VERIFICATION OF SNP MARKERS ASSOCIATED WITH AFTER-COOKING DARKENING IN POTATOES

by

Yuanyuan Wu

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia

in co-operation with

Nova Scotia Agricultural College
Truro, Nova Scotia

December 2011

DALHOUSIE UNIVERSITY

NOVA SCOTIA AGRICULTURAL COLLEGE

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "VERIFICATION OF SNP MARKERS ASSOCIATED WITH AFTER-COOKING DARKENING IN POTATOES" by Yuanyuan Wu in partial fulfilment of the requirements for the degree of Master of Science.

Dated: December 19, 2011

Supervisors: _____

_____

Readers: _____

_____

DALHOUSIE UNIVERSITY

AND

NOVA SCOTIA AGRICULTURAL COLLEGE

DATE: December 19, 2011

AUTHOR: Yuanyuan Wu

TITLE: VERIFICATION OF SNP MARKERS ASSOCIATED WITH AFTER-COOKING DARKENING IN POTATOES

DEPARTMENT OR SCHOOL: Department of Plant and Animal Sciences

DEGREE: M.Sc. CONVOCATION: May YEAR: 2012

_____

Signature of Author

Tables of Contents

List of Tables vii
List of Figures viii
Abstract ix
List of Abbreviations and Symbols Used x
Acknowledgements xiii
Chapter 1. Introduction 1
Chapter 2. Literature Review 3
2.1 The Potato 3
2.1.1 The Potato Tuber 4
2.1.2 Potato Tuber Quality 6
2.2 After-Cooking Darkening 8
2.2.1 Genetic Control of ACD 10
2.2.2 Environmental Factors Related to ACD 12
2.2.3 Genes Associated with ACD 13
2.3 DNA Markers 16
2.3.1 Types of DNA Markers 17
2.3.1.1 Length Polymorphism 17
2.3.1.2 Single Nucleotide Polymorphism 17
2.3.2 Methods for SNP Genotyping 19
2.3.2.1 High-Resolution DNA Melting 20
2.3.2.2 Two Approaches in HRM 21
2.3.3 Applications of Markers 24
2.3.3.1 Quantitative Trait Loci 25
2.3.3.2 Marker-Assisted Selection 26
Chapter 3. Hypotheses and Objectives 30
Chapter 4. Materials and Methods 31
4.1 Potato Samples 31
4.1.1 Field Trial and Storage 32
4.1.2 Evaluation of ACD 32

4.2 Genomic DNA Isolation 34
4.3 Candidate SNP Markers and Primers 35
4.4 Amplicon Genotyping in the Diploid Population 38
4.4.1 Amplification of the Markers Using PCR 38
4.4.2 HRM Scanning 38
4.4.3 Melting Curve Analysis 39
4.5 SNP Identification by Amplicon Sequencing 42
4.6 Unlabelled Probe Assay for Tetraploid Germplasm 45
4.6.1 Unlabelled Probe Design 45
4.6.2 PCR Amplification with Unlabelled Probe 45
4.6.3 Unlabelled Probe Genotyping 46
4.7 Marker Association Analysis 49
4.7.1 Phenotypic Analysis of the Diploid Population 49
4.7.2 The Correlation Analysis of ACD in the Six Trials 49
4.7.3 Single-Marker Analysis in Diploid and Tetraploid Populations 49
4.7.4 Identification of QTLs 50
4.7.5 Marker Analysis in the Individual Trials 50
4.7.6 Marker Genotype Corresponding to the Degree of ACD 51
4.7.7 Marker Contribution to the ACD 51
4.8 Functional Analysis of Different Marker Genotypes 53
Chapter 5. Results 54
5.1 Phenotypic Variance in the Diploid Population 54
5.2 The Correlation of the Degree of ACD in the Diploid Population 55
5.3 The Significant SNP Markers in the Diploid Population 57
5.3.1 Marker Segregation in the Diploid Population 57
5.3.2 SNP Positions of PCR Amplicons 66
5.3.3 Markers Significantly Associated with ACD in the Diploid Population 14946 70
5.3.4 ACD QTLs in the Diploid Population 14946 71
5.3.5 Marker Analysis in the Individual Trials 72
5.4 Marker Genotypes Corresponding to the Degree of ACD 74
5.5 Percentage of ACD Variance Explained by Markers 77

5.6 Amino Acid Sequences of the Three Most Significant Markers 78
5.7 Markers Significantly Associated with ACD in the Tetraploid Population 14945 . 80
5.7.1 Design of Probes 80
5.7.2 Tetraploid Scanning and Genotyping 82
5.7.3 Single-Marker Analysis in the Tetraploid Population 87
Chapter 6. Discussion 89
6.1 Validation of Genetic Markers 89
6.1.1 Population Size 90
6.1.2 Phenotypic Variance 91
6.2 Efficiency of HRM Genotyping 91
6.2.1 Utilization of HRM for Diploid Potato Genotyping 92
6.2.2 Application of HRM on Tetraploid Potatoes 94
6.3 Phenotypic Variance Explained by QTLs 95
Chapter 7. Direction of Future Research 97
Chapter 8. Conclusion 99
References 100
Appendix I 109

List of Tables

Table 1. Description of the 14 SNP markers associated with ACD in population 13610.36

Table 2. The arrangement of samples of the diploid population 14946 in the 96-well plate for PCR amplification. 40

Table 3. The DNA sequences of 12 pairs of M13-tailed primers used for sequencing. 44

Table 4. The arrangement of samples in PCR amplification with unlabelled probes... 47

Table 5. The variance of ACD in the diploid population 14946. 56

Table 6. The correlation coefficient (r) between ACD of each clone in the six trials... 56

Table 7. Summary of marker genotype segregation in population 14946 for the 14 SNP markers associated with ACD in population 13610..... 65

Table 8. Genotypes at each SNP position and derived haplotypes of 10 amplicons based on sequencing results. 67

Table 9. The effects of the significant markers on ACD in the six trials 70

Table 10. Markers significantly associated with ACD in populations 13610 and 14946. 71

Table 11. The effects of the significant markers on ACD in the individual trials..... 73

Table 12. Mean ACD value of each marker genotype. 75

Table 13. The relationship between allele copy number and ACD value for each marker. 76

Table 14. Percentage of ACD (R^2) explained by the six markers in the diploid population 14946..... 77

Table 15. Probe sequences designed for the 11 HRM markers..... 81

Table 16. The effects of the significant markers on ACD in the tetraploid clones of 14945..... 88

Table 17. Mean ACD value of each marker genotype in the progeny clones of 14945.. 88

List of Figures

Figure 1. The chemical reaction of the dark color formation of ACD in potatoes 8

Figure 2. Melting profile of the PCR product. 23

Figure 3. Melting temperature shift indicated by the presence of mismatched bases at the SNP loci. 23

Figure 4. Asymmetric PCR performed in the presence of the unlabelled probe. 24

Figure 5. The locations in the potato genome of 14 SNP markers associated with ACD. 28

Figure 6. Details of the markers mapped on chromosome 2. 29

Figure 7. The origin of the potato populations studied. 32

Figure 8. Amplicon genotyping in the diploid population 14946. 41

Figure 9. Unlabelled probe genotyping in the tetraploid population 14945. 48

Figure 10. The distribution of ACD in the diploid population 14946. 56

Figure 11. Grouping results of markers LS057 and LS097.. 61

Figure 12. Grouping results of markers LS030, LS124, and LS173. 62

Figure 13. Grouping results of marker LS163.. 63

Figure 14. Grouping results of markers BCH100, LS092, LS079, and LS161. 64

Figure 15. Grouping results of markers GP180, LS037, LS112, and LS201. 65

Figure 16. Alignment of DNA sequences of the *c4h* marker in the six marker genotypes (*ab*, *cd*, *bd*, *ad*, *ac*, and *bc*) and corresponding amino acid sequences. 79

Figure 17. Probe melting profiles for the two parents and eight clones of the tetraploid population 14945.. 86

Abstract

After-cooking darkening (ACD), a gray-black discoloration, is one of the key quality defects of the potato. Previous work using a diploid population 13610 identified 14 single nucleotide polymorphism (SNP) markers, located on eight chromosomes, which have the strongest association with ACD. Seven of these markers are located in four major quantitative trait loci (QTL) on chromosomes 2, 4, 6, and 10. It is hypothesized that these 14 markers would also have significant relationship to ACD in other breeding populations. Therefore, their effects in another diploid population 14946 were analyzed by amplicon genotyping using high-resolution DNA melting (HRM). Eleven of the 14 SNP markers were confirmed to be significantly associated with ACD. Six of the 11 markers were located in the four major QTL regions. In addition, unlabelled probe assays using HRM were performed on a group of clones in the tetraploid population 14945. Five from the 14 markers were analyzed, and three showed significant relationship to ACD. The methodology established and the markers identified in this study could benefit breeding programs for developing lower ACD varieties.

