft of the rice genome indicated that rice had between 30 to 50 thousand genes (Goff, et al. 2002) which was much higher than the estimated 25 thousand genes in Arabidopsis and also higher than the estimated 30 thousand genes in human. The reason why plants generally have a larger genome and more genes is not known. However, there are some possible explanations. Unlike animals, plants are autotrophs and need to convert air, water and minerals into nutrients. Plants also produce vast numbers of secondary metabolites which are substances that contribute to functions such as defence and disease resistance. The number of secondary metabolites produced by all plants is estimated to be as high as 100,000 (Arabidopsis Genome initiative 2000; Poethig 2001). As Poethig (2001) put it, "animals are structurally more complex than plants, but plants probably do a lot more biochemistry".

1.3 Plant breeding

The gene pool of modern cereals is essentially derived from landraces and their wild progenitors by selection. The objective of plant breeding is to create new allele combinations followed by the selection of desirable phenotypes such as yield, pest resistance and protein content. Vavilov (1926) described plant breeding as "evolution directed by the will of man". Although modern plant breeding has benefited from a wide range of techniques including gene-manipulation tools, cultivar breeding is still achieved by the same basic mechanism that has been practiced for thousands of years, namely, directional selection for desirable trait(s). The raw material in any breeding program is

genetic variation. But in order to adequately utilize genetic variation, it must be assessed and measured for developing new cultivars, and suitable traits or combination of traits must be identified (Cooke 1992). There is a saying that breeders "cross the best with the best and hope for the best". However, modern day plant breeding is a multibillion dollar, competitive industry; therefore, breeders need more than just luck to develop plant cultivars with superior agronomic traits. In the words of N.W. Simmonds, the plant breeder is "an applied evolutionist, working towards defined objectives by tolerably well understood methods" (Cooke 1992). Breeding programs normally start by crossing two parents containing the desired traits followed by selection and screening of the progeny. In order to enhance the choice of parents and select offspring, the traits under selection must be identifiable by morphological or molecular markers (protein or DNA). Although many morphological markers are still being used in breeding programs, more and more breeders are using molecular markers for the following reasons: 1) Molecular markers are more abundant than their morphological counterpart; 2) with molecular markers genotype can be determined for the whole plant in tissue, or at the cellular level, whereas phenotypes of most morphological markers can only be distinguished for the whole plant; 3) phenotypic plasticity that is associated with morphological markers has no effect on molecular markers; and 4) with morphological markers, strong epistatic effects limit the number of segregating markers that can be unequivocally scored in the same segregating population. Epistasis does not affect molecular markers.

1.4 Identification of plant cultivars

Characterization of genetic variability between cultivars is the basis for their commercialization, the protection of intellectual property such as plant breeders' rights (PBR), and the grain handling system. In many countries including the United States and the European Union, the controlling legislation complies with the Convention of the International Union for the Protection of New Varieties of Plants (UPOV 1991). Canada has not ratified the UPOV (1991) convention. The current Canadian PBR act conforms to the UPOV (1987) and amendments to the current PBR act are required for ratification of UPOV (1991) convention. Compliance with these national and international treaties implies testing for distinctness, uniformity and stability (DUS) of plant cultivars. Cultivar-specific morphological or molecular markers need to be measured and tested to determine the above-mentioned parameters.

In Canada, to enhance grain handling and grading of cereals including barley, new technologies based on molecular markers are being implemented by the Canadian Grain Commission (CGC) and private industry. Various types of PCR-based cultivar identification techniques have already been developed for cereals including simple sequence repeats (SSRs) for durum wheat (Perry 2004), sequence tagged sites (STSs) for barley (Baum et al. 2000), amplified fragment length polymorphism (AFLP) for durum wheat (Soleimani et al. 2002b) and single nucleotide polymorphisms (SNP) for barley (Soleimani et al. 2003). Identification techniques based on molecular markers are gradually replacing traditional morphological characterization of grains based on kernel visual distinguishability (KVD). The main reason for this trend in the case of modern

wheat and barley cultivars is that KVD is insufficient to discriminate among all cultivars due to the limited morphological variation that is present among many cultivars.

Most modern breeding programs use a limited number of cultivars and elite lines from local germplasm sources. As a result the genetic base of modern crops has been reduced (Horsely et al. 1995). For example, analysis of pedigree data has shown that the gene pool of malting barley has been reduced over the last few decades due to the reliance of barley breeders on a small number of original ancestors (Martin et al. 1991; Allard 1992). Genetic diversity analysis using molecular markers has also shown a substantial reduction in diversity from landraces to modern cultivars (Archak et al. 2002; Soleimani et al. 2002a; Brantestam et al. 2004). Since modern cultivars have been developed from landraces which in turn have been selected from wild populations by primitive agriculture, there is a progressive reduction in the genetic base from wild progenitors to present day cultivars (Fig. 1.3). Against this backdrop, plants scientists including breeders, germplasm curators and population geneticists can use the enormous amount of mostly cryptic variation that is hidden in the DNA sequences of individuals within each population for efficient identification.

1.5 Genomes and evolution

The early 20th century view of evolution was shaped before we knew about DNA as the genetic material. This perspective combined the Darwinian concept of gradualism with natural selection and random mutations with the Mendelian principles as the mechanisms of evolutionary genetics (Mayr 1980). Avery et al (1944) showed that DNA was the “transforming principle” and Hershey and Chase (1952) showed that genetic

information was encoded in the sequence of DNA. Watson and Crick (1953) deciphered the double helical structure of DNA. Meanwhile in 1950, McClintock discovered that extensive chromosome rearrangements as well as alterations in the expression of various loci were common place in maize. The Operon theory of Jacob and Monod (1961) showed that genomic loci can form integrated functional systems and established the basic structural organization and regulation of protein coding genes. Finally, Britten and Kohn (1968) showed that higher eukaryotes had abundant amounts of repetitive DNA (Reviewed by Shapiro 1999). Taken together, these discoveries changed our view of genomes and of evolution. The genome was no longer viewed as static and the concepts of gradualism and Mandelian genetics were moulded together into the Neo-Darwinian synthesis of evolution

Long before the advent DNA sequencing projects, it was apparent that a large genome did not necessarily contain more genes (the so-called C-value paradox). In other words, a larger genome was not correlated with more biological complexity. For example, the genome of barley is 10 times larger than that of rice, a related grass. Since barley and rice are both members of the grass family, their gene content is expected to be similar and the difference in genome size between these two species is due to variation in the amount of non-genic DNA, i.e., various classes of repetitive DNA. This assertion was supported by recent comparative analysis of cereal genomes (San Miguel et al. 1996; Gale and Devos 1998; Wicker et al. 2001) which has shown large differences in the amount of repetitive elements among various grasses.

1.6 Transposable genetic elements

The first concrete study of transposable elements was conducted by the maize geneticist Barbara McClintock while studying the color of the aleurone layer of the endosperm in corn (McClintock 1950). The Ac/Ds DNA elements that were involved in phenotypic instability of corn were termed "jumping genes". As a result of this discovery the prevailing notion that genome was static, changing only on an evolutionary time scale was challenged and a new concept of genomic fluidity slowly gained momentum. With the rapid advances in molecular biology techniques during the 1970s and 1980s, transposable elements were found to be present in a wide rage of species (Flavell 1992: Kumar and Bennetzen 1999).

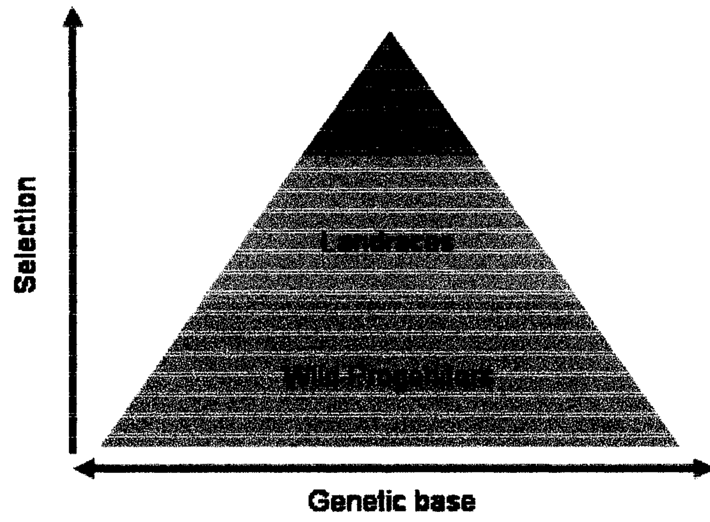
Transposition is defined as the movement of mobile genetic material – transposable genetic elements from one chromosomal location to another. There are two types of transposition: non-replicative transposition in which the element is excised from a genomic location and reinserted into another location and replicative transposition in which the element is copied, and one copy remains in the original site while the second copy transposes to a new location in the genome. There are two main classes of transposable elements, Type I and type II. Type II elements constitute DNA-mediated elements while type I transpose via an RNA intermediate and are hence named retroelements.

1.7 Retrotransposons

Retrotransposons are ubiquitous throughout all eukaryotic lineages (Flavell et al. 1992; Kumar and Bennetzen 1999) and are present in all plant species studied to date. Eucaryotic retrotransposons consist of the long terminal repeat (LTR) and the non-LTR elements. The LTR retrotransposons are further divided into the Ty1-copia and Ty3-gypsy group (Fig. 1.4) which differ from each other by their degree of sequence similarity and structural organization. It is a common belief that modern day retroviruses, LTR retrotransposons, and non-LTR retrotransposons all share a common ancestor (Flavell 1999). The basic difference between retroviruses and LTR-retrotransposons is the latter's lack of an envelope glycoprotein. The inability of retrotransposons to produce an envelope has reduced their life cycle to within the intracellular compartments of their host cells. An LTR-containing retrotransposon's mode of replication is via an RNA intermediate that is converted into extra-chromosomal DNA by encoded reverse transcriptase/RNaseH enzymes followed by integration into a genomic location mediated by an element-encoded integrase enzyme (Heslop-Harrison et al. 1997; Kumar 1996). This type of transposition can rapidly increase the copy number of the element and lead to a rapid expansion of the host genome (Kumar 1996; Kumar and Bennetzen 1999; Kidwell 2002).

Recently, two new classes of LTR retrotransposons have been identified in plants. The first group is terminal repeat retrotransposons in miniature (TRIM) (Witte et al. 2001). TRIMs are composed of 100 to 250 bp terminal direct repeats, which appear to be

Figure 1.3. Progressive reduction in the level of genetic diversity from wild progenitors to the present day cultivars. Narrowing of the genetic base is mainly due to selection pressure and the use of local and limited germplasm in breeding programs.


Selection
Genetic base

LTRs or LTR derivatives, and the priming motifs found internal to the LTRs in functionally intact retrotransposons and retroviruses. Aside from the priming site and a small intervening segment, the TRIMs appear to lack the rest of the internal domain that is required for their mobilization. The second group is large retrotransposon derivative (LARD) which was found in barley (Kalendar et al. 2004). Members of LARD are composed of two large LTRs flanking a variable internal domain with no apparent function as indicated by the lack of an open reading frame (ORF).

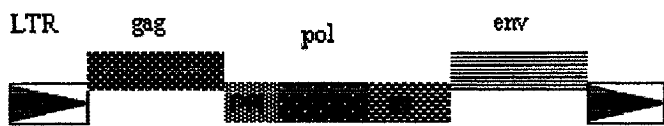
The non-LTR group of retrotransposons is represented by LINEs (long interspersed repetitive elements) and SINEs (short interspersed repetitive elements) which are also found in high copy number in plants (Leeton and Smyth 1993; Bennetzen 1996; Kumar et al. 1997; Kunze et al. 1997). The LINE elements are simpler than LTR-transposons in their structural organization. Studies on human and mouse LINE (L1) sequences show that L1 has two open reading frames. In L1, ORF1 encodes a protein with RNA binding capability and ORF2 encodes a protein with endonucleases (EN) and reverse transcriptase (RT) activity (Moran 1999; Ostertag and Kazazian 2001; Ergun et al. 2004) which code for reverse transcriptase and an RNA binding protein. Most LINE elements are truncated at the 5' region due to premature termination of reverse transcription and are therefore rendered inactive (Ostertag and Kazazian 2001). DNA sequence analysis suggests that LINE elements may be the oldest class of eucaryotic retrotransposons. Cladistic analysis showed that the first LTR-retrotransposons may have evolved from the acquisition of LTRs by a LINE (Xiong and Eickbush 1990). The SINE elements are different from other classes of retrotransposons in that they do not encode any *trans*-acting component required for transposition (Kumar and Bennetzen 1999). All known

SINES are derived from reverse transcription of RNA polymerase III transcripts. These elements appear to have acquired transposition capability using proteins encoded by LINES and/or LTR-transposons (Boeke 1997; Schmid 1998).

Research carried out within the past 20 years indicates that retrotransposons can expand genomes by increasing the amount of repetitive DNA sequences without affecting genome complexity, thus providing an explanation for the C-value paradox (Moffat 2000). In general, a large difference of 5,000-fold in haploid DNA content among angiosperms has been estimated (Petrov et al. 2000). *Arabidopsis* has the smallest known plant genome, consisting of 1.5×10^8 bp. Approximately 30 retrotransposon families have been identified in *Arabidopsis* genome each with one to three members (Hirochika et al. 2000). Among cereals, rice has the smallest genome comprising 4.3×10^8 bp which contain 10^3 copies of retrotransposons (Hirochika et al. 1992).

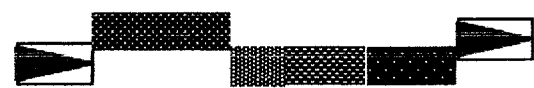
It is now believed that large differences in the genome size variation among angiosperms are mainly the result of polyploidization and transposition (SanMiguel et al. 1996; Wendel 2000; Devos et al. 2002). In grasses, which includes cereals such as wheat and barley comparative analyses have shown that most of the observed variation in genome size is due to repetitive DNA (Springer et al. 1994; Gale and Devos 1998; Heslop-Harrison, 2000). Genetic mapping of wheat, maize, rice and other grasses with common DNA probes has revealed a remarkable conservation of gene order and gene content over the 60 million years of radiation of the Poaceae. The linear organization of genes in nine different grass genomes differing in the nuclear DNA content of 400 to 6000 Mb was described in terms of only 25 rice linkage blocks (Reviewed by Gale and Devos 1998). However, despite major conservation of macro colinearity recent

Figure 1.4. Structural organization of retroelements and retroviruses. A major difference between retrotransposons and retroviruses is the lack of an envelope gene in the former. There are structural and sequence differences between Ty3/gypsy and Ty1-copia retrotransposons. Non-LTR retroposons lack LTR sequences.



Retrovirus (Retroviridae)

Ty3/gypsy (Metaviridae)



Ty1/copia (Psedoviridae)

LARDs

TRIMs

Non-LTR or retroposon

comparative studies indicate major breakdowns in micro colinearity between wheat and rice (Sorrells et al. 2000).

Smaller genome size among some plants may be a consequence of lack of transposition (Wright et al. 1996; Kumar and Bennetzen 1999). A small genome maybe helpful because it can replicate faster, resulting in a faster cell cycle and generation time (Moffat 2000). On the other hand, Kalendar et al (2000) showed that a large genome had its own advantages. The authors studied a population of wild barley in the Evolution Canyon in Israel and found that the plants grown at the rim of the canyon had a larger genome than the plants grown at the bottom. The fact that the plants at the top of the canyon had gained more copies of BARE-1 retroelement and lost fewer copies suggested that the element may confer some advantage.

Comparative analysis of maize and sorghum *adh* region has shown that 70 % of maize *adh* locus is occupied by about two dozen retrotransposons but that the same orthologous region in sorghum is devoid of these elements (Avramova et al. 1996; SanMiguel and Bennetzen 1998). Overall, the maize nuclear genome is 3-4 times larger than the sorghum genome. Given that the divergence time between maize and sorghum is about 15-20 million years, and both have 10 chromosome pairs, most of the increase in genome size between the two species was shown to be the result of increased level of retrotransposition in maize (Bennetzen et al. 1994; SanMiguel et al. 1996). On the other hand, retrotransposon-mediated genome expansion can be counteracted by the DNA loss through recombination between retrotransposon copies to generate solo LTRs and other types of deletions (Petrov 1977; SanMiguel et al. 1996; Chen et al. 1998; Vicient et al. 1999; Vitte and Panaud 2003) (Fig. 1.6). Analysis of a large contiguous sequence (66-kb)

in barley has shown evidence for genome shrinkage by intrachromosomal recombination between LTRs (Shirasu et al. 2000). Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) resulted in the identification of only 5 putative genes, with the remaining 70 % of the contig consisting of various types of retrotransposons (Wicker et al. 2001). The study also showed that major evolutionary processes such as deletions and duplications had taken place at that locus. Further evidence for DNA loss through recombination events involving retrotransposons came from the studies of the BARE-1 retroelement in wild barley, *H. spontaneum* (Vicient et al. 1999). In that study, the ratio of LTRs versus intact elements was greater than predicted ratio of 2 to 1. Given that each LTR retrotransposon is composed of two LTRs flanking an internal domain, a ratio of > 2:1 indicated that the internal region of retrotransposon has been lost as a result of recombination between LTRs. Solo LTRs have been found in many organisms including yeast (Hug and Feldman 1996; Goodwin and Poulter 2000), maize (SanMiguel et al. 1996), rice (Vitte and Panaud 2003) and barley (Vicient et al. 1999; Kalendar et al. 2004). Although the presence of multiple copies of retrotransposons dispersed throughout the genome could in theory trigger a dramatic and large scale loss of DNA including the host genes that are nested between retrotransposon sequences, via homologous recombination, in reality most such deletions have deleterious effects and reduce host fitness. Therefore, most deletions are confined to retrotransposon sequences but not the host DNA sequences (Bennetzen and Kellogg 1997; Kumar et al. 1997).

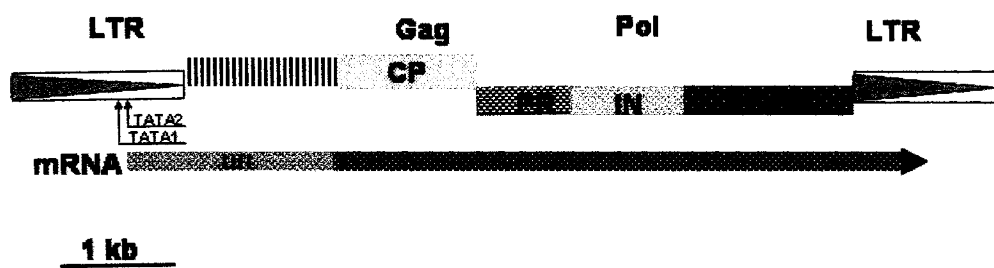
The first reported structure in barley resembling a retrotransposon LTR was made by Moore et al (1991). However, the first complete retrotransposon described for barley was BARE-1, discovered serendipitously (Manninen and Schulman 1993). It is an active 8.9

kb element, composed of two LTRs of 1.8 kb and encoding domains for GAG, AP, IN, RT and RH (Fig. 1.5). Active BARE-1 elements produce a single mRNA transcript which is then translated into a polyprotein in cytoplasm giving rise to trans-acting components that are required for retrotransposon life cycle (Suoniemi et al. 1996). The mRNA molecule also acts as a substrate for the reverse transcription leading to the synthesis of a cDNA copy with subsequent integration into a chromosomal location.

By their response to stress (Kalendar et al. 2000), invasiveness, and function as mobile promoters or transcriptional enhancers (Gribbon et al 1999; Suoniemi et al. 1996) retrotransposons cause many changes in the genomes of their host ranging from mutations to gene activation to transcriptional activities to pleiotropic effects. Retrotransposons can therefore serve as major contributors to genome diversification both within and between populations. Some examples of retrotransposon-mediated genomic mutations are listed in Table 1.1.

Recent development of rapid techniques for isolation of LTR retrotransposon sequences together with our ability to recover retrotransposon sequences from the vast amount of genomic data have extended the application of retrotransposon-based marker systems in plants (Pearce et al. 1999; Kumar and Hirochika 2001). Transposon-based markers have several advantages over other marker systems for the detection of sequence diversity. These include their ubiquitous distribution across species, their presence in high copy numbers within species, and their wide distribution across the genome. Furthermore, retrotransposon-related genome rearrangements lead to new insertion/deletions resulting in detectable polymorphism. These new insertions could be used to temporally order insertional events in a lineage and therefore help establish

Figure 1.5. Structural organization of BARE-1 retrotransposon. The element is flanked by 1.8 kb LTRs on both sides. LTRs contain initiation and termination signals for transcription. The polycistronic mRNA is transcribed from either P1 or P2 promoter corresponding to TATA1 and TATA2 respectively. Translation of the mRNA in cytoplasm results in a polyprotein product which is cleaved to provide all trans-acting components that are needed for retrotransposition.


LTR
Gag
Pol
LTR
CP
IN
TATA2
TATA1
mRNA
1 kb

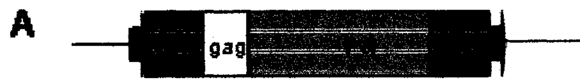
phylogenies. These properties of transposable elements have recently been used to study biodiversity in pea (Ellis et al. 1998), barley (Kalendar et al. 1999) and maize (Purugganan and Wessler 1994).

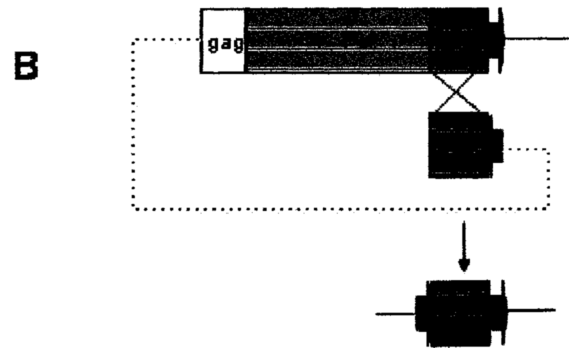
1.8 Retrotransposons and Host Genome evolution

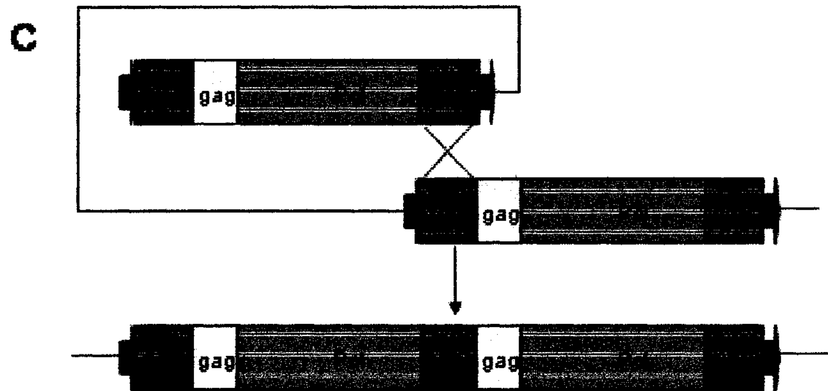
Despite a wide variation in copy number, distribution and type from one species to another, retrotransposons constitute a large fraction of many plants and animal genomes. Retrotransposons play a significant role as mutators in host genome evolution because of the broad spectrum of mutations produced by their activities (Kidwell and Lisch 1997; Kashkush et al. 2003). Evidence for the role of transposable elements in genome reorganization continues to accumulate. Chromosomal reorganization can occur with high frequency during meiosis in plant and animal genomes as a result of inter-chromosomal translocation between multiple copies of retrotransposons

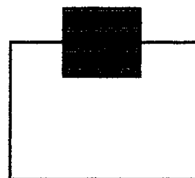
Ectopic recombination between retrotransposons of the same kind which are in a direct orientation (direct repeats) on the same chromosome can cause reciprocal duplications and deletions (Fig. 1.6). On the other hand, recombination between elements on the same chromosome which are in opposite orientation (inverted repeats) can cause a chromosomal inversion of the sequences between the two retrotransposons. Ectopic recombination between two elements on different chromosomes can cause reciprocal translocations. These types of retrotransposon-mediated genome rearrangements have been extensively documented in yeast and Drosophila (reviewed by Kumar and Bennetzen 1999). However, the extent of these mutations and the underlying mechanism

Figure 1.6. Genome contraction as a result of retrotransposon-mediated recombination. A) structural features of an intact retrotransposon, B) recombination mediated by LTRs of the same retrotransposon result in the loss of internal domain flanked by the LTRs. C) recombination between two different copies of retrotransposons results in the loss of DNA between the two elements.


A
gag


B
gag


C
gag
gag
gag
gag



of such retrotransposon-mediated chromosomal rearrangements have not yet been studied in plants.

1.9 Regulation of Transposition

The evolution of control mechanisms for the regulation of transposition is crucial for countering the deleterious effects of transposition. Transcriptional regulation is the simplest regulatory mechanism. Many retrotransposons show unique patterns of developmental/environmental regulation (Suoniemi et al. 1996; Turcich et al. 1996, Kumar and Bennetzen 1999). Tissue specific transcription of tobacco *Tnt*1, barley BARE-1 and maize PREM-2 have been previously documented (Pouteau et al. 1991; Suoniemi et al. 1996; Turcich et al. 1996). A direct correlation between transcription and transposition of retrotransposons was demonstrated for tobacco *Tto*1 and rice *Tos*17 retroelements (Hirochika 1993; Hirochika et al. 1996). Transposition of both *Tto*1 and *Tos*17 increased with an increase in the levels of their RNAs, suggesting that the transposition of these elements may be regulated at the transcriptional level. On the other hand, the BARE-1 element was found to be highly transcribed in leaves where no transposition was observed (Suoniemi et al. 1996). The latter finding suggests that although transcription is a prerequisite, other forms of regulation may control transposition. Since the replication cycle of LTR retrotransposons includes four steps, i.e. transcription, translation, reverse transcription, and integration of the cDNA, regulation at any of these steps can limit the rate of transposition (Feschotte et al. 2002). In a study of transposition in yeast for every 14,000 Ty-1 transcripts only one Ty-1 cDNA copy was made (Curcio and Garfinkel 1999). In rice, transcription of Karma, a LINE-type retrotransposon did not result in retrotransposition (Komatsu et al. 2003). These findings

indicated that regulation at translation or reverse transcription or integration steps are important in the control of transposition. The limited data available on the regulation of plant LTR retrotransposons indicates that transposition is regulated at the level of transcription initiation (Melayah et al. 2001). The genomic copy number of rice *Tos*17 increased from 2 to 30 copies in some strains following stress-induced transcription as revealed by Southern analysis (Hirochika et al. 1996). In a similar study by Melayah et al (2001) using "transposon display" (equivalent to sequence-specific amplification polymorphism, or S-SAP), 38 out of 41 new DNA bands corresponded to *de novo* insertion of *Tnt*1. Recent studies also point to the role of methylation in the regulation of transposition. It was previously assumed that both nucleosomal chromatin formation and cytosine methylation of DNA evolved to regulate host gene expression. However, new evidence indicates that the primary role of DNA methylation is to regulate transposon activity (Kidwell and Lisch 2000). The role of DNA methylation in the regulation of transposition was demonstrated in inter-specific marsupial hybrids where it was found that demethylation was associated with retroelement activation and amplification (O'Neill et al. 1998). In a similar study involving the production of rice lines introgressed with genetic material from wild relatives extensive genomic DNA methylation changes occurred in these lines (Liu et al. 1999) suggesting that DNA methylation may be used as a defence mechanism against invading foreign DNA. These regulatory pathways link transposable element activity to homology-dependent gene silencing. Homology-dependent gene silencing has been associated with plant transgenes and there is recent evidence for the presence of the same phenomenon in mammals. On the other hand in the Drosophila and yeast genomes which generally lack cytosine methylation of DNA an

inactive chromatin state can be established without the need for DNA methylation (Dorer and Henikoff 1994). An even more extreme case involves the chordate sea squirt where most of the functional genes are hypermethylated and the transposable elements are hypomethylated (Simmen et al. 1999). These studies show that despite the crucial role of DNA methylation for genetic silencing, it is not obligatory for this process.

1.10 Interaction between the host and retrotransposons

Genomes with their complex structure could be viewed as an ecosystem in which the ability to replicate is necessary for the survival of all residing sequences. The ability of retrotransposons to increase their copy number within a genome is a selectable attribute required for their survival. Experiments with yeast and bacteria have shown that populations with functioning transposable elements generally out-compete populations lacking such elements (Wilke and Adams 1992). In a chemostat competition experiment with *E. coli*, Chao et al (1983) showed that the transposon *Tn10* was able to increase the growth rate of *E. coli* by increasing the mutational rate of the host genome. High copy number transposable elements may manage to avoid much of the detrimental effects on their host genome by transposing into non-active parts of genome or within the existing elements (nested transposition). Previous studies have shown that grass genomes are largely syntenic in which the order of mapped genes is conserved against a great variation in genome size. Grass genes appear to occur in clusters – gene islands separated by repetitive DNA (Barakat et al 1997; Panstruga et al. 1998). Much of this repetitive DNA in barley and maize is composed of retrotransposons or their solo LTRs (Flavell et al 1997; Barakat et al. 1997).

It seems appropriate to view the genome dynamics both from the perspective of the host and the elements themselves. Although there is mounting evidence that transposition may lead to a modified pattern of gene expression, little is known about the dynamics of transposable element copy-number evolution within and between populations, or the significance of this process with respect to natural selection. Therefore, any possible connection between genome size variation and adaptive genic evolution remains unresolved. Activation of retrotransposons in response to various biotic and abiotic factors such as tissue culture (Horochika et al. 1996), climatic changes (Kalendar et al. 2000) or pathogen attack (Grandbastien et al. 1997) indicate that they may play a major role in the evolution of their host by contributing to host genome ability to adapt to the new environment.

Retrotransposons can rearrange genomes with or without transposition competency. Bs1 retrotransposon of maize has acquired a part of another gene and subsequently amplified it throughout the allotetraploid genome of maize (Bureau et al. 1994). Additionally, the action of reverse transcriptase on normal cellular mRNAs has created intronless pseudogenes in plants (Drouin and Dover 1987; Loguercio and Wilkens 1998).

These amplified genomic loci can enhance the evolution of other genomic sequences. Clustering of large numbers of retrotransposons in centromeric, telomeric, and other heterochromatic regions of eukaryotic organisms suggests that these elements may have acquired some structural/functional role in these regions of genome. Studies have shown that transposition of the HeT-A and TART retrotransposons of *Drosophila melanogaster* may have a role in buffering the shortening of chromosome ends analogous

to the function of telomerase (Pardue et al. 1997) by transposing themselves into the telomeric regions of *Drosophila* chromosomes and therefore capping their end points.

1.11 Maintenance of retrotransposons in genomes

Retrotransposons share many similarities with retroviruses, but they are not by themselves infectious. The basic difference between a retrotransposon and a retrovirus is the retrotransposon's lack of an envelope protein. This critical difference inhibits the formation of an extracellular infectious viral particle and reduces retrotransposons to intracellular entities (Garfinkel et al. 1989). Although there is evidence that LTR retrotransposons are capable of forming viral-like particles (VLPs) (Suoniemi et al. 1997), there is no experimental evidence that supports the possibility of VLP movement between adjacent cells. Phylogenetic analysis based on the sequences of reverse transcriptase and integrase genes from both Ty1-*copia* and Ty3-*gypsy* retrotransposons in plants has provided evidence for vertical transmission of retrotransposons (Falvell et al. 1992; Varagona et al. 1992; Matsuoka and Tsunewaki 1999). Even though there is no experimental evidence that supports horizontal transfer of retroelements in plants a hypothesis of horizontal gene transmission (HGT) also exists (Kumar 1998; Jordan1999). Proponents of this hypothesis argue that divergent lineages of plants often have members of the same retrotransposon family. For instance, plants belonging to Cruciferae family such as *Brassica* sp. and Gramminea (grasses) share the same family of retrotransposons (Flavell 1999). Since the divergence time between Crucifereae and Gramminea is estimated to be 200 million years ago, vertical transmission cannot explain the degree of sequence similarity between their retrotransposon elements. Some evidence to support this hypothesis was provided by the finding of envelope-gene-like sequences in insect

copia retrotransposon (Song et al. 1994) and experimental evidence for the presence of an identical Ty1-*copia* retrotransposon in two diverse species of Drosophila (Jordan et al. 1999). Although, the mode of this horizontal transfer is unknown since the *copia* element lacks the envelope gene and therefore lacks the ability for inter-cellular movement, it has been proposed that retrotransposons could be passively transmitted by retroviruses or by other plant pathogens (Kumar 1998; Bennetzen 1999). For a horizontally transmitted retrotransposon to be able to propagate in a population it must first be transmitted into the germline cells. Given the small ratio of these cells to other somatic cell types in plants the possibility of such transfer is rare. However, if a successful transmission occurs and the retrotransposon is able to amplify to thousands of copies per nucleus, it will have tremendous effects on the host genome. In plants, one possible mode of horizontal transfer would be through wide crosses or interogressive hybridization. Many plant species use non-specific gametic interchange processes, but in rare cases such crosses between different plant species give rise to an allopolyploid which brings two retrotransposon populations into one cross (Leitch and Bennet, 1997). In some exceptional cases, the two genomes of very different species (e.g. maize and oat) may be able to co-exist in the same nucleus for a few plant generations (Ananiev et al. 1997). However, this transient hybrid condition is unstable and eventually, one parental set of chromosomes maybe eliminated. The transfer of retrotransposons from one genome to another may occur in the process.

1.12 Retrotransposons as DNA markers for genetic studies

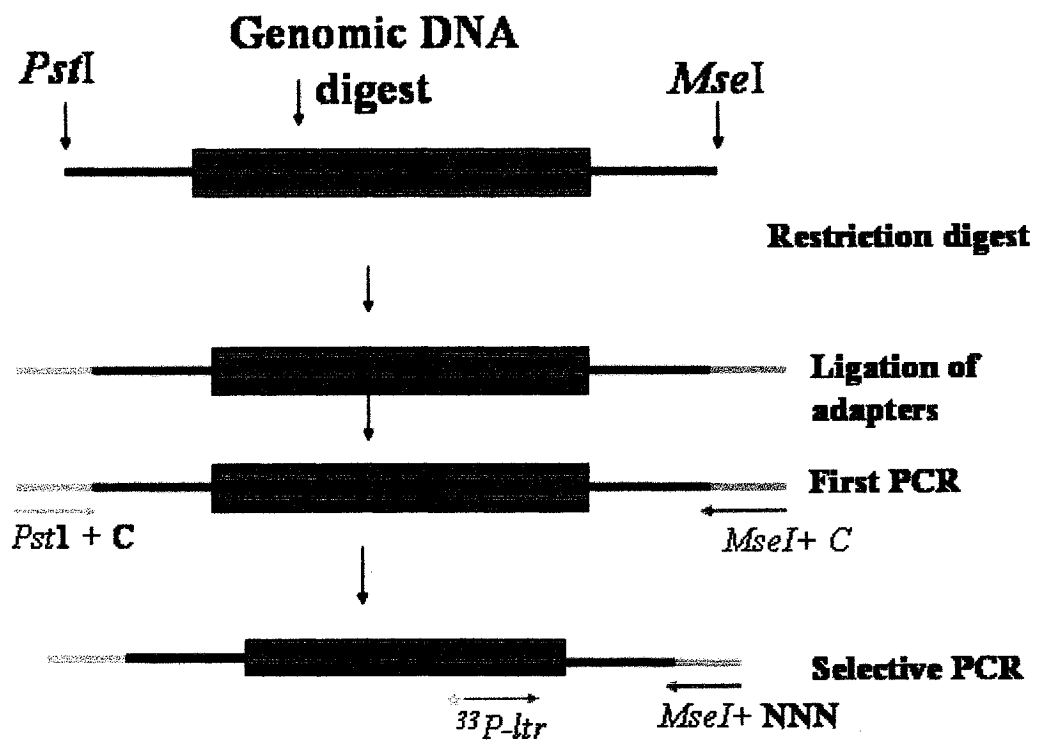
One of the first DNA-based marker systems was restriction fragment length polymorphism (RFLP), a hybridization-based method that was initially used for mapping studies in human (Botstein et al. 1980) and later for similar studies in plants ((Burr et al. 1983). The advent of PCR (Saiki et al. 1988) gave rise to a new generation of markers based on differential amplification of DNA. Currently, several different types of DNA marker systems are available for the study of plant genome evolution and cultivar fingerprinting including; amplified fragment length polymorphism (AFLPs) (Vos et al. 1995; Barrett et al. 1998; Qi and Lindhout 1997; Reamon-Buttner et al. 2000; Soleimani et al. 2002a), random amplified polymorphic DNA (RAPDs) (Williams et al. 1990; Thorman and Osborn, 1992; Tinker et al. 1993), simple sequence repeat (SSRs) (Tautz and Renz 1984; Akkaya et al. 1992), inter-retrotransposon amplified polymorphism (IRAP) (Kalendar et al. 1997), retrotransposon-microsatellite amplified polymorphism (REMAP) (Kalendar et al. 1997) sequence tagged site (STSs) (Olsen et al. 1989) and single nucleotide polymorphism (SNPs). Each system differs in many respects, such as the distribution of the detectable polymorphisms throughout the genome, the number of detected loci per assay (multiplex ratio), dominance (differentiating heterozygous from homozygous genotypes), ease of automation, reproducibility and the costs associated with the marker system. More detailed descriptions of various classes of molecular markers and their utility in the study of genomes can be found at the following URL: http://www.nal.usda.gov/pgdic/tutorial/lesson4.htm. The main applications of molecular markers in cereal genetics include: 1) Assessment of genetic diversity and characterization of germplasm; 2) Marker assisted selection; 3) Identification of cultivars

and genotypes; 4) Study of monogenic and quantitative trait loci (QTL); 5) Gene identification and isolation; 6) Construction of genetic maps.

The most commonly used transposon-based marker system is sequence specific amplified polymorphism (S-SAP) (Fig. 1.7) (Purugganan and Wessler 1995; Waugh et al. 1997; Flavell et al. 1998). Studies have shown that S-SAP-based markers appear to estimate phylogenetic relationships better than AFLPs and other dominant multiplex systems (Ellis et al. 1998). A multi-transposon approach has been applied to estimate phylogenetic relationships within and between species of legumes (Pearce et al. 1990) and cereals (Gribbon et al. 1999; Leigh et al. 2003). This strategy is important because each transposable element has a unique history of transpositional activity. For instance, retrotransposons that have been active in the recent past should be ideal for genetic studies within species and between closely related species. On the other hand those elements that have only been active in the distant past may be used for studies involving distant species or even genera.

Plant retrotransposons have unique features that make them ideal genetic tools for the analysis of genetic diversity and evolution: 1) Retrotransposon mediated insertions are stable because they transpose by a replicative mode. This ensures that the original copy stays in a fixed location while the newly made copy inserts into a new genomic location. 2) Retrotransposons target sites are unlinked with the site of the original copy, making it easy to generate a large number of random insertions for saturation mutagenesis. 3) They are highly mutagenic and many transpose into gene-rich regions (Reviewed by Kumar and Hirochika 2001), potentially affecting gene expression pattern.

Figure 1.7. A schematic outline of S-SAP procedure. Genomic DNA is cut with restriction enzymes followed by ligation of synthetic dsDNA adapters. The product of ligation reaction is used as template for the first PCR amplification using primers complementary to the sequence of the adapters plus a single additional nucleotide at the 3' end of each primer. The resulting amplification product is then used as template for the final PCR product using a primer that is complementary to either of the two adapter sequence which includes three selective nucleotides at its 3' end in combination with a primer designed from the sequence of a transposable element. The S-SAP procedure is identical to AFLP except for the last PCR amplification step.


Genomic DNA
digest
PstI
MseI
Restriction digest
Ligation of
adapters
First PCR
Pst1 + C
MseI+ C
Selective PCR
33P-ltr
MseI+ NNN

These types of markers can be useful for marker assisted selection (MAS) in breeding programs and for the study genetic relationships in plants.

1.13 Single nucleotide polymorphism (SNP)

SNPs are nucleotide variations among individuals in a population and represent the most abundant types of molecular markers. Only a small fraction of SNPs are within the coding regions. Of the SNPs in protein-coding genes, few change the amino acid sequence of the protein. In human, more than 99% of SNPs are not associated with phenotypic changes (Paabo 2001). SNPs within the protein-coding regions are of two types. The degeneracy of the genetic code means that some SNPs are synonymous, e.g., a change from CCA to CCG does not result in the alteration of an amino acid sequence. Synonymous substitutions occur with time as a result of random mutation and are not under selective pressure. Non-synonymous substitutions do alter the amino acid sequence and are therefore subject to selective pressure. Those non-synonymous substitutions that survive the selective pressure are the foundation of adaptive evolution (Liberles et al. 2001). SNPs seem to comprise the largest class of functional polymorphism, i.e., those producing phenotypic effects. This marker system opens the way to the development of ultra-high density maps. The data for maize indicates the presence of one SNP for every 104 bp between randomly selected genomes (Tenaillon et al. 2001). The estimated frequency of SNPs in barley was one per 300 bp based on the sequence analysis of set of 48 genomic loci representing 23 kb of barley genomic DNA (Table 6.2), thus in theory markers could be developed for as many as 15 million potential SNPs in the barley genome.

Basic strategies for SNP detection include base extension using dideoxynucleotide chain termination (mini-sequencing), hybridization (allele-specific oligonucleotide probes), ligation (allele-specific ligation probes), PCR amplification (allele-specific primers), and cleavage (restriction enzyme and invader assay). In human, SNP markers have been used for association studies in which a single nucleotide polymorphism in one gene is responsible for the manifestation of a disease phenotype (Paabo 2001). These studies were aimed at the production of genotype specific drugs (pharmacogenetics) for the treatment of inherited diseases. SNPs applications in cultivated crops such as barley include variety identification for plant property right protection, cultivar purity testing, marker assisted selection, and genetic diversity estimation.

1.14 Measuring genetic diversity

Genetic diversity is the sum of all genetic variation – both cryptic and manifest that is present in a population. Traditionally, genetic diversity measurements were carried out using phenotypic (morphological) markers. However, since a plant's genotype is distinct from its phenotype, the relationship between genetic variation and phenotypic traits observed in plants are difficult to discern. The environment in which a plant grows can influence the expression of its genes. Many agronomically important traits are polygenic in nature. Examples of these traits include grain quality, yield and pest resistance (Rogers et al. 1983; Lindhout 2002). Furthermore, morphological markers are limited in numbers and can mostly be used only at the whole plant level.

Pedigree analysis has also been used to assess genetic diversity in crop species with known genealogy (Cox and Murphy 1990; Tinker et al. 1993; van Beuningen and Busch

1997; Soleimani et al. 2002a). Assumptions that are made in the calculation of genetic diversity from pedigree data – that both parents make equal contribution to the progeny, that there is no linkage and that cultivars with unknown pedigree are unrelated (Cox and Murphy 1990; Tinker et al. 1993) – can introduce a significant bias in the measurements. For example, an RFLP analysis of 55 wheat landraces from Iran, Afghanistan and Turkey which were unrelated by pedigree showed a high level of genetic similarity (mean = 0.91) between accessions (Kim and Ward 1997).

The introduction of DNA-based markers such as such as RFLPs, SSRs, AFLPs etc. coupled with DNA sequence data from various genome sequencing projects has revolutionized the study of plants biodiversity and biosystematics over the past two decades. In this age when classical breeding has been pushed to its limit, molecular biology offers new tools for manipulation and exploitation of largely hidden genetic resources for plant improvement.

Molecular markers have allowed detailed study of species relationships. One of the earliest such studies revealed striking similarities in genome organization between tomato (*Lycopersicon esculentum*) and potato (*Solanum tuberosum*) (Bonierbale et al. 1988). In recent years many different types of marker systems have been developed and applied to the study of various crop species (Saghai Maroof et al. 1994; Tohme et al. 1996; Van Toal et al. 1997; Waugh et al. 1997; Barrett et al. 1998; Soleimani et al. 2003). These studies have been very useful in estimating genetic diversity within and between species and also addressing the evolutionary relationships between them. Understanding the level of genetic diversity in crops, its origin and the evolutionary processes involved in the acquisition of that diversity are important for the improvement and efficient utilization of

genetic resources. There is a general consensus that domestication and breeding have contributed to a progressive narrowing of the genetic base in many crop species relative to their wild progenitors. Genetic analysis with molecular techniques such as S-SAP and SNP can lead to a great increase in our knowledge of cereal genetics. The abundance of these markers and their wide distribution in the genomes can provide realistic measures of the amount of diversity present in the populations.

1.15 Rational for the present study

In the absence of whole genome sequencing, molecular markers are indispensable tools for the study of genomes. Although the completion of the rice genome sequencing project has provided a first glimpse into the structure and function of a cereal genome, such large scale sequence information from other cereal species with larger genomes such as barley is limited. To date, the largest sequences that are publicly accessible from barley are the 60 and 66 kb intervals from *mlo* and *rar*1 loci respectively (Panstruga et al. 1998; Shirasu et al. 2000) and a recent 300 kb contiguous region at the *Ha* locus (Caldwell et al. 2004). These studies have shown that transposable elements and specifically retrotransposons play a major role in the composition and restructuring of barley genome. Parallel to these studies, recent sequencing projects involving cDNA libraries from Triticeae have yielded a large amount of sequencing data in the form of ESTs (http://wheat.pw.usda.gov/genome). The EST databases provide enormous amount of information not only about the makeup of genomes in terms of the number of genes and their expression profile, but can also serve as a large reservoir of allelic variation in organisms which can in turn be used for characterization of genetic resources, studies of

plant genetics and their ultimate improvement by breeding. Recent studies in grasses have shown that retrotransposon sequences occur frequently within *Triticeae* EST databases (Echenique et al. 2002), providing further evidence that grass lineage has active transposition system(s).

In this study, two independent molecular marker systems, S-SAP and SNP were used to assess genetic diversity in the gene pool of Canadian barley cultivars. In total, 103 cultivars belonging to various morphological/agronomical groups were analyzed. Genetic diversity levels were estimated both within and between cultivars. Although S-SAP methodology is well established for barley research, this is the first comprehensive study of S-SAP in cultivated barley that involves a large sample of plants and looks at analyzing genetic diversity at the cultivar level. A previous study had quantified BARE-1 elements among a group of barley cultivars and samples of *Hordeum spontaneum* that were collected from different eco-geographic regions in the centre of its origin (Israel) (Vicient et al. 1999; Kalendar et al. 2000). In this thesis we have used real-time PCR to determine variation in BARE-1 copy numbers and infer the contribution of the retrotransposon to dynamic micro evolutionary processes that are the by product of directional selection in barley breeding programs at the cultivar level. Variation in BARE-1 copy number reflects the differential transpositional activity of the element during breeding and selection, contributing to variation in genome size among closely related cultivars.