List of Abbreviations and Symbols Used

AA	Ascorbic Acid
aa	Amino Acid
AAFC	Agriculture and Agri-Food Canada
ACD	After-Cooking Darkening
AFLP	Amplified Fragment Length Polymorphism
ANOVA	Analysis of Variance
bp	Base Pair
c4h	Cinnamic Acid 4-Hydroxylase
CA	Citric Acid
CGA	Chlorogenic Acid
cM	Centimorgan
cop1	Constitutive Photomorphic1
CPGP	Canadian Potato Genome Project
CTAB	Cetyltrimethylammonium Bromide
dfr	Dihydroflavonol Reductase
DNA	Deoxyribonucleic Acid
dNTP	Deoxy Nucleotide Triphosphate
dsDNA	Double-Stranded DNA
EB	Enzymatic Browning
EDS	Energy Dispersive X-ray Spectrometry
EDTA	Ethylene Diamine Tetra-Acetic Acid
EST	Expressed Sequence Tag
Fe	Iron
Fe^{2+}- CGA	Ferrous-Chlorogenic Acid
ft	Ferritin
g (rcf)	Relative Centrifugal Force
g/L	Gram Per Liter
galLDH	L-Galactone-1,4-Lactone Dehydrogenase
GLM	General Linear Model
hqt	Hydroxycinnamoyl CoA Quinate Transferase

HRM	High-Resolution DNA Melting
ICAP	Inductively Coupled Argon Plasma
K	Potassium
$MgCl_2$	Magnesium Chloride
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization Time of Flight
MAS	Marker-Assisted Selection
min	Minute
mL	Milliliter
mM	Millimolar
MRD	Mean Raw Density
N	Nitrogen
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
ng	Nanogram
nm	Nanometer
P	Phosphorus
PCR	Polymerase Chain Reaction
PPO	Polyphenol Oxidase
PRC	Potato Research Centre
qRT-PCR	Quantitative Reverse Transcription-PCR
QTL	Quantitative Trait Loci
RAPD	Randomly Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RNAi	RNA Interference
rpm	Revolutions Per Minute
SAPP	Sodium Acid Pyrophosphate
sec	Second
SNP	Single Nucleotide Polymorphism
SolCAP	Solanaceae Coordinated Agricultural Project
ssDNA	Single-Stranded DNA

SSR	Simple Sequence Repeat
sss	Soluble Starch Synthase
Taq	*Thermophilus Aquaticus*
TDF	Transcript-Derived Fragments
µL	Microliter
µM	Micromolar
VP-SEM	Variable Pressure Scanning Electron Microscope

Acknowledgements

I would like to thank my supervisor, Dr. Gefu Wang-Pruski for giving me this great opportunity to do my Masters study at Nova Scotia Agricultural College. I really appreciate her guidance and help not only on my project, but also on my life. I also thank my co-supervisor, Dr. David De Koeyer for his supervision and encouragement throughout my whole project. I thank my committee member, Dr. Hossein Farid, for his constructive ideas and comments on my research.

I am grateful to Kathy Douglass, for her help and comment during my research, as well as Agnes Murphy of the Potato Research Centre, Agriculture and Agri-Food Canada, Fredericton, for providing the potato samples to my project. I am grateful to Dr. Andrew Schofield and Dr. Tudor Borza, for reading my writing and giving me valuable comments. I am grateful to all the people who supported and helped me during my study. Also, I would like to thank the funding from the NSERC grant awarded to Dr. Gefu Wang-Pruski and from AAFC grant to Dr. David De Koeyer.

Last but not least, I would like to thank my parents and my friends for their love, support and encouragement in my study.

Chapter 1. Introduction

After-cooking darkening (ACD), a gray-black discoloration of the potato tubers, is one of the most undesirable quality defects of potatoes. This phenomenon occurs after potato tubers are cooked using methods such as boiling, baking, frying, and dehydration. In the French fry industry, ACD has a direct impact on the quality and marketability of the products, because consumers prefer the fries with a light golden flesh colour to those with a gray-black colour. Since 1970s, the processing industries have used sodium acid pyrophosphate (SAPP) as an additive during processing to prevent ACD. However, this treatment not only increases the cost of processing, but also could create an environmental hazard if SAPP is not recovered from the wastewater. Therefore, reducing ACD in processing varieties is still considered an important task for the industries.

After-cooking darkening is known as a highly heritable trait (Wang-Pruski, 2007). Over ten years' study, Dr. G. Wang-Pruski at NSAC has demonstrated that genetic factors dominate the trait, whereas most environmental factors do not significantly affect ACD, except season and storage duration. Previous work done by Dr. Wang-Pruski in collaboration with Dr. De Koeyer at the Potato Research Centre, Agriculture and Agri-Food Canada (AAFC), generated a linkage map using 28 simple sequence repeat (SSR) markers and 115 single nucleotide polymorphism (SNP) markers in a F_1 diploid potato population (13610) segregating for ACD severity. They identified four major quantitative trait loci (QTL) that contribute to ACD in this mapping population. These four QTLs are located on chromosomes 2, 4, 6, and 10, together they explained 39.9% of the total variation of ACD.

However, the result from one mapping population cannot automatically be considered

to be applicable to other germplasm. For example, a marker may not be segregating in all populations, or a marker may not be linked to the desirable allele. In the other words, one population's genetic structure may affect the importance of the QTLs. Therefore, validation of the identified QTLs in other mapping populations is essential. The verification of significant markers associated with the major QTLs will not only validate the important QTLs but also help to gain a better understanding of the genetics controlling ACD. This research was, therefore, aimed at validating previously identified SNP markers associated with ACD in a second diploid population (14946) and selected clones of a tetraploid population (14945).

Chapter 2. Literature Review

2.1 The Potato

The cultivated potato (*Solanum tuberosum* L.) is the fourth most important food crop in the world. Its global production is exceeded only by rice, wheat, and maize (FAO, 2010). It is one of the most valuable food crops that are successfully cultivated in almost all geographic locations around the world. The net farm value of production per year is ranked fifth worldwide, behind rice, wheat, soybean, and maize (FAO, 2010). Because of its adaptability, this crop is potentially one of the best food sources for an increasing world population. In fact, shifting diets of developing countries increased global production of potatoes from 285 to 329 million tonnes from 1985 to 2009. China and India have become the top two producers, making up one third of the world's potatoes (FAO, 2010).

The potato belongs to the *Solanaceae* family, one of the largest families in the plant kingdom (Storey and Davies, 1992). This family contains many important agronomic species, such as eggplant (*S. melongena*), pepper (*Capsicum* spp.), and tomato (*S. lycopersicum*) (Hawkes, 1992). The potato originated in South America, mainly in the Andes of Peru and Bolivia regions. They were domesticated over 10,000 years ago in this region (Hawkes, 1988). The ploidy level of this plant includes diploids (2n = 2x =24), triploids (2n = 3x = 36), tetraploids (2n = 4x = 48), pentaploids (2n = 5x = 60), and hexaploids (2n = 6x = 72). *Solanum tuberosum*, the most commonly cultivated species, is tetraploid (Hawkes, 1994). According to Salaman (1949) and Glendinning (1983), the cultivated potato was introduced into Europe on two occasions, to Southern Spain in 1570s, and then to England in 1590s. It was then spread throughout Europe in the 17th century and to North America in the 18th century. Nowadays, it has become one of the most important

food crops in the world and is grown in more than 140 countries (FAO, 2010).

Canada is the 13th largest potato producer and second largest exporter of frozen potato products in the world. The production in Canada is around 4.5 million tonnes per year. Approximately 40% of Canadian potatoes are produced in Atlantic Canada, mostly in Prince Edward Island (26%) and New Brunswick (15%). Another 30% of the potatoes are grown in Western Canada, mostly in Manitoba (19%) and Alberta (14%) (STATCAN, 2011). Moreover, approximately 50% of the potatoes are utilized by the processing industry (FAO, 2007), with 85% of these potatoes used for French fries, while the others used for chips, dehydration and frozen products, chilled-peeled potatoes, and canned potatoes (Vreugdenhil and Bradshaw, 2007). The cultivars grown for processing are different than those destined for other uses. The main varieties used for French fries are Russet Burbank and Shepody, whereas Atlantic, Superior, and Norchip are mainly used for chipping.

2.1.1 The Potato Tuber

As a food crop, the edible portion of the potato plant is the tuber. A potato tuber is a modified underground stem and has all of the internal structures that are characteristic of a stem tissue (Stark and Love, 2003). The energy stored in the tuber is used for the propagation of the plant through vegetative reproduction during the following growing season (Thomson, 1987). A tuber includes tuber skin, cortex, vascular ring, perimedullar zone, eye bud, pith, and two ends (stem end and bud end). The stem end is where the tuber attaches to the plant, while the bud end is the apical bud of the tuber. The tuber skin is a corky layer of cells whose major role is to protect the tuber from losing water and from

infection (Talburt et al., 1987). Under the tuber skin are the cortex and vascular ring, which contain the vascular tissue. Next to the vascular ring is the perimedullar zone; this portion contains primarily starch. At the center of the tuber moving from the stem end to the bud end, is the pith, which contains more water and less starch than the perimedullar zone (Talburt et al., 1987).

Tubers sprout by developing shoots (stems) from eyes that grow upward to emerge from the soil (Cutter, 1992). Leaves and branch stems are developed from above-ground nodes along with emerged buds. The stolons are developed from underground buds at the base of the main stem. After vegetative growth, tuber initiation begins with an elongation of the stolons underground. Tuber initiation is considered the first visible swelling of the apical part of the modified underground stem. Booth (1963) stated that the initial swelling was a result of extensive cell division and a rapid increase in starch formation in the apical part of the stolon (Reeve et al., 1969). The period when tuber cells expand with the absorption of water, nutrients and carbohydrates, is also called tuber bulking (Cutter, 1992). In the above-ground part, flowering occurs before tuber bulking. The berry becomes mature after the abscission of the flower and true seeds are made (Cutter, 1992). After that, the vines turn yellow and lose leaves, and the photosynthesis of the plant decreases, which results in maximizing the tuber dry matter content and setting the tuber skins (Stark and Love, 2003). When the above-ground plant has died, it is the right time to harvest the tubers from underground.

In Canada, producing potatoes is a complex process. Seed tubers should be warmed to 10-15^{o}C for two weeks prior to planting. Normally, seed performance is affected by seed piece size. In most cases, planting seed pieces averaging 1.5 to 2.5 ounces in size will

provide optimum returns. But acceptability of this size range also depends on the number of eyes per seed tuber. The number of eyes per seed piece influences the number of stems per plant (Stark and Love, 2003). The planting should be conducted under proper conditions. Soil temperature should be above 7^{o}C for rapid emergence and to minimize seed piece decay; the fields should have adequate moisture before planting; and the planting depth should be 6 inches or deeper to reduce the amount of green tubers (Stark and Love, 2003). Other factors influencing the potato production include fertility, water availability, pest and disease management, and harvest management. In Canada, potato vines are top killed by chemical desiccant at least 10 days prior to harvest to mature the tubers and make harvesting easier. Top killing also reduces late-season virus spread into fields by aphid vectors. Growers normally dig tubers about 2 to 3 weeks after killing the vines. Harvesting at this time makes tubers less susceptible to shatter bruise and skinning damage (Stark and Love, 2003).

2.1.2 Potato Tuber Quality

Potato production has played an important role in countries with developed economies since there is an increasing market for fresh potatoes as well as processed potato products, such as chips and French fries (FAO, 2010). The consumer's requirements for these market associated qualities have significant influences on the growers and processing industries (Cutter, 1992). The important qualities of the potato tuber for the industry include yield, tuber size, specific gravity, starch content, after-cooking darkening, skin colour, fry colour, browning, dormancy, and disease resistance. However, all elements that contribute to the tuber qualities are affected by various factors. As a result, it is often difficult to produce a

particular type of variety that possesses all the required attributes (Cutter, 1992).