Diagnostic SNP markers were used to generate identification keys to distinguish Canadian barley cultivars as well as their biotypes. This portion of data is part of a Trade

Secret and is being evaluated by Agriculture and Agri-Food Canada for commercialization.

The thesis is an attempt to understand the extent of genome evolution and micro-evolutionary processes in response to breeding and selection in barley breeding programs. It examines the role and the contribution of the retrotransposon BARE-1 to the extent of genetic variation in the cultivated barley. Comparative analysis using S-SAP, SNP and kinship coefficients provide important information on the underlying evolutionary mechanisms of each marker system and their utility for the study of genetic diversity.

Results presented in the thesis are in the form of five papers that are either published, in press or submitted with the exception of chapter six which cannot be published at the present time due to the Intellectual Property Rights. Each chapter represents an individual paper unchanged from its form at submission. Thus, there may be some repetition in the Materials and Methods section. The last chapter of the thesis is a general discussion that provides an overall analysis and interpretation of the combined results in the thesis.

1.16 HYPOTHESES AND OBJECTIVES

Hypothesis 1

Selection pressure and associated breeding practices coupled with transpositional activity of BARE-1 provide mechanisms for rapid genome evolution in barley cultivars.

Objective 1.1

Determine genetic similarity/diversity estimate between barley cultivars using BARE-1 derived markers (S-SAP) method. Analyze patterns and the extent of genetic similarity/distance using retrotransposon-based markers by population genetic analysis tools. Obtain pedigree information on barley cultivar and calculate the extent of allele sharing between cultivars using kinship coefficient tests.

Objective 1.2

Develop protocols for identification and validation of single SNPs. Obtain informative SNPs (i.e. those with common allele in the population). Screen cultivars for SNP polymorphisms. Perform comparative analyses between genetic similarity estimates derived from S-SAP, SNP and pedigree analysis.

Hypothesis 2

Barley cultivars are not homogeneous and uniform. There is a substantial amount of intra-cultivar variation (biotypes) in barley cultivars despite selection pressure aimed at cultivar purity and homogeneity.

Objective 2

Analyze multiple plants per cultivar using S-SAP and SNP markers. Use population genetic analysis tools such as cluster analysis and Analysis of Molecular Variance

(AMOVA) to infer the variance components for within and between cultivars. Determine the source of observed variation.

Hypothesis 3

There is significant variation in BARE-1 copy number among barley cultivar indicating a major role for this retrotransposon in the genome evolution of cultivated barley.

Objective 3

Determine the copy number of full length as well as truncated retrotransposons (solo LTRs) using Real-time PCR. Determine BARE-1 copy number within and between barley cultivars. Determine the ratio of intact versus truncated retrotransposon. Estimate the amount of DNA loss through homologous recombination.

Hypothesis 4

Barley cultivars and their biotypes can be uniquely identified by single nucleotide polymorphism (SNP) markers.

Objective 4

Use SNP markers to distinguish between cultivars and their biotypes using a key generating program (DELTA).

Table 1.1. Examples of plant transposable genetic elements, their genomic distribution and mobility.

Retrotransposon	Host	Genomic distribution	Evidence for recent mobility	Reference
Ty1/copia type			Autonomous	
BARE-1	Barley	Dispersed	Insertion site polymorphism	Kalendar et al 1999; Waugh et al 1997
B5	Maize		Transposition into the waxy gene	Varagona et al 1992
Opie-1	Maize	Dispersed intergenic	Active within the last 2 to 6 million years	SanMiguel et al 1996
PREM-2	Maize	Dispersed	Copy with identical LTRs	Turcich et al 1996
Prt	Potato		None	Pearce et al 1996
SIRE-1	Soybean		None	Bi and Laten 1996
Tnt1	Tobacco	Dispersed euchromatin	Transposition into the *nia* gene in protoplast culture	Grandbastien et al 1997
Tos17	Rice	Dispersed	Transposition into coding sequences (tissue culture)	Hirochika et al 1996
Wis-2	Wheat		Polymorphism in regenerated plants (tissue culture)	Moore et al 1991
Ty3/gypsy type			Autonomous	
Athila	Arabidopsis	Paracentromeric	None	Pelissier et al 1996

Retrotransposon	Host	Genomic distribution	Evidence for recent mobility	Reference
		region		
Cinful	Maize		None	Avramova et al 1995 SanMiguel et al 1996
Zeon-1	Maize		Transposition into the *zein* A gene in somatic tissue	Hu et al 1995
Grande-1	Maize	Dispersed	None	Martinez-Izquierdo et al 1997
Gret1	Grape vine		Transposition into *Myb*-related gene	Kobayashi et al 2004
LARDs	Widespread		None (non-autonomous)	Kalendar et al 2004
TRIMs	widespread		None (Non-autonomous)	Witte et al 2001
LINEs				
Cin4	Maize		None	Schwarz-Sommers et al 1987
Del2	Lily	Dispersed	None	Leeton and Smyth 1993
Tal1	Arabidopsis	Dispersed	None	Wright et al 1996
SINEs			Non-autonomous	
p-SINE1	Rice		None	Iwamoto et al 1999

Retrotransposon	Host	Genomic distribution	Evidence for recent mobility	Reference
TS	Tobacco	Dispersed	None	Yoshioka et al 1993
Ts	Bell Pepper	Dispersed	Transposed within the PAP2 gene	Pozueta-Romero et al 1998

CHAPTER TWO

Genetic diversity among barley cultivars assessed by sequence-specific amplification polymorphism

The following chapter is a reproduction of a paper published in the Journal of Theoretical and Applied Genetics by Vahab D. Soleimani, Bernard R. Baum and Douglas A. Johnson 2005 Theor. Appl. Genet. 110:1290-1300. The chapter fulfills objective 1.1 of the thesis. Retrotransposon-based primers were designed from the LTR region of BARE-1 retroelement of barley for S-SAP analysis to reveal DNA sequence polymorphisms resulting from the BARE-1retrotransposition. 103 barley cultivars representing 824 individual plants (8 plants per cultivars) were screened with four primer pairs. Polymorphic S-SAP markers were used to generate GS values between groups (i.e. two-rowed versus six-rowed, malting versus feed, winter versus spring and hulless versus hulled), between cultivars within each group, and within cultivars. Population genetic tools were used to partition variance among different hierarchical levels and delineate groups based on preset agronomical traits.

2.1 Abstract

We analyzed the genetic structure and relationships among barley cultivars (*Hordeum vulgare* L.) with Sequence-Specific Amplification Polymorphisms (S-SAPs). Polymorphisms were identified in 824 individual barley plants representing 103 cultivars (eight plants per cultivar) widely grown in Canada and the United States, using PCR primers designed from the long terminal repeat (LTR) of the barley retrotransposon BARE-1 and a subset of four selective *Mse*I primers. From the 404 bands scored, 150 were polymorphic either within or between cultivars. Genetic structure assessed with Analysis of Molecular Variance attributed the largest component of variation to the within groups of cultivars (69 to 86%). Within-cultivar genetic variation was estimated as average genetic diversity over loci and ranged from 0 (completely homogenous) to 0.076 (most heterogeneous cultivar). Only 17 out of 103 cultivars (16%) were judged to be homogenous by this criterion. Relationships among cultivars were analyzed by cluster analysis using unweighted pair-groups using arithmetic averages and found groups similar to those determined by agriculturally significant phenotypic traits such as spike morphology (two-rowed or six-rowed), cultivar type (malting or feed), seed characteristic (hull-less or hulled), and growth habit (winter or spring) with minor overlaps. Discriminant Analysis of groups determined by these phenotypic traits fully supported the different groups with minor overlaps between the malting/feed. S-SAP markers generated from retrotransposons such as BARE-1 are invaluable tools for the study of genetic diversity in organisms with a narrow genetic base such as barley. In this study, S-SAP analysis revealed significant amounts of cryptic variation in closely related cultivars including somaclonal variation, which could not be inferred by the pedigree analysis.

Key words Barley, genetic diversity, molecular markers, polymorphism, S-SAPs, retrotransposons

2.2 Introduction

Assessment of the extent and nature of genetic variation in crop species has important implications in breeding, plant improvement, and conservation of plant genetic resources. In barley breeding programs, emphasis was placed on developing cultivars for specific end use, i.e., malting and brewing; thus, over time the genetic base of barley has become narrower compared to that of other crops (Martin et al. 1991).

The aim of breeding new barley cultivars is to produce new allele combinations. Heritable variation is created by controlled crosses in breeding programs between adapted high-yielding cultivars and breeding lines. Cultivar breeding is a directional selection and mainly based on elite germplasm, although specific germplasm may be incorporated by introgression from wild relatives and landraces in backcrossing programs (Nevo 1992). Selection for desirable traits is made in the field or greenhouse. These traits may include lodging resistance, disease resistance, yield, straw strength, and protein content of the grain. In addition, malting properties including extract yield, viscosity of grain, and malt may be selected. Selection is applied at the F_2 generation for highly heritable traits and at F_3 or F_4 for most traits in a pedigree-type breeding program. The early generations are heterozygous, which makes selection difficult, but an acceptable level of homozygosity is achieved following the sixth or seventh generation of selfing. The raw material in any breeding program is the amount of genetic diversity available in the gene pool that can be subsequently used for selection.

Several different types of DNA marker systems are available for measurements of genetic diversity in plants including, amplified restriction fragment length polymorphism (AFLP) (Vos et al. 1995; Barrett et al. 1998; Qi and Lindhout 1997; Reamon-Buttner

and Jung. 2000; Soleimani et al. 2002), random amplified polymorphic DNA (Williams et al. 1990;Thormann and Osborn, 1993; Tinker and Mather. 1993), Simple Sequence Repeat (Tautz and Renz 1984; Akkaya et al. 1992), and retrotransposon-based marker systems including sequence-specific amplification polymorphism (S-SAP) (Waugh et al. 1997), inter-retrotransposon amplified polymorphism, and retrotransposon-microsatellite amplified polymorphism (Kalendar et al. 1999). Each of these marker systems differs in many respects, such as the distribution of target sequences throughout the genome, number of detected bands per assay (multiplex ratio), level of detected polymorphism, dominance/co-dominance nature (differentiating heterozygous from homozygous genotypes), ease of automation, assay reproducibility, and the cost associated with the marker system.

Retrotransposon-based marker systems rely on two principles. First, integration of a retrotransposon in a genomic location is fixed and behaves in a Mendelian fashion, and second, multiple integrations of the element in various genomic locations provide substrates of known DNA sequence for various PCR-based screening assays.

In barley, S-SAP markers (Waugh et al. 1997; Gribbon et al. 1999) were developed from BARE-1 element. BARE-1 is a member of the Ty1-*copia* group of retroelements (Manninen and Schulman 1993). An average of 14 x 10^3 retrotransposons belonging to the BARE-1 family was found in the genomes of *Hordeum* species (Vicient et al. 1999). Members of this family were found to be transcriptionally and translationally active, encoding a polyprotein and processing signals (Suoniemi et al. 1999; Jaaskelainen et al. 1999; Suoniemi et al. 1996). Active transposition of this element, coupled with its wide

distribution throughout the barley genome, makes it a good candidate as potential source of molecular markers for the study of the structure and evolution of barley genome.

S-SAP markers have been used for linkage and genetic diversity analysis in wheat (Queen et al. 2004), grapevine (Labra et al. 2004), pea (Ellis et al. 1998), oats (Yu and Wise 2000), maize (Casa et al. 2000), and *Medicago* (Porceddu et al. 2002). The ubiquitous presence of retrotransposons (Flavell et al. 1992; Hirochika et al. 1992; Voytas et al. 1992; Suoniemi et al. 1998; Noma et al. 1999) and their wide genomic distribution makes S-SAPs valuable tools for the study of genetic diversity in plants (Waugh et al. 1997; Labra et al. 2004).

The aims of this study were to assess genetic diversity levels within and between barley cultivars and to infer the utility of the S-SAP method for investigation of genetic diversity among a large group of barley cultivars with various agronomic characteristics. The information obtained could be used for future breeding purposes and to further improve barley cultivars.

Table 2.1 Material used in sequence-specific amplification polymorphisms (S-SAP) analysis of Canadian barley cultivars. R Row number, G growth habit (s spring, w winter), S seed characteristic (h hull-less, d hulled), U use (f feed, m malting) NH number of haplotypes, NP number of polymorphisms within cultivar, NPD number of pair-wise differences, $\hat{H}$ gene diversity over loci, SD: standard deviation of $\hat{H}$, / primary cross, // secondary cross, number preceding, * number of backcrosses.

	Cultivar Name	R	G	S	U	Parentage	NH	NP	NPD	$\hat{H}$	SD of $\hat{H}$	Released
1	Ac Albright	6	s	d	f	Otra/6/(Weibull1514-64) Morgenrot/5/Tammi/4/Maja/3/Opal//Hanna/Svanhals	3	2	0.5000	0.0030	±0.0038	1992
2	Ac Alma	6	s	d	f	Chapais/Leger	4	4	1.8210	0.0131	±0.0096	1996
3	Ac Bacon	6	s	h	f	Tupper/Johnston//Conquest/3/Abee/4/Ellice/Bedford	7	14	6.4640	0.0598	±0.0361	1998
4	AC Bountiful	2	s	d	m	Wpg843-234/Manley//AC Oxbow/Manley	2	1	0.5714	0.0041	±0.0041	1999
5	Ac Burman	6	s	d	f	Leger/Bruce/2/2*Leger	2	1	0.2500	0.0018	±0.0025	1991
6	Ac Hamilton	6	s	d	f	(Trent/Vanier)/(York//CI 10853/Parkland)/Perth	8	8	3.7140	0.0269	±0.0172	1994
7	Ac Harper	6	s	d	f	49-125/BT364/6* Galt//BT201/6* Galt/3/BT364	2	1	0.2500	0.0018	±0.0025	1996
8	AC Hawkeye	6	s	h	f	GO1-1/Tupper/3/Virden//Conquest/Post	2	1	0.7514	0.0045	±0.0046	1996
9	Ac Klinck	6	s	d	f	Cadette/Chapais	4	4	1.5000	0.0122	±0.0093	2000
10	Ac Lacombe	6	s	d	f	Klondike//Galt/Unitan	8	21	9.2850	0.0750	±0.0442	1991
11	Ac Legend	6	s	d	f	Chapais/CIMMYT-6	7	7	3.5000	0.0284	±0.0184	1998
12	Ac Malone	6	s	d	f	Callus culture of Leger	4	4	1.7850	0.0095	±0.0077	1999
13	Ac Maple	6	s	d	f	Chapais/CIMMYT-6	4	7	1.9280	0.0156	±0.0113	2000
14	AC Metcalf	2	s	d	m	Oxbow/Manley	7	10	4.6420	0.0329	±0.0205	1997
15	Ac Nadia	6	s	d	f	Leger/QB 173.26	7	9	4.0357	0.0328	±0.0208	1993
16	AC Oxbow	2	s	d	m	TR223 TR222 WPG8020 WPG823	1	0	0.0000	0.0000	±0.0000	1991

	Cultivar Name	R	G	S	U	Parentage	NH	NP	NPD	$\hat{H}$	SD of $\hat{H}$	Released
17	Ac Rosser	6	s	d	f	Galt/Johnston/3/Steptoe//BT 351/Heartland	6	12	4.3214	0.0311	±0.0195	1997
18	Ac Stacey	6	s	d	f	Otal/Melvin	2	3	1.2850	0.0093	±0.0073	1989
19	Ac Westech	6	s	d	f	82RCBB.13/Etienne	5	10	4.2850	0.0308	±0.0194	1998
20	ACCA	6	s	d	f	QB730.2/UL0072/Leger	4	6	2.3570	0.0170	±0.0118	1996
21	Argyle	6	s	d	m	Herta/UM 570//conquest/3/Bonanza	3	3	1.0350	0.0074	±0.0062	1981
22	B1202	2	s	d	m	RPB 70-268/2B75-1223//Klages	2	1	0.5357	0.0038	±0.0039	2000
23	B1602	6	s	d	m	Bumper/6B78-628//Morex/6B78-628	4	2	0.8571	0.0060	±0.0054	1991
24	Banner	6	s	d	f	OB907-33/TBC891-6	8	13	5.8210	0.0412	±0.0251	2000
25	Bedford	6	s	d	f	Keystone/4/Vantage/Jet//Vantmore/3/2* Husky/5/Cree	6	18	8.0714	0.0572	±0.0339	1979
26	Bella	6	s	d	f	Meldugres/Carlsberg 63199	5	14	5.7850	0.0410	±0.0250	1992
27	Belluga	6	s	d	f	TBB 773-6 Mingo OB 339-1	3	3	0.7500	0.0056	±0.0052	1995
28	Blankeney	6	s	d	f	OB907-33/GB8901	3	4	1.9280	0.0146	±0.0105	2000
29	Blitz	6	s	d	f	TBC51-89/AB99-13	1	0	0.0000	0.0000	±0.0000	2000
30	Bonanza	6	s	d	m	Vantage/Jet//Vantmore/3/2* Parkland/4/Conquest	4	4	1.9640	0.0148	±0.0106	1970
31	Brier	6	s	d	f	Leduc///Galt/York//Dickson/Galt	2	1	0.4285	0.0032	±0.0036	1989
32	Bronco	6	s	d	f	Vanier/Laurier//Perth/3/Leger	3	11	2.9285	0.0271	±0.0180	1993
33	Brooke	6	s	d	f	Sandrine/TBC51-89	1	0	0.0000	0.0000	±0.0000	2000
34	Brucefield	6	s	d	f	Maskot/Chapais	3	4	1.9285	0.0178	±0.0128	1997
35	BT954	6	s	d	m	NA	6	14	4.8920	0.0347	±0.0215	2001
36	Cadette	6	s	d	f	QB 139.7/(Min660102/Bonanza)	8	19	6.7140	0.0621	±0.0373	1986
37	CDC Battleford	6	s	d	m	M67/Bt411	4	11	5.1428	0.0369	±0.0228	2001
38	CDC Copeland	2	s	d	m	WM861-5/TR118	2	1	0.5357	0.0039	±0.0040	1999
39	CDC Earl	6	s	d	f	Duke/Heartland	7	19	8.2850	0.0767	±0.0453	1993
40	CDC Kendall	2	s	d	m	Manley SM85221	4	2	1.1428	0.0083	±0.0068	1995

	Cultivar Name	R	G	S	U	Parentage	NH	NP	NPD	$\hat{H}$	SD of $\hat{H}$	Released
41	CDC Silky	6	s	h	f	Duke/(Nordic/BT 413)(M718/B4r6705-15-1)	1	0	0.0000	0.0000	±0.0000	1994
42	CDC Sisler	6	s	d	m	M34/Argyle	2	1	0.4285	0.0032	±0.0036	1996
43	CDC Springside	6	s	d	m	M76/SM93067	1	0	0.0000	0.0000	±0.0000	2001
44	CDC Stratus	2	s	d	m	Manley/ID 810279	1	0	0.0000	0.0000	±0.0000	1994
45	CDC Thompson	2	s	d	m	Nairn/Manley	5	4	1.2850	0.0094	±0.0074	1994
46	CDC Tisdale	6	s	d	m	BT409/Foster	3	2	0.6785	0.0051	±0.0049	2001
47	CDC Yorkton	6	s	d	m	M67/BT411	3	2	0.7857	0.0059	±0.0054	1999
48	Chapais	6	s	d	f	(QB 58.14/Beacon)/BT 904	2	1	0.2500	0.0019	±0.0026	1988
49	Codac	6	s	d	f	Diamond/Duke	7	6	2.7142	0.0204	±0.0137	1993
50	Duel	6	s	d	m	Morex//6B75-1374/M31	2	1	0.2500	0.0019	±0.0026	1992
51	Duke	6	s	d	f	(BrYG3-3/Fedak)//(M65-220/Bonanza)/UA1851	2	1	0.4285	0.0032	±0.0036	1986
52	Etienne	6	s	d	f	Perth/(64-76/(OB595/Algerian//Parkland5))	2	2	1.0714	0.0075	±0.0063	1988
53	Excel	6	s	d	m	Cree/Bonanza//Manker/3/2* Robust	8	13	5.2850	0.0375	±0.0230	1994
54	Falcon	6	s	h	f	11012.2/Tern//Tulelake	5	4	1.6780	0.0132	±0.0098	1992
55	Foster	6	s	d	m	Robust/6/Glenn/4/Nordic//Dickson/Trophy/3/Azure/5/Glenn/Karl	3	2	0.8214	0.0058	±0.0052	2000
56	Gamine	6	s	d	f	Maskot/Chapais	1	0	0.0000	0.0000	±0.0000	1996
57	Grant	6	s	d	f	P885-4/P854-35	1	0	0.0000	0.0000	±0.0000	1996
58	Harrington	2	s	d	m	KlagesX(Gazelle/Betzes//Centennial)	2	1	0.5714	0.0042	±0.0042	1981
59	Jackson	6	s	d	f	Fjola/Husky (H64-35 BT607)/Pomo	1	0	0.0000	0.0000	±0.0000	1985
60	Jaeger	6	s	h	f	Nopal'S "- Ager [(F10.14/Mona-EmirxBco.Mr-Gvs) Api-CM67xOre]	3	2	0.6785	0.0053	±0.0051	1999
61	Johnston	6	s	d	f	Klondike/5/Nord/3/Vantage/Jet//Vantmore/4/Bonanza	2	1	0.5357	0.0040	±0.0042	1980
62	Jolly	6	s	d	f	Laurier/(BR.M. 45-680 X (Montcalm X Byng)	2	1	0.5357	0.0040	±0.0042	1987
63	Kasota	6	s	d	f	Celaya//Mezquita/Godiva/3/Trompillo	2	1	0.4285	0.0032	±0.0036	1995
64	Klondike	6	s	d	f	Galt/NDB133 Vantage/Jet//Vantmore/3/2* Parkland/4/Dickson	2	2	0.5000	0.0038	±0.0040	1976

	Cultivar Name	R	G	S	U	Parentage	NH	NP	NPD	$\hat{H}$	SD of $\hat{H}$	Released
65	Labelle	6	s	d	f	Loyola/Laurier	3	3	0.9285	0.0070	±0.0060	1991
66	Lacey	6	s	d	m	M44/Excel//M46/M44/3/M44/Excel//Stander	8	14	6.3570	0.0457	±0.0276	2001
67	Leduc	6	s	d	f	Steptoe/Klondike	2	1	0.5357	0.0043	±0.0044	1982
68	Legacy	6	s	d	m	6B86-3517/Excel	5	4	1.6785	0.0119	±0.0088	2001
69	Leger	6	s	d	f	Trent/Vanier	2	1	0.5714	0.0045	±0.0046	1982
70	Lucky	6	s	d	f	NA	2	2	0.5000	0.0040	±0.0042	2000
71	Mahigan	6	s	d	f	Celaya//Mesquita/Godiva/3/Trompillo	1	0	0.0000	0.0000	±0.0000	1998
72	Manley	2	s	d	m	Norbert//Hecter/Kleges	5	5	2.3928	0.0169	±0.0117	1991
73	Maskot	6	s	d	f	QB 167.21/OB193.11	2	1	0.2500	0.0020	±0.0028	1989
74	McDiarmid	6	w	d	f	Tapir/Wysor	3	2	0.5000	0.0040	±0.0042	1996
75	McGregor	6	w	d	f	Tapir/Wysor	1	0	0.0000	0.0000	±0.0000	1995
76	Merit	2	s	d	m	TR490/2B80-350	2	1	0.2500	0.0018	±0.0025	1999
77	Mistral	6	s	d	f	NA	3	4	2.2500	0.0159	±0.0111	2000
78	Myriam	6	s	d	f	NA	5	4	1.6428	0.0117	±0.0087	1994
79	Nellygan	6	s	d	f	CQ-CM/Apan//RM508/3/DL69/Hiproly	4	8	3.1428	0.0222	±0.0146	1999
80	Niska	6	s	d	f	BrY63-4/Galt	1	0	0.0000	0.0000	±0.0000	1999
81	Noble	6	s	d	f	Dickson/3/Clho 4738//Traill/UM 570	2	1	0.4285	0.0030	±0.0034	1987
82	Nord	6	s	d	f	Olli/Byng	1	0	0.0000	0.0000	±0.0000	1956
83	OAC Baxter	6	s	d	f	Chapais (male sterile)//OAC Kippen/Leger	3	2	0.6785	0.0048	±0.0046	2000
84	OAC Elmira	6	w	d	f	WB74-69//Wb74-69/WB55-1	6	7	1.9280	0.0153	±0.0110	1987
85	OAC Kippen	6	s	d	f	(York//CI 10853/Parkland)/Perth	2	1	0.2500	0.0018	±0.0024	1987
86	Otal	6	s	d	f	Otra/breeding line of Weibullsholm/3/Hanna/Svanhals//Opal/4/Tammi/5/Morgenrot	2	2	1.0714	0.0075	±0.0063	1982
87	Peregerine	6	s	h	f	H12-4816/R181//M69.77-SHI.R.KCI.NO.87/CEL-5106	1	0	0.0000	0.0000	±0.0000	1999
88	Robust	6	s	d	m	Morex/Manker	2	1	0.2500	0.0018	±0.0024	1994

	Cultivar Name	R	G	S	U	Parentage	NH	NP	NPD	$\hat{H}$	SD of $\hat{H}$	Released
89	Sabina	6	s	d	f	QB 167.21/OB193.11	2	1	0.5714	0.0040	±0.0041	1989
90	Samson	6	s	d	f	Olli/M64-69//R72-181	3	6	2.6428	0.0186	±0.0126	1985
91	Sandrine	6	s	d	f	Mingo OB3391-1 QB203-4	2	11	2.7500	0.0200	±0.0135	1995
92	Sophie	6	s	d	f	BRM 45-680 X (Montcalm/Byng)/BRM 45680/QB 113.1	7	20	7.8920	0.0576	±0.0341	1980
93	Stander	6	s	d	m	Excel//Robust/Bumper	2	2	0.5000	0.0037	±0.0039	1999
94	Stetsen	6	s	d	f	Westbred 501/Gustoe	2	1	0.2500	0.0018	±0.0025	1995
95	Stien	2	s	d	m	Norbert//Hecter/Kleges	5	6	2.5350	0.0179	±0.0122	1987
96	Tankard	6	s	d	m	Argyle/Minnesota M-34	3	2	0.6780	0.0050	±0.0047	1992
97	TB891-6	6	s	d	f	(Vanier/Keystone//BT 421(S6525/Galt))(Vanier/Laurier//Perth)	1	0	0.0000	0.0000	±0.0000	1992
98	Trochu	6	s	d	f	Noble//DL69/DL70///Mari-Coho/Nackta//TR219	1	0	0.0000	0.0000	±0.0000	2000
99	Tukwa	6	s	d	f	I74161/Hiproly	3	6	1.6785	0.0122	±0.0090	1993
100	Virden	6	s	d	f	WA6415-66//Bonanza/NDB136/3/UM67-739R//Bonanza/Dickson	3	3	1.4285	0.0104	±0.0080	1987
101	Vivar	6	s	d	f	Leduc//DL69/DL70/3/Noble/4/CM67-U.Saak1800//Promesa/CM67//DL70	1	0	0.0000	0.0000	±0.0000	2000
102	Viviane	6	s	d	f	NA	3	5	2.4640	0.0178	±0.0122	1999
103	Westford	6	s	d	f	NA	3	3	1.5000	0.0108	±0.0083	2000

2.3 Materials and Methods

2.3.1 Plant materials and S-SAP reactions

Breeders' seed from 103 barley cultivars (Table 2.1) was obtained from the Canadian Grain Commission (CGC, Winnipeg, Man., Canada) and the Canadian Food Inspection Agency (Ottawa, Ont., Canada).

S-SAP procedure is described extensively by Waugh et al. (1997). Briefly, barley seeds were germinated on moist Petri dishes at room temperature for one week. Genomic DNA from 1-week-old seedlings was extracted using DNeasy® 96 plant kit (Qiagen). Half µg of genomic DNA from each plant was incubated with 5 units of *MseI* and 2.5 units of *Pst*I restriction endonucleases at 37 °C for 2 hours. *Pst*I is a methylation-sensitive enzyme that does not contain any restriction site within the BARE-1 retrotransposon (GenBank accession no. Z17327). Both enzymes were used in excess to ensure complete digestion of the genomic DNA. *Pst*I linkers (*Pst*I-1 linker: 5'-CTCGTAGACTGCGTACATGCA-3', *PstI*-2 linker: 5'-TGTACGCAGTCTAC-3') and *Mse*I linkers (*Mse*I-1: 5'-GACGATGAGTCCTGAG-3', *MseI*-2: 5'-TACTCAGGACTCAT-3') were ligated to the ends of restricted fragments using T4 DNA ligase (Invitrogen). The ligation reactions were carried out overnight at room temperature. The products of the ligation reaction were diluted tenfold in TE buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA). S-SAP-PCR amplifications were carried out in two steps: a preamplification PCR step with *Pst*I+C and *Mse*I+C primers (where the symbols +C indicates the presence of the selective nucleotide "C" at the 3' end of the primers) was carried out to reduce the number of amplified restricted fragments and to

ensure that only *Pst*I/*Mse*I fragment were amplified. Lack of a *Pst*I restriction site within the BARE-1 element as described earlier implies that *Pst*I/*Mse*I fragments contain both host and the element DNA. The product of preamplification-PCR was subsequently used as template for selective amplification with an *Mse*I primer containing three selective bases at the 3' end in combination with a [^{33}P]-labeled LTR-derived primer. The PCR products were mixed with an equal volume of denaturing dye (98% de-ionized formamide, 0.025% bromo-phenol-blue, 0.025% xylene cyanol and denatured at 94 °C for 3 min. Amplification products were resolved in a 5% denaturing polyacrylamide gel and electrophoresed at 80 W (constant power) for 2-3 hours. The gels were dried and exposed to Kodak X-Omat film for 3-5 day at -80°C. DNA fingerprints were evaluated by visual inspection of autoradiographs.

2.3.2 BARE-1 primers

LTR-derived PCR primers used in this study were designed from the LTR region of the BARE-1 retrotransposon. Sequences were retrieved from GenBank at National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). Initially, a large set of primers was designed covering most of the BARE-1 LTR sequence. Published LTR-derived primers (Waugh et al. 1997; Kalendar et al. 2000) were also synthesized and used in a prescreening S-SAP assay on barley cultivars in this study. Only two LTR-derived primers, R1 and R4, designed in this study, were retained for further analysis. All other primers especially those that were designed from the more 3' regions of the LTR sequences produced suboptimal amplification products. R1 and R4 (Table 2.2) were combined with various *Mse*I selective primers (Tables 2,3) for final S-SAP analysis.

Table 2.2. Sequences of PCR primers used in this study

Primer	Sequence 5'to 3'
R1	GTTATGTAGTGGGCGAGCGAG
R4	GCATACATGTGTCCTTCGTTGCC
M2	GATGAGTCCTGAGTAACAC
M3	GATGAGTCCTGAGTAACAG
M4	GATGAGTCCTGAGTAACAT
M7	GATGAGTCCTGAGTAACTG

Table 2.3. S-SAP primer pair combinations, number of amplified bands and percentage polymorphism detected among barley cultivars

Primer pairs	No. amplified bands	No. polymorphic bands	Percent polymorphism
R1M4	106	54	50.9
R4M3	102	25	24.5
R1M7	111	36	32.4
R1M2	85	35	41.2
Total	404	150	37.1

2.3.3 Data analysis

DNA fingerprints were scored as binary variables, 1 for the presence and 0 for the absence of bands. The total number of bands that were scored included those that were clearly polymorphic, and those that were clearly monomorphic. Bands whose presence could not be clearly discerned were ignored.

2.3.3.1 Clustering

The binary data were used to generate a genetic similarity matrix with SIMQUAL routine using the DICE formula (Dice 1945), which is equivalent to Nei and Li (1979) genetic similarity coefficient from the NTSYS-pc statistical package (Rohlf 2000). Clustering of genotypes was performed using SAHN, also in NTSYS-pc, based on the genetic similarity matrix with the unweighted pair-group using arithmetic average (UPGMA) method.

2.3.3.2 Gene diversity estimates

The mean gene diversity ($\hat{H}$) and gene diversity over loci (π) were calculated from the binary data based on Nei's (1987) formula using Arlequin program (Schneider et al. 1997). $\hat{H}$ is defined as the probability that two randomly chosen haplotypes are different in the sample.

2.3.3.3 Analysis of molecular variance

Variance components for barley populations and the relationships among haplotypes determined by a minimum spanning tree were calculated using analysis of molecular variance (AMOVA) with the Arlequin program (Schneider et al. 1997).

2.3.3.4 Regression analysis

A linear regression analysis was carried out to determine the effects of selfing/inbreeding on the level of within cultivar variation. Inbreeding levels were determined, where possible, from variety descriptions. These levels refer to the level of inbreeding in the breeders' seed source. Seed samples used in this study are assumed to have been derived from unselected bulks derived from breeders' seeds. They are therefore more highly inbred, but residual heterozygosity would be manifested as heterogeneity of haplotypes within the seed source.

2.3.3.5 Discriminant Analyses

Group determination, i.e., classes, was preset according to the following sets of agronomic traits (or use): two-row/six-row, malting/feed, winter/spring, hulled/hull-less. The analyses were carried out separately for each trait set. Binary variables need not disqualify the various discriminant analyses (Rao 1952, Cochrane and Hopkins 1961, Gilbert 1968, Krzanoswski 1975). Canonical discriminant analysis (DA) (Kshirsagar 1972), an ordination method, was used to assess if there was justification to support each of the agronomic (or use) trait classifications. Canonical DA (in this case, using S-SAP variables) assumes prior knowledge of the grouping of the individuals and does not use

the variables (in these cases the class variables) by which those individuals were grouped for the analysis. Classificatory DA was performed to obtain the appropriate discriminant functions to be used to identify an unknown sample based on its S-SAP profile, using a reduced set of band patterns. The approach taken was a conventional (Anderson 1984, Kshirsagar 1972) or parametric method, i.e. using the linear classification function. Mahalanobis distances based on the pooled covariance matrix were used, in which variables with zero variance, i.e., invariant, were deleted. Based on the pooled covariance matrix squared distances, between class means were computed. The test of the distances between these class means, i.e., the Mahalanobis distances was ignored, because the variables that were generated were binary and therefore do not comply with the assumption of normality. Therefore, the analyses were used as a guide only without relying on the various test statistics. To find the nature of the group differences, the within-canonical structure was examined, and the variables most highly associated with each of the first and only axis were summarized. Plots of the first and only canonical variable were made for visual examination to determine the degree of overlap between the two classes, separately for each set of agronomic traits. Cross validation (Lachenbruch and Mickey 1968) was also available as part of the SAS program and used to assess how the discriminant function classifies the observation that it excludes from the main data matrix. All computations were done using SAS (SAS Institute).

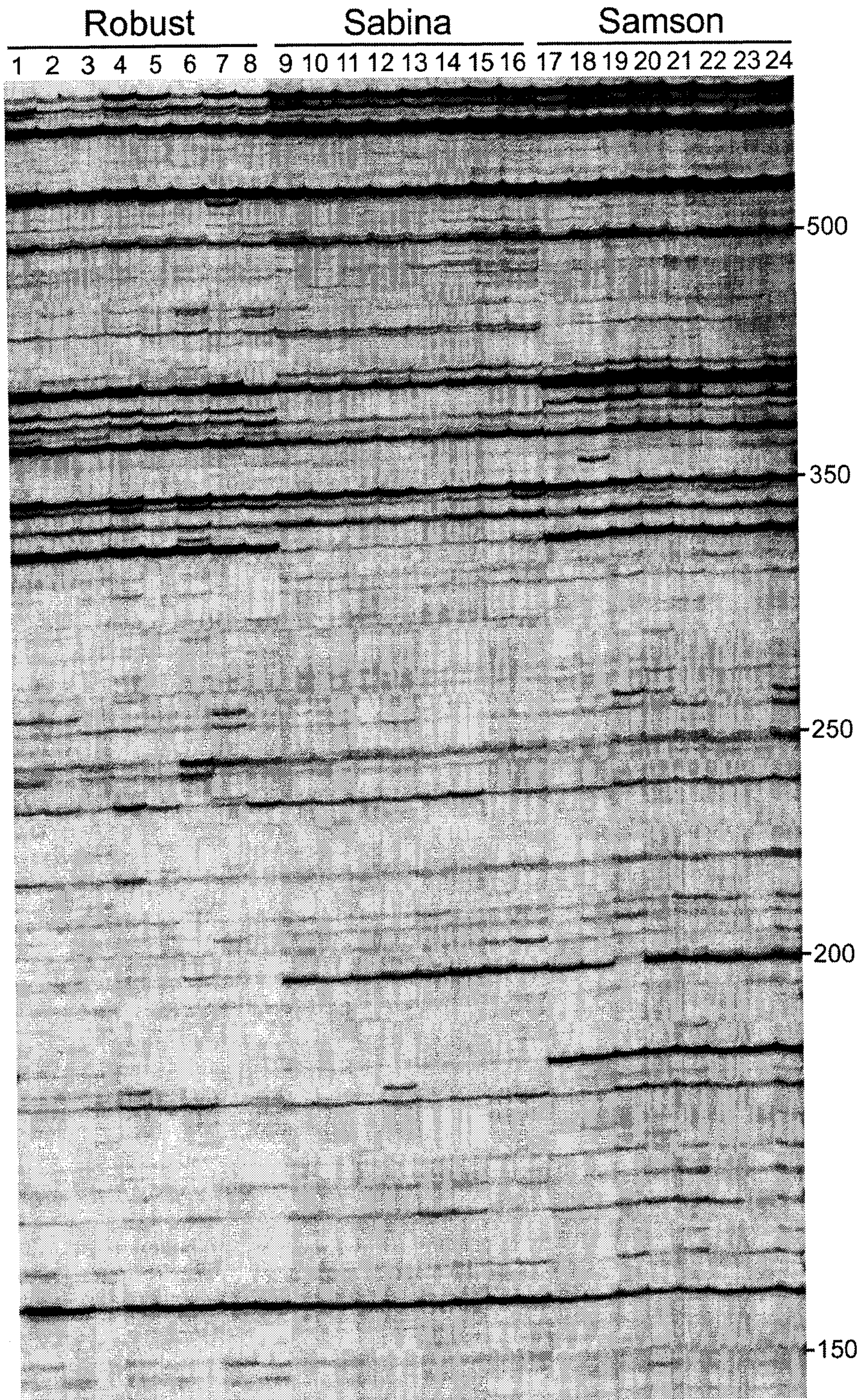
2.4 Results

DNA fingerprints of 824 barley plants were obtained using two specific primers derived from the LTR region of BARE-1 element in combination with four selective *Mse*I primers (Table 2.1). Two LTR-derived primers, R1 and R4, were selected for the final S-SAP analysis based on the result of a prescreening S-SAP assay to select the most optimal primer combinations. R1 primer sequence extends from nucleotide position 70 on the anti-sense DNA strand of BARE-1 LTR, and the R4 primer sequence extends from nucleotide position 384 of the sense strand of the BARE-1 LTR sequence. Generally, primers that were designed from the more 3' regions of the LTR sequences yielded suboptimal amplification products and were therefore eliminated from the final analysis. An example of S-SAP profile for three barley cultivars, eight plants per cultivar, is shown in Figure 2.1.

2.4.1 Genetic diversity/similarity estimates

Four primer pair combinations generated 404 amplified bands, of which 150 were polymorphic either within or between cultivars (Table 2.3). The highest and the lowest multiplex ratios were observed for the R1M7 and R1M4 primer pairs ("R" stands for the LTR-derived primer and "M" for *Mse*I selective primer), 111 and 85, respectively. However, there was no correlation between high multiplex ratio and the level of polymorphism resulting from each primer pair. Indeed, the R1M7 fingerprint had the lowest percentage of polymorphism despite having the highest number of amplified bands

Figure 2.1. Portion of S-SAP profile of three barley cultivars with R4/M3 primer pair combination. For each cultivar, 'Robust', 'Sabina' and 'Samson', DNA was isolated from eight individual plants and S-SAP profiles were generated as described in Materials and Methods. Numbers on the right indicate DNA fragment size in base pair.



Robust
Sabina
Samson
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
500
350
250
200
150

2.4.2 Clustering and DA

UPGMA clustering produced distinct groups for two-rowed, winter, and hull-less cultivars (not shown), but there was some overlap between malting and non-malting six-rowed spring cultivars. This trend was more obvious with DA (Fig 2.2). Furthermore, DA resulted in a clear separation of the phenotypes winter versus spring, two-rowed versus six-rowed and hulled versus hull-less. However, a small amount of overlap between six-rowed malting and six-rowed feed cultivars was seen (Fig 2.2). The grouping of barley cultivars obtained by DA was largely supported by the minimum spanning tree obtained from Arlequin analysis (not shown). Both DA and cluster analysis showed spring six-rowed cultivars to be the most diverse group in this study.

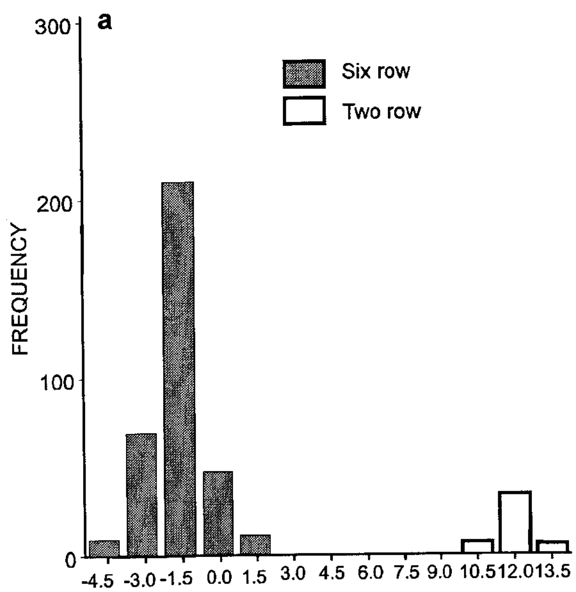
2.4.3 AMOVA analysis: Partitioning of genetic variance

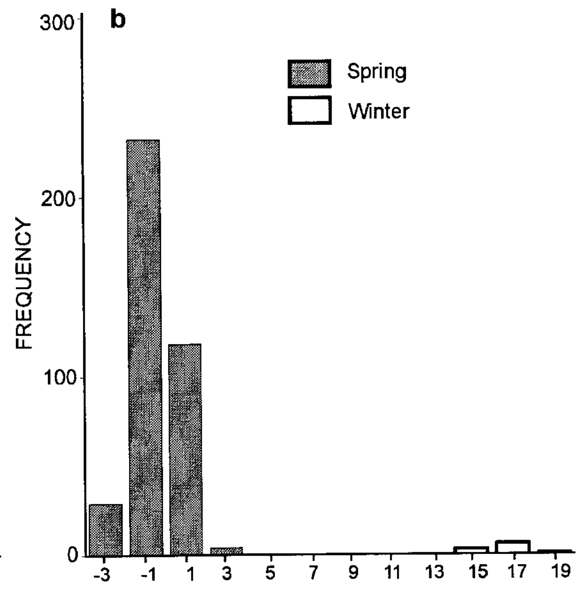
AMOVA (Table 2.4) was performed separately for four different groups as indicated in Table 2.1. The 824 barley plants used in this study were grouped into 441 haplotypes after removing genotypes with common DNA profiles. A haplotype is defined as a genotype with a defined set of alleles. Haplotype sharing was observed exclusively in intra-cultivar hierarchy, (i.e., common genotypes belonging to the same cultivar). However, no inter-cultivar haplotype sharing was observed. Therefore, all individual haplotypes clustered within their own cultivar. In all cases (Table 2.4), AMOVA resulted in highly significant ($P< 0.001$) genetic variance within and between groups. The variance among cultivars in the same group accounted for the largest percentage of the total variance (69 to 86%). In the analysis of six-rowed cultivars based on the malting versus non-malting grouping, the smallest variance component (4%) was obtained

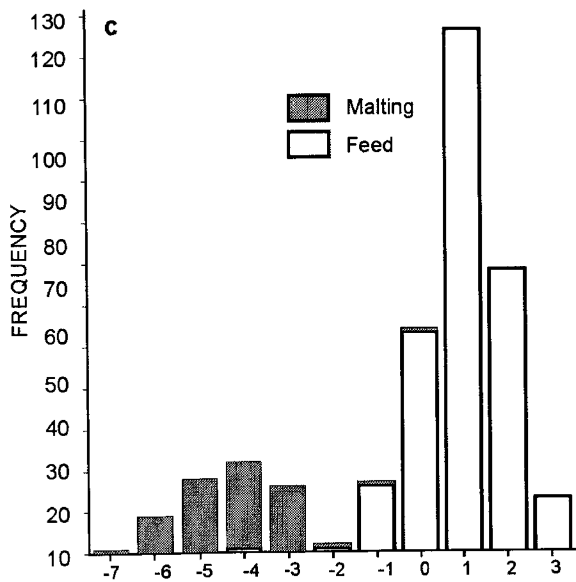
Table 2.4. Summaries of AMOVA analyses among various agronomic groups of barley cultivars

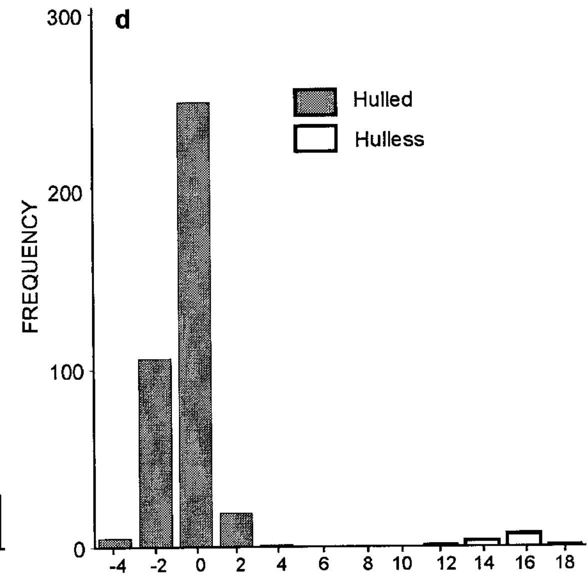
Level	df	Sum of squares	Variance	Percent variation
		Two-rowed/six-rowed		
Among groups	1	299.25	3.61	22.6
Among cultivars	89	3167.43	11.12	69.57
Within cultivars	193	241.21	1.24	7.82
		Winter/spring		
Among groups	1	116.59	3.98	23.22
Among cultivars	78	2914.13	11.87	69.27
Within cultivars	165	212.53	1.28	7.52
		Hull/hulless		
Among groups	1	94.96	2.52	16.04
Among cultivars	78	2935.76	11.93	75.79
Within cultivars	165	212.53	1.28	8.18
		Feed/malting		
Among groups	1	97.68	0.57	4.13
Among cultivars	78	2933.04	11.96	86.55
Within cultivars	165	212.53	1.28	9.32

Figure 2.2. Discriminant analysis (DA) of barley cultivars based on preset morphological criteria. Plot of canonical variables on axis 1: (a), two-rowed versus six-rowed group, (b), spring versus winter group, (c), feed versus malting group, and (d), hulled versus hull-less group. Groups were completely delineated by DA analysis except in "c" where there was a minor overlap between malting and feed cultivars.


a
Six row
Two row
FREQUENCY
300
200
100
0
-4.5
-3.0
-1.5
0.0
1.5
3.0
4.5
6.0
7.5
9.0
10.5
12.0
13.5


b
Spring
Winter
FREQUENCY
300
200
100
0
-3
-1
1
3
5
7
9
11
13
15
17
19


c
Malting
Feed
FREQUENCY
130
120
110
100
90
80
70
60
50
40
30
20
10
-7
-6
-5
-4
-3
-2
-1
0
1
2
3

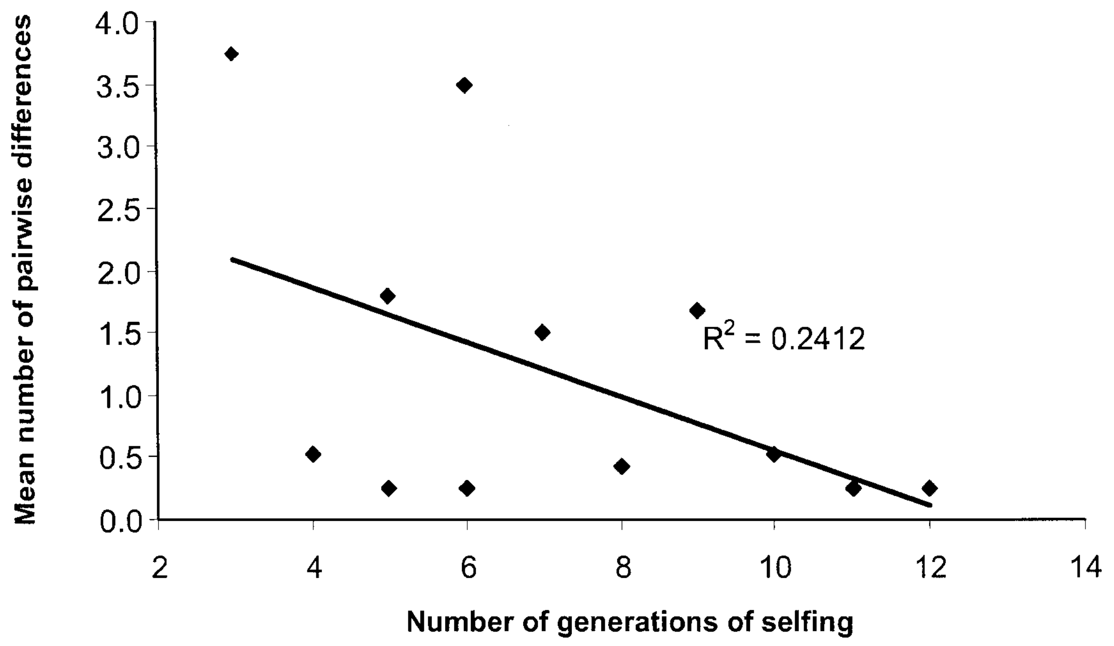

d
Hulled
Hulless
FREQUENCY
300
200
100
0
-4
-2
0
2
4
6
8
10
12
14
16
18

between groups indicating the absence of a major genetic structure on this level. Within-cultivar component of variation ranged from 7.5% to 9.3%.

2.4.4 Genetic heterogeneity within cultivars

Genetic heterogeneity was detected in 84% of all barley cultivars in this study (Table 2.1). The number of haplotypes within each cultivar ranged from one (no genetic variation within cultivar) for 17 homogenous cultivars to eight (maximum heterogeneity within cultivar in which each plant within the cultivar could be uniquely distinguished) for six completely heterogeneous cultivars (Table 2.1). Within cultivar average genetic diversity over loci was estimated to be 0.0134 ± -0.009. When-within cultivar genetic variation, measured as the mean number of pair-wise differences between genotypes of the same cultivar, was plotted against the level of inbreeding, i.e., number of generations of selfing, a negative correlation was observed (Fig 2.3) indicating an inverse relationship between increased number of generations of selfing and the level of genetic heterogeneity within cultivars. Six-rowed spring type (comprising malting and non-malting cultivars) constituted the most diverse group in this study. Average genetic similarity within this overall group was estimated to be 0.61 (not shown) with no significant difference between mean genetic similarity between feed and malting cultivars. The two-rowed group formed a separate subgroup with an average genetic similarity of 0.85 (not shown), the highest value of mean genetic similarity for a group of cultivars in this study. Six-rowed winter and six-rowed spring hull-less cultivars also formed two unique subgroups.

Figure 2.3. Plot of number of generations of selfing versus genetic diversity within cultivars measured as the mean number of pair-wise differences among individual plants in the cultivar. A subset of 13 cultivars ranging from F_3 to F_{12} generations of selfing was used.



Mean number of pairwise differences
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2
4
6
8
10
12
14
Number of generations of selfing
R2 = 0.2412

2.5 Discussion

In this study, a substantial level of intra- and inter-cultivar genetic diversity was detected by S-SAP analysis. The ubiquitous nature of retrotransposons in plant genomes (Flavell et al 1992; Kumar and Bennetzen1999; Suoniemi et al. 1999) and their broad genomic distribution makes retrotransposon-based markers ideal tools for a wide range of applications in plants including measurements of biodiversity, genome evolution, linkage analysis, and mapping. Once a retrotransposon is stably integrated into a genomic location, it behaves in a Mendelian fashion. Therefore, integration sites shared between two plants are likely to have been present in their last common ancestor. S-SAP-based polymorphism may result from transpositional activity of retroelements and/or a restriction site polymorphism, as is the case for AFLP and some restriction length polymorphisms.

In general, most plant retrotransposons studied to date have not demonstrated mobility within the recent past. In contrast, the BARE-1 element of barley is actively transposing (Jaaskalainen et al 1999; Suoniemi et al 1999) and furthermore, both biotic and abiotic stresses cause retroelement-induced mutations in wild barley (Kalendar et al. 2000). Liu and Wendel (2000) used Southern transfers to analyze introgressed rice lines and found a significant increase in the transpositional activity of the rice retrotransposon Tos17. Tissue culture was also shown to activate retrotransposition in rice (Hirochika et al. 1996). Most retrotransposon have been found in the non-coding region of the plants, often within a locus harboring other repetitive elements in a nested fashion (SanMiguel et al 1996; Shirasu et al. 2000). However, many insertions have also been found in and

around normal plant genes (White et al. 1994) including *bm3* mutation in maize (Vignols et al. 1995).

2.5.1 Measurements of genetic diversity, cluster and DA analysis

The objective of breeding new crop cultivars is to create new allele combinations, followed by selection of desirable phenotypes during selfing generations. In an inbreeding crop such as barley, the level of heterozygosity for unlinked loci is reduced by approximately half following each generation of selfing. The trend continues until a pure line is obtained in which almost complete homozygosity over all loci is achieved. Selection pressure is also directed towards achieving a homogenous and stable population of plants exhibiting minimum within-cultivar variation.

Biotypes are variant plants that belong to the same cultivar and possess different DNA profiles. In this study, 84% of all cultivars had biotypes. Although the aim of a breeding program is to produce a homogenous population of plants for a given cultivar, biotypes have been detected in many crop cultivars studied to date (Olufowote et al. 1997; Soleimani et al. 2002; Soleimani et al. 2003; Perry, 2004). The possibility of contamination of seed samples as the cause of this finding was ruled out because all the biotypes of the same cultivar fell within the same cultivar, and because the genetic distances among biotypes within-cultivars were substantially lower compared to the distances between cultivars. Therefore, despite selection pressure aimed at cultivar homogeneity, a substantial level of genetic (molecular) diversity still exists within each cultivar. Regression analysis between the level of inbreeding and within cultivar biodiversity level showed that inbreeding/selfing had a negative effect on the level of

genetic diversity as shown in figure 2.3. The regression analysis was limited to a selected group of cultivars from table 2.1 for which a complete cultivar description was available. The level of selfing in this subgroup ranged from F_3 to F_{12} generation. Other factors in a breeding program that may also affect the extent of genetic heterogeneity within a cultivar, such as the choice of parents and the numbers of backcrosses among others, were not factored in. As described earlier, selection in inbred crops such as barley results in a heterogeneous population of homozygous plants. Although the exact sources of this genetic variation are difficult to predict, a multitude of factors may be responsible for maintaining within-cultivar variation. These may include breeding practices, spontaneous mutations, and induction of the transposition system. Retrotransposon insertions in or near genes can play an important role in species evolution (Shapiro 1999) and many retrotransposon insertions have been found in association with normal plant genes (White et al. 1994). Studies have also shown that environmental stress can induce transposition as described earlier. The effects of transposition on genetic variation are more likely to be profound in plants with an active transposon system such as barley. In order to partially address the source of genetic variation within cultivars, we analyzed families with these lines (parent1-parent2-progeny). Four cultivars were found in this category: ‘AC Alma’ (‘Chapais’ x ‘Leger’), ‘AC Klinck’ (‘Cadette’ x ‘Chapais’), ‘Gamine’ (‘Maskot’ x ‘Chapais’) and ‘AC Metcalfe’ (‘AC Oxbow’ x ‘Manley’). Comparison of gene diversity (from Table 2.1) between parents and progeny indicates that there are two likely sources for the observed genetic variation within each cultivar. Some of the variation may be residual, which is carried over from parents to progeny as in ‘AC Klinck’ and ‘AC Metcalfe’, in which gene diversity in the progeny is not significantly different from that

of the parents. In other cases such as 'AC Alma', the level of genetic variation in the progeny was significantly higher than those of the parents indicating that retrotransposition and recombination may have played a major role in these observations. We also analyzed band-sharing data between parents and progeny of the above mentioned cultivars (not shown) and found that between 2 to 5% of bands that were present in progenies could not be assigned to either parent.

This study demonstrated the usefulness of S-SAP markers for the study of cultivated barley. S-SAPs were found informative in revealing genetic variations both at the intra- and inter-cultivars level. These markers could be used for future breeding purposes and assessing cultivar purity testing in addition to their use for intellectual property right protection (such as Plant Breeders' Rights in Canada). Recent S-SAP-based mapping studies in wheat (Queen et al. 2004) and barley (Waugh et al. 1997) and barley (Manninen et al. 2000) have shown that these markers had wide genome distribution along all chromosomes in these organisms.

Studies using pedigree data have shown that the gene pool of the malting barley cultivars is narrow (Martin et al 1991; Horsely et al 1995). The reduction in the genetic base of cultivated barley is thought to be the result of using local germplasm with reduced genetic base. However, these studies were based on the estimation of the coefficient of parentage and genetic measurements involving phenotypic traits such as malting characteristics. Estimates of genetic diversity based on the coefficient of parentage may be misleading due to various assumptions made regarding the relatedness of ancestors with unknown genealogies, linkage, and unequal contribution of parents to progenies (Barrett and Kidwell 1998; Soleimani et al. 2002).

In this study, many cultivar pairs with close pedigrees were found to be substantially diverse from each other genetically. Example of such cultivar pairs included, 'CDC Yorkton' and 'CDC Battleford', and 'AC Malone' and 'Leger'. Indeed, 'AC Malone' which was developed via somaclonal variation from callus culture of 'Leger' had only 0.73 (73%) genetic similarity to 'Leger' based on the analysis of 404 S-SAP markers. In other words, 27% of S-SAP markers were polymorphic between 'Leger' and 'AC Malone'. This is a very substantial amount of genetic diversity between two cultivars in which one is derived from the other by means of tissue culture.

In conclusion, an objective assessment of the extent of genetic diversity in the genome depends on the ability to detect the most prevalent and widely distributed types of genomic mutations and rearrangements. The S-SAP markers reported here were not mapped and therefore no conclusions can be made regarding their genomic locations and their possible linkage to agronomically important traits in barley. However, the results of the discriminant analysis (Figure 2.2) indicates that at least some of these markers may potentially be linked to agriculturally significant traits such as malting/feed. In this study the emphasis was placed on BARE-1 related polymorphisms. There are various other Ty1-*copia* retroelements in barley genome with their unique history of transposition and genome distribution (Shirasu et al. 2000). Simultaneous analysis of a diverse group of retroelements with S-SAP in a multi-retrotransposon approach (Leigh et al. 2003) may reveal additional polymorphisms. Whether the diversity in DNA sequence measured here could be translated into phenotypic diversity is beyond the scope of this study. However, there is a growing body of evidence implicating the role of retrotransposition in altering the phenotype. Insertion of retrotransposon in and around genes (Miyao et al. 2003;

Kidwell and Lisch, 1997) could alter their expression pattern, resulting in a change in phenotype. Whether or not these mutants survive in the subsequent generations will in turn depend on the fitness of individuals bearing the mutations. Breeding may be viewed in part as a man-made directional selection in which individual plants are subjected to various biotic and abiotic stresses. In this scheme, there are various macro- and micro-evolutionary processes that could potentially affect the genetic makeup of the resulting population. In the presence of an active transposition system, a high degree of genome plasticity is expected to result during the breeding process and associated selection pressure. Therefore, retrotransposon-based marker systems such as S-SAPs can be a useful tool for crop genetic studies, as they target retrotransposon insertion sites in genome and can be used as powerful tools for crop improvement.

2.6 Acknowledgments

The authors would like to thank Ajit Singh Sahota from the Canadian Food Inspection Agency (CFIA) and Canadian Grain Commission (CGC) for providing plant material. We would also like to thank Dr. Thin-Meiw (Alek) Choo of Agriculture and Agri-Food Canada for providing cultivar description of barley material. We are indebted to him for his advice and discussions. This study was partially funded by the Automated Quality Testing (AQT). This paper benefited from comments by Dr. Steve Molnar and Dr. Nick Tinker of Agriculture and Agri-Food Canada, and we thank them for their valuable comments and advice on an earlier draft of the manuscript.

CHAPTER THREE

Efficient validation of Single Nucleotide Polymorphisms in plants by allele-specific PCR, with an example from barley

The following chapter is a reproduction of a paper published in Plant Molecular Biology Reporter (2003) 21:281-288. The chapter fulfils objective 1.2 of the thesis. It describes a new and simple method for SNP validation, Allele-specific PCR with three primers, and compared it to a widely used protocol (the primer extension method). SNPs are variations in DNA sequences in a population. There are large numbers of SNPs imbedded in EST databases for barley and other cereals. However, before they can be used for population genetic studies or diagnostic tests, they need to be validated experimentally and transformed into a genotyping assay. AS-PCR is a one step genotyping assay that uses crude genomic DNA as template with a substantially reduced cost and labour. The method was optimized, and worked equally well with DNA extracted from either seeds or leaves. Amplicons generated by the AS-PCR method in this study are analogous to the sequence tagged sites (STS) markers and have been used for cultivar identification and the gene diversity measurements.

3.1 Abstract

Although an increasing number of ESTs (Expressed Sequence Tags) from the public domain has facilitated SNPs detection, further validation is needed prior to their use as markers. For SNP validation we have compared two independent methods, namely, the primer extension method followed by capillary electrophoresis on an ABI PRISM® 3100 Genetic Analyzer and nested PCR followed by agarose gel electrophoresis and ethidium bromide staining for visualization. An assessment of the efficiency and costs associated with these two methods for SNP validation, based on a sample of barley cultivars, is presented.

Key words: SNP validation, allele-specific PCR (AS-PCR)

Abbreviations: SNP, single nucleotide polymorphism; AS-PCR, allele-specific PCR

3.2 Introduction

Single nucleotide polymorphisms (SNPs) are nucleotide variations in the DNA sequence of individuals in a population and constitute the most abundant molecular markers in the genome. They are suitable for automation and can be used for a wide range of purposes, including rapid identification of crop cultivars, construction of ultra high density genetic maps and association studies related to genetic disorders (Douabin-Gicquel et al., 2001). SNPs are widely distributed throughout genomes (Halushka et al., 1999) although several studies have shown that the occurrence and distribution of SNPs varies among different species. In maize there is one SNP per 60-120 bp (Ching et al., 2002) while in humans, the estimate was one SNP per 1000 bp (Sachidanandam et al., 2001). The frequency of SNP distribution within each genome also varies. In most organisms studied to date SNPs are more prevalent in the non-coding regions of the genome. Theoretically, such mutations should only affect the phenotype if they cause a change in a regulatory region of a gene, thereby changing the expression pattern. Within the coding regions, an SNP is either non-synonymous resulting in a change in amino acid sequence or synonymous in which the SNP does not alter amino acid sequence and therefore is neutral. However, this type of synonymous change can potentially create a splice site resulting in phenotypic changes (Richard and Beckman, 1995).

Prior to its utilization, an SNP needs to undergo detection and validation. Detection of SNPs can be done experimentally by DNA sequencing or by using *in-silico* protocols (Buetow et al., 1999; Sunyaev et al., 1999), both of which are facilitated by an increasing number of expressed sequence tags (EST) sequences available in the public domain. The validation step can be carried out using various techniques such as primer extension,

hybridization, ligation, PCR amplification, and restriction enzyme digestion. There are various rapid PCR-based assays for validation of SNPs such as tetra primer ARMS PCR (Ye, et al., 1992), bi-PASA (Liu et al., 1997), bi-directional AS-PCR (Karhukorpi and Karttunen, 2001), and PCR-CTPP (Hamajima, 2001). Our method is an AS-PCR in a nested format which gives higher level of specificity of detection than the conventional PCR using two primers and it is very close to PCR-CTPP, which was applied to BAR-2 gene in humans. In this paper we present the application of this method for SNP validation in barley, useful to plants (Fig. 3.1). The specificity of this assay was enhanced by combining the high specificity of nested PCR with a hot start Taq DNA polymerase (Titanium Taq DNA polymerase BD Biosciences) and the use of a "touch down" protocol. Finally, we compare the cost and efficiency of this method to that of primer extension using capillary electrophoresis on an ABI Genetic analyzer.

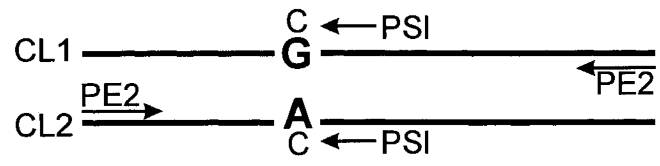
3.3 Materials and Methods

3.3.1 Plant Material and DNA extraction

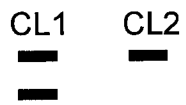
Breeders' seeds from 23 Canadian six-rowed barley cultivars were obtained from the Canadian Grain Commission, (Grain Laboratory Research, Winnipeg, Manitoba). 30 seeds per cultivar were germinated on a petri dish and the leaf material was harvested from each plant one week after germination. Leaf material was ground using the MM 300 Mixer Mill (Retsch®) and DNA extraction was carried out using the DNeasy™ 96 Plant Kit (Qiagen) according to the manufacturer's recommendations.

Figure 3.1. Schematic diagram showing the AS-PCR approach used to validate SNPs. (a), alignment of EST sequences from two cultivars (CL1 and CL2) with a G/A transition (SNP). PE1 and PE2 primers are designed from the consensus sequence of the EST contig. The PS1 primer is designed to detect the SNP polymorphism. The PS1 primer that has a C at its 3' end amplifies the G allele but not the A allele due to the mismatch. To detect the A allele a different primer than PS1 may be designed. (b in Fig.) DNA banding pattern resulting from the AS-PCR amplification. CL1 has two bands indicating the presence of G at the SNP locus, whereas the CL2 has only one band showing the lack of the G allele at that locus

(a)



(b)



3.3.2 Detection and identification of SNPs

Publicly available EST sequences from various sources including the Montana State University, Bozeman, Montana, USA (*http://hordeum.oscs.montana.edu/locus/index.html*), National Centre for Biotechnology Information (NCBI) (*http://www.ncbi.nlm.nih.gov/dbEST/index.html*) and the Clemson University Genomic Institute (*http://www.genome.clemson.edu/projects/barley*) were retrieved and used for the primer design for sequencing of the plant material, the data mining, and the subsequent identification of SNPs in the Canadian barley cultivars. A specific PCR primer was designed for each putative SNP in such a way that the last nucleotide at the 3' end of this primer was complementary to only one allele of the putative SNP. Two outer primers were designed from the consensus sequence of the contig on either side of the SNP marker (Fig. 3.1). For the sequencing procedure, the selected EST loci were PCR amplified from the genomic DNA of 8 six-rowed cultivars. The resulting PCR products were either sequenced directly or alternatively, when multiple amplification products were present, were first gel purified. DNA sequencing was done using the Sanger 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 following the manufacturer's recommendation. Multiple sequence alignments were performed to identify putative SNPs among barley cultivars.

3.3.3 SNP validation

Validation of putative SNPs was carried out by two independent methods, AS-PCR and Primer extension.

3.3.3.1 AS-PCR

Direct AS-PCR amplification on genomic DNA was carried out using three primers simultaneously in a reaction tube by a touchdown protocol followed by visualization on agarose gel and ethidium bromide staining. PCR reactions were carried out in a PTC 200 (MJ Research) thermocycler using Titanium Taq DNA polymerase (BD Biosciences) as follows: 2 min. of initial denaturation to activate Taq DNA polymerase at 95°C, 30 sec. of annealing at 65 °C, 1 min. of extension at 72 °C, for 12 cycles with a reduction of 0.7 °C/per cycle of annealing temperature. This program was followed by 22 more cycles of amplification at the constant annealing temperature of 56 °C.

3.3.3.2 Primer extension

For comparison with AS-PCR above, primer extension experiments were carried out by the incorporation of a dideoxy nucleotide using the SNaPshot™ Multiplex Kit (Applied Biosystems) following the manufacturer's recommendations. The products were separated by capillary gel electrophoresis on an ABI 3100 Genetic Analyser using ABI PRISM®. SNaPshot primers were tailored for multiplexed reactions by the addition of a random DNA sequence (with no known sequence similarity to barley DNA based on the BLAST results) to their 5' end to ensure size separation.

3.4 Results

Putative SNPs were detected using a combination of experimental and *in-silico* methods. Of the 214 SNPs identified, 71% were transition mutations and the remaining 29% were transversions (results not shown). An example of the detection of a G/A transition between two malting cultivars is provided in Fig. 3.2 where one cultivar had the G allele and the other had the A allele in the homologous locus as detected by AS-PCR amplification. Furthermore, this particular SNP marker showed a cultivar-specific pattern in which all individual plants from cultivar 'CDC Yorkton' had the G allele and those from cultivar 'Duel' had the alternative allele (see Discussion below). Intra-cultivar genetic variation was observed in 10 out of 23 (43%) cultivars based on a sample size of 8 individuals per cultivar (not shown). 13 out of 27 cultivars showed no intra-cultivar variation even when the sample size was increased to 32 individual plants per cultivar (results not shown). The intra cultivar variation may be either the result of impurities in seed material or an inherent genetic variation within cultivars. Although the possibility of impurities in breeders' seed cannot be completely ruled out here, it is apparent from this study and previous reports that a substantial level of genetic variation may exist within each cultivar despite rigorous selection pressure aimed at cultivar purity and homogeneity (Soleimani et al., 2002).

The robustness and reproducibility of AS-PCR approach was tested against a second method, namely the primer extension (Table 3.1). Nested PCR using three primers in a simultaneous reaction is a one-step PCR reaction on the total genomic DNA followed by agarose gel visualization. This procedure generally results in the presence of two bands.

Table 3.1 Comparison between the AS-PCR and the primer extension methods

Parameters	AS-PCR	Primer extension
Template	Genomic DNA	Purified EST Locus
Purification	None	PCR purification kit or SAP and Exo I
Post PCR clean up	None	SAP or CIP treatment
Gel analysis	Agarose/ethidium bromide	ABI genetic Analyzer
Multiplexing	Single-plex or duplex	Multiplex (decaplex)
Allelic discrimination	Two*	Four
Heterogygote detection	No	Yes
Reliability	High	High
Relative cost	Low (~$0.4/reaction)**	High ($4.0/reaction)***

* Two by implication, since one allele is amplified by the internal oligo with nested PCR whereas the absence of amplification indicates the alternative allele.

** Titanium Taq DNA polymerase and other ingredients for PCR reaction

*** Single nucleotide primer extension kit, cleanup reagents (Exo/SAP IT)

The higher molecular weight band which is uniformly present in all samples is the result of amplification of the corresponding EST locus. This band also serves as a positive control, while the lower molecular weight band is indicative of an SNP marker (Fig.3.2). The primer extension method requires two rounds of PCR amplification as well as two purification steps of the PCR products before the samples can be analyzed. The cost of this process could be lowered by multiplexing SNP primers. To achieve this, the length of each additional primer must be extended by a minimum of 4 bp of a random sequence at the 5' end of oligos for size separation (Lindblad-Toh et al., 2000). However, as the length of each subsequent oligo increases beyond approximately 40 bp, this requires HPLC or PAGE purification of the oligo after synthesis to separate the resulting mixture of incomplete oligos synthesized in the process, this in turn increases the costs.

3.5 Discussion

Various techniques of SNP validation have been described including mini-sequencing (Chen et al., 1999; Ross et al., 1998), ligase-based techniques (Grossman et al., 1994), molecular beacons (Giesendorf et al. 1998), cleavage with restriction endonucleases, and various PCR-based amplification methods. Although all of the above techniques have been demonstrated to be efficient for validation of SNPs, they differ in their experimental design, associated costs and time needed to carry out the validation process. In this study we found AS-PCR with three primers to be an efficient, cost effective and reliable way for SNP validation.

Figure 3.2. AS-PCR profile of two barley cultivars (Breeder's seed) using three primers: EST forward 5'ATCAAGATGAGGAAGCTGATGCAGT3',
EST reverse 5'TCTTGTCTGAAGCATGGAACCAGAT3', and SNP020
5'TTTGATGGATAATTAATCCGC3'. The upper band (a) is the amplification of the EST locus and serves as positive control. The lower band (b) is an allele-specific amplification. Lanes 1-8 are different samples from 'CDC Yorkton'; lanes 9-16 are different samples from 'Duel'.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
a
b

We first identified 214 SNPs by a combination of sequencing and mining through public barley EST databases. Most of these polymorphisms (71%) were transitions (i.e. A/G or C/T) the remaining 29% were represented by transversions (i.e. A/T, C/A, G/C, G/T), suggesting a non-random nucleotide substitution pattern with a strong bias towards transition mutations. A high level of transition mutations is thought to be the result of 5-methylcytosine deamination reactions that are known to occur at the CpG islands (Fryxell and Zuckerkandl 2000).

A summary of comparisons between AS-PCR and the primer extension approach is provided in Table 3.1. AS-PCR detection of SNPs is performed on genomic DNA extracted from plant leaves, seedlings or a single seed. Employing a one-step PCR reaction simplifies the experimental design and eliminates the need for the removal of PCR primers and dNTPs which are inherent prerequisites for the primer extension experiments. The high specificity of AS-PCR for SNP detection is due to an increase in the number of target loci during PCR which are then used as templates for the simultaneous amplification of the SNP locus. However, unlike the primer extension approach, AS-PCR is less suited for multiplexing. The data obtained from AS-PCR amplification is binary (i.e. presence or absence of a band). Since the internal primer is designed to detect only one nucleotide out of four possible bases, the exact allelic identity of a locus cannot be assigned in the absence of amplification product in one genotype. For example if the internal primer has a G at its 3' end it will only amplify the C allele. The remaining three possible alleles will be marked by the absence of amplification. This limit of allelic discrimination is a disadvantage compared to the multiplex primer extension method in which all four allelic states can be distinguished by fluorescent

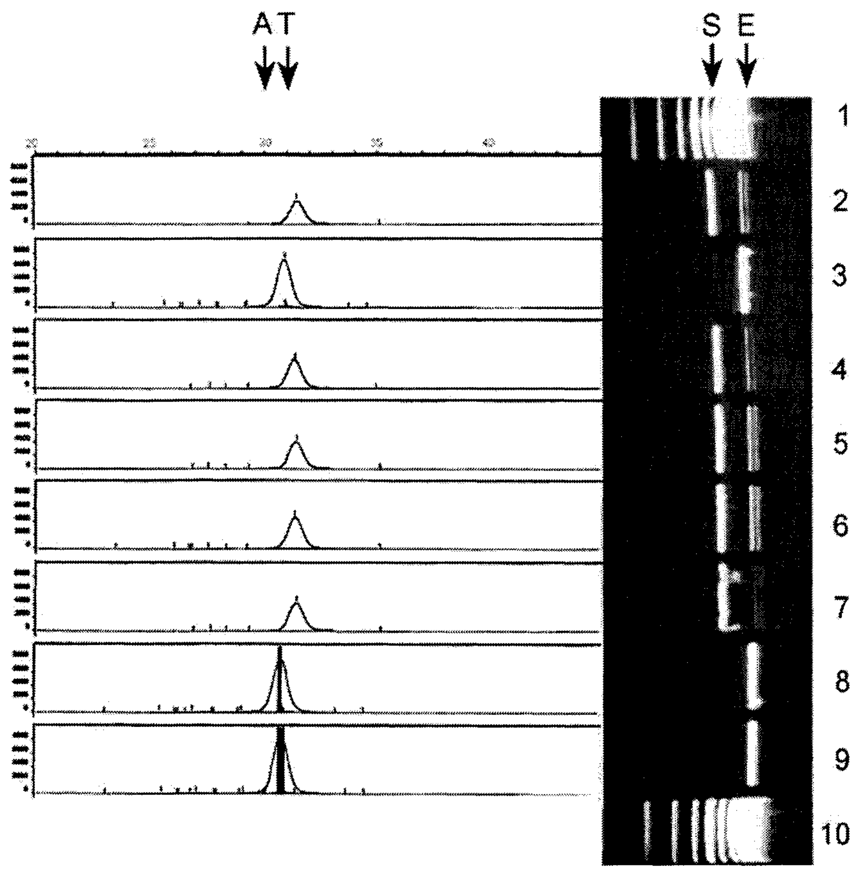
ddNTPs. This could be overcome by simultaneously designing another primer for the other allelic state at that locus. Furthermore, to detect the four possible nucleotide states in the population, one could design four primers with differentiating tail ends, which would increase the cost.

In summary, AS-PCR and agarose gel electrophoresis were as efficient as single nucleotide primer extension with capillary gel electrophoresis for the validation of SNPs (Fig. 3.3). However, with AS-PCR there was a significant reduction in time and resources required due to a reduction in the number of steps needed for SNP validation. This low cost method can be used for high throughput SNP validation.

3.6 Acknowledgments

The authors would like to thank Jean Mellish and Tigest Demeke of the Grain Research Laboratories, Canadian Grain Commission Winnipeg, Manitoba for providing some plant material. We would also like to thank Doug Procunier and Mark Gray (Agriculture & Agri-Food Canada, Cereal Research Centre Winnipeg, Manitoba) for the use of the ABI PRISM® 3100 Genetic Analyzer. This project was supported by a grant from Automated Quality Testing Inc. (AQT) to BRB. The experiments comply with the laws of Canada.

Figure 3.3 Comparison of genotyping with primer extension (left) and AS-PCR (right) for SNP066 on eight barley cultivars using EST21L 5'ATCAATGGAGATTTGCTTAC3', EST21R 5'GTGTTTACATGCTTGTCATA3'. The nested primer was 5'TGAAGCTGTTCAAACTAGAGCA3', and primer extension oligo was 5'TGAAGCTGTTCAAACTAGAGC3'. Lanes 1 and 10 are DNA ladders, lanes 2-9 are individual samples from eight different barley cultivars of breeder's seed origin. A and T are alleles of SNP066, S indicates presence of the "T" allele whereas the absence indicates the other three possible alternatives, E indicates the amplification of the EST locus and also serves as positive control.



A T
S E
1
2
3
4
5
6
7
8
9
10

CHAPTER FOUR

Analysis of genetic diversity in barley cultivars reveals incongruence between S-SAP, SNP and pedigree data

The following chapter is a reproduction of a paper accepted for publication in the Journal of Genetic Resources and Crop Evolution by Vahab D. Soleimani, Bernard R. Baum and Douglas A. Johnson. The chapter fulfills objective 1.2 of the thesis. The extent of genetic diversity/similarity among a group of barley cultivars was investigated by a comparative study of S-SAP, SNP and kinship coefficient. Polymorphic S-SAPs and SNPs were used to construct genetic similarity matrices using population genetics tools. Retrotransposon-based markers were found to be more polymorphic and more useful than SNPs for revealing genetic relationships between closely related cultivars. S-SAPs also showed significantly higher level of intra-cultivar variation than SNPs. The results of this paper indicate that the BARE-1 retrotransposition in the barley genome has a major role in the generation of DNA sequence variation among cultivars.

4.1 Abstract

Accurate assessment of genetic similarity is important for plant breeding, germplasm enhancement and conservation of plant genetic resources. A comparative analysis of genome diversity among a group of six-rowed spring barley (*Hordeum vulgare* L.) cultivars was carried out using sequence-specific amplified polymorphism (S-SAP) and single nucleotide polymorphism (SNP), with the results compared to the kinship coefficients derived from the pedigree data. Mean pair-wise GS values were estimated to be 0.0957 ± 0.144 (Kinship), 0.491 ± 0.189 (SNPs), and 0.602 ± 0.098 (S-SAPs). S-SAP and SNP-based genetic similarity (GS) values were normally distributed but kinship values had a non-normal and skewed distribution. Pair-wise correlation of GS values were lowest for the S-SAP and the SNP matrices ($r = 0.040$, $p < 0.230$) and highest for the SNP and pedigree matrices ($r = 0.240$, $p < 0.001$). AMOVA attributed about 90.4% of observed variation to the cultivars within each of the malting and feed groups. Variance component between malting and feed groups was 6.6% for both SNP and S-SAP data suggesting lack of a significant genetic differentiation along this agronomic division. The remaining 3% of variation was attributed to genetic diversity within cultivars. Although both DNA-based marker systems were in agreement with regard to the partitioning of the variation, significant differences were observed in the pattern of genetic relationships obtained by the two marker systems and the pedigree data.

Key words Barley, genetic similarity, genome evolution, kinship coefficient, S-SAP SNP

4.2 Introduction

In the absence of whole genome sequencing molecular markers remain as indispensable tools for the studies of genome. Given the specific underlying evolutionary mechanism responsible for each marker system, the pattern and extent of genetic variation resulting from different markers system is expected to vary. Some comparative analyses between different classes of marker systems have been carried out in the past. Powell et al (1996) studied the utility of AFLP, RAPD, RFLP, and SSRs in soybean (*Glycine max* L. Merrill) and concluded that all four markers systems had different properties and the genetic relationships derived from them were different. A comparative analysis of RFLPs, RAPDs, SSRs, and AFLPs in maize (*Zea mays*) (Pejic et al. 1998) found that with the exception of RAPDs, the GS trees found with other marker systems were highly correlated. In barley (*Hordeum vulgare* L.), comparison of random amplified microsatellite polymorphisms (RAMP) and SSRs showed that the two marker systems predicted different genetic relationships among *Hordeum* accessions (Davila et al. 1999).

Molecular marker-based genetic diversity estimates may be affected by several factors such as the number of markers used, their genomic distribution, and the nature of evolutionary mechanisms giving rise to the polymorphism (Powell et al.1996). Comparative analysis of genetic diversity between various classes of molecular markers and pedigree data have also been carried out for many cereals crops including barley (Tinker and Mather 1993; Davila et al. 1997; Russell et al. 1997), maize (Smith et al. 1990; Smith and Smith, 1992), and wheat (*Triticum aestivum* L.) (Autrique et al. 1996; Kim and Ward 1997; Barrett et al. 1998; Soleimani et al. 2002a). In crop species studied to date, correlations between molecular markers and pedigree data have varied widely

depending on the organism and the marker system used. Pedigree-based genetic diversity estimates are rough measures of relatedness. Their accuracy depends partly on the availability of a comprehensive pedigree record for each cultivar. Also, the various assumptions used in the pedigree analysis regarding selection pressure, genetic drift, linkage and relatedness of parents with unknown pedigrees could greatly underestimate GS level among plant cultivars based on pedigree analysis (Cox et al. 1985; Graner et al. 1994; Kim and Ward 1997). In barley, comparative analysis between microsatellite polymorphic DNA and pedigree data (Davila et al. 1998) and RFLPs and pedigree data (Graner et al. 1994) showed a low correlation ($r = 0.21$, $p < 0.05$) between the two diversity indices. A similar trend was observed between RFLPs and pedigree data in durum wheat (Autrique et al. 1996). Moderate correlation between AFLPs and pedigree data were reported for durum wheat (Soleimani et al. 2002) whereas a high correlation was detected between RFLPs and pedigree data in maize (Smith et al. 1990).

In this study, we have used SNPs from the non-repetitive genomic regions of barley and S-SAPs (Waugh et al. 1997) derived from BARE-1 retrotransposon of barley to infer the genetic relationships among 35 cultivars of six-rowed barley representing various agronomic groups (Table 4.1). These results were also compared to pedigree information that was compiled from the cultivar description and the published pedigree records (Baum et al. 1985). SNPs are nucleotide variations in the DNA sequences of individuals within a population. They are mainly the result of replication error introduced by DNA polymerase during DNA replication and as such their occurrence is expected to be dependent on the mutation rate for the specific genomic loci from which they are derived. The occurrence of S-SAP polymorphisms is expected to depend in part on the level of

Table 4.1 Spring six-rowed barley cultivars used in this study. D: seed morphology (hulled, d; or hulless, h), U: usage (malting or feed), /: primary cross, //: secondary cross, /3/: tertiary cross, number preceding * indicates backcrosses; S_b, SNP-based number of haplotypes; P_b, S-SAP-based number of haplotypes.

Cultivar	S_b	P_b	D	U	Parentage	Released
AC Albright	1	3	d	f	Otra/6/(Weibull1514-64) Morgenrot/5/Tammi/4/Maja/3/Opal//Hanna/Svanhals	1992
AC Alma	1	4	d	f	Chapais/Leger	1996
AC Bacon	1	7	h	f	Tupper/Johnston//Conquest/3/Abee/4/Ellice/Bedford	1998
AC Burman	1	2	d	f	Leger/Bruce/2/2*Leger	1991
AC Hawkeye	2	2	h	f	GO1-1/Tupper/3/Virden//Conquest/Post	1996
AC Klinck	2	4	d	f	Cadette/Chapais	2000
AC Malone	1	4	d	f	Callus culture of Leger	1999
Argyle	1	3	d	m	Herta/UM 570//conquest/3/Bonanza	1981
B1602	1	4	d	m	Bumper/6B78-628//Morex/6B78-628	1991
Bedford	1	6	d	f	Keystone/4/Vantage/Jet//Vantmore/3/2* Husky/5/Cree	1979
Bonanza	3	4	d	m	Vantage/Jet//Vantmore/3/2* Parkland/4/Conquest	1970
Cultivar	S_b	P_b	D	U	Parentage	Released

Cultivar	S_b	P_b	D	U	Parentage	Released
Brier	1	2	d	f	Leduc///Galt/York//Dickson/Galt	1989
Cadette	3	8	d	f	QB 139.7/(Min660102/Bonanza)	1986
CDC Sisler	2	2	d	m	M34/Argyle	1996
Chapais	1	2	d	f	(QB 58.14/Beacon)/BT 904	1988
Duel	1	2	d	m	Morex//6B75-1374/M31	1992
Excel	1	8	d	m	Cree/Bonanza//Manker/3/2* Robust	1994
Falcon	1	5	h	f	11012.2/Tern//Tulelake	1992
Foster	6	3	d	m	Robust/6/Glenn/4/Nordic//Dickson/Trophy/3/Azure/5/Glenn/Karl	2000
Gamine	1	1	d	f	Maskot/Chapais	1996
Johnston	1	2	d	f	Klondike/5/Nord/3/Vantage/Jet//Vantmore/4/Bonanza	1980
Klondike	1	2	d	f	Galt/NDB133 Vantage/Jet//Vantmore/3/2* Parkland/4/Dickson	1976
Lacey	2	8	d	m	M44/Excel//M46/M44/3/M44/Excel//Stander	2001
Leduc	1	2	d	f	Steptoe/Klondike	1982
Legacy	1	5	d	m	6B86-3517/Excel	2001
Cultivar	S_b	P_b	D	U	Parentage	Released

Cultivar	S_b	P_b	D	U	Parentage	Released
Leger	1	2	d	f	Trent/Vanier	1982
Maskot	1	2	d	f	QB 167.21/OB193.11	1989
Noble	1	2	d	f	Dickson/3/CIho 4738//Traill/UM 570	1987
Nord	1	1	d	f	Olli/Byng	1956
Otal	1	2	d	f	Otra/breeding line of Weibullsholm/3/Hanna/Svanhals//Opal/4/Tammi/5/Morgenrot	1982
Peregrine	2	1	h	f	H12-4816/R181//M69.77-SHI.R.KCI.NO.87/CEL-5106	1999
Robust	1	2	d	m	Morex/Manker	1994
Stander	2	2	d	m	Excel//Robust/Bumper	1999
Tankard	4	3	d	m	Argyle/Minnesota M-34	1992
Virden	1	3	d	f	WA6415-66//Bonanza/NDB136/3/UM67-739R//Bonanza/Dickson	1987

past transpositional activity of the element targeted for S-SAP screening in addition to retrotransposon-related genome rearrangements such as inversions, translocations, and deletions. The BARE-1 family of repetitive elements is a major component of barley genome (Manninen and Schulman, 1993). The family was shown to comprise on average 14 x 10^3 copies in the genomes of *Hordeum* species and cultivars (Vicient et al. 1999) with a wide genome distribution along all barley chromosomes (Waugh et al. 1997). Members of this family have been shown to actively transpose (Jaaskelainen et al. 1999; Suoniemi et al. 1996) resulting in new polymorphisms among cultivars (Gribbon et al. 1999; Waugh et al. 1997; Kalendar et al. 1999). BARE-1 copy number and transposition was shown to be correlated with microclimatic differences among various *H. spontaneum* accessions grown in various habitats along a transect (Kalendar et al. 2000). These results indicate a direct correlation between retrotransposon-mediated genome rearrangements and environmental and abiotic factors, supporting a body of evidence that links transposition to various environmental stresses imposed on the host organisms (Hirochika 1993; Hirochika et al. 1996; Pearce et al. 1996; Wessler, 1996; Tadeka et al. 1998; Bennetzen 2000).

The aims of this study were to compare the extent of population genetic estimates among spring six-rowed barley cultivars using SNPs and S-SAPs, and to compare the results to the kinship coefficients obtained from the pedigree data to test the null hypothesis that there is no difference between diversity estimates derived from the SNPs, the S-SAPs and the pedigree data.

4.3 Materials and Methods

4.3.1 Plant material and DNA extraction

Breeders' seeds of the 35 six-rowed spring malting barley cultivars (Table 4.1) were obtained from the Canadian Grain Commission (CGC) Winnipeg, Manitoba. Additional barley seed samples were obtained from the Canadian Food Inspection Agency (CFIA) Ottawa, Ontario.

Barley seeds were germinated on moist Petri dishes at room temperature for one week. 100 mg of fresh leaf material was harvested from each seedling for DNA extraction. Leaf samples were placed in 1 ml collection tubes (96-well plate format, Qiagen®) containing a 3 mm carbide tungsten bead each. The plates were submerged in liquid nitrogen to freeze the plant material and were then immediately mounted on an MM 300 Mixer Mill (Retsch®). The frozen leaf samples were powdered by shaking the plates at 30 revolutions per minute for 20 seconds. DNA was extracted from individual samples using DNeasy® 96 Plant Kit (Qiagen®) following manufacturer's recommendations.

4.3.2 PCR primers

The long terminal repeat (LTR) based primers that were used in this study were selected from a pool of primers designed from BARE-1-like LTR sequences available at the NCBI (www.ncbi.nlm.nih.gov). Published LTR-derived primers (Waugh et al. 1997; Kalendar et al. 2000) were also synthesized and evaluated for their use on barley material

in this study. A pre-screening assay was carried out to select the best primer combinations for final S-SAP analysis (Table 4.2).

4.3.3 S-SAP analysis

S-SAP procedure was carried out based on Waugh et al (1997). DNA restriction was carried out using *Pst*I and *Mse*I restriction endonucleases. P*st*I restriction enzyme was chosen for two reasons: 1) it is methylation sensitive and therefore it will likely cut hypomethylated sites within euchromatin and 2) it does not have a recognition site within the BARE-1 retrotransposon (based on the sequence analysis of BARE-1 accession Z17327). This ensures that most S-SAP products contain host DNA sequence at the *Pst*I end of the fragment and the retrotransposon sequence at the *Mse*I end. Synthetic double stranded DNA adapter molecules (Table 4.2) were ligated to the ends of the restricted fragments using T4 DNA ligase (Invitrogen™). Ligation reactions were carried out at room temperature overnight. S-SAP PCR reactions were carried out in two steps: a preamplification PCR with primers that were complementary to the adapter sequences and that contained a selective cytosine base at their most 3'end. This reaction was required to reduce the number of amplified restriction fragments. The product of the preamplification reaction was diluted 50 fold in TE buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA) and subsequently used as template for final S-SAP PCR using a selective *Mse*I primer in combination with an LTR primer derived from BARE-1 retrotransposon of barley (Table 4.2). The LTR primer was labelled with γ^{33}PdATP for autoradiography. S-SAP PCR products were mixed with equal volumes of denaturing dye (98% de-ionized formamide, 0.025% bromo-phenol-blue, 0.025% xylene cyanol) and denatured at 94 °C

Table 4.2. DNA sequence of synthetic adapters and PCR primers that were used in the S-SAP analysis and the number of loci that were assayed.