The potato tuber is of rich nutrition and energy (Denny, 1929). It is a good source of carbohydrates, proteins, essential vitamins, minerals, fiber, and antioxidants (Storey and Davies, 1992; Andre et al., 2007). The most common form of carbohydrates in the potato tuber is starch. It contributes about 65-80% of the dry weight of the tuber (Cutter, 1992). Other important tuber carbohydrates include simple sugars and dietary fiber (Talburt et al., 1987). The sugar content of potatoes may vary from only trace amounts to as much as 10% of the dry weight of the tuber (Barker, 1938). In fact, potato produces more protein per hectare than any other crop except soybean, and more dietary energy per hectare than any other crop except sugarcane (Lopez De Rommana et al., 1980). Vitamin C is one of the important antioxidants in potatoes, which is found in concentrations ranging 4 - 40 mg/kg in fresh weight (Fairweather-Tait, 1983). As reported by Augustin (1975), one 150 g potato (medium size) contains about 90% of the daily human requirement for vitamin C. Furthermore, K, P and Mg are the most abundant elements in tubers that are beneficial for human diet (Subar et al., 1998).

Besides the abundant nutrition and appearance characteristics of tubers, tuber quality defects, such as enzymatic browning (EB) and after-cooking darkening (ACD), are also important for customers and producers. The tuber can turn dark through either enzymatic or non-enzymatic processes. Either way is undesirable for fresh market or processing industries. Enzymatic browning is the process in which phenolic compounds are oxidized by the enzyme polyphenol oxidase (PPO) to quinones and the quinones transformed to brown pigments (Friedman, 1997). The phenolic compounds oxidized by PPO are chlorogenic acid and tyrosine. The levels of phenolic compounds and the enzyme PPO

determine the inheritance of susceptibility to the browning (Corsini et al., 1992). Research has shown that the development of EB was blocked by antisense inhibition of PPO (Bachem et al., 1994). However, no correlation was observed between the degree of EB and the amount of chlorogenic acid (Friedman, 1997).

2.2 After-Cooking Darkening

After-cooking darkening is one of the most undesirable quality defects of tubers, and is especially important in table and processing potato cultivars (Wang-Pruski et al., 2007). It is a non-enzymatic gray-black discoloration occurring in every potato growing area in the world. Although ACD does not affect the flavor or nutritional content in the processed tubers, the dark color of the potato product has a strong negative effect on the marketability (Wang-Pruski and Nowak, 2004). After-cooking darkening occurs when the tuber is exposed to air after boiling, baking, frying, or dehydration (Hughes and Swain, 1962a). During cooking, a colorless complex, ferrous-chlorogenic acid (Fe^{2+}- CGA) is formed in the tuber. It is then oxidized to the dark compound of ferri-dichlorogenic acid, when exposed to air (Figure 1). The degree of the darkness is developed to maximum in air 2 hour after cooking (Wang-Pruski, 2007). It ranges from light gray to almost black depending on the cultivar.

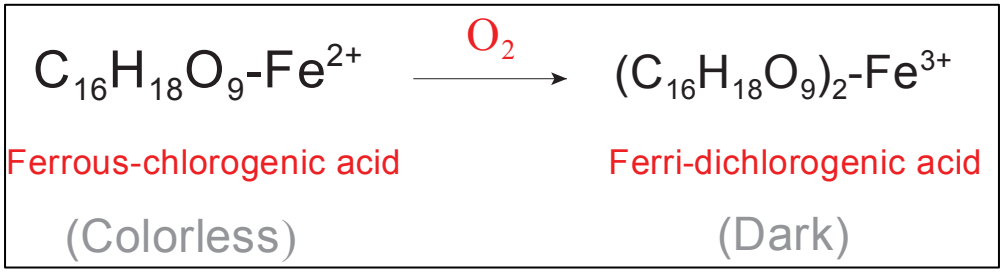


Figure 1. The chemical reaction of the dark color formation of ACD in potatoes (Wang-Pruski and Nowak, 2004).

The distribution of ACD in a tuber is not even. The degree of ACD is always the highest in the cortex tissue directly under the skin. This may occur because the cortex tissue contains as much as 50% of CGA in the tuber (Friedman, 1997). From the cortex towards the center of the tuber, the degree of the darkness gradually decreases with the inner pith showing the least discoloration. Tissue close to the stem end is always darker than that in the bud end (Swiniarski, 1968). This uneven distribution of ACD is correlated with the distribution of CGA content in these tissues (Hughes and Swain, 1962b; Wang-Pruski, 2006).

The degree of the darkness is influenced by the concentrations of chlorogenic acid (CGA), citric acid (CA), ascorbic acid (AA), and iron (Fe) in the tubers (Wang-Pruski and Nowak, 2004). Chlorogenic acid, one of the major secondary metabolites, belongs to a group of phenolic compounds (Ky et al., 1999). One of the key factors in the development of ACD is binding of Fe to chlorogenic acid. In potato tubers, 90% of the total phenolic compounds are CGA (Griffiths and Bain, 1997). Citric acid also binds to Fe to form a colorless compound in potatoes. It reduces the darkening by chelating Fe so that the Fe cannot react with CGA, thereby preventing the formation of the dark-coloured pigment (Mazza and Qi, 1991). Previous studies indicated that a higher ratio of CGA to CA in the tuber corresponds to a higher level of ACD in potato cultivars (Hughes and Swain, 1962b). Ascorbic acid binds to Fe in potato tubers resulting in a light purple compound (Shekhar et al., 1978). However, the complex formed by AA and Fe is not stable in tubers (Friedman, 1997). Therefore, the effects of CGA and CA are more significant than that of AA on ACD development.

Iron is the key factor to work with CGA, CA or AA in the formation of darkening.

Hughes and Swain (1962a) pointed out that cultivars with a higher concentration of Fe do produce tubers with higher ACD susceptibility. They also demonstrated that Fe concentration is not a determinant of ACD distribution in the tuber (Hughes and Swain 1962b). However, the study reported by Heisler et al. (1963) had a different outcome. They examined the concentration of Fe using autoradiography and found that Fe concentration was higher at the stem end than at the bud end. This distribution of Fe was correlated with the degree of ACD, which was higher in the stem end and lower in the bud end. Later, Swiniarski (1968) indicated that the concentrations of Fe in both ends are similar, even when stem end showed greater ACD.

Since previous work demonstrated contradictory results about the involvement of Fe in the development of ACD, the distributions of total Fe and Fe^{2+} in Shepody and Russet Burbank cultivars were investigated using inductively coupled argon plasma (ICAP) spectrometer and energy dispersive X-ray spectrometry (EDS) coupled with the variable pressure scanning electron microscope (VP-SEM) (LeRiche, 2007). Their findings demonstrated that the iron concentration, either total Fe or Fe^{2+}, in a tuber is not directly related to the severity of ACD. This result confirmed the finding reported by Hughes and Swain (1962b) that total iron concentrations vary little among tuber segments.

2.2.1 Genetic Control of ACD

After-cooking darkening is a highly heritable trait that is predominantly controlled by genetic factors (Dalianis et al., 1966; Killick, 1977). The first ACD inheritance study investigated the progeny in a cross of lower susceptible ACD tubers (white) and higher susceptible ACD tubers (dark) (Rieman et al., 1944). The result showed that most of the

progeny were white. This suggests that the genotype representing white is totally or partially dominant. Later, Dalianis et al. (1966) observed a wide variance of ACD amongst ten parental lines. They indicated that selection of parents for their freedom from ACD could be effective in a breeding program. After crossing three commercial cultivars to two parental genotypes and evaluating the degree of ACD for two years in the field, Pika et al. (1984) detected that the broad sense heritability of ACD ranges from 0.60 to 0.68, which means that 60% to 68% of ACD phenotypic variance is contributed by the genetic variance. Also, the narrow sense heritability ranges from 0.33 to 0.63, which means that 33% to 63% of ACD phenotypic variance is contributed by additive gene effects (Pika et al., 1984). When narrow sense heritability is high, it means that breeding a low ACD clone is effective by phenotypic selection (Fairbanks and Andersen, 1999). Moreover, ACD did not appear to be associated with any other tuber characteristics, such as specific gravity, maturity, texture, and yield (Killick, 1977).

In Atlantic Canada, the predominant cultivars used in French fry processing are Shepody and Russet Burbank. Although the ACD levels of both are lower than other cultivars, industry still needs to add more SAPP at later stages of the storage period in order to control elevated ACD. Therefore, breeding a new cultivar with lower ACD under longer storage times is still a challenge for breeders. Wang-Pruski et al. (2003) tested a total of 45 diploid hybrid clones for two years in the field to select parents for ACD breeding. They found that the degree of ACD remains stable from season to season in some of the diploid clones. Some of these clones were used as parents for generating mapping populations in subsequent studies (Wang-Pruski et al., 2003).

2.2.2 Environmental Factors Related to ACD

Although genetic control governs the degree of the darkening, environmental factors also affect the degree of ACD in potato tubers. The environmental conditions that influence the degree of ACD were originally believed to be soil pH, climate temperature, fertilizers, moisture content, and storage (Smith, 1987; Wang-Pruski and Nowak, 2004). Early studies focused more on soil pH, climate, and fertilization regimes (Smith and Nash, 1940; Kaldy and Lynch, 1983; Hughes and Evans, 1967; Hegeny and McPharlin, 2000). Later research identified that these factors are not as significant as season and storage duration (Wang-Pruski et al., 2003; Wang-Pruski et al., 2007). For example, Smith and Nash (1940) found that lower soil pH enhances CGA biosynthetic activity in tubers, which corresponds to higher susceptibility to ACD. Tubers grown under wet and cloudy conditions have higher susceptibility to ACD than that grown under dry and sunny locations (Kaldy and Lynch, 1983). Regarding fertilizer usage, a study by Hughes and Evans (1967) suggested that tubers grown in soils with low potassium (K) levels or low nitrogen (N) levels tend to have higher susceptibility to ACD. However, a later study claimed that an increased application of N resulted in an increased ACD, whereas high levels of K resulted in low levels of ACD (Hegeny and McPharlin, 2000). Most recently, Wang-Pruski et al. (2007) compared seven rates of N fertilization, five rates of K fertilization, and three rates of phosphorus (P) fertilization in two years using Russet Burbank and Shepody cultivars grown in Atlantic Canada. They demonstrated that both N and K reduced ACD in most of the trials tested. In these cases, N fertilization reduced ACD more significantly than K. In addition, no effect of P was found (Wang-Pruski et al., 2007). Wang-Pruski et al. also studied the storage effect on ACD for more than 10 years by using many diploid and

tetraploid samples in several soil types and locations. They concluded that regardless of clone type or growing conditions, the longer the tubers are stored, the darker they become (Wang-Pruski et al., 2003; Wang-Pruski et al., 2007).

2.2.3 Genes Associated with ACD

Since ACD appears to be a genetically dominant trait, searching for genes and gene markers to understand its control mechanisms would be an effective approach to develop a control strategy (Wang-Pruski, 2007). Several approaches to identify the responsible genes have been used in Dr. Wang-Pruski's lab since 1999, including candidate gene identification, gene cloning, gene expression profiling, genome wide gene identification using comparative proteomics, gene functional analysis, and QTL mapping. In order to search for genes related to ACD, the biochemical pathways related to CGA and iron metabolism were initially considered. The first gene studied was *c4h* encoding cinnamic acid 4-hydroxylase, an enzyme involved in early stage of the CGA biosynthesis. It catalyzes the biosynthesis of *p*-coumaric acid from the *t*-cinnamic acid (Cantle, 2004; Wang-Pruski et al., 2011). Besides the *c4h*, a gene *hqt*, encoding hydroxycinnamoyl CoA quinate transferase in the last step of CGA biosynthesis was found (Narayanan, 2005). Also, *ft* encoding for an iron storage protein ferritin was chosen. In the meantime, other candidate genes related to CA and AA biosyntheses became the targets as well. These include *cs*, a gene encoding citrate synthase, an enzyme involved in CA biosynthesis, *galLDH* for L-galactone-1,4-lactone dehydrogenase for AA biosynthesis, and *cop1* for constitutive photomorphic1, a gene involved in decreasing ACD during storage (Narayanan, 2005; Topley, 2004).

After the candidate genes were identified, gene cloning and gene expression analysis were completed using Northern hybridization, reverse transcription-PCR (RT-PCR) and real-time quantitative reverse transcription-PCR (qRT-PCR). The full-length of *c4h* gene has been sequenced in potato cultivar Russet Burbank (Genbank accession number: DQ341174) (Cantle and Wang-Pruski, 2005). The *c4h* gene expression level was confirmed to have a negative correlation with the degree of ACD in potatoes. This means that higher *c4h* gene expression is associated with a higher the degree of ACD (Wang-Pruski et al., 2005). Moreover, Narayanan (2005) found that *hqt* and *galLDH* gene expressions were significantly higher in ACD susceptible clones. These genes were suggested to have moderate correlation with ACD. The *cop1* gene showed consistently higher expression in the ACD resistant clones (Topley, 2004). However, the *cs* and *ft* gene expressions did not show significant differences among all the samples analyzed (Narayanan, 2005). This outcome also agrees with the conclusion that only CGA is controlled genetically, whereas, the levels of citric acid and iron depends primarily on the environmental factors and growth conditions (Heisler et al., 1963; Swiniarski, 1968).

Regarding the work of gene expression profiling, the cDNA-AFLP method was used to search for genes that are up- or down-regulated by storage duration (Topley, 2004). The gene expression at the mRNA level was measured based on the difference of the cDNA abundance. Over 20 transcript-derived fragments (TDFs) were identified (Topley, 2004). These TDFs showed gradually increased or decreased gene expression patterns during the storage duration. After the expressed sequence tag (EST) technology was available, two gene expression profiles were generated in the Canadian Potato Genome Project (CPGP) using Shepody tubers stored for one month and six months at 9^{o}C and 95% relative

humidity (Flinn et al., 2005). Results identified 34 ESTs whose gene expression levels were down-regulated after six months in storage and 104 ESTs whose gene expression levels were up-regulated after six months in storage (Wang-Pruski et al., 2006). Further investigation on their involvement in physiological activities is underway.

A comparative proteomics approach was also used to investigate genetic factors influencing ACD. Protein profiles from ACD-susceptible and ACD-resistant diploid clones in the breeding family 13610 were generated using the LC-MS/MS system (Murphy et al., 2010). In this study, a total of 109 proteins were identified. Within these identified proteins, 38 proteins showed greater abundance in the ACD resistant samples; 71 proteins showed greater abundance in the ACD susceptible samples. Functional annotation analysis of these 109 proteins revealed that most of these proteins are related to wounding stress, storage and defense responses and protease inhibition (Murphy et al., 2010). Real-time qRT-PCR was used to confirm that the ACD control mechanism may be closely related to enzymatic browning, starch metabolism, stress and wounding responses and defense functions in the plant (Murphy et al., 2010). All the above genes play significant roles in guiding cultivar selection processes for ACD in the near future.

Several mapping populations have been established for QTL mapping on ACD using selected diploid hybrid clones from a total of 45 tested diploid hybrid clones (Wang-Pruski et al., 2003). A total of 22 breeding populations using selected diploid parents were generated and three families were evaluated in over 6 growing seasons (Wang-Pruski et al., 2003). These families include 13610 (10908-06 x 11059-01), 14946 (10908-06 x 12115-07), and 14945 (09401-03 x CH72-03) (Wang-Pruski et al., 2003; Sun et al., 2003). Populations were grown from true seeds in the greenhouse to generate seed tubers in both

Fredericton, NB and at NSAC, and these tubers were then used for field trials in Benton Ridge, NB and Truro, NS locations. A genetic map was generated and QTL analysis resulted in the identification of four major QTLs on the chromosomes 2, 4, 6, and 10 in the diploid population 13610 (Wang-Pruski et al., manuscript in preparation). In the mean time, Bradshaw et al. (2008) generated a tetraploid potato population by crossing a processing clone 12601ab1 and table cultivar Stirling. They reported that six QTLs were detected for ACD. Three QTLs have single alleles on chromosomes 2, 4, 6 and C group of 12601ab1, whereas two QTLs were located on homology group 6 and Xlb of Stirling (Bradshaw et al., 2008).

2.3 DNA Markers

DNA markers, which reveal sites of variation in DNA, essentially detect mutations, insertions, deletions or inversions in allelic DNA fragments (Jones et al., 1997). They are widely used for crop improvement in rice (Mackill et al., 1999), wheat (Eagles et al., 2001), maize (Stuber et al., 1999), and many other species (Collard et al., 2005). In potatoes, various types of DNA markers have been used for genetic studies (Brigneti et al., 1997; Meksem et al., 1995; Ronning et al., 1999; Milbourne et al., 1998; Sattarzadeh et al, 2006). In the previous study, simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers were used for generating a genetic map of potatoes (Wang-Pruski et al., manuscript in preparation).

2.3.1 Types of DNA Markers

2.3.1.1 Length Polymorphism

A simple sequence repeat (SSR) is a length polymorphism marker. It is also known as a microsatellite marker, which is a tandemly repeated motif of 1-6 bases (< 100 bp long) (Zane et al., 2002). SSR polymorphisms are detected by PCR amplification with specific primers and subsequent DNA size sieving using ploy acrylamide gels (Feingold et al., 2005). The PCR products from different genotypes frequently show length polymorphisms due to allelic variation of the number of repeat motifs in the microsatellite (Milbourne et al., 1998). Since the motif length is scored, microsatellites are co-dominant, as heterozygotes showing two different lengths and hence can be distinguished from homozygotes. The disadvantages of SSR involve a high cost in time and labour for the discovery of the SSRs and production of the primers (Milbourne et al., 1998).

2.3.1.2 Single Nucleotide Polymorphism

A single nucleotide polymorphism (SNP) is a DNA sequence variation occurring at a single nucleotide position in the genome (Brookes, 1999). SNPs occur at a high frequency in genomes. For example, on average, every 1,000 bp of the human genome contains one SNP (Sachidanandam et al., 2001); every 60-120 bp of the maize genome contains one SNP (Ching et al., 2002; Tenaillon et al., 2001); and in the potato genome, a SNP is present every 15-21 bp (Vreugdenhil and Bradshaw, 2007). The traditional method for detecting SNPs is through DNA sequencing. The process includes PCR amplification and subsequent sequencing of the PCR products. Automated sequencing detects each of the four deoxynucleotides A, T, G, and C with different fluorescent dyes. The four different

nucleotides are represented by four different colours. The deoxynucleotide sequence is recorded as an electropherogram with different nucleotide peaks. The SNPs are detected as multiple coloured nucleotide peaks at a given position in sequence trace files (Vreugdenhil and Bradshaw, 2007).

Older marker systems mainly have SNPs underlying them, such as restriction fragment length polymorphism (RFLP) (Botstein et al., 1980), amplified fragment length polymorphism (AFLP) (Meyer et al., 1998), and randomly amplified polymorphic DNA (RAPD) (Williams et al., 1990). However, their polymorphisms could be due to length differences, too. RFLP is one of the earliest methods to measure DNA variation.
The polymorphism is present at the specific restriction site where the DNA is digested by a restriction enzyme. A mutation within this site results in the loss or gain of a recognition site. When the digested DNA is run on a gel, the restriction fragments are separated out according to different lengths (Beckmann and Soller, 1983). This type of markers proved useful as the first generation of marker technology in potatoes because they are co-dominant, which means that heterozygotes and homozygotes can be distinguished. However, this technique takes a long time for completion and requires a large amount of high quality DNA (Bonierbale et al., 1988).

An amplified fragment length polymorphism (AFLP) is a marker technology based on the polymerase chain reaction (PCR) (Meyer et al., 1998). This technique uses restriction enzymes to digest genomic DNA, followed by ligation of complementary double-stranded adaptors to the ends of the restriction fragments. A subset of the restriction fragments are then amplified using two primers complementary to the adaptor and restriction site fragments. The amplified restriction fragments are radioactive or fluorescent labeled and

size separated by electrophoresis on high-resolution polyacrylamide gels. DNA polymorphisms between different genotypes come from mutations in the recognition sites or the nucleotide extensions. The differences result in either allowing or preventing PCR amplification (Van Eck et al., 1995). AFLP markers were well suited to quickly generate a large number of segregating DNA markers for construction of a dense linkage map for potatoes (Rouppe van der Voort et al., 1997).