Primer/adapter mnemonics	Sequence 5'to 3'
	BARE-1 LTR primers
R1	GTTATGTAGTGGGCGAGCGAG
R4	GCATACATGTGTCCTTCGTTGCC
	*Mse*I selective primers
M2	GATGAGTCCTGAGTAACAC
M3	GATGAGTCCTGAGTAACAG
M4	GATGAGTCCTGAGTAACAT
M7	GATGAGTCCTGAGTAACTG
	Adapters
*Pst*Ia	CTCGTAGACTGCGTACATGCA
*Pst*Ib	TGTACGCAGTCTAC
*Mse*Ia	GACGATGAGTCCTGAG
*Mse*Ib	TACTCAGGACTCAT
	Preamplification primers
*Pst*I primer	GACTGCGTACATGCAGC
*Mse*I primer	GATGAGTCCTGAGTAAC

Primer pair combination	Number of polymorphic bands
R1M4	54
R4M3	25

Primer/adapter mnemonics	Sequence 5'to 3'
R1M7	36
R1M2	35
Total	150

for 3 minutes. The final products were resolved in a 5% denaturing polyacrylamide gel and electrophoresed at 80 W (constant power) for 2.5 hours. The gels were dried and exposed to Kodak X-Omat film for 3-5 days at -80 °C temperature. DNA fingerprints were evaluated by visual inspection of autoradiographs.

4.3.4 SNP analysis

4.3.4.1 Detection

SNPs were discovered by sequencing and in-*silico* (data mining) methods. In the sequencing phase, a set of previously mapped barley genomic loci (Kleinhofs et al. 1993; Langridge et al. 1995) was sequenced from populations of 5-8 barley cultivars. PCR primer sequences for these loci were obtained from the Montana State University web site: (*http://hordeum.oscs.montana.edu*). DNA sequencing was carried out on PCR products using the Sanger dideoxy-chain terminator method (Sanger et al. 1977) on an ABI PRISM 377 DNA sequencer (Perkin Elmer, Applied Biosystems) with ABI Prism Dye terminator Cycle Sequencing Ready Reaction Kit using AmpliTaq DNA polymerase. Multiple alignments of sequences for each locus were carried out using Clustal W (Thompson et al. 1994).

For the *in-silico* SNP discovery phase, barley EST sequences that were available publicly (*www.ncbi.nlm.nih.gov/dbEST/index.html* and *http://www.genome.clemson.edu/projects/barley/HV_CEb/library.html*) were retrieved and assembled into contigs using Contig Assembly Program (CAP3) available at *http://fenice.tigem.it/bioprg/interfaces/cap3.html*. SNP polymorphisms were detected by

visual inspection of aligned sequences as described earlier. Specific primers were designed from the ends of contig and combined with allele-specific primer for each potential SNP for PCR-based validation on barley genomic DNA as descried below.

4.3.4.2 Validation

SNPs were validated by either of the two following methods:

1) Primer extension with Dideoxy-nucleotides using ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystem). Primer extension reactions were multiplexed (up to 5-plex). Allele discrimination was made possible by multiple colours of fluorescence dye with an ABI 3100 Genetic Analyzer (Applied Biosystem). Size separation of the primer extension product in a multiplex reaction was carried out by adding a random sequence to the 5' end of SNP primer. The random sequence did not have similarity to any barley genomic sequences available up to date.

2) Allele-Specific PCR was used with genomic DNA as template for a single PCR reaction with three primers. Two of these primers flanked the SNP of interest and the third primer was an allele-specific for the SNP position (Soleimani et al. 2003). This method is analogous to sequence tagged site (STS).

4.3.5 Data analysis

4.3.5.1 Data acquisition and scoring

DNA fingerprints were detected visually and scored qualitatively as binary variables, 1 for the presence and 0 for the absence of bands at polymorphic sites respectively. For

the S-SAP analysis bands whose presence or absence could not be clearly discerned were ignored in the final analysis.

4.3.5.2 GS estimation

The matrix of binary data were used to generate a GS matrix based on Dice's (1945) formula that is equivalent to Nei and Li's (1979) definition of similarity: $Sij = 2a / (2a + b + c)$ where Sij is the GS between two individuals, i and j, a is the number of bands present in both i and j, b is the number of bands present in i but absent in j, and c is the number of bands absent in i but present in j. Gene diversity (Weir 1990) for the SNP data was calculated as $1 - \Sigma P^2_{ij}$ where P_{ij} is the frequency of the jth allele for i^{th} locus summed across all alleles in the locus. Cluster analysis was carried out on the resulting matrix of GS using NTSYS-pc, ver. 2.1 (Rohlf 2000). In the instances when there were ties in the dendrograms generated from data, cophenetic values were calculated for each tree file and a matrix correlation analysis was performed using the cophenetic and GS matrix values to test the level of association between the two matrices.

4.3.5.3 AMOVA

Variance components for barley cultivars were determined by AMOVA with Arlequin© (Schneider et al. 1997). AMOVA analysis was carried out based on the following hierarchical division: variation among groups (i.e. feed and malting), variation among cultivars within each of the two groups, and variation within each cultivar. Individual genotypes were treated as completely homozygous for all loci based on the

assumption of complete inbreeding and multiple rounds of selfing in barley breeding programs.

4.3.5.4 Matrix comparison and Mantel test

GS matrices of S-SAP, SNP and kinship values were used for a pair-wise comparison in order to assess the degree of association between pairs of matrices using the MAXCOMP routine in the NTSYS-pc (Rohlf 2000). The normalized Mantel statistics Z (Mantel 1967) was used to determine the level of association between the matrices. The association was considered significant if $r \geq 0.50$, $P < 0.01$.

4.3.5.5 Pedigree analysis

Information regarding the known ancestors of barley cultivars was obtained from cultivar descriptions, breeding records, published pedigree databases (Baum et al. 1985), personal communication with breeders and the Germplasm Resource Information Network (http://www.ars-grin.gov/cgi-bin/npgs/html/csr.pl?BARLEY). Kinship coefficients between cultivars were calculated using BigKin87© as described in Tinker and Mather (1993).

4.4 Results

4.4.1 Genetic diversity as revealed by S-SAP

Four primer pair combinations were used to score 150 polymorphic S-SAP bands from a total of 400 amplified bands (Table 4.2). R1 and R4 primers were designed from the LTR region of BARE-1 retrotransposon and combined with various *Mse*I selective primers for an anchored AFLP (S-SAP) (Waugh et al. 1997) analysis (Fig. 4.1). S-SAP-based GS ranged from a minimum of 0.326 to a maximum of 0.915. The values were normally distributed with a mean of 0.601 and standard deviation 0.096 (not shown). UPGMA dendrogram (Fig. 4.2) shows the genetic relationships among cultivars based on the GS values derived from the S-SAP analysis. There is an apparent overlap between feed and malting cultivars. Lack of a clear division between the two groups was also supported by the AMOVA analysis (Table 4.3). This finding maybe partially due to the limited number of loci involved in malting characteristics compared to the total number of genomic loci and also the existence of close kinships between some malting and feed cultivars (Table 4.1). Three of the four hulless feed cultivars, 'Falcon', 'AC Hawkeye' and 'Peregrine', clustered closely in the S-SAP dendrogram (Fig. 4.2) but the fourth 'AC Bacon' formed a separate cluster with 'Cadette', a hulled feed cultivar.

4.4.2 Pedigree analysis

The ancestry of 35 barley cultivars used in this study was traced back to 242 ancestors with a known pedigree plus 92 lines and landraces for which no pedigree record was available. Mean pair-wise GS between pair of cultivars was estimated at 0.097 ±

Table 4.3. AMOVA analysis with S-SAP and SNP data for partitioning the variance. d.f., degrees of freedom; Va, variance among groups; Vb, variance among cultivars within groups; Vc, variance within cultivars.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	**Φ**-Statistics
			SNPs		
Among groups	1	91.419	0.38903 Va	6.67	$F_{CT} = 0.066$
Among cultivars	33	1397.053	5.27076 Vb	90.43	$F_{SC} = 0.968$
Within cultivars	245	41.375	0.16888 Vc	2.90	$F_{ST} = 0.971$
Total	279	1529.846	5.82866	100	
			S-SAPs		
Among groups	1	251.302	1.07958 Va	6.66	$F_{CT} = 0.066$
Among cultivars	33	3838.680	14.51355 Vb	89.50	$F_{SC} = 0.958$
Within cultivars	245	151.911	0.62258 Vc	3.84	$F_{ST} = 0.961$
Total	279	4241.892	16.21572	100	

Figure 4.1. Autoradiogram of a portion of an S-SAP fingerprint using R1/M2 primer pair combination. In this experiment, the LTR-derived R1 primer was radiolabelled with γ^{33}PATP for autoradiographic detection. Columns 1, 2 and 3 represent 'Leduc', 'Leger' and 'Maskot' cultivars respectively. Each column has eight lanes containing fingerprint of a single plant. The last lane on the right is DNA size markers in bp. Cultivar-specific S-SAP polymorphisms can be seen at 247 bp, 190 bp, 165 bp, 157 bp, 152 bp, 145 bp, 142 bp, 140 bp, 132 bp, and 127 bp. The LTR-derived R1 primer was radioalbelled with ^{33}PATP for autoradiographic detection.

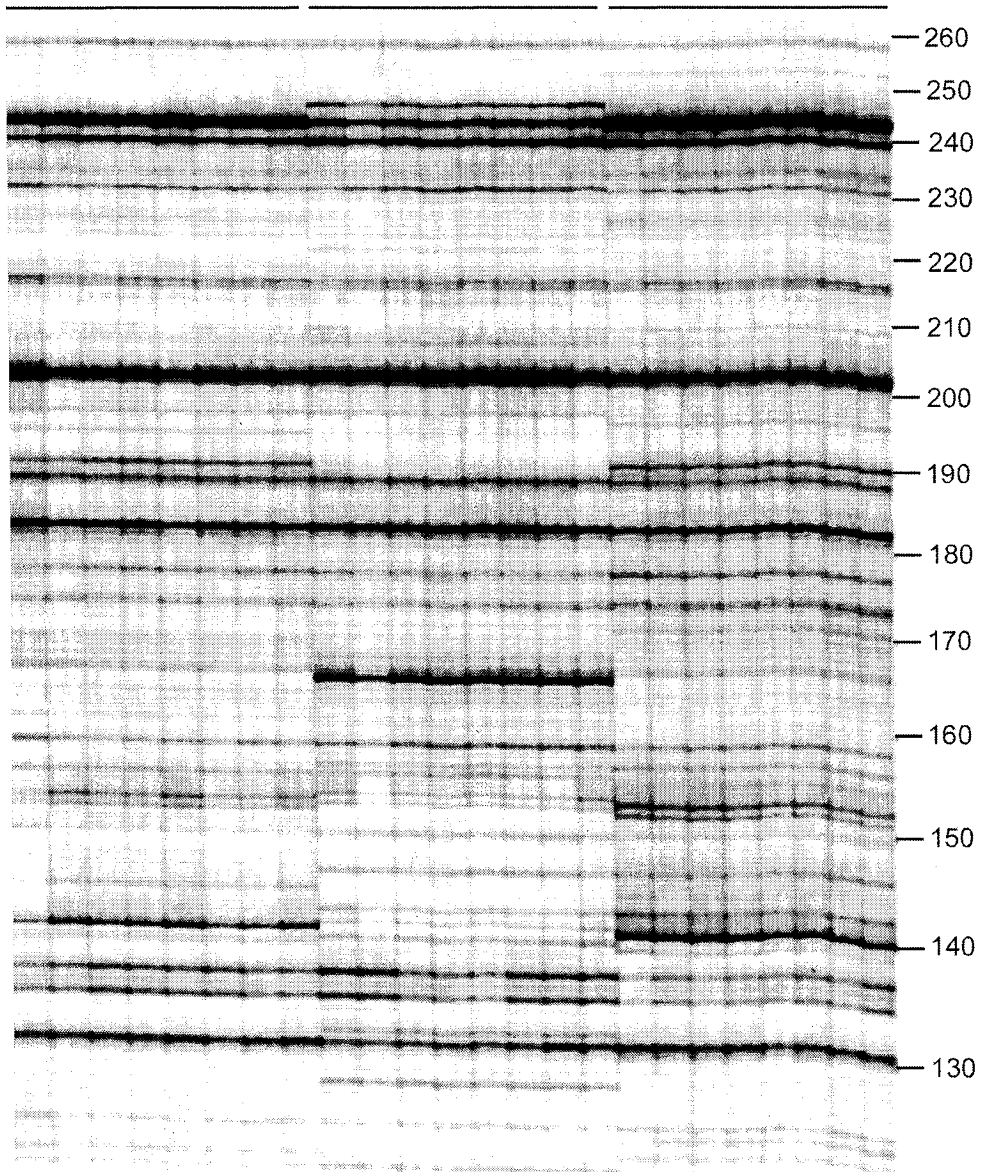
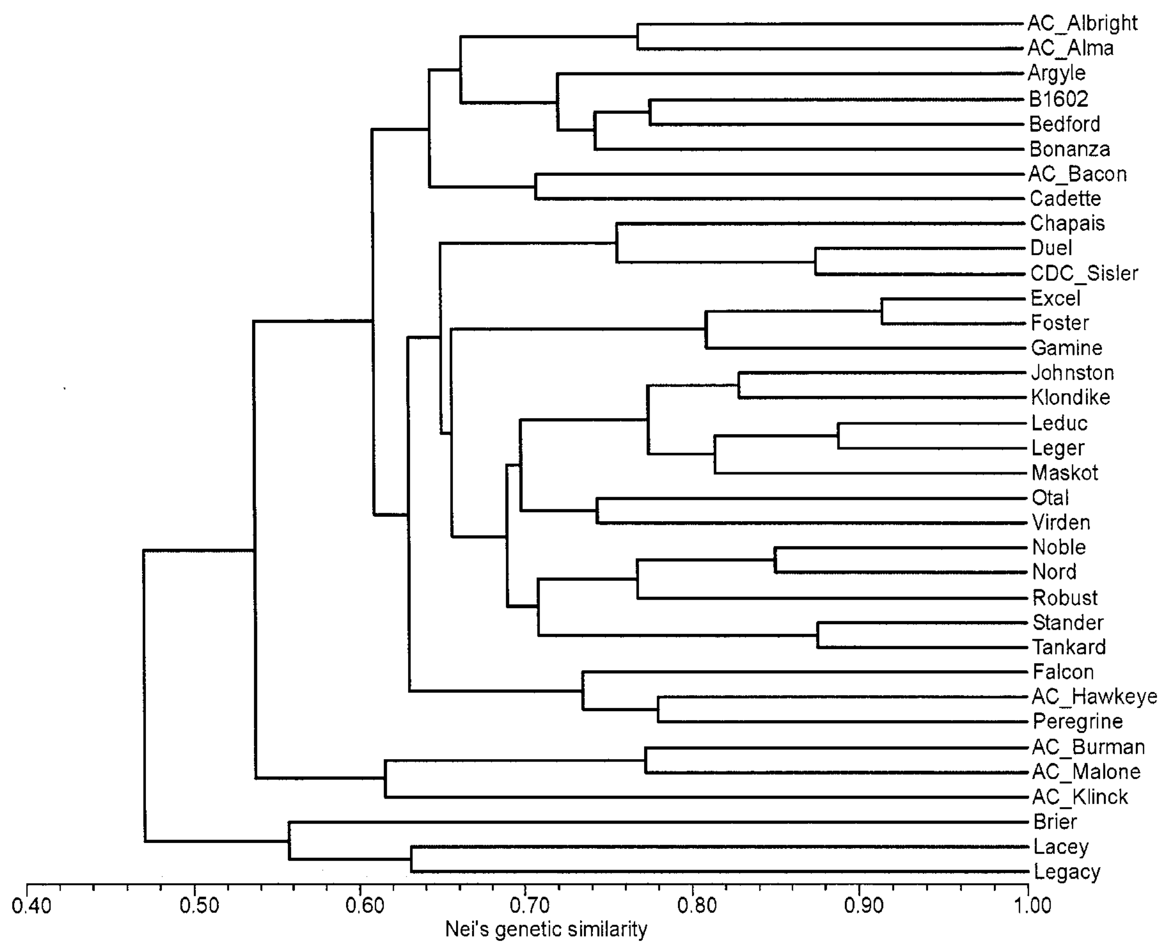

1
2
3
260
250
240
230
220
210
200
190
180
170
160
150
140
130

Figure 4.2. UPGMA dendrogram showing genetic relationship among barley cultivars based on Nei's GS values that were calculated from the S-SAP data.


AC_Albright
AC_Alma
Argyle
B1602
Bedford
Bonanza
AC_Bacon
Cadette
Chapais
Duel
CDC_Sisler
Excel
Foster
Gamine
Johnston
Klondike
Leduc
Leger
Maskot
Otal
Virden
Noble
Nord
Robust
Stander
Tankard
Falcon
AC_Hawkeye
Peregrine
AC_Burman
AC_Malone
AC_Klinck
Brier
Lacey
Legacy
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Nei's genetic similarity

0.144. Kinship values ranged from one between ('Leger' and 'AC Malone', derived from callus culture of 'Leger') to zero (no shared common ancestors). A matrix of kinship coefficients was constructed for the 35 barley cultivars (not shown). A large number of entries with kinship values equal 0 skewed the distribution of kinships towards lower values. The relationship among cultivars was depicted by a UPGMA dendrogram (Fig. 4.3) in which all the malting cultivars with the exception of 'B1602' formed a sub-cluster but the feed cultivars constituted a diverse group with genetic similarities ranging from 0 to 1.

4.4.3 SNP analysis

SNP polymorphisms among barley cultivars were identified by sequencing a set of previously mapped barley genomic loci (Kleinhofs et al. 1993; Langridge et al. 1995). Sequence analysis of these loci (Table 4.4) using BLASTn search showed no similarity to known repetitive elements of barley (not shown) indicating that the detected SNPs were independent of the S-SAPs in this study. From an initial pool of 214 putative SNPs that resulted from sequencing and data mining combined (not shown), we were able to validate 27 SNPs either by allele-specific PCR (Fig. 4.4) or by the primer extension method (Fig. 4.5) as described by Soleimani et al (2003). Preliminary analysis of the 214 putative SNPs showed that 133 SNPs (62%) were transitions with the remaining 38% constituting transversion and small insertions/deletions (indels). However, among the 27 validated SNPs (Table 4.4), transitions accounted for 13 out of 27 (48.2%), transversions 12 out of 17 (44.4%), and indels 2 out of 27 (7.4%). Cluster analysis with SNP data produced a dendrogram (Fig. 4.6). At the lowest level of GS (i.e. at phenon line of 0.40)

Figure 4.3. UPGMA dendrogram showing genetic relationship among 35 barley cultivars. The dendrogram was constructed using GS values inferred from kinship coefficient based on pedigree information.

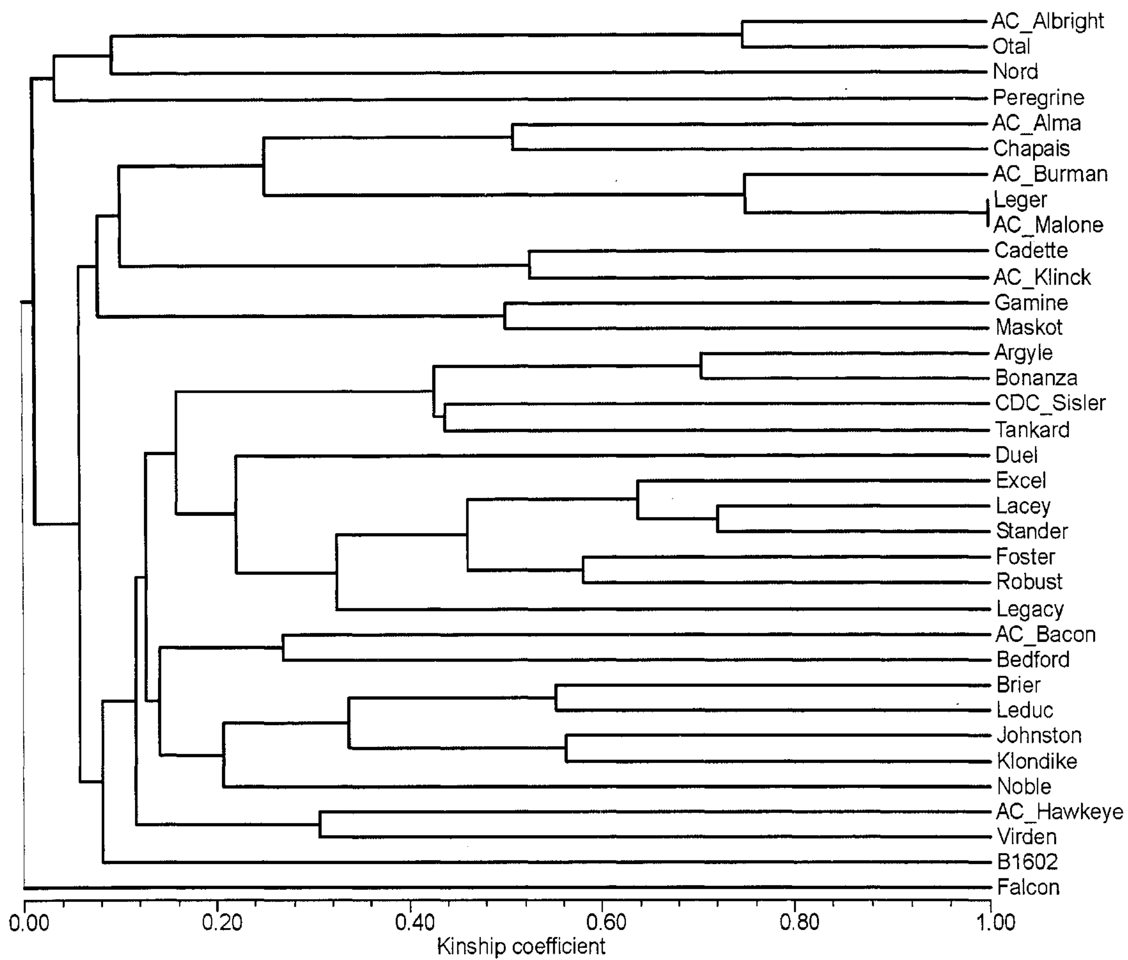

AC_Albright
Otal
Nord
Peregrine
AC_Alma
Chapais
AC_Burman
Leger
AC_Malone
Cadette
AC_Klinck
Gamine
Maskot
Argyle
Bonanza
CDC_Sisler
Tankard
Duel
Excel
Lacey
Stander
Foster
Robust
Legacy
AC_Bacon
Bedford
Brier
Leduc
Johnston
Klondike
Noble
AC_Hawkeye
Virden
B1602
Falcon
0.00
0.20
0.40
0.60
0.80
1.00
Kinship coefficient

Figure 4.4. Agarose gel electrophoresis showing allele-specific PCR detects a G/T transversion between two malting cultivars. Group 1 samples are plants from cv 'Robust' and group 2 samples are from 'Stander'. The higher molecular band is a 700 bp genomic region corresponding to a contig that contained the SNP of interest. The lower molecular weight band (400 bp) is SNP-specific amplicon. The presence of the latter band in 'Stander' indicates T allele and its absence in Robust indicates G or A or C allele.

1
2
1000
500

Figure 4.5. Multiplex detection of SNPs with the primer extension using SNaPShot™ multiplex kit (Applied Biosystems). Allele separation is achieved by adding a random tail of varying length to primers and multiple colours of fluorophor. Panel 1, 2 and 3 represent 'Robust', 'Stander' and 'Foster' cultivars respectively. The scale on top shows the size of primer extension products.

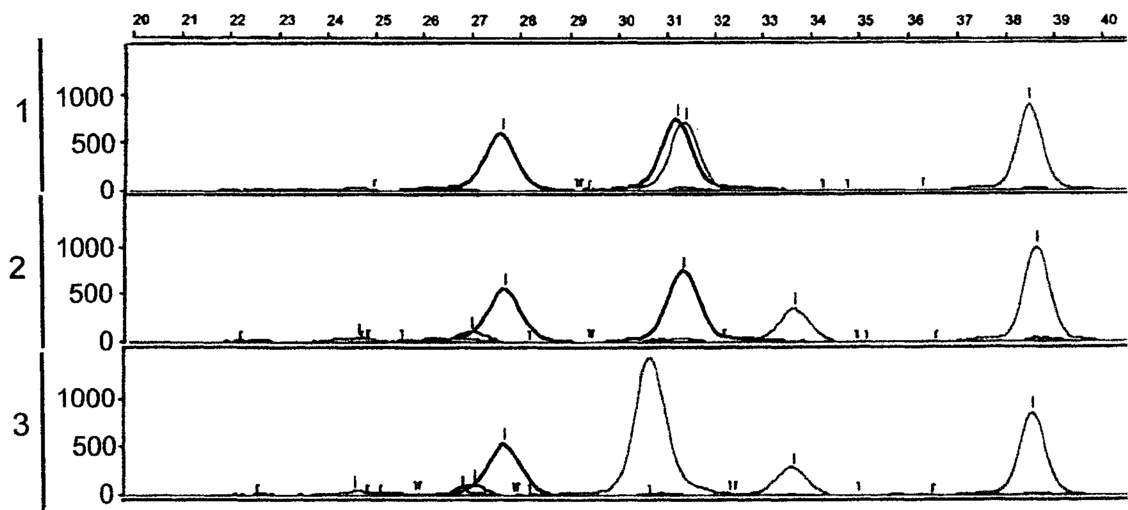

20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
1
2
3
1000
500
0

Figure 4.6. UPGMA dendrogram showing genetic relationship among barley cultivars based on Nei's genetic GS calculated from SNP data.

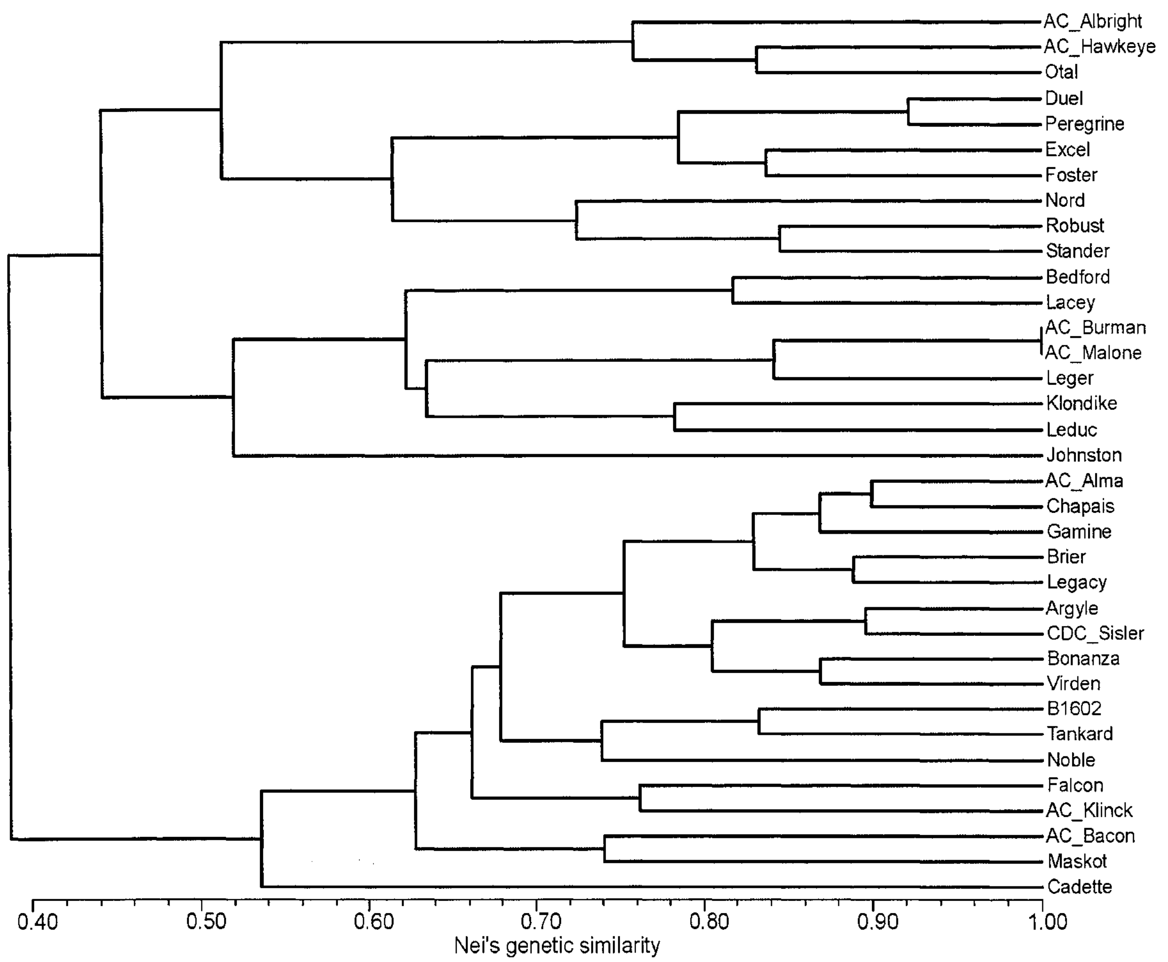

AC_Albright
AC_Hawkeye
Otal
Duel
Peregrine
Excel
Foster
Nord
Robust
Stander
Bedford
Lacey
AC_Burman
AC_Malone
Leger
Klondike
Leduc
Johnston
AC_Alma
Chapais
Gamine
Brier
Legacy
Argyle
CDC_Sisler
Bonanza
Virden
B1602
Tankard
Noble
Falcon
AC_Klinck
AC_Bacon
Maskot
Cadette
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Nei's genetic similarity

Table 4.4. List of SNP alleles found within their respective loci and indices of their gene diversity within cultivars, their allelic nature and method of validation. AP: allele-specific PCR; PE: primer extension.

Locus Name	Allele	Gene diversity	Validation
ABC153-30	C/G	0.454	AP
ABC153-31	G/T	0.494	AP
ABC302-29	C/A	0.472	AP
ABG058-42	C/T	0.495	AP
ABG058-43	C/T	0.492	AP
ABG058-44	T/G	0.477	AP
ABG058-45	Indel #1	0.475	AP
ABG058-46	C/T	0.480	AP
ABG058-47	A/T	0.446	AP
ABG500-01	A/G	0.178	AP
ABG601-17	A/G	0.405	AP
ABG601-18	A/G	0.432	AP
ABG609-66	A/T	0.490	PE
AL503851-81	G/C	0.465	AP
MWG2218-24	A/C	0.495	AP
MWG2218-25	C/G	0.472	AP
MWG2218-26	G/C	0.499	AP
MWG2218-27	C/T	0.475	PE
MWG502-20	C/T	0.461	AP

Locus Name	Allele	Gene diversity	Validation
MWG502-22	A/G	0.457	AP
MWG502-57	A/G	0.363	PE
MWG502-59	C/T	0.479	AP
MWG502-61	Indel #2	0.391	AP
MWG502-75	T/A	0.360	AP
MWG502-77	A/G	0.420	AP
MWG502-78	A/G	0.500	AP
215	G/T	0.498	AP

two distinct clusters were formed with no apparent correspondence to the division of cultivars along any common agronomical characteristics (i.e. feed/malting, hulled/hulless) or breeding locations.

4.4.4 Low rate of SNP validation with PCR analysis

From a total of 214 putative SNPs which included 86 from sequencing and 128 from *in-silico* method only 27 (13%) were found to be valid by PCR analysis. Among the putative SNPs derived from the *in-silico* method only one SNP was validated successfully, indicating a success rate of less than 1%. Among the putative SNPs derived from sequencing, 26 out of 86 (30%) were successfully validated with PCR. Three of the PCR-validated SNPs were analyzed by primer extension whereas the remaining 24 SNP were validated by allele specific PCR – analogous to the sequence tagged site (STS).

4.4.5 AMOVA

The 280 plants representing the 35 barley cultivars were summarized into 117 haplotypes by S-SAP analysis or 57 haplotypes using SNP markers. SNP-based analysis showed that 10 out of 35 (28%) of cultivars had at least two haplotypes. This number increased to 32 out of 35 (91%) for the S-SAP data. Even though the number of haplotypes inferred from the S-SAP data was influenced by the larger S-SAP dataset compared to a smaller SNP dataset, the pattern of haplotypes distribution among cultivars was found to be unrelated between S-SAP and SNP dataset by correlation analysis (not shown).

AMOVA analysis identified the largest component of variation (about 90%) to cultivars within groups for both S-SAP and SNP data (Table 4.3). In other words most of

the observed variation was assigned to differences in the genetic make up of cultivars within the feed and malting groups. Among group percentage of variation was estimated at 6.67 and 6.66 for SNP and S-SAP respectively indicating that there is little genetic differentiation among barley cultivars on the basis of feed/malt division. The percentage of variation for individual plants within each cultivar ranged from 2.9 to 3.84% for SNP and S-SAP data respectively.

4.4.6 Matrix comparison

The degree of relationship between similarity matrices based on S-SAPs, SNPs and kinship coefficient was measured by a pair-wise comparison among the three GS matrices with a normalized Mantel Statistics. The lowest correlation ($r = 0.0410$) was observed between S-SAP and SNP matrices and a mild correlation ($r = 0.240$) between kinship coefficient and SNP matrices (Table 4.5). The goodness of fit, determined by the correlation analysis between matrix of GS values and a cophenetic matrix derived from the UPGMA tree of SNP, S-SAP, and kinship data showed correlation values of $r \geq 0.75$, $P < 0.001$ which was considered significant.

Table 4.5. Pair-wise matrix correlation analysis with Mantel test for kinship, S-SAP and SNP matrices.

		r-values	
		Kinship	S-SAP
1	Kinship		
2	S-SAPs	0.093 $p < 0.0906$	
3	SNPs	0.240 $p < 0.001$	0.041 $p < 0.762$

4.5 Discussion

A multitude of molecular marker systems have been developed over the past decade and used for the study of genetic diversity, to identify plant individuals, cultivars and species and to aid plant breeding (Saghai Maroof et al. 1994; Powell et al. 1996; Russell et al. 1997; Davila et al. 1998; Soleimani et al. 2002). An obvious advantage of molecular markers over traditional morphological characters is their sheer number in the genome. In addition, genetic analysis with molecular markers can be carried out at any stage of the plant development providing a fast and reliable study of the plants. For several reasons, PCR-based marker systems are the most widely used. They are relatively fast to screen, are cost effective and require little genomic DNA per assay. An important consideration in the use and analysis of molecular marker systems is their genome coverage and distribution in addition to their often complex evolutionary mechanisms.

In this study we have shown that the estimates of genetic similarities and/or genome diversity are influenced by the nature of the molecular markers used. We compared the pattern and extent of genetic diversity among a group of six-rowed barley cultivars with SNPs and S-SAPs and compared the findings with the diversity estimates obtained from the pedigree data to test the null hypothesis that there is no difference between diversity estimates derived from the two marker systems. Approximately 37% of S-SAP markers were polymorphic either between or within cultivars (Fig. 4.1). Mean GS values obtained from S-SAP (0.601 ± 0.096) and SNP (0.490 ± 0.187) were close but were significantly higher than pedigree-based GS values (0.098 ± 0.134) (Table 4.6).

Several previous studies have measured GS values to examine genetic relatedness of barley. Tinker et al. (1993) obtained GS values ranging from 0.06 to 0.61 with a mean of

Table 4.6. Maximum, minimum and mean GS estimates from pedigree (kinship), SNP, and S-SAP data for six-rowed spring barley.

Data set	Max	Min	Mean	SD	Distribution
Kinship	1.000	0.000	0.098	± 0.134	skewed
SNP	1.000	0.080	0.490	± 0.187	normal
S-SAP	0.915	0.326	0.601	± 0.096	normal

0.32 in a set of 27 North American barley cultivars using RAPDs. In a comparative analysis involving RFLPs, AFLPs, SSRs, and RAPDs among European barley accessions, Russell et al. (1997) found GS values among accessions to vary from 0.97 (AFLPs) to 0.45 (SSRs) indicating a wide range of variation in the extent and pattern of GS values revealed by different classes of molecular markers.

To test the levels of genetic homogeneity within each cultivar, we analyzed multiple plants per cultivar for both SNP and S-SAP as described earlier. S-SAP analysis assigned the 280 barley genotypes to 117 haplotypes. Only three cultivars 'Nord', 'Gamine' and 'Peregrine' showed complete homogeneity over all loci. The remaining 32 cultivars were composed of at least two haplotypes (Table 4.1). On the other hand, the 280 barley genotypes were reduced to 57 haplotypes based on the SNP analysis. More than 70% of the cultivars were found to be homogeneous with the SNP analysis. This difference between the two markers systems may be due in part of the larger set of S-SAPs compared to the limited number of SNP markers that were used in the analysis. No correlation was found between haplotype frequencies obtained from the SNP and the S-SAP data. For example, the cultivar 'Excel' was found to contain eight haplotypes with S-SAP but only one haplotype with SNP data (Table 4.1). This finding shows that the observed pattern of GS for each marker system may be different. We tested the possibility of seed contamination for the observed level of within cultivar genetic heterogeneity by treating each plant as an independent genotype to calculate all possible pair-wise GS values for cluster analysis and found that all individual plants within each cultivar clustered together (not shown) therefore, eliminating the possibility of seed contamination for the observed genetic heterogeneity within cultivars.

Variance components were estimated using AMOVA analysis. When cultivars were divided into malting and feed groups, only 6.6% of variation separated the two groups. Approximately 90% of the total variation was attributed to the genetic variation among groups (Table 4.3), indicating that there is little support for population differentiation along these groups. Within cultivar variance component was approximately 3% for both SNP and S-SAP data. In a previous study involving 103 barley cultivars representing spring six-rowed, spring two rowed and winter six-rowed type with S-SAP analysis, all major groups were clearly separated into distinct groups with the exception of an overlap between malting and feed cultivars (Soleimani et al. submitted). The lack of a significant separation between feed and malting cultivars may be due to a small number of loci involved in the malting properties over the whole genome. In addition, common parentage between some malting and feed cultivars may play a role on the extent of allele sharing among the groups. The ancestry of barley cultivars that were used in this study was traced back to 334 cultivars, landraces and breeding lines. Pair-wise GS values based on the kinship coefficient indicated that most malting cultivars share a substantial degree of common parentage with the feed counterparts based on the available pedigree information. This finding was also supported by the UPGMA cluster analysis (Fig. 4.3).

The GS values from both SNPs and S-SAP were normally distributed. In contrast pedigree-based GS values were skewed (not shown). Major differences in the genetic relationships among cultivars were found between pedigree-based data and both SNP and S-SAP data with cluster analysis. Discrepancies between pedigree and molecular markers in resolving genetic relationships in crops have been reported previously (Tinker et al. 1993; Barrett et al. 1998; Soleimani et al. 2003). These discrepancies are thought to be

due to various assumption that are made in the pedigree analysis including the assumption of equal parentage contribution to progeny, selection pressure, genetic drift and relatedness of parents with unknown pedigree (Souza and Sorrells 1989; Cox and Murphy 1990; Martin et al. 1991).

Interpretation of GS values from molecular marker data maybe confounded by assuming that adequate genome coverage is achieved, genetic markers segregate independently and that co-migrating fragments are homologous. Although the SNP data meet some of these assumptions similar conclusions cannot be made with respect to the S-SAP data. Each SNP was exclusively targeted by a PCR primer in both allele-specific (Fig. 4.4) and the primer extension assays (Fig. 4.5). Furthermore, the genomic loci containing SNPs were sequenced in multiple cultivars to infer the allelic state of the underlying polymorphism (Table 4.4). An important difference between the primer extension and the allele-specific method of SNP assay was that the former can determine heterozygosity while the latter cannot. This may be important in an out crossing organism or the analysis of F_1 hybrids in general but not for inbred barley cultivars in which it can be assumed that most plants have achieved near complete homozygosity through successive rounds of selfing.

The observed differences between S-SAP and SNP-based dendrograms and lack of correlation between the matrices of their GS values (Table 4.5) shows that the two marker systems target different regions of genome and that each has a unique evolutionary mechanism. S-SAP may result from retrotransposon related genome rearrangements including new insertion events, and deletions as a result of recombination between pre-existing elements. In addition, creation or loss of endonuclease restriction

sites may also give rise to S-SAP polymorphism. The latter component of polymorphism is comparable to AFLPs (Waugh et al. 1997). Data quality of S-SAP was shown to resemble that of AFLP (Queen et al. 2004). However, S-SAP was shown to be more advantageous than AFLP based on the level of detected polymorphism (Waugh et al. 1997; Ellis et al. 1998). Also, in the latter study, it was concluded that the retrotransposon based markers generated data that were more consistent with geographical and morphological criteria.

In our study, we found that SNP-based diversity estimates were more correlated to the pedigree-based estimates despite some major differences as described earlier. But there was no correlation between S-SAP and SNP-based genetic relationships among barley cultivars as inferred from cluster analysis and analysis of GS matrix correlation of the two GS estimates (Table 4.5). Both SNP and S-SAP gave similar results about the overall partitioning of genetic diversity within and between groups as well as the level of diversity within cultivars.

Despite the abundance of SNPs throughout the genome (Halushka et al. 1999; Ching et al. 2002; Zhu et al. 2003), we were able to validate only a small fraction of the putative barley SNPs by PCR analysis. The lowest success rate was observed for the *in-silico* set of putative SNPs in which only one SNP was validated and used subsequently. A higher success rate of 30% was achieved for the SNPs that were discovered by sequencing. Two factors may play a role in this finding: 1) sequence anomalies, especially in the EST data. 2) Duplications and presence of multi-copy loci in genome. Although the extent of genome duplication in barley is largely unknown, less than 5% of the total barley genome was shown to be represented by gene-coding sequences (Kankaanpaa et al 1997). Recent

studies in rice have shown that 15% of the rice genome was found in duplicated blocks (Vandepoele et al. 2003). In a study of soybean with SNPs, Zhu et al (2003) argued that failure to convert SNPs into STSs was likely the result of amplification of multiple sites in the genome.

4.6 Conclusion

Molecular markers have become important tools in the study of crop species. In the absence of whole genome sequencing, they are indispensable tools for elucidating genetic relationship among organisms and in many crops in particular. However, after having demonstrated in this study and other comparative studies in crops cited earlier there is often significant incongruence between the results of genetic diversity estimates derived from different classes of molecular markers. Although SNP-based GS values were more closely related to the pedigree based estimates, S-SAPs were found more useful to infer genetic relationships among very closely related cultivars such as 'AC Albright' and 'Otal' as well as detecting somaclonal variation between cultivars such as 'AC Malone' and 'Leger'. A disadvantage in the use of both SNP and S-SAP is their requirement for prior sequence knowledge. However, with increasing number of new EST and retrotransposon sequences from various organisms in the public domain this constraint is being gradually overcome.

4.7 Acknowledgments

The authors would like to thank Ajit Singh Sahota from the Canadian Food Inspection Agency (CFIA) and Canadian Grain Commission (CGC) for providing plant

material. We would also like to thank Dr. Thin-Meiw (Alek) Choo of Agriculture and Agri-Food Canada for providing cultivar description of barley material. We are indebted to him for his advice and discussions. We would also like to thank Dr. Douglas Procunier and Mark Gray at Cereal Research Centre Winnipeg Manitoba for ABI gel analysis of SNPs. This study was partially funded by the Automated Quality Testing Inc. (AQT).

CHAPTER FIVE

Quantification of the retrotransposon BARE-1 reveals the dynamic nature of the barley genome

The following chapter is a reproduction of a paper submitted to the journal Genome by Vahab D. Soleimani, Bernard R. Baum and Douglas A. Johnson. The chapter fulfills the third objective of the thesis, namely, estimating copy number of BARE-1 retrotransposon among barley cultivars as well as samples of *H. spontaneum* (wild relative of barley) to test the null hypothesis that there is no variation in BARE-1 copy number among barley cultivars. LTR and RT regions of BARE-1 were amplified by PCR, cloned into plasmids and used as external standards for quantification with real-time quantitative PCR. Quantitative analysis of BARE-1 revealed major differences (up to 3000 copies per haploid nucleus) among cultivars. This chapter provides further evidence that BARE-1 retrotransposition has played a major role in reshaping barley genome during breeding and selection.