Randomly amplified polymorphic DNA (RAPD) markers are also PCR based. The sequence polymorphisms are detected by using random short sequences (9 or 10 bp) as primers. A successful PCR reaction requires DNA sequences complementary to the primer binding sites. If the primer binding sites are missing or are too far apart, the PCR reaction fails and no fragments are generated for that region. Successful fragments will amplify with random primers (Williams et al., 1990). The advantages of RAPDs include: 1) no knowledge of DNA sequence of the target gene is required; 2) a particular random primer can reveal several loci at once (Williams et al., 1990). However, they are dominant, as heterozygotes and homozygotes have the same PCR products. In addition, they also have reproducibility problems. So they are not commonly used for potato mapping studies (Razdan and Mattoo, 2005).

2.3.2 Methods for SNP Genotyping

Many methods have been developed for SNP genotyping. All these methods are a combination of distinct methods for allele discrimination and signal detection. Allele discrimination is a method for determining the type of base present at a given SNP locus, whereas signal detection is a method for reporting the presence of the allele(s) (Twyman,

2005). In potatoes, SNPs have been genotyped in tetraploids using pyrosequencing (Rickert et al., 2002), iron-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry (Oberacher et al., 2004), and allele-specific oligonucleotide microarrays (Rickert et al., 2005). Pyrosequencing is a novel method for sequencing short stretches of DNA based on the detection of pyrophosphate, a normal by-product of DNA synthesis (Ronaghi et al., 1996). The analysis of DNA by mass spectrometry requires soft ionization and is usually achieved by matrix-assisted laser desorption/ionization time of flight analysis. This method can accurately discriminate between alternative alleles in DNA fragments of 3-20 nucleotides in length (Twyman, 2005). Allele-specific oligonucleotide microarray is a method of direct fluorescence detection. This approach is to use an intercalating fluorescent dye and to study the melting profile of DNA duplexes (Howell et al., 1999). However, none of these genotyping systems is suitable for a potato breeding program (De Koeyer et al., 2010). The ideal system for mapping tetraploid potatoes would be easy to use, robust, and affordable; however, it should also be dosage-sensitive and able to distinguish various haplotypes combinations (De Koeyer et al., 2010). Currently, a technology, called high resolution DNA melting (HRM), is being used to identify SNP markers rapidly without sequencing the PCR products (Lehmensiek et al., 2008).

2.3.2.1 High-Resolution DNA Melting

High-resolution DNA melting (HRM) is a technology that is useful for genotyping, mutation scanning and sequence matching (Reed et al., 2007). This technique involves a standard PCR reaction performed in the presence of a saturating fluorescent DNA dye (e.g.,

LCGreen Plus). The dye, LCGreen Plus, cannot bind single-stranded DNA (ssDNA) but will bind double-stranded DNA (dsDNA) without inhibiting the PCR reaction (Montgomery et al., 2007; Lehmensiek et al., 2008). After the PCR is complete, the temperature is slowly increased and the fluorescence signal decreases as the PCR product changes from dsDNA to ssDNA (Lehmensiek et al., 2008). A Lightscanner (Idaho Technology Inc.) or other instrument is used for capturing the fluorescent intensity (Figure 2). SNPs are readily distinguishable because the melting temperature of a DNA fragment with a homozygous base pair is higher than the melting temperature of a DNA fragment containing a heterozygous base pair (Lehmensiek et al., 2008). Based on these different melting temperatures, it is easy to distinguish homozygotes from heterozygotes by their melting curves (Figure 3).

The advantages of HRM over other marker techniques is that it is less time-consuming, it does not require electrophoresis separation, it has a low cost, and is highly sensitive (Zhou et al., 2005; Reed et al., 2007; Mackay et al., 2008). HRM enables scanning for mutations and genotyping in a few minutes after the PCR amplification. It provides rapid, cost-effective analysis without reagent additions or electrophoresis separations. When a single unlabeled probe is added to the PCR, HRM can be used to detect multiple alleles (Reed et al., 2007). In this case, HRM can be used for quantifying allele dosage and for variant scanning in tetraploid potatoes (De Koeyer et al., 2010).

2.3.2.2 Two Approaches in HRM

There are two common approaches that employ HRM for genotyping. The first is amplicon genotyping. Amplicon genotyping involves PCR amplification of specific

genomic regions with primers followed by HRM analysis of the amplicon. This approach is commonly applied in diploid genotyping. Heterozygote or homozygote differences in the PCR products are detected by differences in the melting curves (Figure 3). The melting curves can be generated by computer software in the scanner using normalized fluorescence curves, temperature-shifted melting curves, and difference plots, all of which are used to compare with the baseline (Mackay et al., 2008; De Koeyer et al., 2010). Amplicon genotyping has been used to genotype plants, especially for diploids such as grape (*Vitis spp.*) (Mackay et al., 2008), barley (*Hordeum vulgare* L.) (Lehmensiek et al., 2008), and potato (De Koeyer et al., 2010).

The second HRM genotyping method is unlabelled probe genotyping. It has good potential for genotyping tetraploids. The unlabelled probe is a single-stranded oligonucleotide fragment designed to complement the target SNP in either the sense strand or the antisense strand; as well, the probe is blocked at the 3' end in order to prevent the DNA polymerase from extending during the PCR reaction (Zhou et al., 2005). In the case of using a probe that complements the sense strand, the method is as follows. An asymmetric PCR is employed with excess amount of the forward primer and limited amount of the reverse primer, so that limited amplicon products are produced (Figure 4). The excess forward primer produces the single strands of the sense sequence. The excess products of the sense sequences are annealed to the probe sequences (Zhou et al., 2005). If there is a polymorphism (mismatch) between the sense strand and the probe, it will result in an HRM scan with a distinct melting temperature curve, because the melting temperature of a perfectly matched genotype is higher than that of a mismatched one (De Koeyer et al., 2010). In the case of using a probe that complements the antisense strand, the method is

identical except that the reverse primer is in excess.

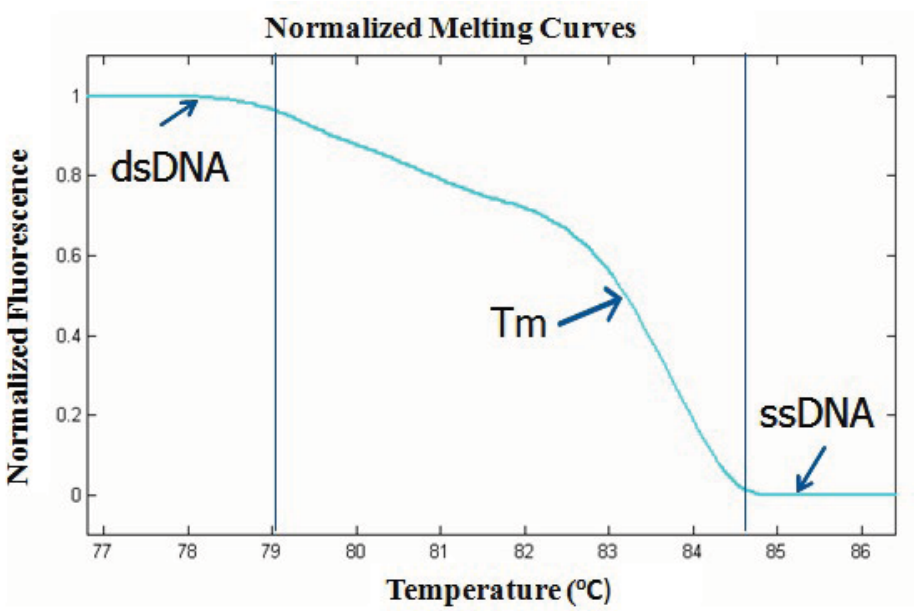


Figure 2. Melting profile of the PCR product. The fluorescence intensity decreases with increasing temperature as double-stranded DNA (dsDNA) melts to single-stranded DNA (ssDNA). Tm: melting temperature.

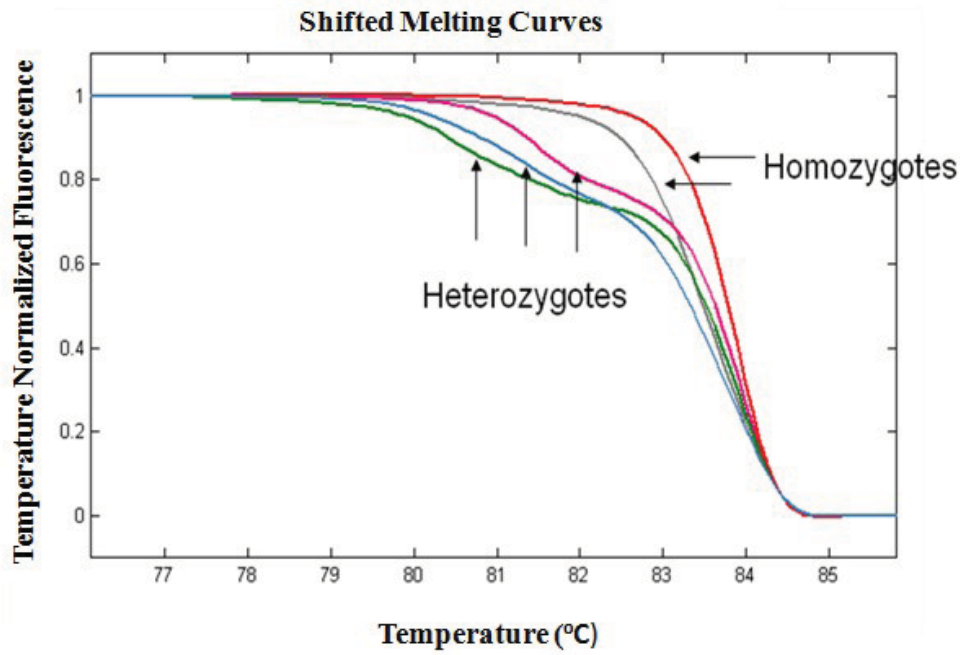


Figure 3. Melting temperature shift indicated by the presence of mismatched bases at the SNP loci. Heterozygous PCR products have lower Tm's than homozygous PCR products.

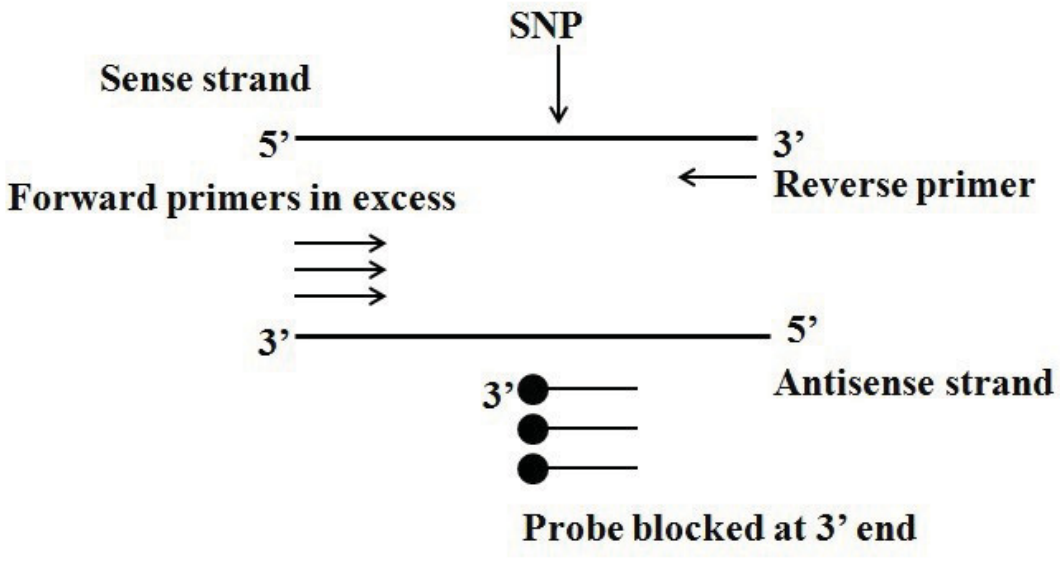


Figure 4. Asymmetric PCR performed in the presence of the unlabelled probe. The excess primer (forward) binds to the lower strand (antisense) to produce excess copies of the upper strand (sense). The probe is blocked on 3' end to prevent extension. The limiting primer (reverse) is complimentary to the upper strand (sense) resulting in limited copies of lower strand (antisense). This scheme allows for increased binding of the probe to its complimentary strand, which is the upper strand (sense).

2.3.3 Applications of Markers

DNA markers can be used for many different applications. These applications include early generation selection, detection of recessive genes, genotyping in the population, and screening of quantitative characters (Razdan and Mattoo, 2005). The most common uses of DNA markers for improving potato breeding are generation of genetics maps for QTL identification and marker-assisted selection (MAS).

2.3.3.1 Quantitative Trait Loci

A QTL is the genomic region on a chromosome that is associated with a quantitative trait (Falconer and Mackay, 1996). The identification of QTLs was initiated by the development of DNA markers in the 1980s (Collard et al., 2005). QTL mapping is the process of locating genetic markers that are associated with target traits on the linkage maps to find the location of QTLs in the genome (Klug et al., 2006). The construction of a QTL map first requires a segregating population (Collard et al., 2005). In order to generate genetic variation, breeders make crosses between parents with complementary traits. Usually, two selected parents should be different at one or more traits of interest (Bradshaw et al., 2008). In addition, accurate quantification of a trait is the key to success. In the previous QTL study for ACD, the diploid F_1 segregating population 13610 was generated by crossing a low ACD hybrid male clone and a high ACD hybrid female clone (Wang-Pruski et al., 2003). The second step is to identify DNA markers that reveal the polymorphisms between the parents. After the markers are confirmed in the parents, they must be screened across the entire population. This is known as marker genotyping of the mapping population (Collard et al., 2005). A total of 28 SSRs and 115 SNPs markers were used to generate a linkage map for the 13610 population. These 28 mapped SSRs markers came from a total of 111 SSRs tested, whereas the 115 SNPs markers came from 168 SNPs tested. The 111 SSRs were selected from over 200 SSRs based on the published data. The 168 SNP markers were selected from over 200 ESTs published in the Canadian Potato Genome Project (CPGP) (http://www.cpgp.ca) database and generated using candidate gene information. Fourteen significant SNP markers were identified in the diploid population 13610. They are located on eight of the 12 chromosomal regions (Figure 5). For

example, two SNP markers representing the genes *dfr* encoding for dihydroflavonol reductase (LS030) and *sss* encoding for soluble starch synthase (LS124) on chromosome 2 are shown in Figure 6.

Generally, QTL analysis is based on the principle of detecting an association between a phenotype and the genotype of the markers. Markers are used to distinguish the mapping population into different genotypic groups and to determine whether significant differences occur in these groups (Young, 1996). The significant differences between mean trait values indicate linkage between the marker and the QTL. The closer a marker is to a QTL, the lower the chance of recombination occurring between the marker and the QTL (Collard et al., 2005). In the other words, the more significant value obtained for differences of trait means the higher chance of inheriting marker and QTL together in the progeny. It also means that the marker is more associated with the trait. Oppositely, if there is no significant difference between the means of the genotype groups, it means the marker is unlinked to a QTL.

2.3.3.2 Marker-Assisted Selection

Marker-assisted selection (MAS) is a process where markers are used for selecting a trait of interest (Collard et al., 2005). Recently, MAS is emerged as a useful tool in potato breeding programs because of the advent of molecular markers and linkage maps developed in potatoes (D'hoop, 2009). The markers in MAS should be easy to use, robust, cost effective and link to the trait of interest (De Koeyer et al., 2010). The application of MAS will have strong impact on improving potato breeding. For example, the identified markers can be used to screen breeding lines, select parents, and follow multiple traits

simultaneously during early generations (Barone, 2004). HRM has become a technique used for MAS in potatoes (De Koeyer et al., 2010) and lettuce (Simko et al., 2009). De Koeyer et al. (2010) demonstrated the utility of HRM for genotyping and polymorphism detection in diploid and tetraploid potatoes. Examples of HRM assays developed for MAS include potato tuber skin colour, tuber flesh colour, tuber quality, and resistance to the golden nematode (De Koeyer et al., 2010).

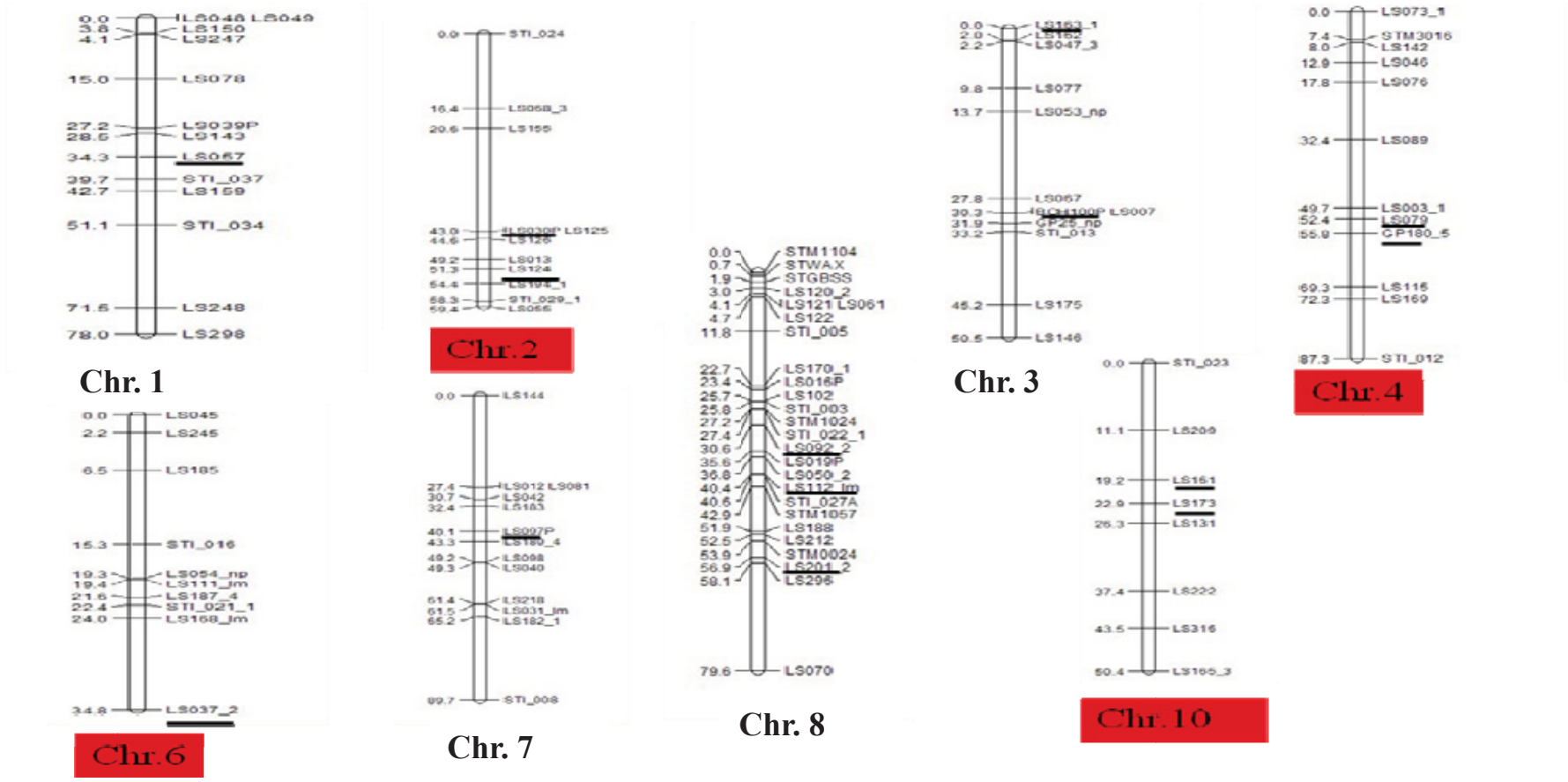


Figure 5. The locations in the potato genome of 14 SNP markers associated with ACD. The 14 SNP markers (underlined) used in this study are distributed on eight chromosomes (1, 2, 3, 4, 6, 7, 8, and 10) out of the 12 chromosomes in the potato genome (Unpublished data generated by Dr. De Koeyer and Dr. Wang-Pruski using 13610). Other markers labeled on the eight chromosomes were used for establishing the linkage maps in these chromosomes (by De Koeyer and Wang-Pruski, manuscript in preparation). The four highlighted chromosomes contained the QTL regions on chromosomes 2, 4, 6, and 10.