5.1 Abstract

We have used quantitative real-time PCR analysis to measure the copy number of BARE-1 retrotransposons in five cultivars of barley *Hordeum vulgare* as well as samples from its wild relative *H. spontaneum*. Two sets of PCR primers were used to amplify regions within the long terminal repeat (LTR) and the reverse transcriptase (RT) gene of BARE-1 (GenBank accession Z17327). The LTR primers detected an average of $2.148 \pm 0.012 \times 10^5$ copies per haploid genome among barley samples while the RT primers detected an average of $1.588 \pm 0.085 \times 10^4$ copies. The average ratio of LTR/RT was estimated to be 13.5/1. This finding indicates that more than 7 % of barley genome is occupied by BARE-1 elements in the form of solo LTRs and another 2.6 % of the genome is occupied by the full length element. Taken together, BARE-1 sequences represent approximately 9.6 % of the barley genome among the barley plants in this study. For the above estimation a genome size of 5.44×10^3 Mb for *H. vulgare* and 5.39×10^3 Mb for *H. spontaneum* were assumed. Our study on quantification results of the BARE-1 for a small group of barley cultivars showed that there are significant differences among cultivars in terms of BARE-1 copy number, providing further evidence that BARE-1 has a major role in shaping the barley genome as a result of breeding and selection. Quantification results also showed that most of the elements (>90 %) are present as truncated copies (solo LTRs). These results show that there is a high level of recombination leading to the formation of truncated elements and a subsequent DNA loss from the genome. Taken together, our study provides an insight into a dynamic micro-evolutionary process that is the by product of directional selection in barley breeding programs

Key words: BARE-1, genome evolution, quantification, Real-time PCR, retrotransposons

5.2 Introduction

Retrotransposons are major contributors to genome size variation among plants. Estimates of the copy number of retrotransposons vary within and between plant species (Pearce et al. 1996a; Vicient et al. 1999; Kalendar et al. 2000). In addition, retrotransposon copy numbers have been shown to be correlated with the genome size (Kumar and Bennetzen 1999; Vicient et al. 1999) providing some explanation for the C-value paradox. Recent sequencing of large stretches of DNA from grasses has also provided strong evidence on the role played by transposable elements in the restructuring of genomes through insertions and intra and inter-element recombination (Panstruga et al. 1998; Tikhanov et al. 1999; Wicker et al. 2001; Gu et al. 2003). In Brassicaceae, comparative analysis between related species such as *Arabidopsis thaliana* and *Brassica oleracea* has shown that the genome size increase in *B. oleracea* is due to the amplification of transposable elements (Zhang and Wessler 2004). The main contributors to the increase are retroposons, both LTR and non-LTR elements (LINES). For most plant retrotransposons studied to date there is no evidence for mobility, at least within the recent past (Horohcika 1993; Vernhettes et al. 1997; Takeda et al. 1998), BARE-1 retrotransposon (Manninen and Schulman 1993) was shown to be an active component of the barley genome (Suoniemi et al. 1996a; Jaaskalainen et al. 1999). Analysis of BARE-1 in barley has shown that the element is widely distributed throughout the genome (Suoniemi et al. 1996b; Waugh et al. 1997). Using Southern hybridization technique, estimates of BARE-1 copy number were shown to vary substantially both within the *Hordeum* genus and within species (Vicient et al. 1999; Kalendar et al. 2000).

Analysis of retrotransposon copy numbers have been mostly based on hybridization technique (Vershinin et al. 1990; Pearce et al. 1996b; Vicient et al. 1999; Kalendar et al. 2000; Kalendar et al. 2004). Hybridization-based estimates may be biased towards lower measures of copy number than the actual estimates as the estimated copy number is expected to be dependent on the washing stringency.

An alternative is real-time PCR (Higuchi et al. 1992; Gibson et al. 1996; Heid et al. 1996), an approach that allows quantification of PCR products using the exponential phase of the PCR amplification. This is achieved by determining the increase in the amount of fluorescence due to the binding of a fluorescent DNA stain such as SYBR green I to the amplification product (Higuchi et al. 1992) or by fluorescence resonance energy transfer (FRET) technology as employed with TaqMan probe or molecular beacons. Real-time PCR offers many advantages over conventional PCR and hybridization-based copy number estimates. It is fast, reliable and requires minutes amount of template DNA. It is performed in a single closed reaction tube and target concentration is measured after each cycle. The measurements that fall in the log linear phase are then used for accurate estimates of the unknown target. Unlike conventional quantitative PCR it does not require any post PCR manipulation.

The aim of this study was to use Real-Time PCR for quantification of BARE-1 retrotransposon in a small group of barley cultivars representing diverse growth habit (i.e. winter versus spring), seed characteristics (two rowed/six-rowed) and in samples of *H. spontaneum* for comparative analysis and to gain insight into the extent of BARE-1 activity on the evolution of the barley genome as a result of breeding and selection.

5.3 Materials and Methods

5.3.1 Plant material

Seed material of barley cultivars representing winter six-rowed, spring six-rowed and spring two-rowed types was obtained from the Canadian Food Inspection Agency (CFIA) Ottawa, ON Canada (Table 5.1). Seed samples of *Hordeum spontaneum* were obtained from the Institute of Evolution, University of Haifa, Israel.

5.3.2 DNA isolation

Seeds were germinated in a Petri dish containing a moist tissue paper at room temperature for one week. 100 mg of fresh leaf tissue was harvested from each seedling and placed in a 1.5 ml Eppendorff tube together with a 3 mm carbide tungsten bead. Leaf samples were submerged in liquid nitrogen to freeze the samples. The frozen leaf material was ground using an MM 300 Mixer Mill (Retsch®). DNA was extracted from individual seedlings using DNeasy™ Plant Mini Kit (Qiagen®). The concentration of DNA was determined using Hoefer DynaQuant™ Fluorometer (Hoefer Pharmacia Biotech). Working solutions of 1.0 ng/ µL were prepared by dilution of the original stock solution with TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 7.5).

5.3.3 PCR primer design and amplification reactions

Primers were designed from the reverse transcriptase (RT) and the long terminal repeats (LTR) of BARE-1 retrotransposon of barley (GenBank accession Z17327) using the Oligos program (Kalendar 2001). Primer pairs were designed to amplify domains

Table 5.1. Barley material used for BARE-1 quantification experiment with real-time PCR

Barley	Parentage	Origin
	Winter six-rowed feed	
'McGregor'	Tapir, Wysor	Blenheim Ontario
'McDiarmid'	Tapir, Wysor	Blenheim Ontario
'OAC-Elmira'	WB74-69//WB74-69/WB55-1	University of Guelph, Ontario
	Spring six-rowed malting	
'Bonanza'	Vantage/Jet//Vantmore/3/2* Parkland/4/Conquest	Brandon Manitoba
	Spring two-rowed malting	
'Harrington'	Klages x (Gazelle/Betzes//Centennial)	Saskatoon Saskatchewan
	Hordeum spontaneum (wild barley)	
Spt_Gln	*H. spontaneum* collected from the Golan Heights	Israel
Spt_JRD	*H. spontaneum* from Israel Jordan border region	Israel

within the RT gene spanning nucleotide position 11-161 and LTR region spanning nucleotide position150-439 of BARE-1 retrotransposon. Each primer pair resulted in the amplification of a single band. PCR reactions were carried out in 20 μL volume using Titanium Taq DNA polymerase (BD Biosciences®) in a PTC 200 thermocycler (MJ Research) according to the following thermocycling profile: 1) two minutes of initial denaturation at 94 °C to activate the polymerase; 2) 20 seconds of primer/template annealing at 55 °C; 3) 1 minute of extension at 72 °C. The cycle was programmed to run 35 times. Ten μl of the PCR product was mixed with 2 μl of gel loading dye and run in a 1.25% agarose gel for 1 hour at 80 volts. DNA bands were visualized by ethidium bromide staining. The remaining product was saved for ligation and cloning as described below.

5.3.4 Cloning and sequencing of RT, and LTR domains for use as external standards

The PCR products were ligated into pDrive vector (Qiagen®) and used to transform Qiagen EZ competent *E. coli* cells. Colonies containing potential recombinant plasmid were selected with ampicillin (final concentration of 100 μg/mL) supplemented with 100 μL of 0.5 mM IPTG and 20 μL of 50 mg/mL X-gal. Three white colonies from each experiment were picked for colony PCR analysis to confirm the presence of the insert. Plasmid DNA was extracted using QIAprep® Miniprep (Qiagen®) following the manufacturer's recommendations. DNA sequencing was done 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 BigDye Terminator

Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase and following the manufacture's recommendations. The identity of the obtained sequences was confirmed by BLAST search (Altschul et al. 1990) available at the National Centre for Biotechnology Information (NCBI, Maryland USA).

5.3.5 Quantitative real-time PCR

The external standards were cloned RT and LTR domains contained within the polylinker region of the pDrive cloning vector (Qiagen®). The purified plasmid was obtained from transformed *E. coli* culture using QIAprep® Spin Miniprep kit (Qiagen®) and linearized with *Not*I restriction enzyme prior to its use as template for real-time PCR analysis. Ten-fold serial dilutions of the stock solution of the plasmid DNA were carried out to produce $5x10^{-1}$, $5x10^{-2}$, $5x10^{-3}$, $5x10^{-4}$, and $5x10^{-5}$ ng/μL solutions of the recombinant plasmid. PCR reactions for both the standards and the unknown samples were prepared from the same master mix using LightCycler FastStart DNA Master^{plus} SYBR Green I (Roche®). SYBR green I is a non-specific DNA binding dye that binds to the minor groove of double stranded DNA resulting in fluorescence. In its free form in solution, the dye does not produce detectable fluorescence. The 20 μL reaction mix in each LightCycler® capillary tube contained 4.0 μL of LightCycler FastStart DNA Master^{plus} SYBR Green I, 2.0 μL of HPLC purified primers (3.2 picomol/μl), 9.0 μL of PCR grade H_2O and 5.0 μL target DNA solution. The PCR protocol consisted of an initial denaturation step at 95 °C for 3 minutes, 45 cycles of amplification, each consisting of 5 seconds of denaturation at 95 °C, 8 seconds of annealing at 55 °C, and 10 seconds of elongation at 72° C followed by cooling at 40 °C for 30 seconds. The

temperature transition rate was 20 °C/cycle. The fluorescence was measured once after each elongation step. The PCR reaction conditions for both RT and LTR specific primers were identical as they were designed to have similar annealing temperature.

5.3.6 Standard curves and copy number estimation

Standard curves were produced by plotting C_T (Fluorescence threshold values) against the logarithm of the concentration in copy number using Microsoft® Excel. The standard curve is a linear line described by $Y = mX + b$ in which m (the slope of the line) is -1/Log (PCR efficiency) and b (intercept) is the log of the amount of amplification product at the threshold divided by the log of the PCR efficiency. The starting copy number of plasmid (standard) was obtained based on weight, concentration and size parameter of the plasmid using a computer program available at the URL: (http://molbiol.edu.ru/eng/scripts/h01_07.html).

Copy numbers of the unknown samples were determined from the standard curve based on the obtained C_T value for each sample using basic computational tools available in Microsoft® Excel software.

5.3.7 Differences between means

Squared Mahalanobis distances between means were computed using SAS/STAT using the procedure CANDISC (SAS Institute).

5.4 Results

5.4.1 PCR amplification and cloning of RT and LTR sequences

Amplification of RT and LTR domains of BARE-1 produced single bands as visualized by ethidium bromide staining in agarose gel (Fig. 5.1). The estimated sizes correspond to the predicted size of approximately 150 bp and 250 bp for the RT and LTR domains respectively. The bands were cloned and confirmed as RT and LTR domains of BARE-1 by sequencing. Purified plasmid DNA were quantified and used as external standards for PCR.

5.4.2 Standard and melting curves

For reliable quantification, the amplification efficiency during PCR must be equal between the standards and the unknown. Both the standard plasmid DNA and the unknown were PCR amplified with the same primer set under identical reaction conditions. Standard curves were generated using plasmid DNA through a range of 5 orders of magnitude ranging from 5×10^{-1} to 5×10^{-5} ng of plasmid DNA (Fig. 5.2). The coefficient of determination (R^2) for the standard curve was estimated to be more than 0.998 with calculated PCR efficiency of 1.998 (Fig. 5.3), the ideal efficiency of PCR is two. A melting curve (Fig. 5.4) plotted as the negative derivative of fluorescence versus temperature showed melting peaks at 83 °C for the BARE-1 product and a peak at lower temperature for non-specific products (i.e. primer dimer).

Figure 5.1. Agarose gel electrophoresis showing PCR amplified product of RT and LTR domains of BARE-1. Lanes: 1 RT, 2 LTR, 3 DNA size markers. DNA from barley cultivar 'Bonanza' was used as template.

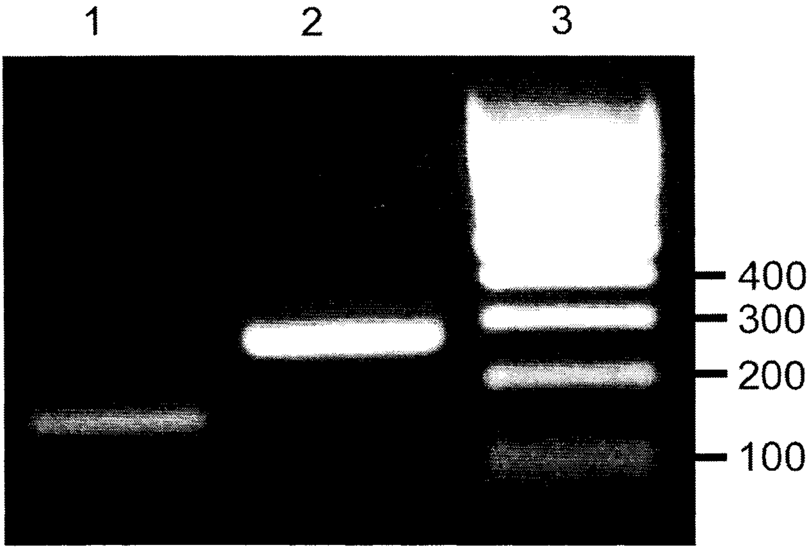
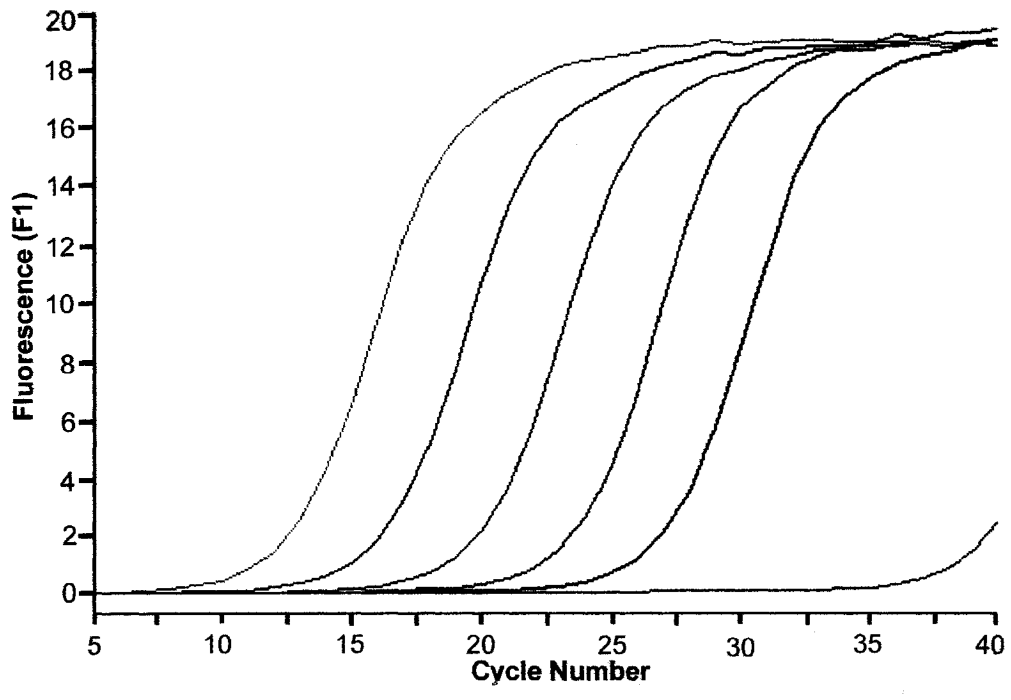
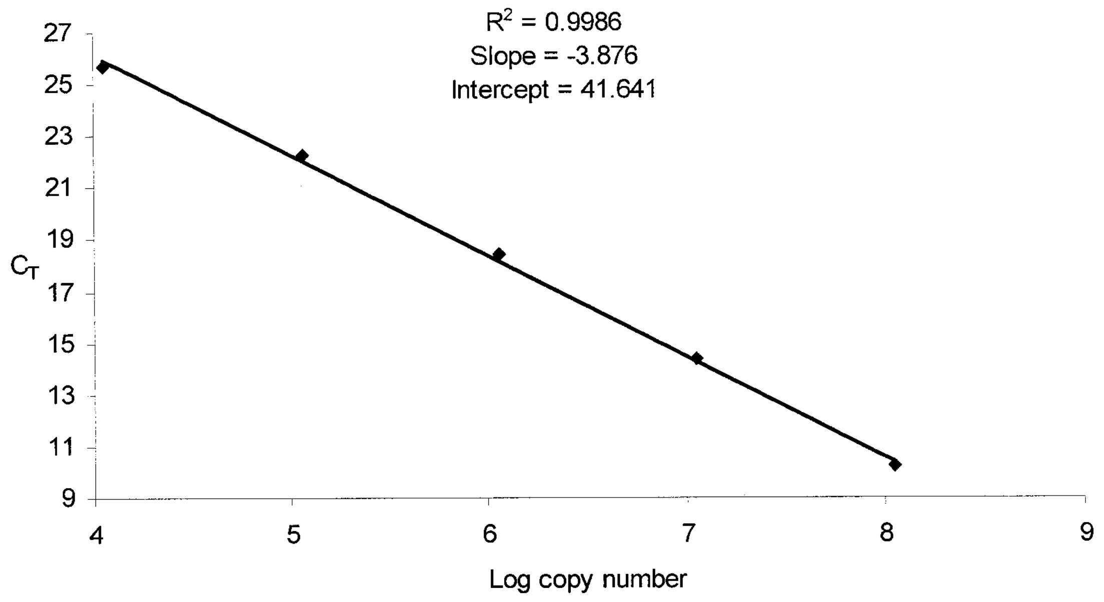

1
2
3
400
300
200
100

Figure 5.2. PCR product detection in real time through five orders of magnitude (5×10^{-5} to 5×10^{-1}) range in template (standard) concentration. Fluorescence was measured during each cycle at the end of the extension phase.



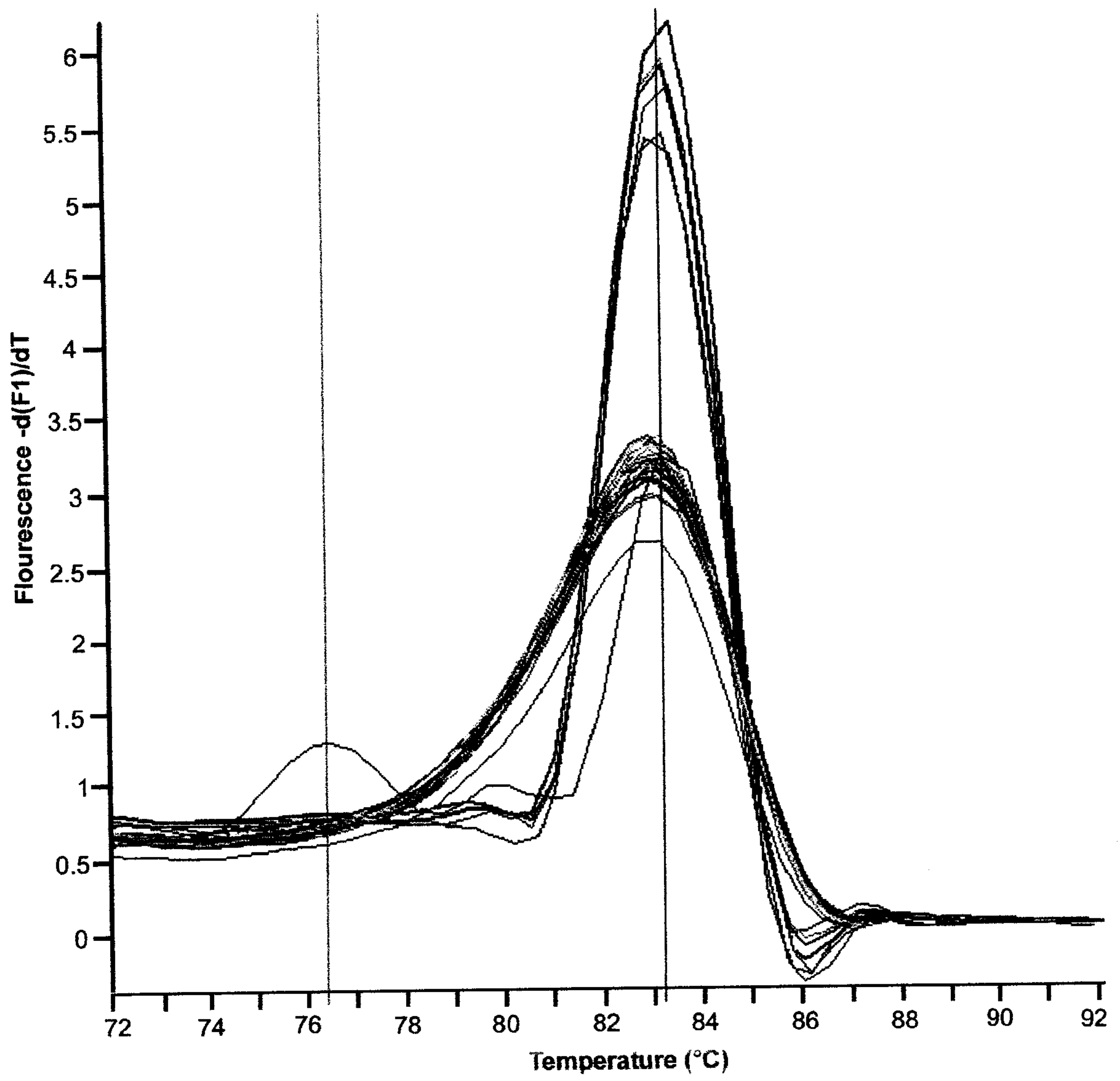
Fluorescence (F1)
20
18
16
14
12
10
8
6
4
2
0
5
10
15
20
25
30
35
40
Cycle Number

Figure 5.3. Standard curve generated from plasmid DNA containing RT domain of BARE-1.Fluorescent threshold values (C_T) were plotted against log copy number to produce a linear function. The slope and intercept of the curve were calculated from the linear equation describing the standard curve.



R² = 0.9986
Slope = -3.876
Intercept = 41.641
27
25
23
21
19
CT
17
15
13
11
9
4
5
6
7
8
9
Log copy number

Figure 5.4. Melting curve analysis of RT domain plotted as the negative derivative of fluorescence against temperature for the standard and the unknown samples. The analysis revealed a major peak at 83 °C for RT. Sample with –d(F1)/dT value of 3.5 are unknown and those with the corresponding value of 6 are the standards.



Flourescence -d(F1)/dT
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
72
74
76
78
80
82
84
86
88
90
92
Temperature (°C)

5.4.3 RT and LTR copy number estimates

Copy number/haploid nucleus of LTR and RT domains of BARE-1 are presented in Table 5.2. Quantification results showed that most BARE-1 elements were in the form of solo LTRs in both cultivated and wild barley. Significant inter-cultivars differences in BARE-1 content were detected among barley cultivars by the squared Mahalanobis distances between their means (Table 5.3). Intra-cultivar variation in BARE-1 content was statistically insignificant for all five cultivars. CDA analysis also showed no significant variation among *H. spontaneum* samples collected from two diverse habitats for both RT and LTR. We therefore treated all *H. spontaneum* genotypes as a single group for comparative analysis with *H. vulgare*. The relationship between LTR and RT copy numbers among genomes were assessed by regression analysis (not shown) and found to be significantly and negatively correlated.

Table 5.2. Copy number of BARE-1 LTR and RT gene among barley samples

Genome	LTR	RT	LTR/RT Ratio
'McDiarmid'	214118 ± 766	16181 ± 489	13.2
'McGregor'	214887 ± 604	15838 ± 181	13.5
'OAC Elmira'	214950 ± 368	16065 ± 149	13.4
'Bonanza'	217371 ± 672	14260 ± 137	15.2
'Harrington'	212866 ± 1381	17166 ± 164	12.4
Spt_Gln	215325 ± 467	15922 ± 234	13.5
Spt_Jrd	214713 ± 936	15756 ± 186	13.6
Average	214890 ± 1258	15884 ± 858	13.5

Table 5.3. Pairwise squared Mahalanobis distances estimated from BARE-1 copy number between genomes

	Genome	1	2	3	4	5
1	'McDiarmid'					
2	'McGregor'	1.850200				
3	'OAC Elmira'	1.119930	1.401150			
4	'Bonanza'	*57.48884	*39.19725	*52.96918		
5	'Harrington'	*15.99010	*27.94043	*18.75672	*133.3107	
6	†*H. spontaneum*	1.958110	0.043820	1.112770	*39.63624	*27.60541

* indicates pair-wise comparison for which there is significant variation between means ($P<0.0001$).

† includes Spt_Gln and Spt_Jrd samples (see Table 5.1).

5.5 Discussion

The importance of transposable elements especially retrotransposons in genome architecture and evolution has been realized through many studies in recent years (Flavell et al. 1977; SanMiguel et al. 1996; Vicient et al. 1999; Shirasu et al. 2000; Kalendar et al. 2004). The replicative nature of retrotransposition offers the potential for an explosive increase in the copy number of the active elements within genomes (Vicient et al. 1999). As a result in plants such as maize more than 50% of the genome is occupied by retrotransposons (SanMiguel et al. 1998). More recent studies have shown that even in rice – with the smallest genome among the cereals more than 20 % of the genome is composed of retrotransposon sequences (Ma et al. 2004). Two possible mechanisms control the effects of retrotransposon-mediated genome size expansion: 1) regulation/suppression of transposition and 2) retrotransposon-mediated recombination leading to the removal of DNA from the genome. For LTR retrotransposons, the footprints for DNA loss through recombination are solo LTRs scattered throughout the genome. In barley, BARE-1 was shown to be actively transcribing in somatic tissue (Suoniemi et al. 1996a), leading to insertional polymorphism (Waugh et al. 1997; Kalendar et al. 1999). Estimates of BARE-1 copy number in the genus *Hordeum* was shown to vary considerably among different species (Vicient et al. 1999) and the estimates were found to be positively correlated with genome size providing evidence for the role of retrotransposons in genome expansion.

Quantitative analysis with real-time PCR using RT domain of BARE-1 retrotransposon in this study showed that approximately 2.6 % of barley genome is occupied by BARE-1 retrotransposon. This estimate is consistent with previous measures

of BARE-1 content in *Hordeum* species with various ploidy levels (Vicient et al. 1999). The proportion of genome that is occupied by solo LTRs was estimated to be 7 % among the plants studied here. This value was obtained by assuming a length of 1.8 kb for solo LTRs – corresponding to the size of BARE-1 (GenBank accession Z17327) LTR. Previous studies involving quantification of BARE-1 in *Hordeum* have found significant variation between species (Vicient et al. 1999) as well as within species of wild barley grown in different microclimates (Kalendar et al. 2000). In this study, we found significant variation in BARE-1 copy number among a group of barley cultivars representing different agronomic classes (Table 5.3). Intra cultivar variation in BARE-1 content was found to be statistically insignificant. Also, no significant variation in BARE-1 content was detected between *H. spontaneum* samples collected from two distinct regions. Inter-cultivar variation in BARE-1 content can have significant implication in genome evolution during breeding and selection as the genetic makeup of cultivars is shaped by mutation, recombination, and genetic drift. In organisms with an active transposition system such as barley, retrotransposition and/or deletions mediated by retrotransposon sequences can provide a rapid mechanism for genome evolution. Because the genetic base of the modern barley cultivars is thought to be narrow (Martin et al. 1991; Horsely et al. 1995) as a result of using local germplasm, gross variations in retrotransposon composition in barley genomes are likely the result of differential transposition rates in lines and cultivars rather than vertical transmission of these elements from parents with diverse genomes. Cultivars are developed for various specific agronomic traits and environmental regions and are therefore subjected to various selection pressures which can trigger retrotransposition and related genome

rearrangements. Activation of retrotransposons has been linked to various biotic and abiotic factors such as tissue culture (Hirochika 1993; Hirochika et al. 1996), climatic changes (Kalendar et al. 2000), methylation and epigenetic factors (Liu and Wendel 2000; Kaeppler et al. 2000; Kubis et al. 2003) and pathogen attack (Grandbastien et al. 1997).

The LTR-based primers showed that most of the BARE-1 copies were represented by solo LTRs lacking the internal domain (i.e. the gag and pol regions). A high ratio of 13.5/1 for solo LTRs versus intact BARE-1 elements in this study and similar findings by Vicient et al (1999) indicates that genome expansion by retrotransposition is reversible by recombination mechanisms. Both unequal recombination (Shirasu et al. 2000) and illegitimate recombination (Bennetzen an Kellogg 1997; Devos et al. 2002) have been implicated as counteracting forces resulting in genome size reduction. The LTR sequences themselves can function as promoter or enhancer to derive the expression of adjacent genes (Michaud et al. 1994; Kashkush et al. 2003). Recent studies in yeast have shown that retrotransposons and interspersed LTR sequences can effect the regulation of gene expression during development via RNAi formation (Schramke and Allshire 2003). This type of gene regulation is initiated by the formation of dsRNA molecules derived from transposable elements and which can subsequently trigger RNAi initiation (Madlung and Comai 2004).

RT primers were expected to detect the full length elements which are also flanked by two LTRs giving rise to potential overestimation of LTR copy numbers. To compensate for this seeming overestimation, we multiplied RT copy number values by a factor of two and subtracted the resulting values from the copy numbers derived from

LTR primers. Assuming that each solo LTR is the result of one integration event, on average 2.3 x 10^5 BARE-1 integrations have occurred in the genome. If all integrations are random in the genome, barley with a genome size of 5.4 x 10^3 Mb there will have one BARE-1 copy (either solo LTR or full length) per 0.02 CM of the genome which can be exploited in mapping projects. The copy number estimates provided here and elsewhere for BARE-1 retrotransposon is likely underestimation of the actual sequence contribution of the element to the barley genome. This is because defective copies accumulate mutations – blend into the genomic sequences and are unlikely to be detected experimentally. In addition, a significant portion of BARE-1 elements may correspond to degraded insertions that are neither complete elements nor solo LTRs as was recently shown for the *Arabidopsis* retroelements (Pereira 2004).

The differential abundance of BARE-1 retrotransposon in the genome of different barley cultivars, and its dynamic nature provides further evidence for its major role in genome evolution on a micro-evolutionary time scale.

5.6 Acknowledgements

We thank Dr. A. Belyayev, Institute of Evolution, University of Haifa, Israel, for having kindly provided the seed material of *H. spontaneum*; Dr. B. Blais, Leah Isaac, and A. Martinez of the Canadian Food Inspection Agency, for the use of the LightCycler. This paper benefited from comments made by Drs. S. Molnar and B. Miki of Eastern Cereal and Oilseed Research Center.

CHAPTER SIX

Identification of six-rowed spring malting barley cultivars by using allele-specific PCR to detect Single Nucleotide Polymorphisms

The following chapter is part of a "Trade Secret" and it is protected. It's a manuscript prepared for eventual publication by Vahab D. Soleimani, Bernard R. Baum and Douglas A. Johnson. The chapter fulfils objective 4 of the thesis. A panel of 15 informative SNP markers was used to identify all registered Canadian spring six-rowed malting cultivars including their various biotypes by allele-specific PCR method as described earlier. There are two identification keys provided in this thesis and are both protected as Trade Secret. The first key (this chapter) is used to identify all registered Canadian six-rowed malting barley cultivars. The key assumes that the unknown cultivars to be identified are drawn from a set of seventeen six-rowed malting cultivars. If the unknown cultivar(s) is drawn from a pool of spring six-rowed malting and/or feed cultivars, then the key in the Appendix should be used.

6.1 Abstract

Modern malting barley cultivars (*Hordeum vulgare* L.) are generally produced from a narrow range of germplasm. As a result it is increasingly difficult to differentiate among them based solely on morphological and biochemical characteristics. In this study, a set of 15 single nucleotide polymorphisms (SNPs) was sufficient to uniquely distinguish all 17 malting six-rowed barley cultivars including several biotypes present within some cultivars commonly grown in Canada and the United States. The PCR conditions optimized for screening allowed discrimination between alleles of all diagnostic SNPs. Using DNA extracted from seed or leaf, the methodology is rapid, lends itself to automation, and can be used for the efficient identification and purity testing of six-rowed malting barley cultivars.

Key words: single nucleotide polymorphism, identification, allele-specific PCR, barley

6.2 Introduction

The ability to rapidly distinguish among barley cultivars is important to the brewing industry due to cultivar to cultivar differences in malting properties. Identification is also important to distinguish malting from food or feed grades of barley. Traditionally, morphological criteria such as kernel visual distinguishability (KVD) were used to discriminate between cereal grains, including barley. There are at least two factors limiting the use of morphological characters for identification purposes: 1) the limited number of such markers that are useful for discrimination; and 2) phenotypic plasticity, where different environmental conditions affect the outcome of plant and seed characteristics.

In Canada, new malting barley cultivars are being regularly introduced. As they are essentially derived from a narrow gene pool, the morphological traits may lack sufficient power to differentiate among them. Thus there is a need for more robust discriminating techniques for barley identification. Within the past thirty years, various DNA- and protein-based molecular techniques have been proposed for genotyping and cultivar identification in barley. These include electrophoretic patterns of seed storage protein (Cook, 1988; Gebre et al. 1986; Nielsen and Johansen, 1986), isozyme electerophoresis (Hoffman and Goates, 1990; Fenwick et al. 1993), restriction fragment length polymorphism (RFLP) (Saghai-Maroof et al. 1994), random amplified polymorphic DNA (RAPD) (Baum et al. 1998), sequence tag sites (STS) (Baum et al. 2000; Hoffman et al. 2002) among others. There are various limitations on the routine use of many of the techniques listed above. Isozyme electrophoresis patterns may be influenced by the growing conditions and environmental effects on cultivars, therefore, complicating

identification. In particular, the analysis of F1 hybrid seeds may present a challenge to the use of protein based identification methods (Cooke, 1988; Konarev and Gavrilyuk, 1988). Reproducibility of RAPDs and their application for genotyping is controversial (Skroch and Nienhuis, 1995; Baum et al. 2000). RFLPs (Saghai-Maroof et al. 1994) and AFLPs (Soleimani et al. 2002a) have been successfully used for cultivar identification but their routine use is limited by the requirement for a high level of technical expertise and costs associated with their use. Single nucleotide polymorphisms (SNPs) are nucleotide variations in the DNA sequence of individuals within a population. They constitute the most abundant class of molecular markers. Estimates of SNPs distribution in genomes have ranged from 1 SNP per 60-120 bp in maize (Ching et al. 2002; Tenaillon et al. 2001) to 1 SNP per 1000 bp in Human (Sachidanandam et al. 2001). In organisms studied thus far, SNPs were more prevalent in the non-coding region of the genome (Douabin-Gicquel et al. 2001; Cargill et al. 1999). Such mutations are expected to be neutral unless they affect the expression profile of genes in their vicinity. Cargill et al (1999) found that within the coding regions of the human genome non-synonymous nucleotide substitutions were more widespread than synonymous substitutions, indicating selection bias against amino acid replacements as a result of non-synonymous substitution. A synonymous nucleotide substitution that results in creation or loss of a splice site may affect the fitness of an individual and therefore be subjected to selection constrains (Richard and Beckman 1995).

The identification of SNPs is achieved either by direct sequencing of known genomic loci among individuals within a population or by using an *in silico* protocol of mining SNPs from EST databases (Beutow et al. 1999; Sunyaev et al. 1999) followed by

confirmation of the polymorphisms by sequencing. The *in silico* technique is facilitated by an increasing number of EST sequences in the public domain. Once a set of appropriate SNPs is obtained, they could be placed on a chip and used for rapid identification purposes. The potential for automation of SNPs and detection is a major advantage of this approach.

The aim of this research was to develop a DNA-based platform for the identification of malting barley cultivars. We found that a panel of only15 SNP markers was needed to distinguish all 17 cultivars of six-rowed malting barley commonly grown in North America including their various biotypes. These markers could be further developed into a DNA chip for rapid and accurate identification of six-rowed spring malting barley cultivars.

6.3 Materials and Methods

6.3.1 Plant Material

Breeders' seed of six-rowed spring malting barley cultivars (Table 6.1) was obtained from the Canadian Grain Commission, Grain Research Laboratory (GRL) Winnipeg, Manitoba and the Canadian Food Inspection Agency (CFIA) Ottawa, Ontario.

6.3.2 DNA Extraction

DNA was extracted from barley seeds and leaves by two independent methods for comparison 1) DNA extraction from leaf material: Seeds were germinated on moist paper towel in Petri dishes at room temperature for one week. 100 mg of leaf material was harvested from each seedling and placed in a 1 ml collection tube in a 96-well plate format (Qiagen®). A 3 mm carbide tungsten bead was placed in each tube, and plates were submerged in liquid nitrogen for one minute to freeze leave material. Two plates containing frozen leaf material were immediately mounted on an MM 300 Mixer Mill (Retsch®) and were shaken at thirty revolutions per minute (RPM) for up to one minute. DNA extraction was carried out using a 96-well-plate format, plant DNA extraction protocol (DNeasy™ from Qiagen®) following the manufacturer's recommendation.
2) DNA extraction from seeds: Individual seeds were placed between two sheets of silicon paper and crushed with a hammer. The pulverized material was placed in an individual 1.5 ml Eppendorff tube containing a 3 mm tungsten carbide bead and ground with an MM 300 Mixer Mill (Retsch®) for 15 to 30 seconds at maximum speed. To each tube containing the ground seed material 500 µl of extraction buffer (200 mM Tris-HCl

Table 6.1. Six-rowed malting barley cultivars used in this study, NDSU, North Dakota State University; "/", primary cross; "//", secondary cross; "*" indicates number of backcrosses

	Cultivar	Biotypes	Parentage	Breeder/location	Released
1	Argyle	1	Herta/UM 570//Conquest/3/Bonanza	University of Manitoba	1981
2	B1602	1	Bumper/6B78-628//Morex/6B78-628	Busch, Fort Collins, Co	1991
3	Bonanza	4	Vantage/Jet//Vantmore/3/2* Parkland/4/Conquest	AC Brandon, Manitoba	1970
4	BT954	1	N/A	Busch Agricultural resources	2001
5	CDC Battleford	1	M67/BT411	University of Saskatchewan	2001
6	CDC Sisler	2	M34/Argyle	CDC Saskatoon SK	1996
7	CDC Springside	4	M76/SM93067	University of Saskatchewan	2001
8	CDC Tisdale	1	BT409/Foster	University of Saskatchewan	2001
9	CDC Yorkton	8	M67/BT411	CDC Saskatoon SK	1999
10	Duel	1	Morex//6B75-1374/M31	Busch, Fort Collins, Co	1992
11	Excel	1	Cree/Bonanza//Manker/3/2* Robust	University of Minnesota	1994
12	Foster	5	Robust/6/Glenn/4/Nordic//Dickson/Trophy/3/Azure/5/Glenn/Karl	NDSU, Fargo, ND	2000
13	Lacey	2	M44/Excel//M46/M44/3/M44/Excel//Stander	Minnesota Agr. Exp. Station	2000
14	Legacy	1	6B86-3517/Excel	Busch, Winnipeg, Manitoba	2001
15	Robust	1	Morex/Manker	University of Minnesota	1994

	Cultivar	Biotypes	Parentage	Breeder/location	Released
16	Stander	2	Excel//Robust/Bumper	University of Minnesota	1999
17	Tankard	3	Argyle/Minnesota M-34	CDC Saskatoon, SK	1992

pH 7.5, 250 mM NaCl, 25 mM EDTA pH 7.5, 0.5% SDS) was added. The tubes were shaken for an additional 15 seconds at the maximum speed on the mixer mill. Subsequent steps in the seed DNA extraction were carried out as described in Edwards (1991).

6.3.3 DNA sequencing and identification of SNPs

Nine genomic loci (Table 6.2) were selected for sequencing on a sample of 5-8 malting cultivars. The sequences of PCR Primer used in this study were publicly available from the Montana State University, Bozeman, Montana (http://horduem.oscs.montana.edu).

DNA sequencing was done using the Sanger Dideoxy-chain terminator method (Sanger et al. 1997) on an ABI PRISM 377 DNA sequencer (Perkin Elmer, Applied Biosystems) with ABI Prism Dye terminator Cycle Sequencing Ready Reaction Kit using AmpliTaq DNA polymerase. Multiple sequences obtained for each locus were aligned using Clustal W (Thompson et al. 1994) for visual detection of nucleotide variation among cultivars. For the i*n silico* detection of SNPs, EST sequences were retrieved from the public domain including NCBI (http://www.ncbi.nlm.nih.gov/dbEST/index.html). EST sequences were assembled and aligned using the Contig assembly program Cap3 (Huang and Madan, 1999) available at: (http://fenice.tigem.it/bioprg/interfaces/cap3.html). The selected contigs containing putative SNPs were used to design PCR primers for amplification of the corresponding region among a group of six-rowed malting cultivars. The resulting PCR products were sequenced and multiple alignments of the sequences were performed to visualize SNPs.

6.3.4 SNP Validation

For each potential SNP obtained from the sequencing experiments, an allele-specific PCR primer was designed (Soleimani et al. 2003). PCR reactions were carried out on total genomic DNA using the locus-specific primer in combination with an SNP-specific primer. PCR reactions were carried out in 20 µl volume containing a final concentration of 0.2 picomole of each primer, 1 unit of Titanium™ Taq DNA polymerase (BD Biosciences®), 0.8 mM of each dNTP and 1 X polymerase buffer containing 1.5 mM $MgCl_2$. Amplifications were carried out on a PTC 200 (MJ Research®) thermocycler using a touch down protocol as follows: 2 minutes of initial denaturation at 95°C to activate the Titanium™ Taq DNA polymerase, 30 seconds of annealing at 65°C, one minute of extension at 72°C with a reduction of 0.7°C/cycle in annealing temperature for 12 cycles. The thermocycling program was followed by another 22 more cycles of amplification as follows: 30 seconds of denaturation at 95°C, 30 seconds of annealing at 56°C and one minute of extension at 72°C. PCR products were mixed with 2 µl of loading dye and run in 1.25% agarose gel followed by staining with ethidium bromide.

6.3.5 Data analysis and cultivar identification

Amplification products were scored as 2s and 1s for the presence and absence of bands respectively. Three separate files were created and used as input files for DELTA program (Dallwitz et al. 1999). An item file that contained a DNA profile for each unique haplotype, a character file that contained the character states for each marker and an specification file that specified the number of characters and their state as well as the

number of items. These files were used with the DELTA program CONFOR and KEY to generate an identification key (Table 6.3).

6.3.6 Cluster analysis

The binary data were used to generate a genetic similarity matrix with SIMQUAL routine using the DICE coefficient (Nei 1972) from the NTSYS-pc statistical package (Rohlf 2000). Clustering of genotypes was performed using SAHN, also in NTSYS-pc, based on the genetic similarity matrix with the UPGMA method.

6.4 Results

Polymorphisms at 15 SNP loci were sufficient to distinguish between all Canadian malting barley cultivars including various biotypes contained within some of them (Table 6.2). These markers were selected from a larger pool of barley SNPs that were obtained by sequencing of selected barley genomic loci as well as mining from EST data bases from public domain. Eight genomic loci encompassing a total of 3.15 kb were selected and analyzed for this study. All of these loci were previously mapped (Kleinhofs et al. 1993; Langridge et al. 1995), except for the EST2 locus which was an unmapped locus and was sequenced based on the primers derived from the *in-silico* method of SNP mining (Table 6.2). Forty six putative SNPs were detected from all loci combined, based on multiple alignments of DNA sequences for each locus. Of these, 26 were transitions and the remaining 20 were transversion. On average one putative SNP was detected for every 68 base pairs of barley genomic sequences over the loci sequenced in this study. Of the 46 putative SNPs, 15 were validated experimentally by allele-specific PCR and found useful for the identification of six-rowed spring malting barley cultivars (Table 6.2). Five of the validated SNPs were located in MWG522 locus alone on chromosome 5.

6.4.1 SNP discovery and confirmation

SNPs were identified by sequencing a set of previously mapped barley genomic loci (Kleinhofs et al. 1993; Langridge et al. 1995). These genomic loci were chosen because they had previously been analyzed and found to contain SNPs among some *Hordeum spontaneum* and *H. vulgare* accessions (Kanazin et al. 2002). The *in-silico* approach of mining SNPs from a public database also resulted in identification of some putative

SNPs. However, only one such marker was chosen by the key generating program as informative for cultivar identification in this study (Table 6.2, 6.3). To confirm the presence of polymorphic loci, and thus the potential utility of these loci for SNP analysis, DNA was isolated from a single seed for each of five to eight barley cultivars and amplified by PCR using primers specific for each locus (Table 6.2). If multiple bands were observed after PCR amplification, gel purification to isolate the band with a size corresponding to the predicted size of the locus was necessary. The re-amplified PCR products were sequenced directly. Multiple alignments of DNA sequences from a set of cultivars were used to visually detect putative SNPs (Fig. 6.1). At least four putative SNPs were found within MWG2218 locus. Two of these, SNP024 at position 186 and SNP 025 at position 224, were subsequently validated by allele-specific PCR and were found to useful for malting barley cultivar identification.

6.4.2 SNP validation with allele-specific PCR

An allele-specific PCR primer was designed for each putative SNP that could be either 5' or 3' relative to the position of the SNP on the DNA sequence. The primer was combined with two other PCR primes flanking the genomic locus that was used for the initial sequencing experiments. Allele-specific validation of SNPs described in this paper works equally well with DNA from leaf and seed material (Fig. 6.2). The prominent band at 400 bp is MWG2218 and the polymorphic band at 250 bp is the allele-specific band.

Figure 6.1 Multiple alignment of DNA sequences from the MWG 2218 locus among 6 spring six-rowed malting barley cultivars. Sequences generated by sequencing of specific PCR products were aligned with CLUSTAL W. Polymorphic sites are located where there is a blank space in the consensus sequence.

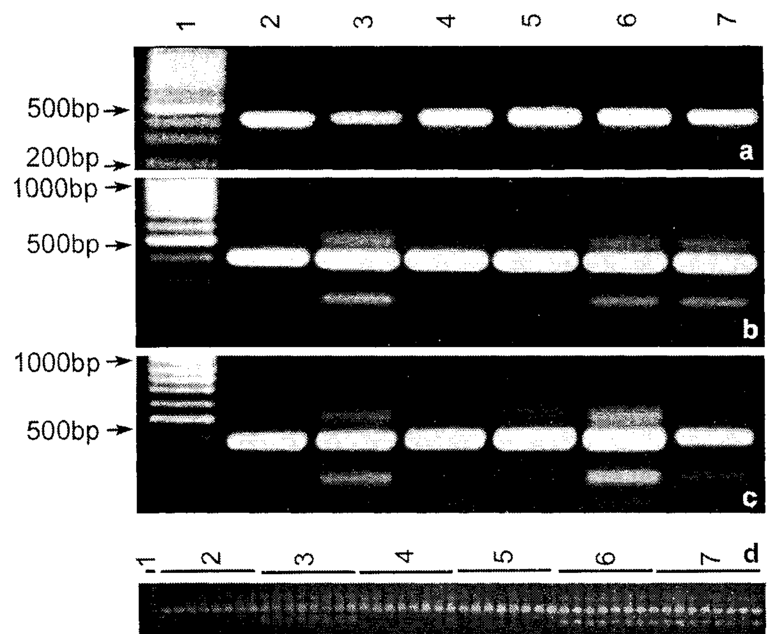
```
                    *        20         *        40         *        60         *        80         *
Argyle     : AAGAGTAGCTCCTCCAAGAAGTCCCAGTCCAGCGACCGAGAATCCTGACCGTGGTGCATTTTTATCATGTTCTGCATGGACCTCCTGAAG :  90
CDC Sisler : AAGAGTAGCTCCTCCAAGAAGTCCCAGTCCAGCGACCGAGAATCCTGACCGTGGTGCATTTTTATCATGTTCTGCATGGACCTCCTGAAG :  90
Bonanza    : AAGAGTAGCTCCTCCAAGAAGTCCCAGTCCAGCGACCGAGAATCCTGACCGTGGTGCATTTTTATCATGTTCTGCATGGACCTCCTGAAG :  90
B1602      : AAGAGTAGCTCCTCCAAGAAGTCCCAGTCCAGCGACCGAGAATCCTGACCGTGGTGCATTTTTATCATGTTCTGCATGGACCTCCTGAAG :  90
CDC Yorkton: AAGAGTAGCTCCTCCAAGAAGTCCCAGTCCAGCGACCGAGAATCCTGACCGTGGTGCATTTTTATCATGTTCTGCATGGACCTCCTGAAG :  90
Duel       : AAGAGTAGCTCCTCCAAGAAGTCCCAGTCCAGCGACCGAGAATCCTGACCGTGGTGCATTTTTATCATGTTCTGCATGGACCTCCTGAAG :  90
             AAGAGTAGCTCCTCCAAGAAGTCCCAGTCCAGCGACCGAGAATCCTGACCGTGGTGCATTTTTATCATGTTCTGCATGGACCTCCTGAAG

                   100         *       120         *       140         *       160         *       180
Argyle     : TCGGTGTAAGGGTCCATGGAGAATAGGACTATGGCCGCATCATCCCTGGTGTCATCACCTGGCCCGGCTTCTCCGGTGTCTTCCGTGACT : 180
CDC Sisler : TCGGTGTAAGGGTCCATGGAGAATAGGACTATGGCCGCATCATCCCTGGTGTCATCACCTGGCCCGGCTTCTCCGGTGTCTTCCGTGACT : 180
Bonanza    : TCGGTGTAAGGGTCCATGGAGAATAGGACTATGGCCGCATCATCCCTGGTGTCATCACCTGGCCCGGCTTCTCCGGTGTCTTCCGTGACT : 180
B1602      : TCGGTGTAAGGGTCCATGGAGAATAGGACTATGGCCGCATCATCCCTGGTGTCATCACCTGGCCCGGCTTCTCCGGTGTCTTCCGTGACT : 180
CDC Yorkton: TCGGTGTAAGGGTCCATGGAGAATAGGACTATGGCCGCATCATCCCTGGTGTCATCACCTGGCCCGGCTTCTCCGGTGTCTTCCGTGACT : 180
Duel       : TCGGTGTAAGGGTCCATGGAGAATAGGACTATGGCCGCATCATCCCTGGTGTCATCACCTGGCCCGGCTTCTCCGGTGTCTTCCGTGACT : 180
             TCGGTGTAAGGGTCCATGGAGAATAGGACTATGGCCGCATCATCCCTGGTGTCATCACCTGGCCCGGCTTCTCCGGTGTCTTCCGTGACT

                    *       200         *       220         *       240         *       260         *
Argyle     : GAGTCAGACAAGGATGATGTTCCCGCGGTGGGCTTTGTTTCGTGGACGACAGACGTGGATGACGTCCCCGGCGATGACCAGAAGCGCACG : 270
CDC Sisler : GAGTCAGACAAGGATGATGTTCCCGCGGTGGGCTTTGTTTCGTGGACGACAGACGTGGATGACGTCCCCGGCGATGACCAGAAGCGCACG : 270
Bonanza    : GAGTCAGACAAGGATGATGTTCCCGCGGTGGGCTTTGTTTCGTGGACGACAGACGTGGATGACGTCCCCGGCGATGACCAGAAGCGCACG : 270
B1602      : GAGTCCGACAAGGATGATGTTCCCGCGGTGGGCTTTGTTTCGTCGACGACAGACGTGGATGACGTCCCCGGCGATGACGAGAAGCGCACG : 270
CDC Yorkton: GAGTCCGACAAGGATGATGTTCCCGCGGTGGGCTTTGTTTCGTCGACGACAGACGTGGATGACGTCCCCGGCGATGACGAGAAGCGCACG : 270
Duel       : GAGTCCGACAAGGATGATGTTCCCGCGGTGGGCTTTGTTTCGTCGACGACAGACGTGGATGACGTCCCCGGCGATGACGAGAAGCGCACG : 270
             GAGTC GACAAGGATGATGTTCCCGCGGTGGGCTTTGTTTCGT GACGACAGACGTGGATGACGTCCCCGGCGATGAC AGAAGCGCACG

                   280         *       300         *       320         *       340         *
Argyle     : TTGTCGTCACGGTCATGGATATAGAGAGACCGGAAGTTGTCGAAGAGGAAGCGGTCGATGTCGGAGAGCGTTGCGGCGTGGCC : 353
CDC Sisler : TTGTCGTCACGGTCATGGATATAGAGAGACCGGAAGTTGTCGAAGAGGAAGCGGTCGATGTCGGAGAGCGTTGCGGCGTGGCC : 353
Bonanza    : TTGTCGTCACGGTCATGGATATAGAGAGACCGGAAGTTGTCGAAGAGGAAGCGGTCGATGTCGGAGAGCGTTGCGGCGTGGCC : 353
B1602      : TTGTCGTCACGGTCATGGATATAGAGAGACCGGAAGTTGTCGAAGAGGAAGCGGTCGATGTCGGAGAGCGTCGCGGCGTGGCC : 353
CDC Yorkton: TTGTCGTCACGGTCATGGATATAGAGAGACCGGAAGTTGTCGAAGAGGAAGCGGTCGATGTCGGAGAGCGTCGCGGCGTGGCC : 353
Duel       : TTGTCGTCACGGTCATGGATATAGAGAGACCGGAAGTTGTCGAAGAGGAAGCGGTCGATGTCGGAGAGCGTCGCGTAGTGGCC : 353
             TTGTCGTCACGGTCATGGATATAGAGAGACCGGAAGTTGTCGAAGAGGAAGCGGTCGATGTCGGAGAGCGT GCGgcGTGGCC
```

Figure 6.2 SNP identification by agarose gel electorphoresis. (a) PCR products from amplification of the MWG2218 locus of 7 cultivars with the locus specific primers. (b) PCR products from amplification of the MWG2218 locus of 7 cultivars with the locus specific primers plus the allele-specific PCR primer SNP25. Genomic DNA was extracted from leaf material. (c) A replicate of "b" but DNA was extracted from seeds. (d) SNP25 polymorphic profile of six spring six-rowed barley cultivars using eight plants per cultivar. 'CDC Sisler' has two biotypes while the rest of the six cultivars are homogenous. Lane 1, DNA size marker; lanes 2-7, PCR products form cultivars 'Argyle','B1602', 'Bonanza', 'CDC Sisler', 'CDC Yorkton', and 'Duel' respectively.



1
2
3
4
5
6
7
500bp
200bp
a
1000bp
500bp
b
1000bp
500bp
c
d

6.4.3 Cultivar identification

The identification scheme uses a convention taxonomic analytical key consisting of subordinated couplets. The first divides all cultivars into two groups with further subdivisions carried out until all cultivars are uniquely identified (Table 6.3). The couplets are sequentially numbered. The numbers to the right refer to the following couplet and the numbers in the bracket refer to the previous couplet. This key is able to identify 40 unique haplotypes among the 17 cultivars. Nine of the cultivars in this study were represented by only a single biotype as judged by this assay, while the remaining eight cultivars contained at least two biotypes (Table 6.1). 'CDC Yorkton' with eight biotypes was found to be the most heterogeneous cultivar among all the six-rowed spring malting barley in this study. Cluster analysis was performed to determine whether this variation was de-novo or the result of contamination of seed samples (not shown). All individual plants of the same cultivar clustered within its own respective cultivar, eliminating the possibility of seed contamination for the observed genetic heterogeneity within cultivars.

6.5 Discussion

All 17 cultivars of six-rowed malting barley were uniquely identified using a set of 15 SNP markers. In the identification key all SNPs were given equal weight in terms of reliability and reproducibility. Almost half of the cultivars analyzed in this study contain at least two or more biotypes (Table 6.1). The presence of biotypes in crops has been reported previously (Olufowote et al. 1997; Soleimani et al. 2002b; Soleimani et al. 2003; Perry, 2004). Biotypes may not be viewed as a problem for cultivar identification if all

biotypes within each cultivar can be assigned to a specific cultivar which was the case in this study.

Intra-cultivar variation was attributed to two factors in this study: 1) Variation related to the cultivar itself, this type of variation was detected among the individual plants within a cultivar irrespective of which locus was analyzed. 2) Variation related to a specific locus under analysis, for example 'CDC Springside' showed significant amount of intra-cultivar variation at the SNP042 locus, but it was completely homogeneous for SNP061 and SNP075 loci. This finding shows that there is a differential scale of selection pressure on various loci in the genome and that some of the observed intra-cultivar variation maybe due to hyper-variable loci which are not subjected to selection pressure. Nucleotide substitutions reported in this study were all point mutations except for an indel for SNP061 (Table 6.2). Analysis of the nature of mutations among the 46 putative SNPs in this study showed that transitions with 56% frequency were dominant types of mutations. Although the likelihood of transversion mutations is twice that of transition from a theoretical standpoint, in reality there is a bias for transitions in all genomes studied to date (Brown et al. 1982; Gojobori et al. 1982; Wakeley 1994; Ziheng and Yoder 1999; Soleimani et al. 2003). The reasons for the observed bias in transitions versus transversion rates are partly due to 5-methylcytosine deamination reactions that are known to occur in the CpG islands (Holliday and Grigg 1993). Also, in a coding DNA sequence transversions are more likely to be non-synonymous than transitions therefore causing a bias in selection for transitions.

The SNP markers that were selected for spring six-rowed malting cultivar identification in this study had allele frequencies ranging from 0.2 to 0.5 with an average

of 0.4 for the rare allele (not shown). Those markers that had frequencies below 0.2 for the rare allele were eliminated by the key generating program and were not reported here. These markers were deemed less informative in identifying cultivars and their biotypes. Determination of biotypes and their frequencies were based on a sample of 272 plants representing 17 malting cultivars. The 272 plants were assigned to 40 unique biotypes all of which were included in the identification key (Table 6.3). There was an uneven distribution of biotypes among cultivars. For example, cultivars such as 'Argyle', 'B1602', 'BT954', 'CDC Battleford', 'CDC Tisdale', 'Duel', 'Excel' and 'Legacy' were completely homogenous for all 15 markers. On the other hand 'Bonanza', 'CDC Yorkton', 'CDC Sisler', 'CDC Springside', 'Foster', 'Lacey', 'Stander' and 'Tankard' contained significant number of biotypes. 'CDC Yorkton' was represented by 8 unique biotypes. This cultivar had the highest level of intra-cultivar heterogeneity (Table 6.1). Although 'CDC Yorkton' and 'CDC Battleford' share common parentage (Table 6.1), the latter was homogeneous for all the 15 SNP markers. This finding may indicate that the observed intra-cultivar variation in the latter two cultivars is unlikely to be carryover from parents to progeny.

An efficient molecular marker-based identification method for grain cultivars including barley should have the following attributes: 1) sufficient numbers of discriminating markers between cultivars including biotypes, 2) Identifying admixtures in a sample of grain for purity testing, 3) low cost of the detection platform for routine testing 4) automation and flexibility for updates and addition of new markers/probes to cope with the introduction of new cultivars. The abundance of SNPs in the barley genome and their ease of automation make them good candidates for barley cultivar

identification. Thus SNP markers are superior to previously described molecular methods used for identification of barley cultivars. Identification schemes that were previously developed for barley (Baum et al. 1998; Baum et al. 2000) have been based on single plant per cultivar from breeders' seeds. Furthermore, the genotyping assays (i.e. concentration of template DNA, $MgCl_2$ concentration in PCR reaction, annealing temperature of primers) in these studies varied for each marker and therefore restricting their routine use for cultivar identification. The genotyping assay described here follows a common and simple protocol for all SNPs. In addition, the issue of biotypes within cultivars was addressed by sampling multiple plants per cultivar.

Only a small fraction of SNPs are informative for identification purposes. Multiple SNPs found within the same genomic locus often possessed the same allelic pattern making some of these SNPs redundant for identification purposes within the set of barley cultivars in this study. For example MWG2218 locus contained four SNPs with the same allelic pattern among the six plants that were used for sequencing (Figure 6.1). Although the DNA extracted from the seed material was crude, the quality of the assay was not affected (Figure 6.2b, 6.2c). This is an important attribute of the assay making it a fast and reliable alternative to previously described methods.

The identification scheme provided here uses a simple method for barley cultivar identification. The method can be used at any stage of plant development including seed sample. Inclusion of multiple plants from each cultivar was important to identify various biotypes. The primer sequences used with allele specific PCR in this study can be easily placed on a microchip in the future for a fully automated malting barley identification procedure.

6.6 Acknowledgments

The authors thank Ms. Jean Mellish and Dr. Tigest Demeke of the Canadian Grain Commission, Winnipeg, Manitoba and Mr. Ajit Singh Sahota of the Canadian Food Inspection Agency, Ottawa, Ontario, for providing plant material. This study was supported by a grant from Automated Quality Testing Inc. (AQT).

Table 6.2. Consensus sequence of barley genomic loci derived from sequencing five to eight cultivars. AS-PCR primers sequence is underlined and SNP position is indicated in bold face.

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
		Abg500 Length: 252
SNP001	ABG500L,5'AAGAAGAACCCGGAGAATC	1 TTTAAGAAGA ACCCGGAGAA TCTGCCTCTC GCGCCGAGGT ARATTAATCC
	T,	51 GACCGTCACT GCATTAGTGA TTCAACTT**R**T TTTGTGTTAA TTCTTGTGAT
	ABG500R,5'GCTAGAACTTGACCAATCT	101 TGTCTTAATG TCTTCTCAGA ATGACGTCAA GGATCTTTAG GAGCACTGGT
	C,	151 GCCTGATCGG CGAGGCCATC CACGACGATC GGCCTTCTCG ATGAACGGAA
	SNP001_AS_PCR,5'CAATCACAAGAAT	201 TAAATAATGG AGGACGGCAG CGAGGAGTTC GTGTCGTGTA TCACKATAGC
	TAACACAAAAC	251 CT
SNP017	ABG601L,5'AGACAGACTATTTGCAGAA	Abg601 Length: 725
	G,	1 CAGAAGCACG GGTTGAAAGT GAAAGAGGCC ACTAAGATCA TCCGATTCTA
	ABG601R,5'AGACAAAGACAAAGGACAT	51 CCCGTGTCAC TCTTTTGTTT ATTAGATAAT GATCGACCTC CTGCATCCCT
	T,	101 GTGAAGAGGA ACAAATATCA AATGCAATTT AAACTCGTAT GCTAATGGAA
	SNP017_AS_PCR,5'GCTCCGCTATTAA	151 CAGAGTACAA CATTGGATTT AAGGAAGGGC TCTTTTAGAG TAACCATTAG
	GTATTCG	201 TTATTATTAT AATAGACTGT CCATCCTAAT AAAAATATCA GATCAGTTAG
		251 GATTTGGTTT GCTCCGCTAT TAAGTATTC**R** ACATATCATC ATGATAATAA
		301 AAATATATGT TTATAGTTTT TTTAGGGCTA CAACAAAATA TCAAGGCTRT
		351 CAACGACATG GAAGTTTAGA CTAAGAAGTC CATAATGACC TTTAGCAATA

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
		401 TACAGTTAAA TTTCTCTCGT ACTCCATCCA CAGTCCGCAT ATAAAGGAGA
		451 ATTGTGTTCT GTGTCTAAAT ACTCAGATAT GTAAAATTTC ATGTATTTAT
		501 AAACATGCTT CTTAGAAAAC AAATTTGGTC AGGACTTATA AACAGAACCT
		551 CTTAAAACAT GAAATCTGGA AGAACTGAAC AGCAGCCCTA TAGTAAGAGC
		601 GTAAACAGAG TTAAGGGGAC TTCCTTTTCT TGTTGAAGTT TTTTATATGT
		651 AAACAGATCA AACAGAATGT AAAAAGTTAG AAGAAATTGT CGACCATACT
		701 TGTCAATAAT ATGACCTCTT TCAAT
SNP018	ABG601L,5'AGACAGACTATTTGCAGAA	Abg601 Length: 725
	G,	1 CAGAAGCACG GGTTGAAAGT GAAAGAGGCC ACTAAGATCA TCCGATTCTA
	ABG601R,5'AGACAAAGACAAAGGACAT	51 CCCGTGTCAC TCTTTTGTTT ATTAGATAAT GATCGACCTC CTGCATCCCT
	T,	101 GTGAAGAGGA ACAAATATCA AATGCAATTT AAACTCGTAT GCTAATGGAA
	SNP018_AS_PCR,5'CTACAACAAAATA	151 CAGAGTACAA CATTGGATTT AAGGAAGGGC TCTTTTAGAG TAACCATTAG
	TCAAGGCTG	201 TTATTATTAT AATAGACTGT CCATCCTAAT AAAAATATCA GATCAGTTAG
		251 GATTTGGTTT GCTCCGCTAT TAAGTATTCR ACATATCATC ATGATAATAA
		301 AAATATATGT TTATAGTTTT TTTAGGGCTA CAACAAAATA TCAAGGCT**R**T
		351 CAACGACATG GAAGTTTAGA CTAAGAAGTC CATAATGACC TTTAGCAATA
		401 TACAGTTAAA TTTCTCTCGT ACTCCATCCA CAGTCCGCAT ATAAAGGAGA

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
		451 ATTGTGTTCT GTGTCTAAAT ACTCAGATAT GTAAAATTTC ATGTATTTAT
		501 AAACATGCTT CTTAGAAAAC AAATTTGGTC AGGACTTATA AACAGAACCT
		551 CTTAAAACAT GAAATCTGGA AGAACTGAAC AGCAGCCCTA TAGTAAGAGC
		601 GTAAACAGAG TTAAGGGGAC TTCCTTTTCT TGTTGAAGTT TTTTATATGT
		651 AAACAGATCA AACAGAATGT AAAAAGTTAG AAGAAATTGT CGACCATACT
		701 TGTCAATAAT ATGACCTCTT TCAAT
SNP020	MWG502L, 5'ATCAAGATGAGGAAGCTGA	MWG502 Length: 650
	TGCAGT,	1 CCGGTTTTAG CAAGCATCGA TCGACRCCGA TATCGTAATT AAATGTGGGG
	MWG502R, 5'TCTTGTCTGAAGCATGGAA	51 GTTTAGGATA AAATAATTGS CATGATTAGT TTGATGGATA ATTAATCCG**Y**
	CCAGAT,	101 TGTTTGGCCG CCTCGGTACG CCATCGTCCT TGGGTCCAAC TAATCGCCGT
	SNP020_AS_PCR, 5'TTTGATGGATAAT	151 TTATGGACTA TAAACATTTC TTTCATRCAA AACAAAACAG AGCGTCAATG
	TAATCCGC	201 AAACTGCCAT CTAACTCCCA ACCACAGTAA GTTAATTACC ATATAAAACA
		251 GACATATGCR CAATGAAAAA CAAGAGGAAA CTGAATTATG CCGAGTTCTG
		301 TTCRCGCGAA TATTCGGTTA GCACGGATAT TAAAGCCCAC TGTTTAGCTA
		351 AGCTGTTGCT GAGGTAGGCG TGACGCATGC CGATTAATTA ATTAGCATCA
		401 CTATATTGGC CACACAAAAT GTAGTAGGGG GGTTAGGGGT TGCTGTCATG

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
		451 GTCCTTGATT TTTAGTGTGG GTTGCTGTTT CTGCTTGCTT ACTCAGCTTT
		501 CCTACATTAA CTACTGGCAT CTTTTTTCCG AGCCATTTGT CGCCATTATA
		551 CGCGTGTTCC TACTTGATTG CTTTCAATCA CGGAGAAGCT TCTCTTTGAA
		601 TTCGGTGGCC GCGGCGCGCC ACGTGGCCGC CCCTGATTGG GCTGTCAAAC
SNP022	MWG502L, 5'ATCAAGATGAGGAAGCTGA	MWG502 Length: 650
	TGCAGT,	1 CCGGTTTTAG CAAGCATCGA TCGACRCCGA TATCGTAATT AAATGTGGGG
	MWG502R, 5'TCTTGTCTGAAGCATGGAA	51 GTTTAGGATA AAATAATTGS CATGATTAGT TTGATGGATA ATTAATCCGY
	CCAGAT,	101 TGTTTGGCCG CCTCGGTACG CCATCGTCCT TGGGTCCAAC TAATCGCCGT
	SNP022_AS_PCR, 5'CATATAAAACAGA	151 TTATGGACTA TAAACATTTC TTTCATRCAA AACAAAACAG AGCGTCAATG
	CATATGCG	201 AAACTGCCAT CTAACTCCCA ACCACAGTAA GTTAATTACC ATATAAAACA
		251 GACATATGC**R** CAATGAAAAA CAAGAGGAAA CTGAATTATG CCGAGTTCTG
		301 TTCRCGCGAA TATTCGGTTA GCACGGATAT TAAAGCCCAC TGTTTAGCTA
		351 AGCTGTTGCT GAGGTAGGCG TGACGCATGC CGATTAATTA ATTAGCATCA
		401 CTATATTGGC CACACAAAAT GTAGTAGGGG GGTTAGGGGT TGCTGTCATG
		451 GTCCTTGATT TTTAGTGTGG GTTGCTGTTT CTGCTTGCTT ACTCAGCTTT

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
		501 CCTACATTAA CTACTGGCAT CTTTTTTCCG AGCCATTTGT CGCCATTATA
		551 CGCGTGTTCC TACTTGATTG CTTTCAATCA CGGAGAAGCT TCTCTTTGAA
		601 TTCGGTGGCC GCGGCGCGCC ACGTGGCCGC CCCTGATTGG GCTGTCAAAC
SNP024	MWG2218L,5'AAGAGTAGCTCCTCCAAG	MWG2218 Length: 362
	AAGTCCC,	1 AAGAGTAGCT CCTCCAAGAA GTCCCAGTCC AGCGACCGAG AATCCTGACC
	MWG2218R,5'CCCACGTACCCCGTCTCT	51 GTGGTGCATT TTTATCATGT TCTGCATGGA CCTCCTGAAG TCGGTGTAAG
	GGACATC,	101 GGTCCATGGA GAATAGGACT ATGGCCGCAT CATCCCTGGT GTCATCACCT
	SNP024_AS_PCR,5'CGGGAACATCATC	151 GGCCCGGCTT CTCCGGTGTC TTCCGTGACT GAGTC**M**GACA AGGATGATGT
	CTTGTCG	201 TCCCGCGGTG GGCTTTGTTT CGTSGACGAC AGACGTGGAT GACGTCCCCG
		251 GCGATGACSA GAAGCGCACG TTGTCGTCAC GGTCATGGAT ATAGAGAGAC
		301 CGGAAGTTGT CGAAGAGGAA GCGGTCGATG TCGGAGAGCG TYGCGGCGTG
		351 GCCTCCGCTG TT

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
SNP025	MWG2218L,5'AAGAGTAGCTCCTCCAAG	MWG2218 Length: 362
	AAGTCCC,	1 AAGAGTAGCT CCTCCAAGAA GTCCCAGTCC AGCGACCGAG AATCCTGACC
	MWG2218R,5'CCCACGTACCCCGTCTCT	51 GTGGTGCATT TTTATCATGT TCTGCATGGA CCTCCTGAAG TCGGTGTAAG
	GGACATC,	101 GGTCCATGGA GAATAGGACT ATGGCCGCAT CATCCCTGGT GTCATCACCT
	SNP025_AS_PCR,5'GTCATCCACGTCT	151 GGCCCGGCTT CTCCGGTGTC TTCCGTGACT GAGTCMGACA AGGATGATGT
	GTCCTCC	201 TCCCGCGGTG GGCTTTGTTT CGT**S**GACGAC AGACGTGGAT GACGTCCCCG
		251 GCGATGACSA GAAGCGCACG TTGTCGTCAC GGTCATGGAT ATAGAGAGAC
		301 CGGAAGTTGT CGAAGAGGAA GCGGTCGATG TCGGAGAGCG TYGCGGCGTG
		351 GCCTCCGCTG TT
SNP026	MWG2218L,5'AAGAGTAGCTCCTCCAAG	MWG2218 Length: 362
	AAGTCCC,	1 AAGAGTAGCT CCTCCAAGAA GTCCCAGTCC AGCGACCGAG AATCCTGACC
	MWG2218R,5'CCCACGTACCCCGTCTCT	51 GTGGTGCATT TTTATCATGT TCTGCATGGA CCTCCTGAAG TCGGTGTAAG
	GGACATC,	101 GGTCCATGGA GAATAGGACT ATGGCCGCAT CATCCCTGGT GTCATCACCT
	SNP026_AS_PCR,5'GACGACAACGTGC	151 GGCCCGGCTT CTCCGGTGTC TTCCGTGACT GAGTCMGACA AGGATGATGT
	GCTTCTC	201 TCCCGCGGTG GGCTTTGTTT CGTSGACGAC AGACGTGGAT GACGTCCCCG
		251 GCGATGAC**S**A GAAGCGCACG TTGTCGTCAC GGTCATGGAT ATAGAGAGAC
		301 CGGAAGTTGT CGAAGAGGAA GCGGTCGATG TCGGAGAGCG TYGCGGCGTG
		351 GCCTCCGCTG TT

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
SNP027	MWG2218L,5'AAGAGTAGCTCCTCCAAG	MWG2218 Length: 362
	AAGTCCC,	1 AAGAGTAGCT CCTCCAAGAA GTCCCAGTCC AGCGACCGAG AATCCTGACC
	MWG2218R,5'CCCACGTACCCCGTCTCT	51 GTGGTGCATT TTTATCATGT TCTGCATGGA CCTCCTGAAG TCGGTGTAAG
	GGACATC,	101 GGTCCATGGA GAATAGGACT ATGGCCGCAT CATCCCTGGT GTCATCACCT
	SNP027_SNAPSHOT,5'GGTCGATGTCG	151 GGCCCGGCTT CTCCGGTGTC TTCCGTGACT GAGTCMGACA AGGATGATGT
	GAGAGCGT	201 TCCCGCGGTG GGCTTTGTTT CGTSGACGAC AGACGTGGAT GACGTCCCCG
		251 GCGATGACSA GAAGCGCACG TTGTCGTCAC GGTCATGGAT ATAGAGAGAC
		301 CGGAAGTTGT CGAAGAGGAA GCGGTCGATG TCGGAGAGCG T**Y**GCGGCGTG
		351 GCCTCCGCTG TT
SNP029	ABC302L,5'ATAAAGGAGAAGATTGAGT	Abc302 Length: 715
	C,	1 TTTACAGCCA TCCATCCCAG TTACTGAAAT GTACTACCCC CMCACCCGCA
	ABC302R,5'ATAAGGAACAGGAACAGAG	51 CCAAAGCCGC TGCCGATGAG GCGATTATTC GTTTATTCTG CTTT**M**AGGAA
	T,	101 TGTCCCCAAA ACTGAAACTG CACAGGGTCG ATTCAATTGC GCAACAGCAC
	SNP029_AS_PCR,5'TTATTCGTTTATT	151 AGGACCCCCA GCCTGGAAAC GGCGCCGGCG GAGCTCATGA CAGACCCATG
	CTGCTTTC	201 ACCTTCCCGG TGGCCGTGTC GATGTCGATT TCATAGAAGC AGGGCCTCGT
		251 TGGGAGACCG GGCATCGTGC TCATGGTGCC GACGAGCGGG TAGATGAAAC
		301 CGGCACCGAT GCTGGCCCGC ACGTCCCTTA TCGGCAGCAC AAAGCCGGAT
		351 GGCGCGCCCT TCATGGACGG CACATGGGAG AAGGAGTACT GAGTCTTCGC
		401 CATGCATATC GGGAGGTTGG AGAAGCCTTG CTTGGTGTAC ATCTCGATCT

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
		451 GCTTTTCAGC CTACAGGAAA GAGATAGAGA AGTCAGCAGA GGCAAAATCT
		501 TATGACTGCA GCTAGACACA GAGTGTTCCA TATTATCCCT CTACTGCTAT
		551 CTCTTAAGCA TTTCTCATAT ATCTCAAAAC ATGAAACTGT GGAGAACTCT
		601 GCCACACCAA ATTATCCTTT GGGAAGATCG CATATTTCTG GAATGTATCA
		651 CAGGCCATCA TTCTCATAAT TAGAACAAAT CTCCCAATCA CAGTAAGTAA
		701 TTACAGCAGA TGATG
SNP030	ABC153L, 5'GCCTCTGCCGCTGGAACTA	Abc153 Length: 189
	C,	1 TGCCTCTGCC GSCGGAACTA CGCGCTCAAG TCCATCCAGC GAGCTGCCAA
	ABC153R, 5'AACACCTCCTGGCTCTCAG	51 GGCCAGGGCC AACGTGCGAG CCAGCGT**S**TC GCAGGGCAAG AAGCAGCTAC
	,	101 CTGCCTCTGC CTCGGCCATG GCGGAGAAGG CCGAGGAKGG GCTGAGGACC
	SNP030_AS_PCR, 5'CAACGTGCGAGCC	151 GTCATGTACC TCAGCTGCTG GGGTCCCAAT TAGTAGAGC
	AGCGTC	
SNP031	ABC153L, 5'GCCTCTGCCGCTGGAACTA	Abc153 Length: 189
	C,	1 TGCCTCTGCC GSCGGAACTA CGCGCTCAAG TCCATCCAGC GAGCTGCCAA
	ABC153R, 5'AACACCTCCTGGCTCTCAG	51 GGCCAGGGCC AACGTGCGAG CCAGCGTSTC GCAGGGCAAG AAGCAGCTAC
	,	101 CTGCCTCTGC CTCGGCCATG GCGGAGAAGG CCGAGGA**K**GG GCTGAGGACC

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
	SNP031_AS_PCR, 5'ATGACGGTCCTCA	151 GTCATGTACC TCAGCTGCTG GGGTCCCAAT TAGTAGAGC
	GCCCA	
SNP042	ABG058L, 5'AGAACTGGAGGAACGAGAT	Abg58
	A,	1 AAGGACTGCA TATCTGTTGT GACAAGTGTA ACTGCGTCCC TTCCGGCACC
	ABG058R, 5'ACCCAGGCTTGATTATTAG	51 GGCCAAGAGA CCCGTAACGA ATGCCCCTGC TATGCCAAAT TGGTTGACTC
	G,	101 AAAGAC**Y**GGC AAGCTCAAGT GCCCATAATC GATCGATCAT AATGCATTCA
	SNP042_AS_PCR, 5'CCAAATTGGTTGA	151 TCTACGTGCC AGACTGCTAC CTGRMTCGAA TATGCTCTCT CTTGTTTTAC
	CTCAAAGACC	201 CTGTCCCAAT GTTCAATAAT CTCTCTTGCC ATcatgccat GTAAACCAAG
		251 AAAARTACGT ACATATATAT TATCGATATA TTGGTATGCT CAATCAAAGT
		301 CTTATAcAAT ACAACCATCA ACCATGGTCT CATTTTGGAT CTAGTTAAAA
		351 AGGKTYTTTT KTTCTTGRGG AAAMAAATCT TTGAAGTATC CAACCCTACA
		401 AAAGGAAAGT TCTCTTTTTC CTTACCGGCC CATC
SNP043	ABG058L, 5'AGAACTGGAGGAACGAGAT	Abg58 Length: 434
	A,	1 AAGGACTGCA TATCTGTTGT GACAAGTGTA ACTGCGTCCC TTCCGGCACC
	ABG058R, 5'ACCCAGGCTTGATTATTAG	51 GGCCAAGAGA CCCGTAACGA ATGCCCCTGC TATGCCAAAT TGGTTGACTC
	G,	101 AAAGACYGGC AAGCTCAAGT GCCCATAATC GATCGATCAT AATGCATTCA
	SNP043_AS_PCR, 5'TCTACGTGCCAGA	151 TCTACGTGCC AGACTGCTAC CTG**R**MTCGAA TATGCTCTCT CTTGTTTTAC

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
	CTGCTACCTGA	201 CTGTCCCAAT GTTCAATAAT CTCTCTTGCC ATcatgccat GTAAACCAAG
		251 AAAARTACGT ACATATATAT TATCGATATA TTGGTATGCT CAATCAAAGT
		301 CTTATAcAAT ACAACCATCA ACCATGGTCT CATTTTGGAT CTAGTTAAAA
		351 AGGKTYTTTT KTTCTTGRGG AAAMAAATCT TTGAAGTATC CAACCCTACA
		401 AAAGGAAAGT TCTCTTTTTC CTTACCGGCC CATC
SNP044	ABG058L,5'AGAACTGGAGGAACGAGAT	Abg58 Length: 434
	A,	1 AAGGACTGCA TATCTGTTGT GACAAGTGTA ACTGCGTCCC TTCCGGCACC
	ABG058R,5'ACCCAGGCTTGATTATTAG	51 GGCCAAGAGA CCCGTAACGA ATGCCCCTGC TATGCCAAAT TGGTTGACTC
	G,	101 AAAGACYGGC AAGCTCAAGT GCCCATAATC GATCGATCAT AATGCATTCA
	SNP044_AS_PCR,5'TAAAACAAGAGAG	151 TCTACGTGCC AGACTGCTAC CTGR**M**TCGAA TATGCTCTCT CTTGTTTTAC
	AGCATATTCGAG	201 CTGTCCCAAT GTTCAATAAT CTCTCTTGCC ATcatgccat GTAAACCAAG
		251 AAAARTACGT ACATATATAT TATCGATATA TTGGTATGCT CAATCAAAGT
		301 CTTATAcAAT ACAACCATCA ACCATGGTCT CATTTTGGAT CTAGTTAAAA
		351 AGGKTYTTTT KTTCTTGRGG AAAMAAATCT TTGAAGTATC CAACCCTACA
		401 AAAGGAAAGT TCTCTTTTTC CTTACCGGCC CATC

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
SNP045	ABG058L,5'AGAACTGGAGGAACGAGAT	Abg58 Length: 434
	A,	1 AAGGACTGCA TATCTGTTGT GACAAGTGTA ACTGCGTCCC TTCCGGCACC
	ABG058R,5'ACCCAGGCTTGATTATTAG	51 GGCCAAGAGA CCCGTAACGA ATGCCCCTGC TATGCCAAAT TGGTTGACTC
	G,	101 AAAGACYGGC AAGCTCAAGT GCCCATAATC GATCGATCAT AATGCATTCA
	SNP045_AS_PCR,5'TAATCTCTCTTGC	151 TCTACGTGCC AGACTGCTAC CTGRMTCGAA TATGCTCTCT CTTGTTTTAC
	CATCATGCCA	201 CTGTCCCAAT GTTCAATAAT CTCTCTTGCC AT**catgccat** GTAAACCAAG
		251 AAAARTACGT ACATATATAT TATCGATATA TTGGTATGCT CAATCAAAGT
	8 bp indel	301 CTTATAcAAT ACAACCATCA ACCATGGTCT CATTTTGGAT CTAGTTAAAA
		351 AGGKTYTTTT KTTCTTGRGG AAAMAAATCT TTGAAGTATC CAACCCTACA
		401 AAAGGAAAGT TCTCTTTTTC CTTACCGGCC CATC
SNP046	ABG058L,5'AGAACTGGAGGAACGAGAT	Abg58 Length: 434
	A,	1 AAGGACTGCA TATCTGTTGT GACAAGTGTA ACTGCGTCCC TTCCGGCACC
	ABG058R,5'ACCCAGGCTTGATTATTAG	51 GGCCAAGAGA CCCGTAACGA ATGCCCCTGC TATGCCAAAT TGGTTGACTC
	G,	101 AAAGACYGGC AAGCTCAAGT GCCCATAATC GATCGATCAT AATGCATTCA
	SNP046_AS_PCR,5'AATATATCGATAA	151 TCTACGTGCC AGACTGCTAC CTGRMTCGAA TATGCTCTCT CTTGTTTTAC
	TATATATGTACGTAC	201 CTGTCCCAAT GTTCAATAAT CTCTCTTGCC ATcatgccat GTAAACCAAG
		251 AAAA**R**TACGT ACATATATAT TATCGATATA TTGGTATGCT CAATCAAAGT
		301 CTTATAcAAT ACAACCATCA ACCATGGTCT CATTTTGGAT CTAGTTAAAA
		351 AGGKTYTTTT KTTCTTGRGG AAAMAAATCT TTGAAGTATC CAACCCTACA

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
		401 AAAGGAAAGT TCTCTTTTTC CTTACCGGCC CATC
SNP047	ABG058L,5'AGAACTGGAGGAACGAGAT A, ABG058R,5'ACCCAGGCTTGATTATTAG G, SNP047_AS_PCR,5'CTGAACTTGAAAA ATTGTTAATGAT	Abg058 Reverse complement Length: 430 1 AGTCTTATAC AATACAACCA TCAACCATGG TCTCATTTTG GATCTAGTTA 51 AAAAGGMTRT TTTTTTTGTC TTGASYAAAC AAT----ACC AGTCTTTCAG 101 GAACAGAAAA TAATAGTATG TGGGTTCTTG GATAATACCG AAGTCATCTT 151 CTTTCAACAT TCACTAAGGA GCGCCCCGAG AAAAGCCCAA TGCAGTAGAA 201 ATGGAACATT TTAGAAAATC ACAAACTCTT TTTTTTC**W**TC ATTAACAATT 251 TTTCAAGTTC AGAAACATTT TTTATTTTTA AAGATATTTT TGAATTTTTG 301 AACAATTCTT AAATGCACAA ACATTCATGA TTTCACCATT TCTTTTCAAA 351 AATTAAAAAC CAAAACCGGT GGCTGGTTAT TTTAGACAGC GAAGCAGCAT 401 AAAACATGCG AGCAAGCCAG CAAGCAAGCG

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
SNP057	MWG522L,5'AACCTCTTGTTACAAATGT CATGTG, MWG522R,5'TGCATGTAGGGGCTAGTTA CTATAA, SNP057_SNAPSHOT_U2,5'ATGGCGCT TATCAGATTTTCTTCATCTCT	MWG522 Length: 718 1 ACTCTCATCC CGGGGTCGAG TCCTAACTCA CCGGTCTGAA GTGCACGCAT 51 GCTATWARRA ACCTGGTGAA TCAGATTTTC TTCATCTCT**R** YATGGATTTG 101 TTGCAGGCTC TTTACCAGAT AGCRGTTCTT CTCRTCTTCA ATTTTGACGG 151 CAAAAGGATT TTCCATTTGC ATAATGAAAG TCGAGAGCAC GCTGACAAAA 201 TTAAGAACAC CTTTGTCTTC AATGCATTTG TCTTCTGCCA AGTAAGATCA 251 CTTGGAGGAA TTTTCTTCCA ACATCCACCG ATTTAATATT CGAAATTYGA 301 AACTATAGTT TGGTGTCTCT TCCTAAGCTA TCATTGATTC CTTCAGCACA 351 TCCTTGCCGC ATACCAATTT TGTCATCGGA TGTGTCATAT ATGTAAGAAA 401 TCATAGAATT CTTTTCTTAT GCCTCATGTT CCATTTATGT CAGATCTTCA 451 ATGAGTTCAA TGCTCGCAAG CCTGAGGAGA AGAATGTCTT CCTAGGAGTT 501 ACAAGCAACC GCCTTTTCAT GGGTATAGTG GGTATAACTA CCATTCTTCA 551 GGTAACAATG CCGTAACATG CTAGTATATG AATGCATACA TGTTTTCACA 601 AGGCACAAGT AGGATCATCA TCTCATCTGG KSGATCTTAA AAGAAACAAC 651 TGAGTTTCTG AGCATCATCC ATCTTGCAGA TCTTGGATAA TKGAATTTCT 701 CGGGAAGTTC TTCGGAAC

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
SNP059	MWG522L,5'AACCTCTTGTTACAAATGT	MWG522 Length: 718
	CATGTG,	1 ACTCTCATCC CGGGGTCGAG TCCTAACTCA CCGGTCTGAA GTGCACGCAT
	MWG522R,5'TGCATGTAGGGGCTAGTTA	51 GCTATWARRA ACCTGGTGAA TCAGATTTTC TTCATCTCTR YATGGATTTG
	CTATAA,	101 TTGCAGGCTC TTTACCAGAT AGCRGTTCTT CTCRTCTTCA ATTTTGACGG
	SNP059_AS_PCR,5'CCACCGATTTAAT	151 CAAAAGGATT TTCCATTTGC ATAATGAAAG TCGAGAGCAC GCTGACAAAA
	ATTCGAAATTC	201 TTAAGAACAC CTTTGTCTTC AATGCATTTG TCTTCTGCCA AGTAAGATCA
		251 CTTGGAGGAA TTTTCTTCCA ACATCCACCG ATTTAATATT CGAAATT**Y**GA
		301 AACTATAGTT TGGTGTCTCT TCCTAAGCTA TCATTGATTC CTTCAGCACA
		351 TCCTTGCCGC ATACCAATTT TGTCATCGGA TGTGTCATAT ATGTAAGAAA
		401 TCATAGAATT CTTTTCTTAT GCCTCATGTT CCATTTATGT CAGATCTTCA
		451 ATGAGTTCAA TGCTCGCAAG CCTGAGGAGA AGAATGTCTT CCTAGGAGTT
		501 ACAAGCAACC GCCTTTTCAT GGGTATAGTG GGTATAACTA CCATTCTTCA
		551 GGTAACAATG CCGTAACATG CTAGTATATG AATGCATACA TGTTTTCACA
		601 AGGCACAAGT AGGATCATCA TCTCATCTGG KSGATCTTAA AAGAAACAAC
		651 TGAGTTTCTG AGCATCATCC ATCTTGCAGA TCTTGGATAA TKGAATTTCT
		701 CGGGAAGTTC TTCGGAAC
SNP066	ABG609L,5'ATCAATGGAGATTTGCTTA	Abg609 Length: 694
	C,	1 TGCCCCACAC GTACCTTCTC TCTTTCTCTG CTATGAACCT TATTTGTCTC
	ABG609R,5'GTGTTTACATGCTTGTCAT	51 ACAGGTCTGT ATCCTCTTCT CATAAAAAGC TATCTTGAAC TCTTCAGCTT

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
	A,	101 CTGCAATTAT TTGGGCGCGT AGCTCCTTCT CCTTCCGTTC TTTTTCCTCG
	SNP066_SNAPSHOT_U1,5'GGCAATGA	151 AGAACTAGAG CATTTTGCCT GAAACCACAA TGAAGCTGTT CAAACTAGAG
	AGCTGTTCAAACTAGAGC	201 C**W**MATAGACA GAGCAGAATT TAGACCATGC AAGGGAAACC AGAAAAAaGG
		251 AAAATTAAAG GAAGAACAAT TGCTACATCA TTTAATTTTA CAGGGTACTG
		301 TKGTAGGCCA GTTATCAGGG TTCAGGTATC AGTACACAAC ACATGGCCAT
		351 TAACCAAGGA CATTGCATAA CTAATTTACT GGGCTCTCTG TATGACAACT
		401 atcacagcaa aaatgaagca aaataCTGCC ATCACAACGC CTAGKTCACA
		451 CAGAACACAC CACAAATTTG GGCAAAATTA TACAAAACAA ATTCATCGAT
		501 TCTAGAACTA ATAATCTTAC CGTCTTCAAT ACAGAACTAT AGTTGCCGAT
		551 AGGCATGCAG CAAATTACAG AAAAAAAAGA ACATGTTCAG AAAAGCAATG
		601 CAAGKGCAGA GTACAAGAAT ATGCCACTCA GACACACACT TCAGAAACAC
		651 AAATAATGCA TGCTCGGCCA CTGGCAGACA AGCGGGCTGC ATTG

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
SNP075	MWG522L,5'AACCTCTTGTTACAAATGT	MWG522 Length: 718
	CATGTG,	1 ACTCTCATCC CGGGGTCGAG TCCTAACTCA CCGGTCTGAA GTGCACGCAT
	MWG522R,5'TGCATGTAGGGGCTAGTTA	51 GCTAT**W**ARRA ACCTGGTGAA TCAGATTTTC TTCATCTCTR YATGGATTTG
	CTATAA,	101 TTGCAGGCTC TTTACCAGAT AGCRGTTCTT CTCRTCTTCA ATTTTGACGG
	SNP075_AS_PCR,5'CTGAAGGCACGCA	151 CAAAAGGATT TTCCATTTGC ATAATGAAAG TCGAGAGCAC GCTGACAAAA
	TGCTATT	201 TTAAGAACAC CTTTGTCTTC AATGCATTTG TCTTCTGCCA AGTAAGATCA
		251 CTTGGAGGAA TTTTCTTCCA ACATCCACCG ATTTAATATT CGAAATTYGA
		301 AACTATAGTT TGGTGTCTCT TCCTAAGCTA TCATTGATTC CTTCAGCACA
		351 TCCTTGCCGC ATACCAATTT TGTCATCGGA TGTGTCATAT ATGTAAGAAA
		401 TCATAGAATT CTTTTCTTAT GCCTCATGTT CCATTTATGT CAGATCTTCA
		451 ATGAGTTCAA TGCTCGCAAG CCTGAGGAGA AGAATGTCTT CCTAGGAGTT
		501 ACAAGCAACC GCCTTTTCAT GGGTATAGTG GGTATAACTA CCATTCTTCA
		551 GGTAACAATG CCGTAACATG CTAGTATATG AATGCATACA TGTTTTCACA
		601 AGGCACAAGT AGGATCATCA TCTCATCTGG KSGATCTTAA AAGAAACAAC
		651 TGAGTTTCTG AGCATCATCC ATCTTGCAGA TCTTGGATAA TKGAATTTCT
		701 CGGGAAGTTC TTCGGAAC

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
SNP078	MWG522L,5'AACCTCTTGTTACAAATGT	MWG522 Length: 718
	CATGTG,	1 ACTCTCATCC CGGGGTCGAG TCCTAACTCA CCGGTCTGAA GTGCACGCAT
	MWG522R,5'TGCATGTAGGGGCTAGTTA	51 GCTATWARRA ACCTGGTGAA TCAGATTTTC TTCATCTCTR YATGGATTTG
	CTATAA,	101 TTGCAGGCTC TTTACCAGAT AGCRGTTCTT CTC**R**TCTTCA ATTTTGACGG
	SNP078_AS_PCR,5'CTTTACCAGATAG	151 CAAAAGGATT TTCCATTTGC ATAATGAAAG TCGAGAGCAC GCTGACAAAA
	CRGTTCTTCTCG	201 TTAAGAACAC CTTTGTCTTC AATGCATTTG TCTTCTGCCA AGTAAGATCA
		251 CTTGGAGGAA TTTTCTTCCA ACATCCACCG ATTTAATATT CGAAATTYGA
		301 AACTATAGTT TGGTGTCTCT TCCTAAGCTA TCATTGATTC CTTCAGCACA
		351 TCCTTGCCGC ATACCAATTT TGTCATCGGA TGTGTCATAT ATGTAAGAAA
		401 TCATAGAATT CTTTTCTTAT GCCTCATGTT CCATTTATGT CAGATCTTCA
		451 ATGAGTTCAA TGCTCGCAAG CCTGAGGAGA AGAATGTCTT CCTAGGAGTT
		501 ACAAGCAACC GCCTTTTCAT GGGTATAGTG GGTATAACTA CCATTCTTCA
		551 GGTAACAATG CCGTAACATG CTAGTATATG AATGCATACA TGTTTTCACA
		601 AGGCACAAGT AGGATCATCA TCTCATCTGG KSGATCTTAA AAGAAACAAC
		651 TGAGTTTCTG AGCATCATCC ATCTTGCAGA TCTTGGATAA TKGAATTTCT
		701 CGGGAAGTTC TTCGGAAC

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
SNP081	AL503851L,5'CAGGCAAGATTGACGCA	AL503851 gb partial Length: 678
	GTA,	1 CAATTCTTCA ATGTATTCAT TTTGCTGACA AAATGTGGCA TCAATTCACC
	AL503851R,5'ATCCGCAGCAGATCGAG	51 GGCTGTTTGC GTAATGAACA TTTTGTGGCA ATTCAATGCC AGAAAAGGGG
	GAC,	101 ACAGAAGACT AAATAATCAT GTAACTAGCT GCGCAAATAT GTAGTGAGAG
	SNP081_AS_PCR,5'GCTTAACAGTTCC	151 GGACCCCAAC CAGTCCAGGA GAAACCAGGA AGTAAGCTCC ATCCCAAAAT
	TCAG	201 GTACCAATTC AGCG**S**TGAGG AACTGTTAAG CAACCAGTGT CCAAAATACA
		251 GTACCCAGCT AGCCTAGCTA ATGCAGTGTT AAGTTGTTCC AAAATCTTAC
		301 TGGCATCAGT AAGTAATTTG GTCAGTGTGA TAAGGGAACT GGAGTAGTGA
		351 AGAGAGAGAG AGAGAGAAGC AGGAACAGAG TGATGATGAT CACCTTCATG
		401 TCGTCCTCGA TCTTGGCGAC CTGGGACTGG ATGCTGGCGA CGGTCTCCTT
		451 GAGCGGGGCG TTGACGGGGT ACTTGCCCTC GTCCCACACG AACCTGCATC
		501 CATCCATCCA TCCTCAGCAG ATAAGCAAGA ACTGGTGGGC AGATGAAGGA
		551 AGAATGTAAG CAGCGACAGG CAGAGGGAGG GGGACCTGGT GAGGTACTTG
		601 TCGACGGGGA ACCCCCACNA AGGTGANGGN GCCNGGGTTT GNANCCCCTT
		651 CNGGGGGNGN TTCCAGGGTC CTCNTTTT
SNP215	EST215L, 5'	No consensus sequence available
	GGTGGGACTGATAACCATTTA,	
	EST215R, 5'	
	GGATGTTGCTGTCGGATTGC,	

SNP	Primers (left, right, nested)	Barley genomic loci - consensus sequence, Nested primer underlined; SNPs in bold face.
	SNP215_AS_PCR, 5' ATCCTCTTGATGTAAGCGCA	

Table 6.3. Identification key of six-rowed spring malting barley cultivars using SNPs. To identify a cultivar, start from couplet 1 and follow the steps given on the right. Each couplet carries two alternatives; it hits a further couplet number to follow or a cultivar name. The numbers after the dashes following the cultivar names indicate one of the biotypes.