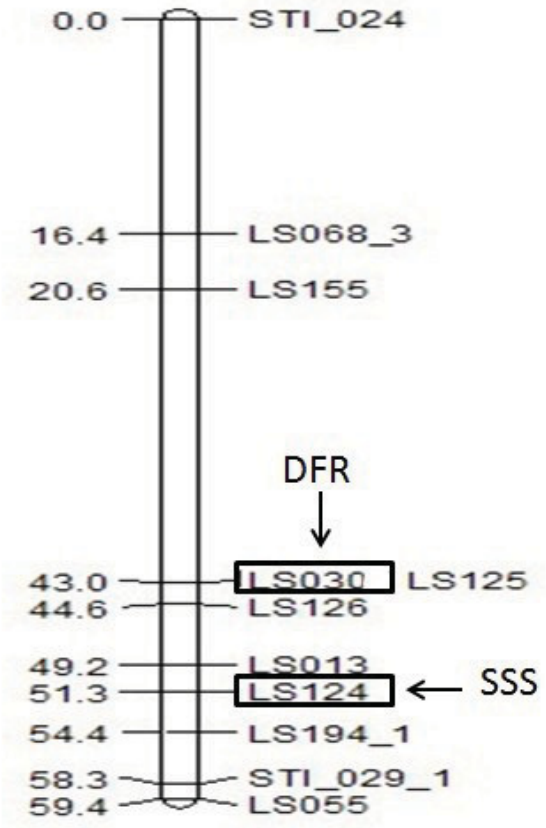


Figure 6. Details of the markers mapped on chromosome 2. Two SNP markers coded as LS030 and LS124 showing the significant relationships to ACD were mapped on the chromosome 2 of the potato genome using the population 13610.

Chapter 3. Hypotheses and Objectives

The first hypothesis of this study was that the identified SNP markers associated with ACD in one diploid mapping population would show the same association in another diploid mapping population. Therefore, the main objective of this research was to validate the previously identified 14 SNP markers associated with ACD from the population 13610 in the second diploid population 14946. Amplicon genotyping was achieved by using HRM.

The second hypothesis of this study was that the HRM methodology could also be employed in tetraploid mapping. The objective was to validate selected SNP markers associated with ACD in eight clones of the tetraploid population 14945. Unlabelled probe genotyping was achieved by using HRM.

Chapter 4. Materials and Methods

4.1 Potato Samples

The plant materials used in this study consisted of a diploid population named as 14946 and a tetraploid population named as 14945. The parents of both populations were diploid hybrid clones. These parental clones were derived by crossing the cultivated South American diploid species *S. stenotomum* and wild diploid potato species with haploids of *S. tuberosum* at the AAFC, Potato Research Centre (PRC), Fredericton, Canada (Sun et al., 2003).

A diploid hybrid female cultivar 10908-06 which had a higher degree of ACD and a diploid hybrid male cultivar 12115-07 which had a moderately lower degree of ACD were bred to generate the diploid population 14946. The diploid population consisted of 85 diploid hybrid progeny. This population shared one of the parents, 10908-06, with the previous diploid mapping population 13610 (Figure 7). The population 14945 was derived from crossing a diploid female cultivar 09401-03 which had a higher degree of ACD and a diploid male clone CH72-03 which had a lower degree of ACD. Because both parents produce un-reduced games, this cross generated a group of tetraploid progeny instead of diploid clones. Because of this meiotic abnormality, the population was not kept for further research. Eventually, only four higher ACD progeny clones and four lower ACD progeny clones were maintained and used in this study.

The breeding lines of 13610, 14946, and 14945 were generated by Dr. Henry De Jong in collaboration with Dr. Wang-Pruski at PRC in 2002. True seeds from the berries from each hybridization were harvested in 2002. They were planted in the growth chamber in 2003 to generate mini tubers. These mini tubers were further grown in a greenhouse to

generate seed tubers for field planting at PRC, Benton Ridge breeding station and at NSAC since 2004.

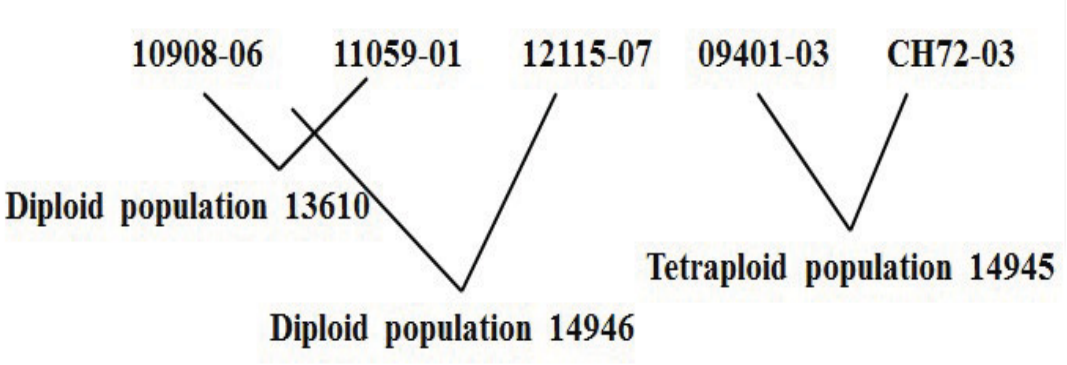


Figure 7. The origin of the potato populations studied. Diploid populations 13610 and 14946 share one of the parents (10908-06). The tetraploid population 14945 is originated from two hybrid diploid parental clones.

4.1.1 Field Trial and Storage

The field tuber samples of each population were planted during late May to early June and harvested in late September to late October each year at both Benton Ridge potato breeding station, AAFC, Benton, New Brunswick and at NSAC research field in Brookside, Nova Scotia locations every year from 2004 to 2009. The newly harvested tubers were stored at 15^{o}C with 95% relative humidity. The temperature decreased 1^{o}C every two weeks until the temperature reached 9^{o}C in early or mid-December. The tubers were stored until evaluated for the degree of ACD in January and February of the following year.

4.1.2 Evaluation of ACD

The field tubers of both populations were evaluated for ACD from 2005 to 2009. ACD evaluation was performed using the digital imaging method (Wang-Pruski, 2006). The

evaluation processes included tuber cooking, image acquisition and ACD measurement. Four stored tubers of each clone were cooked by steaming for 20 to 30 min depending on size. The cooked tubers were cooled down to room temperature and sliced into two halves vertically and exposed to air for one hour to allow the darkening to develop. Four tuber halves were imaged at the same time. The digital images of the cooked tuber surfaces were captured in the dark chamber of the UVP Chemi Imaging System (UVP Inc., Upland, CA, USA) and the level of the darkening of each clone was digitally evaluated by using LabWorks™ Imaging Analysis and Acquisition software (UVP Inc., Upland, CA, USA). The degree of ACD was measured between 0.001 and 255.999 pixel levels with 0.001 as absolute black and 255.999 as absolute white. Mean raw density (MRD) was used to express the level of ACD, which indicated the average pixel density of the selected area. Consequently, a higher MRD value indicated a lower degree of ACD, whereas a lower MRD value indicated a higher degree of ACD.

The ACD data from the diploid population 14946 were collected from 2005 to 2009 in six trials, five from the Benton Ridge, New Brunswick location and one from the Brookside, Nova Scotia location. They were named as 2005 Benton (2005 B), 2006 Benton (2006 B), 2007 Benton (2007 B), 2008 Benton (2008 B), 2008 Truro (2008 T), and 2009 Benton (2009 B). The ACD data of the tetraploid population 14945 were collected from 2005 to 2009 in five trials, all in the Benton Ridge, New Brunswick location. They were named as 2005 Benton, 2006 Benton, 2007 Benton, 2008 Benton, and 2009 Benton. Benton represents Benton Ridge potato breeding station, AAFC, in New Brunswick; and Truro represents Brookside research field of NSAC, in Nova Scotia.

4.2 Genomic DNA Isolation

Genomic DNA from each clone was extracted and used as template for PCR amplification and further genotyping. Seed tubers of the four parents and all the clones of the diploid population 14946 and the tetraploid population 14945 were planted in 6" pots with Pro-Mix in the growth chamber at NSAC in May 2008. After emergence, the plants were fertilized with 10-52-10 (N-P-K), water-soluble fertilizer (Plant-prod, Brampton, Ontario, Canada) at the rate of 1.5 g/L at every second week for six weeks until the leaf quality was good enough for DNA extraction. The growth chamber was set at 22^{o}C for 16 hours in the daytime and at 16^{o}C for 8 hours in the night time with 75% relatively humidity. After six weeks growth, fresh leaf materials of each clone were collected and stored at -80^{o}C freezer for genomic DNA extraction.

The protocol of DNA isolation from leaf issues was based on Doyle et al. (1990) with minor modifications. One gram leaf tissue was ground using a mortar and a pestle along with sand quartz and 5 mL of pre-warmed CTAB buffer (55 mM cetyltrimethylammonium bromide, 1.4 M NaCl, 0.02 M EDTA, pH 8.0, 0.1 M Tris-HCl, pH 8.0, 10 µL of β-mercaptoethanol, 60^{o}C). The ground mixture was transferred to a 15 mL tube and incubated in a 60^{o}C water bath for 30 min with intermittent shaking at every 5 min during the incubation. The mixture was then added with an equal volume of chloroform and shaken vigorously, and then centrifuged at 7,000 g for 10 min. The supernatant was collected and chloroform extraction was repeated 2 to 3 times until the interphase was no longer visible. A 0.6 volume of isopropanol was added to precipitate the genomic DNA for 10 min at room temperature. The sample was spun at 7,000 g for 10 min and then the pellet was washed with 70% ethanol and air dried for 5 min. The dried pellet was dissolved in 1

mL of TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, pH 8.0) containing 10 µL of RNaseA (10 mg/mL) and incubated for 30 min at 37^{o}C. After that, a 0.25 volume of 10 M NH_4Ac (pH 7.3) and 0.6 volume of isopropanol were added and slowly inverted and kept it at -20^{o}C for 30 min. The DNA was centrifuged at 8,000 g for 10 min, and then the pellet was washed with 70% ethanol again. The DNA was suspended in 1 mL of TE buffer and stored at -20^{o}C.