1(0)	SNP081 absent	2
	SNP081 present	22
2(1)	SNP215 absent	3
	SNP215 present	13
3(2)	SNP059 absent	4
	SNP059 present	8
4(3)	SNP061 absent	5
	SNP061 present	7
5(4)	SNP057 absent	23 CDC Yorkton-16
	SNP057 present	6
6(5)	SNP075 absent	16 CDC Yorkton-2
	SNP075 present	22 CDC Yorkton-13
7(4)	SNP024 absent	18 CDC Yorkton-6
	SNP024 present	10 Bonanza-8
8(3)	SNP057 absent	9
	SNP057 present	11
9(8)	SNP018 absent	35 Robust-1
	SNP018 present	10

10(9) SNP077 absent ....19 CDC Yorkton-8

SNP077 present ....17 CDC Yorkton-5

11(8) SNP018 absent 25 Excel-1

SNP018 present 12

12(11) SNP075 absent ....20 CDC Yorkton-9

SNP075 present ....21 CDC Yorkton-12

13(2) SNP018 absent 14

SNP018 present 18

14(13) SNP020 absent 15

SNP020 present 16

15(14) SNP059 absent 30 Foster-16

SNP059 present ....1 CDC Battleford-1

16(14) SNP001 absent 17

SNP001 present 4 Argyle-1

17(16) SNP030 absent 8 Bonanza-1

SNP030 present 11 Bonanza-9

18(13) SNP020 absent 19

SNP020 present 21

19(18) SNP057 absent 20

SNP057 present 24 Duel-1

20(19) SNP025 absent 36 Stander-10

SNP025 present 37 Stander-11

21(18) SNP024 absent 5 B1602-1

SNP024 present 9 Bonanza-2

22(1) NP001 absent 23

SNP001 present 30

23(22) SNP215 absent 24

SNP215 present 27

24(23) SNP057 absent 25

SNP057 present 26

25(24) SNP020 absent 3 Lacey-7

SNP020 present 2 Lacey-6

26(24) SNP024 absent 26 Foster-2

SNP024 present 7 Legacy-1

27(23) SNP020 absent 6 BT954-1

SNP020 present 28

28(27) SNP031 absent .... 12 CDC Tisdale-1

SNP031 present 29

29(28) SNP077 absent 28 Foster-6

SNP077 present 29 Foster-12

30(22) SNP020 absent 31

SNP020 present 35

31(30) SNP061 absent 32

SNP061 present 33

32(31) SNP042 absent .... 34 CDC Springside-11

SNP042 present .... 33 CDC Springside-10

33(31) SNP042 absent .... 32 CDC Springside-8

SNP042 present 34

34(33) SNP215 absent 40 Tankard-5

SNP215 present .... 31 CDC Springside-1

35(30) SNP018 absent 36

SNP018 present 37

36(35) SNP215 absent 38 Tankard-1

SNP215 present 39 Tankard-2

37(35) SNP215 absent .... 14 CDC Sisler-10

SNP215 present .... 13 CDC Sisler-1

CHAPTER SEVEN

7.1 General discussion

This thesis describes the analysis of genetic variation and genome evolution below the species level in barley, an organism with an active transposition system. The use of two independent marker systems and the inclusion of a relatively large number of barley cultivars with varying degrees of shared parentage and representing various agronomic groups has allowed a detailed comparison of the genetic relationships and genome evolution among groups as well as between cultivars. Multiple plants per cultivar were used in all analyses to assess the level of genetic homogeneity within cultivar. In addition the study describes the use of SNPs for cultivar identification.

We chose to study the dynamics of genome evolution in barley cultivars for the following reasons: 1) The genealogy of cultivars is known through their pedigrees making it easy to trace allele sharing between cultivars and parent-progeny relationships; 2) The genetic base of modern barley cultivars is thought to be narrow due to a limited number of ancestral lines and landraces in their pedigree and as such, differences in their retrotransposon content are likely the result of recent transposition/genomic rearrangements; 3) Various agro-ecosystems are likely to have different selection pressures causing the activation of retrotransposon. This can subsequently lead to measurable differences in the genetic composition of barley cultivars.

In this thesis four hypotheses were tested. The first or the main hypothesis was to test the effects of breeding and selection pressure on the evolution of barley genome. Two objectives were set and successfully met. The first objective was met by providing quantitative measures of genome diversity resulting from retrotransposition of BARE-1.

The second objective was met by surveying selected genomic loci for the presence of single nucleotide variations and using the data to make genetic diversity estimates between cultivars. Comparison between the rates of genetic variation obtained from single nucleotide variation and those obtained from retrotransposon-mediated mutations provided evidence for a rapid genome evolution mediated by retrotransposition.

The second hypothesis was to test the level of genetic homogeneity within cultivars. It is assumed that plant cultivars at the breeders' seed level are genetically homogeneous except for some minor contaminations. In this thesis, intra-cultivar genetic variation at the breeders' seed level were quantified and found to be widespread among cultivars. Furthermore, the possibility of seed contamination for the observed intra-cultivar genetic variation was ruled out based on the results of population genetic analyses tools as described before.

The third hypothesis was to test the extent of variation in BARE-1 copy number among barley cultivars. Differences in the BARE-1 copy number among cultivars that share common ancestors indicate recent transpositional events and can have profound effects on genome's ability to cope with different biotic and abiotic stresses emanating from different agro-ecosystems. In this study, major differences in the BARE-1 copy numbers among cultivars were detected by the quantitative real-time PCR analysis suggesting a role for the BARE-1 retrotransposon in the restructuring of barley genome during breeding and selection.

The fourth hypothesis was to test the feasibility of single nucleotide polymorphisms for identification of cultivars and their various biotypes. SNP markers were discovered,

validated and transformed into allele-specific amplicons for unequivocal identification of barley cultivars.

The Results section of the thesis has been organized into chapters reflecting published, in press or submitted manuscripts and as such each chapter has its own publication-specific conclusions and discussions. In this chapter our goal is broader, and more general, to bring together the information from each chapter and to draw general conclusions concerning cultivated barley and the effects of breeding on the evolution of its genome.

7.1.2 Survey of barley genome with S-SAP and SNP markers

The first objective of this thesis was successfully met by using S-SAP method to analyze the patterns and the extent of genetic variation within and between cultivars. Two enzymes, *Pst*I (a six base pair) and *Mse*I (a four base pair) cutter were used for this analysis. *Pst*I is a methylation sensitive enzyme that does not cleave within BARE-1 element (GenBank accession no. Z17327). Use of the above mentioned enzyme pairs ensure that the amplified restricted product contains DNA from both the host and the element and minimizes the possibility of targeting polymorphisms within the retrotransposon itself. However, there may be exceptions where the retrotransposon has accumulated mutations over time and has acquired a recognition site for the *Pst*I enzyme. Separating the latter from the former is not possible without sequencing every fragment which was not an objective of this thesis. The drawback in using a methylation sensitive enzyme such as *Pst*I is that some of the detected polymorphism might be epigenetic in nature and not directly related to variation in DNA sequences between cultivars. Kubis et

al (2003) compared the effects of methylation versus genome rearrangements mediated by transposable elements and concluded that the mantle phenotype of oil palm was likely the result of methylation rather than transposition. In the same study, only minor retrotransposon-mediated genome rearrangements were detected by Southern transfer and fluorescence in situ hybridization (FISH) approaches. However, the study also revealed that normal levels of methylation were restored after the regeneration of normal trees from callus culture. DNA methylation in plants as in mammals is implicated in tissue-specific gene regulation (Cocciolone et al. 2001; Kloti et al. 2002) and it is thought to have evolved as a mechanism to control and reduce transposable elements ability to insert into the host DNA (Finnegan et al. 1998). To partially address the issue of the methylation sensitive nature of the *Pst*I enzyme, a pilot project involving *Eco*RI/*Mes*I restriction enzyme pair was carried out using a small group of plants representing several cultivars. The level of polymorphism and the extent of genetic similarity between the two enzyme pairs were compared (not shown). Although the pattern of polymorphism for the two systems was different as expected, no significant differences were observed for the overall genetic similarity among plants for the two enzyme systems indicating that the contribution of methylation to genetic variation in this study was minimal. In addition, to minimize the possibility of contribution of tissue-specific methylation to polymorphisms, plant material for all S-SAP analyses was obtained from barley leaves grown under identical conditions.

BARE-1 transposition cannot be considered the only source of S-SAP polymorphism in barley. S-SAPs could potentially arise from sequence variations at the restriction sites. This portion of polymorphism is analogous to the polymorphism measure by the AFLP

technique. Retrotransposon-specific variation in S-SAP falls into two categories: 1) Sequence variation at the LTR sequences of BARE-1 element and 2) Insertional polymorphism resulting from integration of a new copy of retrotransposon into a genomic location. Waugh et al (1997) addressed this by studying the increase in the frequency of polymorphism measured by BARE-1 LTR-derived S-SAP in comparison to that seen by the AFLPs in a segregating population. Assuming that all AFLPs were derived from single nucleotide polymorphism (with a mutation rate of ca $10^{-7} - 10^{-9}$/nucleotide per generation), his study showed that BARE-1 transposition in barley occurred at 25% of the frequency of point mutation fixation. This conclusion is likely biased as it is difficult to prove that AFLPs and S-SAPs are independent and mutually exclusive markers. In theory, AFLP polymorphisms could easily arise from the insertion of transposable elements into genome. Conversely, some S-SAPs may arise from point mutations rather than transpositions.

Contrary to the anonymous S-SAPs, SNPs were derived from sequencing previously mapped genomic loci (Kleinhofs et al. 1993; Langridge et al. 1995) and data mining from public domain as described in Chapter Two and Four of this thesis. An objective of the study was to validate and transform these polymorphisms into allele-specific amplicons that were analogous to STSs for rapid screening of plants (Chapter three). Due to a low success rate with the validation of SNPs and their subsequent transformation into AS-PCR, only 27 SNP markers were successfully used for genetic diversity analysis and cultivar fingerprinting. In addition to their use for genetic diversity measurements, these markers were also used for cultivar identification (Chapter 6 and Appendix).

Comparative analysis of genetic diversity in barley was important because each marker system provides a unique window on the extent and pattern of genetic variation. SNPs are derived from the expressed and non-repetitive regions of the genome and as such they are expected to be under evolutionary constraints. These constraints are especially important when the mutation results in the alteration of amino acid sequences in the resulting protein. S-SAP markers on the other hand, were derived from the repetitive regions, making them less responsive to selection and evolutionary constraints.

7.1.3 Comparative analysis of genetic diversity using S-SAP, SNP and KIN

While molecular markers such as S-SAPs and SNPs provide a direct measure of the extent of allele sharing between genotypes, kinship coefficient is only a theoretical measure of the extent of genetic similarity. Kinship values are based on the assumption of equal contribution of each parent to the progeny. In addition, KIN assumes that there is no linkage between genes and those parents with unknown pedigrees are unrelated. These assumptions cause a large bias towards low GS values (Cox and Murphy 1985; Tinker et al. 1993). Comparison between the distribution of GS values obtained from S-SAP, SNP and KIN is shown in figure 7.1. While both S-SAP and SNP values were normally distributed, KINs were skewed towards lower values. The distribution of GS values indicates that both S-SAP and SNP markers have better utility in breeding programs than KINs as there is more flexibility in choosing parental genotypes with maximum allelic variation with the former two markers.

S-SAPs were found to be more informative than SNPs in revealing intra-cultivar variation as well as revealing genetic variation between closely related cultivars (Chapter

4). This is in agreement with results from previous studies. For example, In a comparative analysis of S-SAP, AFLP and SSR on genetic diversity in tomato and pepper, Tam et al (2005) found that S-SAP showed up to nine-fold more polymorphism than AFLP, and Ellis et al (1998) observed that S-SAP exhibited a three-fold improvement over AFLP in discriminating accessions. The results of matrix correlation analysis using the Mantel test showed that GS_{SNP} values were more correlated to the pedigree data than $GS_{S\text{-}SAP}$, while there was no correlation between GS_{SNP} and $GS_{S\text{-}SAP}$ or between $GS_{S\text{-}SAP}$ and GS_{KIN} matrices. Therefore, Kinship data was a better predictor of GS_{SNP}. This was not surprising as breeders rely on selectable morphological markers for their crosses, and the SNPs that were used in this analysis were those in the expressed regions of barley genome and were therefore expected to affect phenotype.

7.1.4 Analysis of genetic variation in the population

AMOVA analysis was carried out to partition variance and to test the genetic differentiation among groups, among cultivars and within cultivars. Groups were set dichotomously based on end use (feed/malt) and seed morphology (covered or naked kernel). AMOVA results of S-SAP and SNP datasets were very similar. A hierarchical analysis of variance partitioned the total variance into covariance components based on intra-cultivar differences, inter-cultivar differences and/or inter-population (group) differences. The covariance components were used to compute fixation indices as defined by Wright (1965) in terms of inbreeding coefficients, and later in terms of coalescent time by Slatkin (1991). There was no statistical support for any genetic differentiation on the basis of feed versus malt or covered versus naked kernel divisions. Lack of a genetic

differentiation along these agronomic divisions was not surprising because they are controlled by a small number of genes. For example, seed morphology, i.e., covered versus naked kernel is believed to be regulated by a single recessive gene (Bothmer et al. 2003). The largest covariance component (about 90%) was attributed to the variation among cultivars within each group. The remaining 10% of variation was attributed to variation between groups (about 7%) and variation within each cultivar (about 3%).

Although the extent of genetic variation that was revealed by S-SAP and SNP was similar in the population as indicated by AMOVA analysis, the pattern of the observed variation was significantly different as shown by cluster analysis and the Mantel test (Chapter 5). Mean pair-wise genetic similarity values were highest for the S-SAP dataset (0.602) and lowest for kinship coefficient dataset (0.0957) that was obtained from pedigree data. Pedigree-based GS values did not accurately reflect the extent of genetic relationships between cultivars due to the assumptions that were made in their calculations.

7.1.5 Incongruence between S-SAP and SNP diversity estimates

Lack of a significant correlation between S-SAP and SNP-based diversity estimates is likely due to the differences in their respective genome coverage and underlying evolutionary mechanisms. SNPs are similar to AFLPs. They are single nucleotide variation in DNA sequences including small indels. The main mechanism giving rise to SNPs is the introduction of errors during DNA replication in germ cells (Cargill et al. 1999; Sachidanandam et al. 2001). Under the neutral theory of molecular evolution, the majority of DNA variation observed in a population is due to random drift of neutral or

nearly neutral mutations (Kimura 1985). Under this model, purifying selection eliminates mutations with deleterious effects and results in the reduction of the deleterious mutations compared to neutral mutations. In addition, it may also lead to a high proportion of mutations with rare alleles in a rapidly expanding population (Zhao et al. 2000). A survey of human genome showed that out of the predicted 3 million SNPs only 240,000 to 400,000 qualified as common mutations, the majority of the rest of SNPs were found to be rare alleles (Cargill et al. 1999). A common allele was defined as one having an allele frequency of greater than 1% in the population. In our study on SNPs (Chapters 2, 4 and 6), out of 214 markers only 27 were found to qualify as common alleles. On the other hand, S-SAPs are mainly the result of retrotransposition and related genomic rearrangements, independent of the rates of point mutations in genome. A high level of genetic polymorphism that was detected with the S-SAP is in agreement with previous studies using retrotransposon-based marker systems (Waugh et al.1997; Ellis et al. 1998; Tam et al. 2005). This finding is also supported by recent studies of comparative sequencing of closely related organisms including members of the same species that showed drastic differences between orthologous loci both in terms of sequence composition and structural organization (Fu and Dooner 2002; Brunner et al. 2005). Comparative sequence analysis of two inbred maize line stretching over 2.8 Mb showed that more than 50% of the compared sequence was non-collinear (Brunner et al. 2005). Furthermore, the same study showed that the main reason for the breakdown of colinearity between the two maize lines was due to line-specific retrotransposon insertion in which two third of all the retrotransposon insertions were line-specific.

7.1.6 Retrotransposition and the evolution of barley genome

The effects of BARE-1 retrotransposons on the evolution of barley genome as a result of breeding and selection were assessed by measuring the copy number of the element among different barley cultivars. All the cultivars that were chosen for this analysis have all been released within the past 30 years and share many common ancestors. Significant differences in BARE-1 copy number were detected between cultivars as described in chapter 5. Also, consistent with previous studies, notably Vicient et al (1999), most of the retrotransposon copies were present in truncated form as solo LTRs. These findings have important implications not only for barley but also for other organisms with an active transposition system. A direct way to identify recent BARE-1 transposition events is to analyze parental and progeny populations for the occurrence of novel bands in the progeny where those same bands are absent in the parent population. In this study, at least 5% of bands in the progeny population could not be assigned to either parent as described in chapter two.

The concept of dynamic genomes, once obscure, has gained much support within the past two decades as a result of comparative genome analysis and completion of various large scale sequencing projects. Comparative genetic mapping has also shown a significant degree of synteny – conservation of gene order (macrocolinearity) among related species (Devos and Gale 1998; Paterson et al. 2000). However, sequence analysis of orthologous regions in rice, sorghum and maize has shown that there are major breakdowns in microcolinearity (Tikhanov et al. 1999; Fu and Dooner 2002; Lai et al. 2004; Brunner et al. 2005). Sequence analysis of 100 kb from the *bz* genomic region showed dramatic differences between two maize lines (Fu and Dooner 2002), mainly due

to cluster of nested retrotransposons. More importantly, the genes within the *bz* locus themselves differed between the lines. These studies together with the results of this thesis provide strong evidence for the fluidity of genomes in cereal species. Quantification results for BARE-1 retrotransposon in barley (Chapter 5) showed that retrotransposon copy number between cultivars can vary by up to 3000 copies, equivalent to 26 Mb difference in nuclear genome content. These findings have important implications in plant breeding and genetics. Differences in the quantity of nuclear DNA content may affect vigor and stability of cultivars over a range of environmental conditions. This is particularly important for barley which is grown over a larger environmental range than any other cereal species. BARE-1, by virtue of its recent activity has played an important role in the enhancing of the evolutionary dynamism of barley genome.

7.1.7 Intra cultivar Genetic variation

Intra cultivar variation was assessed by S-SAP and SNP markers. S-SAP showed considerably higher level of intra cultivar genetic heterogeneity than SNP. To test the possibility that some of the observed intra cultivar variation may be the result of admixture, we treated each individual plant as an independent genotype and calculated all pair wise genetic similarities between individual plants in the entire plant population in this study. Cluster analysis was carried out on the resulting matrix of genetic similarities. Each individual plant fell within its own cultivar with no cross cultivar placement of any plant, eliminating the possibility of contamination for the observed levels of intra cultivar variation.

In many countries including Canada, registration of new plant cultivars requires DUS test as described in the introduction. Cultivar uniformity and stability are key components of DUS test and cultivars are often grown in different locations to evaluate these criteria. Although DUS tests have traditionally relied mainly on morphological markers, DNA-based markers such as S-SAPs and SNPs offer many advantages over traditional morphological markers for evaluation of crop cultivars including DUS test. DNA markers are more numerous than morphological counterparts; they can be applied at any stage of the plant development and unlike morphological markers they are not subjected to epistatic and pleiotropic effects.

The final objective of this thesis was to develop an identification scheme for barley cultivars. Both S-SAP and SNP method produced cultivar-specific polymorphisms. However, S-SAP is a multi locus and labor intensive assay. On the other hand SNP is a single locus assay and it is prone to automation, making it more appropriate for cultivar identification. Unambiguous identification of barley cultivars is a challenge to the grain handling system and cultivar purity is a core requirement in the brewing industry. Although morphological and biochemical markers are useful for identification of some cultivars, unequivocal identification of closely related cultivars is not possible with the traditional markers. In this study, SNP markers were used for efficient identification of all six-rowed malting cultivars (Chapter 6) and combined six-rowed malting and feed cultivars (Appendix). The cultivar diagnostic SNPs are part of a "Trade Secret" which is currently being evaluated by the Agriculture and Agri-Food Canada for commercialization.

7.2 Conclusion

Breeding and selection during cultivar development are the major driving forces behind an accelerated genome evolution in cultivated barley. Major differences in BARE-1 content between different cultivars were detected by quantitative PCR analysis, suggesting that different cultivars or lines of the same species can have drastic differences in their genomic DNA content and sequence composition. Retrotransposons, equipped with strong promoters and other regulatory sequences may contribute significantly to differential patterns of gene expression resulting in the phenotypic plasticity that is often observed in breeding programs.

Due to their large copy number and wide distribution throughout genome, retrotransposon-based marker systems are invaluable tools for inferring genetic diversity estimates between cultivars and breeding lines. This is especially important in the study of crop cultivars such as barley that are derived from a limited gene pool. S-SAPs were found to be more useful for the study of closely related cultivars. Significant levels of genetic variation were detected between sister lines, i.e., cultivars derived from the same parents and clones by using this marker system. Although SNPs were also found to polymorphic in the population, they were less efficient than S-SAP for the study of closely related cultivars and revealing intra cultivar variation. Comparative analysis using S-SAP, SNP and kinship coefficient showed that both S-SAP and SNP were better than kinship in resolving the extent of allele sharing between cultivars. GS values of both S-SAP and SNP were normally distributed while kinship values were skewed indicating that both S-SAPs and SNPs are better tools for selecting parental combinations with maximum allelic diversity over a range of genotypes in breeding programs. Although

both S-SAP and SNP markers exhibited cultivar-specific polymorphic profile, only SNPs were used for generation of identification keys for two reasons: 1) SNPs were found to be more stable across genotypes within each cultivars compared to S-SAPs. The latter showed a high degree of intra cultivar variability. 2) SNPs were derived mostly from single copy loci in the genome while retrotransposon sequences are present in multiple copies. The repetitive nature of S-SAP is a disadvantage in the transformation of these markers into hybridization-based platforms such as DNA chip for rapid cultivar identification. Taken together, the results of this thesis and recent studies on the comparative sequence analysis of cereals provide valuable information and new insight into the dynamic nature of cereals genomes. Simple DNA base substitutions maybe adequate for diversification and optimization of local protein space (Kidwell and Lisch 1997), but a cascade of mutational events is required for rapid generation of protein diversity in evolution (Bogarad and Deem 1999). It is commonly known that transposable elements and viruses are more capable in generating complex types of mutations.

7.3 Future work

This thesis has answered some of the questions pertaining to the levels, sources and causes of genetic variation among barley cultivars and the effects that breeding and selection may play in the restructuring of the barley genome as mediated by retrotransposition. The thesis has also provided a methodology for efficient identification of barley cultivars. However, many fundamental and general questions regarding genome plasticity and integrity in barley in particular and plants in general remain unanswered.

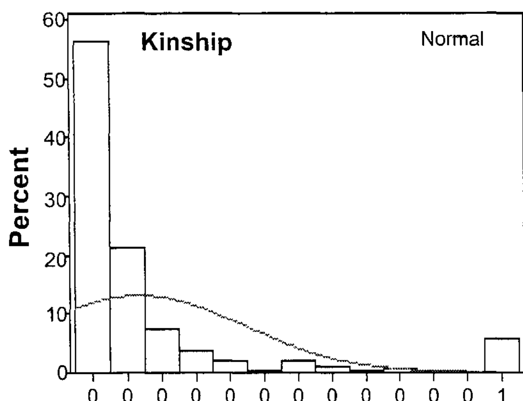
Questions regarding the function of transposable elements and their contribution to genome evolution remain to be answered. Are transposable elements "selfish DNA" that are concerned only with their proliferation within the genome without notable contribution to phenotype (Orgel and Crick 1980), "junk DNA", considered as intracellular obligatory parasite (Doolittle and Sapienza 1980) or do they provide essential elements for genome evolution (Jurka 1998) and have an important role in eukaryotic genome function (Fedoroff 2000; Jiang et al. 2004)? Most plant retrotransposon are inactive, but many have gone through bursts of retrotransposition in the past (SanMiguel et al. 1996; Kumar and Bennetzen 1999). What factors were involved in the initial activation and subsequent suppression of these elements? What roles have transposable elements played in plant speciation? Did speciation events follow a punctuated equilibrium mediated by retrotransposition bursts? Have transposable elements contributed to the adaptation of plants to diverse ecological niches? Is polyploidy, widespread among plants, a means to neutralize the effects of deleterious transposon insertion? What contributions have transposable elements made to the complex mechanisms of gene regulation in eukaryotes?

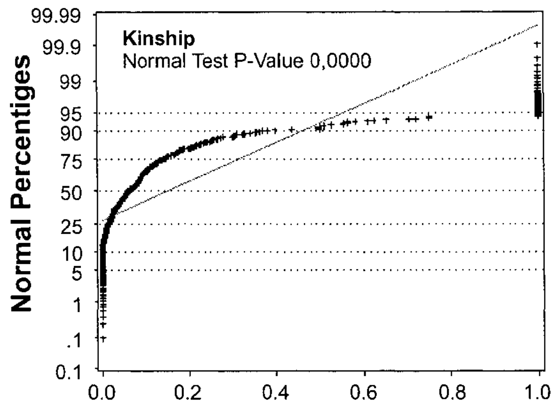
While these questions and many more await answers, evidence on the important role of transposable elements in shaping the architecture of genome in eukaryotes and their co-evolution with the host is slowly emerging. Mechanism of exon shuffling by LINE elements in humans (Moran et al. 1999) and fusion of diverse host genomic loci that have open reading frames by mutator-like transposable elements in plants (Jiang et al. 2004) have provided direct evidence that transposable elements can have an important role in the evolution of new gene functions in their hosts.

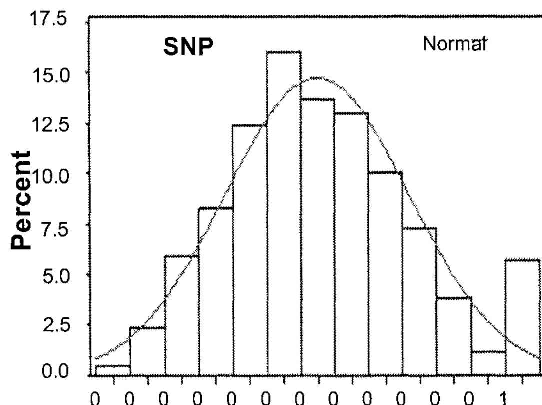
The application of retrotransposon-based marker techniques such as S-SAP and IRAP to closely related genotypes, i.e., sister lines in crop cultivars and cultured plant cells, can reveal new insertional events and provide a quantitative measure of the transpositional activity of the element. Future large-scale sequencing of complex genomes such as barley could also provide valuable information about the genome and the effects of retrotransposon on its architecture. Genome-wide analysis of retrotransposons expression could provide important information about the transcriptional activity of the elements and their regulation. S-SAP and SNP markers that were generated in this thesis can be mapped in a segregating population and be used to assess possible linkage to important agronomic traits in barley. In addition, cultivar-specific S-SAP markers can be transformed into retrotransposon-based insertion polymorphism (RBIP) (Flavell et al.1998) for high throughput cultivar identification. Identification of cultivars with SNPs in this work was gel-based, however these SNP markers can be used to produce a DNA chip for rapid non-gel based identification. Additional cultivar-specific SNPs need to be discovered to cope with the introduction of future cultivars and the identification platform will require regular updates.

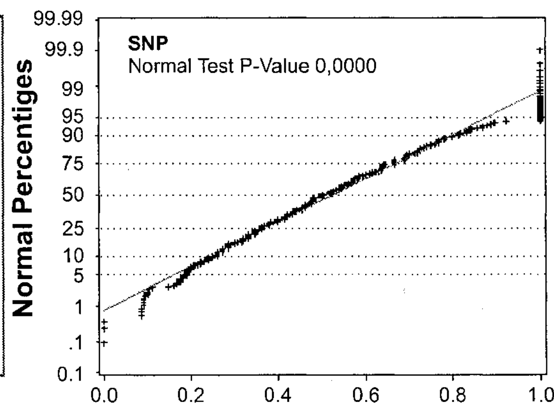
Figure 7.1 Distribution of GS values around the mean. GS values of S-SAP and SNP were normally distributed while those of KIN were skewed indicating that both S-SAP and SNP were better indicators genetic relationships among cultivars.

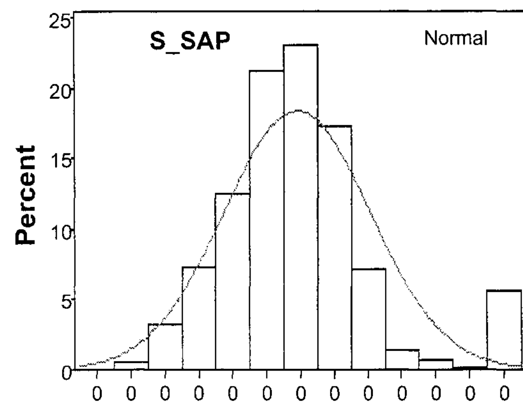
Histogram | **Normal probability**

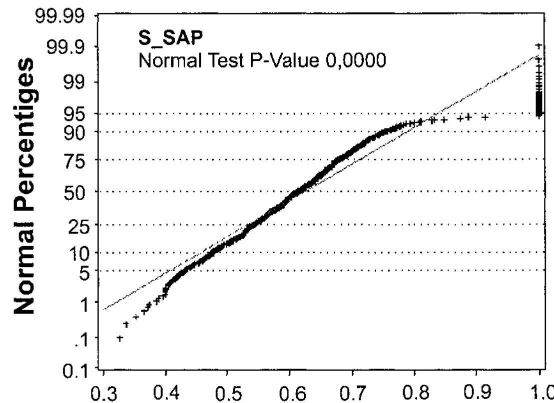












LITERATURE CITED

Akkaya, M.S., Bhagwat, A.A., and Cregan, P.B. 1992. Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

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. J. Mol. Biol. 215: 403-410.

Anderson, T.W. 1984. An introduction to multivariate statistical analysis, 2nd ed. JohnWiley & Sons, New York. USA.

Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815

Archak, S., Karihaloo, J.L., and Jian, A. 2002. RAPD markers reveal narrowing genetic base of Indian tomato cultivars. Current Sci. 82:1139-1143.

Autrique, E., Miloudi, M.N., Monneveux, P., Tanksley, S.D., and Sorrell, M.E. 1996. Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage Crop Sci. 36:735-742.

Avery, O.T., MacLeod, C.M., and McCarty, M. 1944. Studies on the chemical nature of the inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. 79:137-157.

Avramova, Z., Tikhonov, A., SanMiguel, P., Jin, Y.K., Lui, C., Woo, S.S., and Bennetzen, J.L. 1996. Gene identification in a complex chromosomal continuum by local genomic cross referencing. Plant J. 10:1163-1168.

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.

Baum, B.R., Bailey, L.G., and Thompson, B.K. 1985. Barley register, Agriculture Canada Publ. 1783B, Ottawa, Canada.

Baum, B.R., Mechanda, S.M., and Soleimani, V.D. 2000. Identification of Canadian six-row barley (*Hordeum vulgare* L) cultivars with primers derived from STSs obtained from RAPD diagnostic bands. Seed Sci & Technol 82:445-466.

Bennetzen, J.L. 2000. Transposable element contribution to plant genes and genome evolution. Plant Mol. Biol. 42:251-269.

Bennetzen, J.L., and Kellogg, E.A. 1997. Do plants have a one-way ticket to genome obesity? Plant Cell 9:1509-1514.

Bennetzen, J.L., Schrick, K., Springer, P.S., Brown, W.E., and SanMiguel, P. 1994. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome 37:565-576.

Bi, Y-A., and Laten, H.M. 1996. Sequence analysis of a cDNA containing the *gag* and *prot* regions of the soybean retrovirus-like element, SIRE-1. Plant Mol. Biol. 30:1315-1319.

Bogarad, L.D., and Deem, M.W. 1999. A hierarchical approach to protein molecular evolution. Proc. Natl. Acad. Sci. USA 96:2591-2595.

Bonierbale, M.W., Plaisted, R.L., and Tanksley, S.D. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095-1103.

Bothmer, R. von, van Hintum, T.J.L., Knupffer, H., Sato, K. (eds.) 2003. Diversity in barley (*Hordeum vulgare*). Elsevier Science.

Botstein, D., White, R.L. Skolinck, M., and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum. Genet. 32:314-331.

Brantestam, A.K., von Bothmer, R., Rashal, I., and Weibull, J. 2004. Changes in the genetic diversity of barley of Nordic and Baltic origin, studied by isozyme electrophoresis. Plant Genet. Resources 1:143-149.

Britten, R.J., and Kohne, D.E. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 161:529-540.

Brown, W.M., Prager, E.M., Wang, A., and Wilson, A.C. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol. Evol. 18: 225-239.

Brunner, S., Fengler, K., Morgante, M., Tingey, S. and Rafalski, A. 2005. Evolution of DNA sequence non-homologies among maize inbreds. Plant Cell 17:343-360.

Buetow, K.H., Edmonson, M.N., and Cassidy, A.B. 1999. Reliable identification of large number of candidate SNPs from public EST data. Nat. Genet. 21: 323-325.

Bureau, T.E., White, S.E., and Wessler, S.R. 1994. Transduction of a cellular gene by a plant retroelement. Cell 77:479-480.

Burr, B., Evola, S.V., Burr, F.A., Beckman, J.S. 1983. The application of restriction fragment length polymorphism to plant breeding. *In* Setlow J & Hollaender A (eds.): Genetic engineering 5, pp 45-59, New York.

Caldwell, K.S., Langridge, P., and Powell, W. 2004. Comparative sequence analysis of the region harbouring the hardness locus in barley and its collinear region in rice. Plant Physiol. 136:3177-3190.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., and Lander, E.S. 1999. Characterization of single nucleotide polymorphisms in the coding regions of human genes Nature Genet. 22:231-238.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S., and Wessler, S.R. 2000. The MITE family Heartbreaker (Hbr): Molecular markers in maize. Proc Natl Acad Sci USA 97 10083–10089.

Casacuberta, J.M., and Puigdomenech, P. 2000. Genome fluidity, The case of plants Contributions to Science 1 (3): 313-322.

Chao, L., Vargas, C., Spaear, B.B., and Cox, E.C. 1983. Transposable elements as mutator genes in evolution. Nature 303:633-635.

Chen, X., Levine, L., and Kwok, P.Y. 1999. Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res 9: 492-498.

Ching, A.D.A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O.S., Tingey, S., Morgante, M., and Rafalski, A. 2002. SNP frequency, haplotypes structure and linkage disequilibrium in elite maize inbred lines. MBC Genetics 3:19 (http://www.biomedcentral.com/1471-2156/3/19).

Cocciolone, S.M., Chopra, S., Flint-Garcia, S.A., McMullen, M.D., and Peterson, T. 2001. Tissue-specific pattern of maize Myb transcription factor are epigenetically regulated. Plant J. 27:467-478.

Cochrane, W.G., and Hopkins, C.E. 1961. Some classification problems with multivariate qualitative data. Biometrics 17:10-32.

Cooke, R.J. 1988. Electrophoresis in plant testing and breeding. Adv. Electrophor. 2:711-261.

Cooke, R.J. ed. 1992. Electrophoresis handbook: Variety identification. Pp 0-50: International Seed Testing Association, Zurich, Switzerland.

Cox, T.S., and Murphy, J.P. 1990. The effects of parental divergence on F_2 heterosis in winter wheat crosses Theor. Appl. Genet. 79:241-250.

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.

Dallwitz, M.J., Paine, T.A., and Zurcher, E.J. 1993. User's guide to DELTA system: a general system for processing taxonomic descriptions. 4th Ed. http://bioidversity.uno.edu/delta.

Davila, J.A., Loarce, Y., Ramsay, L., Waugh, R., and Ferrer, E. 1999. Comparison of RAMP and SSR markers for the study of wild barley genetic diversity Hereditas 131: 5-13.

Davila, J.A., Sanchez, 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.

Devos, K.M., and Gale, M.D. 1992. the use of random amplified polymorphic DNA markers in wheat. Theor. Appl. Genet. 84:567-572.

Devos, K.M., Brown, J.K.M., and Bennetzen, J.L. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res. 12:1075-1079.

Dice, L.R. 1945. Measure of the amount of ecological association between species. Ecology 26:297-302.

Doolittle, W.F., and Sapienza, C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601-603.

Douabin-Gicquel, V., Soriano, N., Ferran, H., Wojcik, F., Palierne, E., Tamin, S., Jovelin, T., McKei, A.T., Le, Gall, J.Y., David, V. and Mosser, J. 2001. Identificaiton of 96 single nucleotide polymorphisms in eight genes involved in iron metabolism: efficiency of bioinformatic extraction compared with a systematic sequencing approach. Human Genet109: 393-401.

Echenique, V., Stamova, B., Wolters, P., Lazo, G., Carolla, L., and Dubcovsky, J. 2002. Frequencies of the Ty1-*copia* and Ty3-*gypsy* retroelements within the Triticeae EST databases. Theor. Appl. Genet. 104:840-844.

Edwards, K., Johnston, C., and Thomson, C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis Nucl. Acids Res. 19: 1349.

Ellis, T.H.N., Poyser, S.J., Knox, M.R., Vershinin, A.V., and Ambrose, M.J. 1998. Ty1-*copia* class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. Mol. Gen. Genet. 260:9-19.

Ergun, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., and Stratling, W.H. 2004. Cell type-specific expression of LINE-1 open reading frame 1 and 2 in fetal and adult human tissue. J. Biol. Chem. 279:27753-27763.

Fedoroff, N.V. 2000. Transposons and genome evolution in plants. Proc. Natl. Acad. Sci. USA 97:7002-7007.

Fenwick, A., Blake, T.K., Kanazin, V., Ananiev, E., Dahleen, L., Kudrna, D., Bollinger, J., Knapp, S.J., Liu, B., Sorrels, M., Heun, M., Franckowiak, J.D., Hoffman, D., Skadsen, R., and Steffenson, B.J. 1993. A molecular, isozyme and morphological map of barley (*Hordeum vulgare*) genome. Theor. Appl. Genet. 86: 705-712.

Finnegan, E.J., Genger, R.K., Peacock, W.J., and Dennis, E.S. 1998. DNA methylation in plants. Ann. Rev. Plant Physiol. & Plant Mol. Biol. 49:223-247.

Fischbeck, G. 2003. Diversification through breeding. In: von Bothmer, T.J.L. van Hintum, H. Knupffer and K. Sato (Eds.), diversity in barley (*Hordeum vulgare*). Elsevier Science, pp. 29-52.

Flavell, A.J., Knox, M.R., Pearce, S.R., and Ellis, T.H. 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J. 5:643-650.

Flavell, A.J., Dunbar, E., Anderson, R., Pearce, S.R., Hartley, R., and Kumar, A. 1992. Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants Nucleic Acids Res. 20:3639-3644.

Fryxell, K.J., and Zuckerkandl, E. 2000. Cytosine deamination plays a primary role in the evolution of mammalian isochors. Mol. Biol. Evol. 17:1371-1383.

Fu, H., and Dooner, H. 2002. Intraspecific violation of genetic colinearity and its implications in maize. Proc. Natl. Acad. Sci. USA 99:9573-9578.

Gebre, H., Khan, K., and Foster, A.E. 1986. Barley cultivar identification by polyacrylamide gel electrophoresis of hordein proteins. Crop Sci. 26: 454-460.

Gibson, U.E., Heid, C.A., and Williams, P.M. 1996. A novel method for real time quantitative RT-PCR. Genome Res. 6: 995-1001.

Giesendorf, B.A., Tyagi, V.S., Mensink, E.J., Trijbels, F.J., and Bloom, H.J. 1998. Molecular beacons: a new approach for semi-automated mutation analysis. Clin Chem 44: 482-486.

Gilbert, E.S. 1968. On discrimination using qualitative variables. J Am Stat Assoc 63:1399-1412.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchinson, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszhkowski, U., Zhang, S., Colbert, M., Sun, W-L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. 2000. A draft sequence of the rice genome (*Oryza sativa* L. ssp. Japonica). Science 296:92-100.

Gojobori, T., Li, W-H., and Graur, D. 1982. Patterns of nucleotide substitutions in pseudogenes and functional genes. J. Mol. Evol. 18: 360-369.

Goodwin, T.J.D., and Poulter, T.M.R. 2000. Multiple LTR retrotransposon families in the asexual yeast *Candida albicans* Genome Res. 10:174-191.

Grandbastien, M.A., Lucas, H., Morel, J.B., Mhiri, C., Vernhettes, S., and Casacuberta, J.M. 1997. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense response. Genetica 100: 241-252

Graner, A., Bjornstad, A., Konishi, T., and Ordon, F. 2003. Molecular diversity of the barley genome. In: R. von Bothmer, T.J.L. van Hintum, H. Knupffer and K. Sato (Eds.), Diversity in barley (*Hordeum vulgare*) Elsevier Science, pp. 121-141.

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.

Gribbon, B.M., Pearce, S.R., Kalendar, R., Schulman, A.H., Pauline, L., Jack, P., Kumar, A., and Flavell, A.J. 1999. Phylogeny and transcriptional activity of Ty1-copia group retrotransposons in cereal genomes. Mol. Gen. Genet. 261:883-891.

Grossman, P.D., Bloch, W., Brinson, E., Chang, C.C., Eggerding, F.A., Fung, S., Iovannisci, D.M., Woo, S., Winn-Deen, E.S. and Iovannisci, D.A. 1994. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. Nucleic Acids Res 22: 4527-4534.

Gu, Y.Q., Anderson, O.D., Londeore, C.F., Kong, X., Chbbar, R.N., and Lazo, G.R. 2003. Structural organization of the barley D-hordein locus in comparison with its orthologous region in wheat genomes. Genome 46: 1084-1097.

Halushka, M.K., Fan, J.B., Bently, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. and Chakravarti, A. 1999. Patterns of single nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22: 239-247.

Hamajima, N. 2001. PCR-CTPP: a new genotyping technique in the era of genetic epidemiology. Expert Review in Molecular Diagnostics. 1: 119-123.

Heid, C.A., Stevens, A.J., Livak, K.J., and Williams, P.M. 1996. Real-time quantitative PCR. Genome Res. 6: 986-994.

Hershey, A.D., and Chase, M. 1952. Independent function of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. 36:39-56.

Higuchi, R., Dollinger, G., Walsh, P.S., and Grifith, R. 1992. Simultaneous amplification and detection of specific DNA sequences. Bio/Technology 10: 413.

Hirochika, H. 1993. Activation of tobacco retrotransposons during tissue culture. EMBO J 12:2521-2528.

Hirochika, H., Fukuchi, A., and Kikuchi, F. 1992. Retrotransposon families in rice. Mol. Gen. Genet. 233:209-216.

Hirohcika, H., Okamoto, H., and Kakutani, T. 2000. Silencing of retrotransposons in Arabidopsis and reactivation by the ddm1 mutation. Plant Cell 12:357-368.

Hirochika, H., Sugimoto, K., Otsuki, Y., and Kanda, M. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93: 7783-7788.

Hoffman, D., Hang, A., Larson, S., and Jones, B. 2002. Conversion of a RAPD marker to an STS markers for barley variety identification. Plant Mol. Biol. Reporter 21: 81-91.

Hoffman, D.L., and Bregitzer, P. 1990. Polymorphism and genetic basis of β-N-acetylglucosaminidase, hexokinase and basic β-D-galactosidase in barley (*Hordeum* spp.). Biochem. Genet. 28: 487-493.

Holliday, R., and Grigg, W.G. 1993. DNA methylation mutation. Mutation Res 285: 61-67.

Horsely, R.D., Schwarz, P.B., and Hammond, J.J. 1995. Genetic diversity in malt quality of North American six-rowed spring barley Crop Sci. 35: 113-118.

Hu, W., Das, O.P, and Messing, J. 1995. *Zeon*-1, a member of a new retrotransposon family. Mol. Gen. Genet. 248:471-480.

Huang, X., and Madan, A. 1999. CAP3: A DNA Sequence Assembly Program. Genome Res. 9: 868-877.

Hug, A.M. and H. Feldmann. 1996. Yeast retrotransposon Ty4: The majority of the rare transcripts lack a U3-R sequence. *Nucleic Acids Res.* 24: 2338-2346.

Iwamoto, M., Nagashima, H., Nagamine, T., Higo, H., and Higo, K. 1999. *pSINE1*-loke intron of the *CatA* catalse homologs and phylogenetic relationships among AA-genome *Oryza* and related species. Theor. Appl. Genet. 98:853-861.

Jaaskelainen, M., Mykkanen, A.H., Arna, T., Vicient, C., Souniemi, A., Kalendar, A., Savilahti, H., and Schulman, A.H. 1999. Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells. Plant J 20: 413-422.

Jacob, F., and Monad, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3:318-356.

Jiang, N., Bao, Z., Zhang, X., Eddy, S., and Wessler, S.R. 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature Genet. 431:569-573.

Jordan, K.I., Matyunina, L.V., and McDonald, J.F. 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposons. Proc. Natl. Acad. Sci. USA 96:12621-12625.

Jurka, J. 1998. Repeats in genomic DNA: Mining and meaning. Curr. Opin. Struc. Biol. 8:333-337.

Kaeppler, S.M., Kaeppler, H.F., and Rhee, Y. 2000. Epigenetic aspects of somaclonal variation. Plant Mol. Biol. 43: 179-188.

Kalendar, R. 2001. Oligos. A computer program for the design of PCR primers. Institute of Biotechnology University of Helsinki, Finland. Available at http://www.biocenter.helsinki.fi/bi/bare-1_html/download.htm.

Kalendar, R., Grob, T., Regina, M., Suoniemi, A., and Schulman, A. 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques Theor. Appl. Genet. 98:704-711.

Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., and Schulman, A. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97:6603-6607.

Kalendar, R., Vicient, C.M., Peleg, O., Anamthawat-Jonsson, K., Bolshoy, A., Schulman, A.H. 2004. Large retrotransposon derivatives: abundance, conserved but non-autonomous retroelements of barley and related genomes. Genetics 166: 1437-1450.

Kanazin, V., Talbert, H., See, D., DeCamp, P., Nevo, E., and Blake, T. 2002. Discovery and assay of single nucleotide polymorphisms in barley (*Hordeum vulgare*) Plant Mol. Biol. 48:529-537.

Kankaanpaa, J., Mannonen, L., and Schulman, A.H. 1997. The genome size of *Hordeum* species show considerable variation. Genome 39:730-735.

Karhukorpi J and Karttunen R (2001) Genotyping interleukin-10 high and low producers with single-tube bi-directional allele-specific amplification. Exp Clin Immunogenet 18: 67-70.

Kashkush, K., Feldman, M., and Levy, A.A. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nature Genetics 33:102-106.

Kidwell, M.G. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115:49-63.

Kidwell, M.G., and Lisch, D. 1997. Transposable elements are sources of variation in animals and plants. Proc Natl Acad Sci USA 94:7704-7711.

Kim, H.S., and Ward, R.W. 1997. Genetic diversity in Eastern US soft winter wheat (*Triticum aestivum* L. em. Thell.) based on RFLPs and coefficient of parentage. Theor. Appl. Genet. 94:472-479.

Kimura, M., 1985. The neutral theory of molecular evolution, Cambridge University Press, Cambridge, UK.

Kleinhofs, A., Kilian, A., Saghai-Maroof, M.A., Biyashev, R.M., Hayes, P., Chen, F.Q., Lapitan, N., Konarev, V.G., and Gavrilyuk, I.P. 1988. Biochemical identification of

varieties. International Seed Testing Association and N.I. Vavilov Institute of plant industry, St. Petersburg, Russia.

Kleinhofs, A., Kilian, A., Saghai-Maroof, M.A., Biyashev, R.M., Hayes, P., Chen, F.Q., Lapitan, N., Fenwick, A., Blake, T.K., Kanazin, V., Ananiev, E., Dahleen, L., Kudrna, D., Bollinger, J., Knapp, S.J., Liu, B., Sorrels, M., Heun, M., Franckowiak, J.D., Hoffman, D., Skadsen, R., and Steffenson, B.J. 1993. A molecular, isozyme and morphological map of barley (*Hordeum vulgare*) genome. Theor. Appl. Genet. 86: 705-712.

Kloti, A., He, X., Potrykus, I., and Futterer, J. 2002. Tissue-specific silencing of a transgene in rice. Proc. Natl. Acad. Sci. USA. 99:10881-10886.

Kobayashi, S., Goto-Yamamoto, N., and Hirochika, H. 2004. Retrotransposon-induced mutations in grape skin colour. Science 304, p. 982.

Komatsu, M., Shimamoto, K., and Kyozuka, J. 2003. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. The Plant Cell. 15:1934-1944.

Krzanowski, W.J. 1975. Discrimination and classification using both binary and continuous variables. J Am Stat Assoc 70:782-790.

Kshirsagar, A.M. 1972. *In*: Multivariate analysis. Marcel Dekker, New York.

Kubis, S.E., Castilho, M.M.F., Vershinin, A.V., and Heslop-Harrison, J.S. 2003. Retroelements, transposons and methylation status in the genome of oil palm (*Elaeis guineensis*) and the relationship to somaclonal variation. Plant Mol. Biol. 52:69-79.

Kumar, A. 1996. The adventure of Ty1-*copia* group of retrotransposons in plants. Trends Genet. 12:41-43.

Kumar, A., and Bennetzen, J.L. 1999. Plant retrotransposons. Ann Rev Genet 33:479-532.

Kumar, A., and Hirochika, H. 2001. Application of retrotransposons as genetic tools in plant biology. Trend in Plant Sci. 6:127-134.

Kumar, A., Pearce, S.R., McLean, K., Harrison, G., and Heslop-Harrison, J.S. 1997. The Ty1-*copia* group of retrotransposons in plants: genome organization, evolution and use as molecular markers. Genetica 100:205-217.

Kurata, N., and Nomomura, K-I 2002. Rice genome organization: the centromere and genome interactions. Ann. Bot. 90:427-435.

Labra, M., Imazio, S., Grassi, F., Rossoni, M., and Sala, F. 2004. vine-1 retrotransposon-based sequence-specific amplified polymorphism for Vitis vinifera L. genotyping Plant Breeding 123: 180-185.

Lachenbruch, P.A., and Mickey, M.A. 1968. Estimation of error rates in discriminant analysis. Technometrics 10:1-10.

Lai, J., Ma, J., Swigonova, Z., Ramakrishna, W., Linton, E., Llaca, V., Tanyolac, B., Park, Y-J., Jeong, O-Y., Bennetzen, J.L., and Messing, J. 2004. Gene loss and movement in the maize genome. Genome Res. 14:1924-1931.

Langridge, P., Karakousis, A., Collins, N., Kretchmer, J., and Manning, S. 1995. A consensus linkage map of barley. Mol Breeding 1: 389-395.

Leeton, P.R., and Smyth, D.R. 1993. An abundant LINE-like element amplified in the genome of *Lilium speciosum*. Mol. Gen. Genet. 237:97-104.

Leigh, F., Kalendar, R., Lea, V., Lee, D., Donini, P., and Schulman, A.H. 2003. Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol Gen Genomics 269:464-474.

Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Laviolette, J-P., Ardlie, K., Reich, D.E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Bing Fan, J-B., Gingeras, T., Warrington, J., Patil, N., Hudson, T.J., and Lander, E.S. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse Nature Genetics 24: 381-386.

Lindhout, P. 2002. The perspective of polygenic inheritance in breeding for durable disease resistance. Euphytica 124:217-226.

Liu, B., and Wendel, J.F. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43:874-880.

Liu, Q., Thorland, E.C., Heit, J.A. and Sommers, S.S. 1997. Overlapping PCR for bidirectional PCR amplification of specific alleles: A Rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. Genome Res 7: 389-398.

Ma, J., Devos, K.M., and Bennetzen, J.L. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. 14: 860-869.

Madlung, A., and Comai, L. 2004. The effects of stress on genome regulation and structure. Ann. Bot. 94: 481-495.

Manninen, I., and Schulman, A. 1993. BARE-1, a *copia*-like retroelement in barley (*Hordeum vulgare* L.) Plant Mol. Biol. 22:829-846.

Manninen, I., Kalendar, R., Robinson, J., and Schulman, A.H. 2000. Application of BARE-1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. Mol. Gen. Genet. 264:325-334.

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 American spring barley based on coefficient of parentage. Crop Sci 31:1131-1137.

Martinez-Izquierdo, J.A., Garcia-Martinez, J. And Vicient, C.M. 1997. What makes Grande-1 retrotransposon different? Genetica 100:15-28.

Mayr, E. 1980. How I became a Darwinian. In: Mayr E, Provine WB (eds) The evolutionary synthesis. Perspectives on the unification of biology. Harvard University Press, Cambridge Mass., pp 413-423.

Melayah, D., Bonnivard, E., Chalhoud, B., Audeon, C., and Grandbastien, M-A. 2001. The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. The Plant J. 28:159-168.

McClintock, B. 1950. The origin and behaviour of mutable loci in maize. Proc. Natl. Acad. Sci. USA 36:344-355.

Michaud, E.J., van vugt, M.J., Bultman, S.J., Sweet, O.H., Davidson, M.T., and Woychik, R.P. 1994. Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes & Development 8: 1463-1472.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., and Hirochika, H. 2003. Target site specificity of Tos17 retrotransposon

shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. The Plant Cell 15:1771-1780.

Moore, G., Cheung, W., Schwarzacher, T., and Flavell, A. 1991 BIS-1, a major component of the cereal genome and a tool for studying genomic organization. Genomics 19:469-476.

Moran, J.V., DeBernardinis, R.J., and Kazazian, H.H. Jr. 1999. Exon shuffling by L1 retrotransposition. Science 283:1530-1534.

Nei, M. 1972. Genetic distance between populations Amer. Naturalist 106:283-292.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, NY, USA.

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.

Nevo, E. 1992. Origin, evolution, population genetics and resources for breeding of wild barley, *Hordeum spontaneum*, in the Fertile Crescent. *In* Shewry PR (ed.): Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, p19-43, CAB International Oxford.

Nielsen, G., and Johansen, H.B. 1986. Proposal for the identification of barley varieties based on the genotypes of 2 hordein and 39 isozyme loci of 47 reference varieties. Euphytica 35: 717-728.

Noma, K., Ohtsubo, E., and Ohtsubo, H. 1999. Non-LTR retrotransposon LINEs are ubiquitous components of plant genomes. Mol. Gen. Genet. 261:71-79.

Olsen, M., Hood, L., Cantor, C., and Botstein, D. 1989. A common language for physical mapping of the human genome. Science 245:1434-1435.

Olufowote, J.O., Xu, Y., Chen, X., Park, W.D., Beachell, H.M., Dilday, R.H., Goto, M., and McCouch, S.R. 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellites and RFLP markers. Genome 40:370-378.

Orgel, L.E., and Crick, H.C. 1980. Selfish DNA: The ultimate parasite. Nature 284:604-607.

Ostertag, E.M., and Kazazian, H.H., Jr. 2001. Biology of mammalian L1 retrotransposon. Annu. Rev. genet. 35:501-538.

Panstruga, R., Buschges, R., Piffanelli, P., and Schultz-Lefert, P. 1998. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res. 26: 1056-1062.

Paterson, A.H., Bowers, J.E., Burow, M.D., Draye, X., Elsik, C.G., Jiang, C-X., Katsar, C.S., Lan, T-H., Lin, Y-R., Ming, R., and Wright, R.J. 2000. Comparative genomics of plant chromosomes. Plant Cell 12:1523-1539.

Pearce, S.R., Harrison, G., Heslop-Harrison, J.S., Flavell, A.J., and Kumar, A. 1997. Chracterization and genomic organisation of Ty1-copia group retrotransposons in rye (*Secale cereale*). Genome 40: 1-9.

Pearce, S.R., Harrison, G., Li, D., Heslop-Harrison, J.S., Kumar, A., and Flavell, A.J. 1996b. The Ty1-*copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localization. Mol. Gen. Genet. 250: 305-315.

Pearce, S.R., Kumar, A., and Flavell, A.J. 1996. Activation of the Ty1-*copia* group retrotransposons of potato (*Solanum tuberosum*) during protoplast isolation. Plant Cell Rep. 15: 949-953.

Pejic, I., Ajmone-Marsan, P., Morgante, M., Kozumplick, V., Castiglioni, P., Taramino, G., and Motto, M. 1998. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs Theor. Appl. Genet. 97:1248-1255.

Pelissier, T., Tutois, S., Tourmente, S., Deragon, J. M., and Picard, G.1996. DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in *Athila* retroelement sequences. Genetica 97:141-151.

Pereira, V. 2004. Insertion bias and purifying selection for retrotransposons in the *Arabidopsis thaliana* genome. Genome Biology 5: R79.

Perry, D.J. 2004. Identification of Canadian durum wheat varieties using a single PCR. Theor Appl Genet 109:55-61.

Petrov, D. 1977. Slow but steady: Reduction of genome size through biased mutation. Plant Cell 11:1900-1901.

Poethig, R.S. 2001. Life with 25,000 genes. Genome Res. 11:313-316.

Porceddu, E., Albertini, G., Barcaccia, G., Marconi, F., Bertoli, F., and Veronesi, F. 2002. Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol. Gen. Genet. 267:107-114.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalski, A. 1996. The comparison of RFLPs, RAPDs, AFLPs, and SSR (microsatellite) markers for germplasm analysis Mol. Breeding 2:225-238.

Pozueta-Romero, J., Houlne, G., and Schantz, R. 1998. Identification of a short interspersed repetitive element in partially spliced transcript of the bell pepper

(*Capsicum annuum*) PAP gene: new evolutionary and reuglatory aspects on plant tRNA related SINEs. Gene 214:51-58.

Qi, X., and Lindhout, P. 1997. Development of AFLP markers in barley. Mol Gen Genet 254:330-336.

Queen, R.A., Gribbon, B.M., James, C., Jack, P., and Flavell, A.J. 2004. Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat Mol. Gen. Genomics 271:91-97.

Rao, C.R. 1952. Advanced statistical methods in biometric research. Wiley, New York. USA.

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.

Richard, I., and Beckman, J.S. 1995. How neutral are synonymous codon mutations? Nat. Genet 10: 259.

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. 2000. NTSYS-pc. Numerical taxonomy and multivariate analysis system. Ver. 2.1 Exeter software, Setauket, New York.

Ross, P., Hall, L., Smirnov, I. and Haff, L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol 16: 1347-1351.

Russell, J.R., Fuller, J.D., Macaulay, M., Hatz, B.G., Jahoor, A., Powell, W., and Waugh, R. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs Theor. Appl. Genet. 95:714-722.

Sachidanandam, R.,Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., and Altshuler, D. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933.

Saghai-Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q., and Allard, R.W. 1994a. Extraordinary polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. USA 91:5466-5470.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., and Allard, R.W. 1994b. Ribosomal DNA spacer-length polymorphism in barley : Mendelian inheritance, chromosomal location, and population dynamics Proc. Natl. Acad. Sci. USA 81:8014-8018.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467.

SanMiguel P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, JL. 1998. The palaeontology of integrase retrotransposons of maize. Nature Genet. 20: 43-45.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765-764.

SAS Institute Inc. (2002) SAS/STAT user's guide, version 8.2 [computer program] SAS Institute Inc, Cary, N.C.

Schneider, S., Kueffer, J.M., Roessli, D., and Excoffier, L. 1997. Arlequin, a software for population genetic data analysis. Ver. 1.1, URL: http://anthropologie.unige.ch/arlequin.

Schramke, V., and Allshire, R. 2003. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science 301: 1069-1074.

Schwarz-Sommer, Z., Leclercq, L., Geobel, E., and Saedler, H. 1987 *Cin4*, an insert altering the structure of *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J. 6:3873-3880.

Shapiro, J.A. 1999. Transposable elements are the key to the 21st century view of evolution. Genetica 107:171-179.

Shirasu, K., Schulman, A.H., Lahaye, T., and Schulze-Lefert, P. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908-915.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. Genet. Res. Camb. 58:167-175.

Smith, O.S., and Smith, J.S.C. 1992. Measurements of genetic diversity among maize inbreds, a comparison of isozymes, RFLP, pedigree, and heterosis data. Maydica 37:53-63.

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, F_1 heterosis, and RFLPs. Theor Appl. Genet. 80:833-840.

Soleimani, V.D., Baum, B.R., and Johnson, D.A. 2002a. AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat [*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.]. Theor. Appl. Genet. 104: 350-357.

Soleimani, V.D., Baum, B.R., and Johnson, D.A. 2002b. Identification of Canadian durum wheat [*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.] cultivars using AFLP and their STS markers. Can. J. Plant Sci. 82: 35-41.

Soleimani, V.D., Baum, B.R., and Johnson, D.A. 2003. Efficient validation of single nucleotide polymorphisms in plants by allele-specific PCR, with an example from barley Plant Mol. Biol. Reporter 21:281-288.

Sorrells, M.E., La Rota, M., Bermudez-Kandianis, C.E., Greene, R.A., Kantety, R., Munkvold, J.D., Mahmoud, M.A., Ma, X., Gustafson, P.J., Qi, L.L., Echalier, B., Gill, B.S., Mathews, D.E., Lazo, G.R., Chao, S., Anderson, O.D., Edwards, H., Linkiewicz, A.M., Dubcovsky, J., Akhunov, E.D., Dvorak, J., Zhang, D., Nguyen, H.T., Peng, J., Lapitan, N.L.V., Gonzalez-Hernandez, J.L., Anderson, J.A., Hossain, K., Kalavacharla, V., Kianian, S.F., Choi, D.W., Close, T.J., Dilbirligi, M., Gill, K.S., Steber, C., Walker-Simmons, M.K., McGuire, P.E., and Qualset, C.O. 2000.

Comparative DNA sequence analysis of wheat and rice genomes. Genome Res. 13:1818-1827.

Souza, E., and Sorrells, M.E. 1989. Pedigree analysis of North American oat cultivars released from 1951 to 1985 Crop Sci. 29:595-601.

Sunyaev, S., Hanke, J., Aydin, A., Wirkner, U., Zastrow, I., Reich, J. and Bork, P. 1999. Prediction of nonsynonymous single nucleotide polymorphisms in human disease-associated genes J Mol Med 77: 754-760.

Suoniemi, A., Anamthawat-Jonsson, K., Arna, T., and Schulman, A.H. 1996b. Retrotransposon BARE-1 is a major, dispersed component of barley (*Hordeum vulgare* L.) genome. Plant Mol. Biol. 30: 1321-1329.

Suoniemi, A., Narvanto, A., and Schulman, A. 1996a. The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol Biol 31:295-306.

Suoniemi, A., Schmidt, D., and Schulman, A.H. 1999. BARE-1 insertion site preferences and evolutionary conservation of RNA and cDNA processing sites. Genetica 100:219-230.

Suoniemi, A., Tanskanen, J., and Schulman, A.H. 1998. Gypsy-like retrotransposons are widespread in the plant kingdom. Plant J. 13:699-705.

Tadeka, S., Sugimoto, K., Otsuki, H., and Hirochika, H. 1998. Transpositional activation of the tobacco retrotransposon Tto1 by wounding and methyljasmonate. Plant Mol. Biol. 36:365-376.

Tam, S.M., Mhiri, C., Vogelaar, A., Kerkveld, M., Pearce, S.R. 2005. Comparative analyses of genetic diversities within tomato and pepper collections detected by

retrotransposon-based SSAP, AFLP and SSR. Theor. Appl. Gene. 8 February 2005, online version.

Tautz, D., and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes Nucl. Acids Res. 12:4127-4138.

Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F., Gaut, B.S. 2001. Pattern of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.) Proc. Natl. Acad. Sci. USA 98:9161-9166.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.

Thormann, C.E., and Osborn, T.C. 1993. Use of RAPD and RFLP markers for germplasm evaluation, In application of RAPD technology to plant breeding, Neff M, Ed, ASHS publishers, St. Paul, MN.

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.

Tinker, N.A., and Mather, D.E. 1993a. KIN: software for computing kinship coefficients. J. Hered. 84:238.

Tinker, N.A., and Mather, D.E. 1993b. Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor Appl Genet 85:976-984.

Tohme, J., Orlando Gonzalez, D., Beebe, S., and Duque, M. 1996. AFLP analysis of gene pools of a wild bean core collection. Crop Sci. 36:1375-1384.

Turcich, M.P., Bokharirriza, A., Hamilton, D.A., He, C.P., and Messier, W. 1996. PREM-2, a *copia*-type retroelement in maize is expressed preferentially in early microspores. Sexual Plant Reprod. 9:65-74.

UPOV, 1991. International convention for the protection of new varieties of plants, Publication No. 221 (E), March 19, Geneva.

Van Toal, T.T., Peng, J., and St Martin, S.K. 1997. Using AFLP markers to determine the genomic contribution of parents to populations. Crop. Sci. 37:1370-1373.

Vandepoele, K., Simillion, C., and Van de Peer, Y. 2003. Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15:2192-2202.

Varagona, M.J., Purigganan, M., and Wessler, S.R. 1992. Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. Plant Cell 4:811-820.

Vavilov, N.I. 1926. Studies in the origin of cultivated plants. St. Petersburg: Institute of Applied Botany and Plant Breeding, 248 p.

Vernhettes, S., Grandbastien, M-A., and Casacuberta, J.M. 1997. In vivo characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco *Tnt1*. Plant Mol. Biol. 35: 673-679.

Vershinin, A.V., Salina, E.A., Solovyov, V.V., and Timofeyeva, L.L. 1990. Genomic organization, evolution and structural peculiarities of high repetitive DNA of *Hordeum vulgare*. Genome 33: 441-449.

Vicient, C.M., Sioniemi, A., Anamthawat-Jonsson, K., Tanslanen, J., Beharav, A., Nevo, E., and Schulman, A.H. 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. The Plant Cell 11:1769-1784.

Vignols, F., Rigau, J., Torres, M.A., Capellades, M., and Puigdomenech, P. 1995. The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. Plant Cell 7:407-416.

Vitte, C., and Panaud, O. 2003. Formation of solo LTRs through unequal homologous recombination counterbalances amplification of retrotransposons in rice *Oryza sativa* L. Mol. Biol. Evol. 20:528-540.

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.

Voytas, D.F., Cummings, M.P., Konieczny, A., Ausubel, F.M., and Rodermel, S.R. 1992. *Copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA 89:7124-7128.

Wakeley, J. 1994. Substitution rate variation among sites and the estimation of transition bias. Mol. Biol. Evol. 11: 436-442.

Watson, J.D., and Crick, F.H.C. 1953. A structure for deoxyribonucleic acid. Nature 171:737-738.

Waugh, R., Mclean, K., Flavell, A.J., Pearce, S.R., and Kumar, A. 1997. Genetic distribution of BARE-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplified polymorphism (S-SAP). Mol Gen Genet 253:687-694.

Weir, B.S. 1990. Genetic data analysis methods for discrete genetic data. Sinauer Associates , Sunderland MA.

Wessler, S. 1996. Plant retrotransposons: turned on by stress. Curr. Biol. 6:959-961.

White, S.E., Habers, L., and Wessler, S.R. 1994. Retrotransposons in the flanking regions of normal plant genes. A role for *copia*-like elements in the evolution of gene structure and expression. Proc. Natl. Acad. Sci. USA 91:11792-11796.

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E., and Keller, B. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. The Plant Journal 26:307-316.

Wilke, C. M. and Adams, J. 1992. Fitness effects of Ty transpositions in *Saccharomyces cerevisiae*. Genetics 131:31-42.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers Nucl. Acids Res. 18:6 Baum, B.R., Mechenda, S., and Soleimani, V.D. 2000. Identification of Canadian six row barley (*Hordeum vulgare* L.) cultivars with primers derived from STSs obtained from RAPD diagnostic bands Seed. Sci. & Technol. 28: 445-466.

Witte, C.P., Le, Q.H., Bureau, T., and Kumar, A. 2001. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc. Natl. Acad. Sci. USA. 98:13778-13783.

Wright, D.A., Ke, N., Smalle, J., Hauge, B.M., Goodman, H.M., and Voytas, D.F. 1996. Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. Genetics 142:569-578.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evol. 19:395-420.

Ye, S.., Humphries, S., and Green, F. 1992. Allele specific amplification by tetra-primer PCR. Nucleic Acids Res. 20: 1152-1156.

Yoshioka, Y., Matsumoto, S., Kjima, S., Ohshima, K., Okada, N., and Nachida, Y. 1993. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNA. Proc. Natl. Acad. Sci. USA 90:6562-6566.

Zhang, X., and Wessler, S.R. 2004. Genome-wide comparative analysis of the transposable elements in related species *Arabidopsis thaliana* and *Brassica oleracea*. Proc. Natl. Acad. Sci. USA 101: 5589-5594.

Zhao, Z., Li, J., and Fu, Y. 2000. Worldwide DNA sequence variation in a 10-kilobase non-coding region on human chromosome 22. Proc. Natl. Acad. Sci. USA 97:11354-11358.

Zhu, Y.L., Song, Q.J., Hyten, D.L., Van Tassell, C.P., Matukumalli, L.K., Grimm, D.R. Hyatt, S.M., Fickus, E.W., Young, N.D., and Cregan, P.B. 2003 Single nucleotide polymorphisms in soybean. Genetics 163:1123-1134.

Ziheng, Y., and Yoder, A.D. 1999. Estimation of the transition/transversion rate bias and species sampling. J. Mol. Evol. 48: 274-283.

Zohary, D., and Hopf, M. 1993. Domistication of plants in the old world. The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Clarendon Press. Oxford, UK.

APPENDIX

An identification key that can be used to distinguish spring six-rowed barley cultivars that are currently registered in Canada. The key was generated using SNP markers and includes confirmatory characters for reliable identification. Barley cultivars included in the key are listed in tables 2.1 and 6.1. SNP markers that are used in the identification key are described in table 6.2.

```
1(0).    SNP024 absent.................................................. 2
         SNP024 present................................................. 94

2(1).    SNP029 absent.................................................. 3
         SNP029 present................................................. 42

3(2).    SNP043 absent.................................................. 4
         SNP043 present................................................. 25

4(3).    SNP020 absent.................................................. 5
         SNP020 present................................................. 18

5(4).    SNP215 absent.................................................. 6
         SNP215 present................................................. 12

6(5).    SNP018 absent.................................................. 7
         SNP018 present................................................. 10

7(6).    SNP061 absent; SNP078 absent; SNP081 absent........ 104. AC Rosser-8
         SNP061 present; SNP078 present; SNP081 present.................. 8

8(7).    SNP001 absent..................................... 721. Sophie-1
         SNP001 present................................................. 9
```

9(8). SNP030 absent.. 757. Tankard-5

SNP030 present... 724. Sophie-4

10(6). SNP022 absent; SNP025 absent........................ 815. Westford-2

SNP022 present; SNP025 present................................. 11

11(10). SNP031 absent; SNP081 absent........................ 336. Cadette-8

SNP031 present; SNP081 present........................ 723. Sophie-3

12(5). SNP031 absent.. 13

SNP031 present... 16

13(12). SNP018 absent; SNP081 present................................. 14

SNP018 present; SNP081 absent................................. 15

14(13). SNP001 absent.. 647. Noble-7

SNP001 present....................................... 641. Noble-1

15(13). SNP078 absent...................................... 329. Cadette-1

SNP078 present..................................... 331. Cadette-3

16(12). SNP025 absent; SNP001 absent......................... 805. Vivar-8

SNP025 present; SNP001 present................................. 17

17(16). SNP030 absent...................................... 755. Tankard-3

SNP030 present....................................... 727. Sophie-7

18(4). SNP030 absent.. 19

SNP030 present... 22

19(18). SNP081 absent; SNP018 present......................... 172. B1602-4

SNP081 present; SNP018 absent.................................. 20

20(19). SNP042 absent; SNP022 present; SNP001 absent........ 48. AC Harper-8

SNP042 present; SNP022 absent; SNP001 present................... 21

21(20). SNP215 absent....................................... 753. Tankard-1

SNP215 present...................................... 760. Tankard-8

22(18). SNP022 absent.. 23

SNP022 present... 24

23(22). SNP025 absent; SNP026 present; SNP031 present......... 797. Virden-8

SNP025 present; SNP026 absent; SNP031 absent......... 92. AC Nadia-4

24(22). SNP025 absent; SNP026 present; SNP078 present.......... 789. Tukwa-8

SNP025 present; SNP026 absent; SNP078 absent...... 152. AC Westech-8

25(3). SNP081 absent.. 26

SNP081 present... 31

26(25). SNP018 absent.. 27

SNP018 present... 29

27(26). SNP075 absent; SNP077 present; SNP059 absent.................... 28

SNP075 present; SNP077 absent; SNP059 present.128. CDC Battleford-16

28(27). SNP022 absent.................................. 887. Peregerine-2

SNP022 present................................ 901. Peregerine-16

29(26). SNP030 absent; SNP057 present; SNP059 absent........... 420. Duel-12

SNP030 present; SNP057 absent; SNP059 present................... 30

30(29). SNP031 absent; SNP044 absent; SNP045 present............ 655. Nord-7

SNP031 present; SNP044 present; SNP045 absent........ 495. Jackson-7

31(25). SNP042 absent; SNP044 absent; SNP045 present.................... 32

SNP042 present; SNP044 present; SNP045 absent.................... 39

32(31). SNP025 absent.. 33

SNP025 present... 34

33(32). SNP031 absent; SNP026 absent; SNP078 present.......... 774. Trochu-1

SNP031 present; SNP026 present; SNP078 absent........ 810. Viviane-5

34(32). SNP022 absent.. 35

SNP022 present... 38

35(34). SNP077 absent..................................... 462. Foster-6

SNP077 present... 36

36(35). SNP215 absent..................................... 458. Foster-2

SNP215 present... 37

37(36). SNP066 absent.................................... 470. Foster-14

SNP066 present................................... 467. Foster-11

38(34). SNP077 absent; SNP078 present; SNP215 absent........... 437. Etien-5

SNP077 present; SNP078 absent; SNP215 present..... 147. AC Westech-3

39(31). SNP018 absent.. 40

SNP018 present... 41

40(39). SNP031 absent................................. 281. CDC Tisdale-1

SNP031 present.................................. 285. CDC Tisdale-5

41(39). SNP030 absent................................. 294. CDC Tisdale-14

SNP030 present................................. 291. CDC Tisdale-11

42(2). SNP030 absent.. 43

SNP030 present.. 72

43(42). SNP020 absent; SNP022 present.................................... 44

SNP020 present; SNP022 absent..................................... 58

44(43). SNP042 absent.. 45

SNP042 present.. 55

45(44). SNP059 absent.. 46

SNP059 present.. 52

46(45). SNP001 absent.. 47

SNP001 present.. 49

47(46). SNP018 absent; SNP057 present; SNP075 absent..................... 48

SNP018 present; SNP057 absent; SNP075 present........ 734. Stander-6

48(47). SNP081 absent....................................... 472. Foster-16

SNP081 present...................................... 471. Foster-15

49(46). SNP061 absent.. 50

SNP061 present............................... 541. CDC Springside-5

50(49). SNP044 absent.............................. 552. CDC Springside-16

SNP044 present.. 51

51(50). SNP043 absent; SNP047 absent................ 551. CDC Springside-15
SNP043 present; SNP047 present.............. 550. CDC Springside-14

52(45). SNP025 absent..................................... 740. Stander-12
SNP025 present... 53

53(52). SNP018 absent; SNP057 present.................................. 54
SNP018 present; SNP057 absent...................... 743. Stander-15

54(53). SNP031 absent; SNP081 present........................ 463. Foster-7
SNP031 present; SNP081 absent........................ 464. Foster-8

55(44). SNP078 absent... 143. Lacey-15
SNP078 present... 56

56(55). SNP018 absent; SNP001 present; SNP043 absent.................... 57
SNP018 present; SNP001 absent; SNP043 present........ 226. Bedford-2

57(56). SNP045 absent; SNP046 absent; SNP061 present.. 538. CDC Springside-2
SNP045 present; SNP046 present; SNP061 absent. 546. CDC Springside-10

58(43). SNP057 absent.. 59
SNP057 present... 65

59(58). SNP042 absent; SNP043 present; SNP044 absent......... 693. Robust-13
SNP042 present; SNP043 absent; SNP044 present................... 60

60(59). SNP077 absent.. 61
SNP077 present... 63

61(60). SNP018 absent; SNP081 present....................... 142. Lacey-14
SNP018 present; SNP081 absent................................. 62

62(61). SNP078 absent.................................. 380. CDC Yorkton-4

SNP078 present................................. 391. CDC Yorkton-15

63(60). SNP059 absent; SNP078 present.................. 386. CDC Yorkton-10

SNP059 present; SNP078 absent................................... 64

64(63). SNP075 absent.................................. 387. CDC Yorkton-11

SNP075 present................................. 381. CDC Yorkton-5

65(58). SNP018 absent; SNP042 absent; SNP043 present.......... 456. Excel-16

SNP018 present; SNP042 present; SNP043 absent.................... 66

66(65). SNP059 absent.. 67

SNP059 present... 70

67(66). SNP077 absent.. 68

SNP077 present... 69

68(67). SNP061 absent.................................. 378. CDC Yorkton-2

SNP061 present................................. 382. CDC Yorkton-6

69(67). SNP075 absent.................................. 377. CDC Yorkton-1

SNP075 present................................. 389. CDC Yorkton-13

70(66). SNP061 absent.. 71

SNP061 present................................. 390. CDC Yorkton-14

71(70). SNP075 absent; SNP078 present.................. 385. CDC Yorkton-9

SNP075 present; SNP078 absent.................. 388. CDC Yorkton-12

72(42). SNP057 absent; SNP075 present.................................. 73

SNP057 present; SNP075 absent.................................... 81

73(72). SNP061 absent.. 74

SNP061 present... 77

74(73). SNP025 absent; SNP045 present; SNP046 present.......... 581. Leger-5

SNP025 present; SNP045 absent; SNP046 absent..................... 75

75(74). SNP042 absent; SNP081 present................................... 76

SNP042 present; SNP081 absent..................... 249. Blankeney-1

76(75). SNP026 absent...................................... 80. AC Malone-8

SNP026 present..................................... 24. AC Burman-8

77(73). SNP018 absent; SNP031 absent; SNP215 present.................... 78

SNP018 present; SNP031 present; SNP215 absent.................... 80

78(77). SNP081 absent...................................... 528. Klondike-8

SNP081 present... 79

79(78). SNP020 absent...................................... 632. Nellygan-8

SNP020 present..................................... 630. Nellygan-6

80(77). SNP022 absent.. 617. Myriam-1

SNP022 present....................................... 621. Myriam-5

81(72). SNP081 absent.. 82

SNP081 present... 88

82(81). SNP061 absent.. 83

SNP061 present... 85

83(82). SNP031 absent; SNP043 absent; SNP044 present.................... 84

SNP031 present; SNP043 present; SNP044 absent........ 612. Mistral-4

84(83). SNP059 absent.. 512. Jolly-8

SNP059 present.. 506. Jolly-2

85(82). SNP215 absent... 312. Brooke-8

SNP215 present.. 86

86(85). SNP043 absent.. 87

SNP043 present.................................... 357. CDC Earl-5

87(86). SNP042 absent; SNP044 absent; SNP045 present........ 353. CDC Earl-1

SNP042 present; SNP044 present; SNP045 absent....... 359. CDC Earl-7

88(81). SNP020 absent.. 89

SNP020 present.. 90

89(88). SNP215 absent...................................... 713. Sandrine-1

SNP215 present....................................... 745. Stetsen-1

90(88). SNP018 absent.. 91

SNP018 present.. 93

91(90). SNP031 absent; SNP078 present; SNP215 absent.................... 92

SNP031 present; SNP078 absent; SNP215 present....... 714. Sandrine-2

92(91). SNP022 absent.................................... 40. AC Hamilton-8

SNP022 present.................................. 37. AC Hamilton-5

93(90). SNP022 absent.................................... 38. AC Hamilton-6

SNP022 present.................................. 34. AC Hamilton-2

94(1). SNP030 absent.. 95

SNP030 present... 127

95(94). SNP018 absent.. 96

SNP018 present... 114

96(95). SNP042 absent.. 97

SNP042 present... 103

97(96). SNP029 absent.. 98

SNP029 present... 100

98(97). SNP059 absent; SNP215 absent; SNP026 present.................... 99

SNP059 present; SNP215 present; SNP026 absent......... 861. Falcon-8

99(98). SNP081 absent..................................... 49. AC Klinck-1

SNP081 present.................................... 53. AC Klinck-5

100(97). SNP077 absent... 101

SNP077 present.. 102

101(100). SNP026 absent; SNP031 present; SNP043 absent............. 32. ACCA-8

SNP026 present; SNP031 absent; SNP043 present........... 425. Duke-1

102(100). SNP215 absent.. 200. BT954-16

SNP215 present....................................... 185. BT954-1

103(96). SNP022 absent; SNP020 present................................. 104

SNP022 present; SNP020 absent................................. 108

104(103). SNP001 absent.. 105

SNP001 present.. 158. Argyle-6

105(104). SNP029 absent; SNP043 absent.................................... 106

SNP029 present; SNP043 present.......................... 484. Grant-4

106(105). SNP031 absent; SNP078 absent; SNP081 present......... 400. Chapais-8

SNP031 present; SNP078 present; SNP081 absent.................... 107

107(106). SNP027 absent.. 265. Bonanza-1

SNP027 present.. 268. Bonanza-4

108(103). SNP215 absent.. 109

SNP215 present.. 111

109(108). SNP078 absent....................................... 70. AC Legend-6

SNP078 present.. 110

110(109). SNP029 absent.. 84. AC Maple-4

SNP029 present.. 222. Banner-6

111(108). SNP026 absent; SNP031 absent; SNP077 absent............ 408. Codac-8

SNP026 present; SNP031 present; SNP077 present.................. 112

112(111). SNP029 absent.. 113

SNP029 present.. 223. Banner-7

113(112). SNP001 absent; SNP059 absent; SNP081 present......... 81. AC Maple-1

SNP001 present; SNP059 present; SNP081 absent........ 881. Jaeger-12

114(95). SNP022 absent.. 115

SNP022 present.. 122

115(114). SNP001 absent.. 116

SNP001 present.. 119

116(115). SNP029 absent; SNP081 absent; SNP215 present.................... 117

SNP029 present; SNP081 present; SNP215 absent.................. 118

117(116). SNP042 absent; SNP043 present; SNP044 absent..... 833. AC Hawkeye-12

SNP042 present; SNP043 absent; SNP044 present........ 269. Bonanza-5

118(116). SNP031 absent.. 297. Brier-1

SNP031 present.. 209. Legacy-9

119(115). SNP020 absent; SNP059 present; SNP081 absent......... 885. Jaeger-16

SNP020 present; SNP059 absent; SNP081 present.................. 120

120(119). SNP215 absent................................. 375. CDC Sisler-15

SNP215 present.. 121

121(120). SNP031 absent................................. 371. CDC Sisler-11

SNP031 present................................. 365. CDC Sisler-5

122(114). SNP057 absent; SNP077 absent.................................. 123

SNP057 present; SNP077 present................................ 126

123(122). SNP029 absent; SNP042 absent; SNP044 absent.................... 124

SNP029 present; SNP042 present; SNP044 present......... 233. Bella-1

124(123). SNP043 absent; SNP045 absent; SNP075 present....... 8. AC Albright-8

SNP043 present; SNP045 present; SNP075 absent.................. 125

125(124). SNP020 absent.................................. 822. AC Hawkeye-1

SNP020 present................................. 831. AC Hawkeye-10

126(122). SNP026 absent; SNP029 present; SNP059 absent....... 112. AC Stacey-8

SNP026 present; SNP029 absent; SNP059 present...... 846. CDC Silky-9

127(94). SNP043 absent.. 128

SNP043 present... 144

128(127). SNP215 absent.. 129

SNP215 present... 139

129(128). SNP022 absent; SNP020 present................................. 130

SNP022 present; SNP020 absent................................. 134

130(129). SNP078 absent.. 131

SNP078 present... 133

131(130). SNP031 absent; SNP042 absent........................ 16. AC Alma-8

SNP031 present; SNP042 present............................... 132

132(131). SNP026 absent; SNP029 absent; SNP018 present......... 272. Bonanza-8

SNP026 present; SNP029 present; SNP018 absent.......... 585. Lucky-1

133(130). SNP025 absent...................................... 566. Legacy-22

SNP025 present..................................... 575. Legacy-31

134(129). SNP026 absent; SNP029 absent.................................. 135

SNP026 present; SNP029 present................................. 138

135(134). SNP059 absent.. 136

SNP059 present.................................... 60. AC Lacombe-4

136(135). SNP042 absent; SNP057 absent; SNP025 absent........ 58. AC Lacombe-2

SNP042 present; SNP057 present; SNP025 present.................. 137

137(136). SNP018 absent.. 728. Sophie-8

SNP018 present.. 725. Sophie-5

138(134). SNP031 absent; SNP081 present........................ 313. Bronco-1

SNP031 present; SNP081 absent....................... 248. Belluga-8

139(128). SNP045 absent; SNP046 absent.................................. 140

SNP045 present; SNP046 present................................ 143

140(139). SNP029 absent.. 141

SNP029 present.. 142

141(140). SNP026 absent; SNP031 absent; SNP022 present........... 304. Brier-8

SNP026 present; SNP031 present; SNP022 absent....... 280. Bonanza-16

142(140). SNP001 absent; SNP081 present; SNP020 absent......... 773. TB891-6-8

SNP001 present; SNP081 absent; SNP020 present.......... 257. Blitz-1

143(139). SNP029 absent; SNP031 present; SNP044 present.......... 640. Niska-8

SNP029 present; SNP031 absent; SNP044 absent........ 337. AC Bacon-1

144(127). SNP215 absent.. 145

SNP215 present.. 151

145(144). SNP031 absent.. 146

SNP031 present.. 150

146(145). SNP026 absent.. 147

SNP026 present... 149

147(146). SNP025 absent................................... 670. OAC Kippen-6

SNP025 present... 148

148(147). SNP018 absent................................... 671. OAC Kippen-7

SNP018 present.................................. 666. OAC Kippen-2

149(146). SNP078 absent... 703. Sabina-7

328. Brucefield-8

SNP078 present.................................. 321. Brucefield-1

150(145). SNP022 absent; SNP018 absent; SNP044 present.......... 473. Gamine-1

SNP022 present; SNP018 present; SNP044 absent........... 680. Otal-8

151(144). SNP022 absent.. 152

SNP022 present... 156

152(151). SNP026 absent; SNP078 present................................... 153

SNP026 present; SNP078 absent.................................... 155

153(152). SNP029 absent; SNP061 present................................... 154

SNP029 present; SNP061 absent........................ 520. Kasota-8

154(153). SNP001 absent...................................... 593. Mahigan-1

SNP001 present..................................... 594. Mahigan-2

155(152). SNP029 absent; SNP031 present; SNP045 present..... 664. OAC Baxter-8

SNP029 present; SNP031 absent; SNP045 absent........ 499. Johnston-3

156(151). SNP081 absent.. 157

SNP081 present.. 158

157(156). SNP029 absent; SNP031 present; SNP078 present......... 706. Samson-2

SNP029 present; SNP031 absent; SNP078 absent......... 533. Labelle-5

158(156). SNP026 absent.. 159

SNP026 present.. 160

159(158). SNP025 absent; SNP031 absent; SNP044 present........... 560. Leduc-8

SNP025 present; SNP031 present; SNP044 absent......... 710. Samson-6

160(158). SNP031 absent; SNP045 absent; SNP046 absent........... 601. Maskot-1

SNP031 present; SNP045 present; SNP046 present........ 712. Samson-8