4.3 Candidate SNP Markers and Primers

A total of 14 SNP markers were selected based on their significant relationships to ACD in the diploid mapping population 13610 (Table 1). They were LS057, LS030, LS124, LS163, BCH100, LS079, GP180, LS037, LS097, LS092, LS112, LS201, LS161, and LS173. The markers were located on chromosomes 1, 2, 3, 4, 6, 7, 8, and 10. Seven of the 14 SNP markers came from four major ACD QTL regions that had been identified on chromosomes 2 (LS030 and LS124), 4 (LS079 and GP180), 6 (LS037) and 10 (LS161 and LS173) in the population 13610. The other seven markers were chosen because of their significance levels ($P<0.05$). The 14 SNP markers and their primers information are listed in Table1. Fourteen pairs of primers were designed by OMP software (De Koeyer et al., 2010) and used for amplifying the PCR product of each marker (Table 1). The target melting temperature (Tm) of each primer was 60^{o}C and a PCR product size between 100 bp and 400 bp. The marker sequences are provided in the Appendix I.

Table 1. Description of the 14 SNP markers associated with ACD in population 13610. Given in the table are marker name, significant relationship to ACD, accession number from the database which the sequence was retrieved, marker in the protein name corresponding to the gene, the chromosomal location, and nucleotide sequences of the forward and reverse primers of each marker.

Marker name	P-value	Accession number	Database	Protein name
LS057	<0.005	05Mar10.CPGP.5947.C1	CPGP[a]	protein phospatase 2A catalytic subunit
LS030	<0.1	AY289923 and AY289924	NCBI[b]	dihydroflavonol 4-reductase
LS124	<0.0001	05Mar10.CPGP.3773.C1	CPGP	soluble starch synthase
LS163	<0.01	05Mar10.CPGP.2730.C1	CPGP	unnamed protein product
BCH100	<0.1	05Mar10.CPGP.4697.C1	CPGP	beta-carotene hydroxylase
LS079	<0.05	05Mar10.CPGP.615.C1	CPGP	Patatin-like protein 1
GP180	<0.0001	05Mar10.CPGP.1573.C1	CPGP	NRC1(Solanum lycopersicum)
LS037	<0.0005	05Mar10.CPGP.3329.C3	CPGP	cinnamic acid 4-hydroxylase
LS097	<0.005	05Mar10.CPGP.1119.C1	CPGP	pescadillo-like protein
LS092	<0.01	05Mar10.CPGP.2276.C1	CPGP	dehydration-responsive family protein
LS112	<0.05	05Mar10.CPGP.1752.C2	CPGP	Polyphenol oxidase
LS201	<0.005	05Mar10.CPGP.1088.C1	CPGP	expressed protein
LS161	<0.01	05Mar10.CPGP.6527.C1	CPGP	F-box family protein
LS173	<0.0005	05Mar10.CPGP.8270.C1	CPGP	hypothetical protein

a. Canadian Potato Genome Project (http://www.cpgp.ca/);
b. Marker LS030 is given by using accession numbers from NCBI (http://www.ncbi.nlm.nih.gov/).

Table 1. (Cont'd)

Chromosomal location	Forward primer	Reverse primer
1	5'-CTCTATGTGATCAGGCGAGG-3'	5'-CAACATAATCGCCCATGAAG-3'
2	5'-AGGAGTATTTCATGTGGCTAC-3'	5'-GCTTTAGCACATGATTCTATGATAC-3'
2	5'-CTCCTGATCCTAGGGTACAAA-3'	5'-AGGTGAGAAAGTGGCTCG-3'
3	5'-CACCATTCCTCGAATCATCA-3'	5'-AAGCCTGTTAAACAAACGTC-3'
3	5'-AGTTCTTGTTCTTCTCTCCGTTA-3'	5'-AGTAAACCTCTCCGATTTCTTC-3'
4	5'-GGAGAGGATGCTAGGCTC-3'	5'-CTTAGGACCATGTTCAAGATAGAA-3'
4	5'-TCCACCCCTGGATTCAAGAA-3'	5'-TCAAACTCCTTCACAAAGCA-3'
6	5’-CATAAGCTGCAATCTTTTCCTC-3’	5’-TTTTACAGGTAAGGGTCAAGA-3’
7	5’-TCTCAAAGAAGAACCGGGAT-3’	5’-GGTTTACCGGTATCTTTTCTCTC-3’
8	5’-TATGCATCTGCTCCTCTAGGG-3’	5’-CTGGATTTTGAGTCACACCATAATT-3’
8	5’-CCAATAATAACGGTGACCAAAAC-3’	5’-GCAACAACTGTACGGCACCA-3’
8	5’-CGGTCGTGAATCCATTGGTA-3’	5’-GCAAAACCTCCAACTCATTC-3’
10	5'-GTTGAACATGATGATCTAAAGAGGT-3'	5'-TCAGCTAGTTTCTTCCGGC-3'
10	5'-ATGTTCGTCTACAGTCAAGATAG-3'	5'-GAACATTGAGTTGAAGAGAAGTTC-3'

4.4 Amplicon Genotyping in the Diploid Population

4.4.1 Amplification of the Markers Using PCR

The primer pairs of the 14 SNP markers were used to amplify the DNA fragments using the genomic DNA of the 85 clones and their parents. All PCR reactions were carried out using PTC 200 MJ Research thermocycler with hard shell 96-well plates (Bio-Rad Cat. No. HSP9665). The position of each sample on a 96-well plate is shown in Table 2. Each plate was used for one SNP marker. Each marker was tested three times in three individual plates, which were considered three replications. Each PCR reaction included 1 μL of 10 ng/μL template genomic DNA, 1 μL of 10 x PCR buffer, 0.2 μL of 10 mM dNTPs, 0.2 μL for each 10 μM specific forward and reverse primers, 0.1 μL of 5U/μL Taq DNA polymerase, 1 μL of 10 x LCGreen Plus (Bio-Chem), 0.95 μL of 25 mM $MgCl_2$ and the total volume was made to 10 μL. For a 10 μL PCR reaction, one drop (15–20 μL) of mineral oil was added on to the top of PCR mix in order to avoid evaporation of the samples during HRM. After that, the plates were covered with adhesive film (Sarstedt Cat. No.95.1994). A thermal cycling for diploid samples was conducted with 40 cycles. The conditions were as follows: incubation at 94^{o}C for 2 min, 40 cycles and each cycle contained 60 sec at 94^{o}C, 60 sec at 60^{o}C, 60 sec at 72^{o}C, and final extension for 5 min at 72^{o}C. The PCR samples were stored at 4^{o}C at the end.

4.4.2 HRM Scanning

After PCR amplification, the melting curve of the PCR products was recorded in each sample well. First, the plates were briefly centrifuged at 1,000 g for 1 min. They were subsequently transferred to a 96-well Lightscanner (Idaho Technology Inc.) for HRM

scanning. In the Lightscanner, the plate was heated from 65ºC to 95ºC at the default Lightscanner melting rate (0.1ºC/sec) (De Koeyer et al., 2010). The fluorescent intensity was automatically collected through the scanning. After the scanning was completed, the melting temperature profile was saved on the computer for further analysis.

4.4.3 Melting Curve Analysis

Analysis of the melting curves was based on the protocols described by Montgomery et al. (2007). The curve analysis included four steps: 1) negative filtering, 2) normalization, 3) curve shifting, and 4) grouping. Negative filtering involved removing any samples that did not amplify or those with low fluorescent intensity (less than 600 nm). The rest of samples were considered positive and were used for the analysis (Figure 8A). In Figure 8A, all samples were positive, with a high fluorescent intensity (between 800-1300 nm). In order to see any consistent differences between genotypes, the melting curves are normalized to have the same beginning and ending fluorescence. Therefore, in the step 2, normalization involved selecting 1ºC intervals above and below the melting region that would define 100% fluorescence and a 0% baseline (Figure 8B). In Figure 8B, the melting region was located between 79.6ºC and 86.6ºC. Appropriate 1ºC intervals were selected above and below the melting region range from 78.6ºC to 79.6ºC and from 86.6ºC to 87.6ºC, respectively. After the suitable interval regions were selected, the normalized data were displayed in Figure 8C. Ideally, the initial plateau and the lower plateau should be flanked by 1-3 degrees of a flat line on the normalized plot (Figure 8C). This means that the background fluorescence is dropping in a consistent way over all the samples. The step 3 was to shift the individual melting curves to align at a point where all of the dsDNA are

completely denatured. It was performed by selecting the 5% fluorescence level (default) where all the samples were brought together (Figure 8D). In this step, samples with heterozygous SNPs can be easily distinguished from the homozygous SNPs by the change in the shape of their melting curves. For example, the red curves shown in Figure 8D are homozygous because of the higher melting temperature while the grey curves are heterozygous. The final step, step 4 of the analysis was to group the samples into groups based on the shapes of the melting curves. The data was presented in a subtractive difference plot of the fluorescence in comparison to a single reference sample (Figure 8E). In Figure 8E, the grey group was chosen as the reference. The shape of red curves was identified by using the fluorescence of red curves minus that of the grey curves from Step 3. Therefore, the fluorescence values of red curves are negative in Figure 8E. After grouping, the population can be divided into different groups. The number of groups represents the different genotypes. For example, two groups in the difference plot would represent two different genotypes in the population (Figure 8E).

Table 2. The arrangement of samples of the diploid population 14946 in the 96-well plate for PCR amplification.

	A	B	C	D	E	F	G	H	I	J	K	L
1	P1[a]	P1	P2[b]	P2	1[c]	2	3	4	5	6	7	8
2	10	12	13	14	15	16	17	18	19	20	21	22
3	23	24	25	26	27	28	29	30	31	32	33	34
4	35	36	37	38	39	40	41	42	43	44	45	46
5	47	48	49	50	51	52	53	54	55	56	57	58
6	59	60	61	62	63	64	65	66	67	68	69	70
7	71	72	73	74	75	76	77	78	79	80	81	82
8	83	84	85	86	87	2	3	4	5	6	7	8

a. P1 indicated parental clone 10908-06;
b. P2 indicated parental clone 12115-07;
c. Numbers 1-87 indicated the 85 progeny clones.

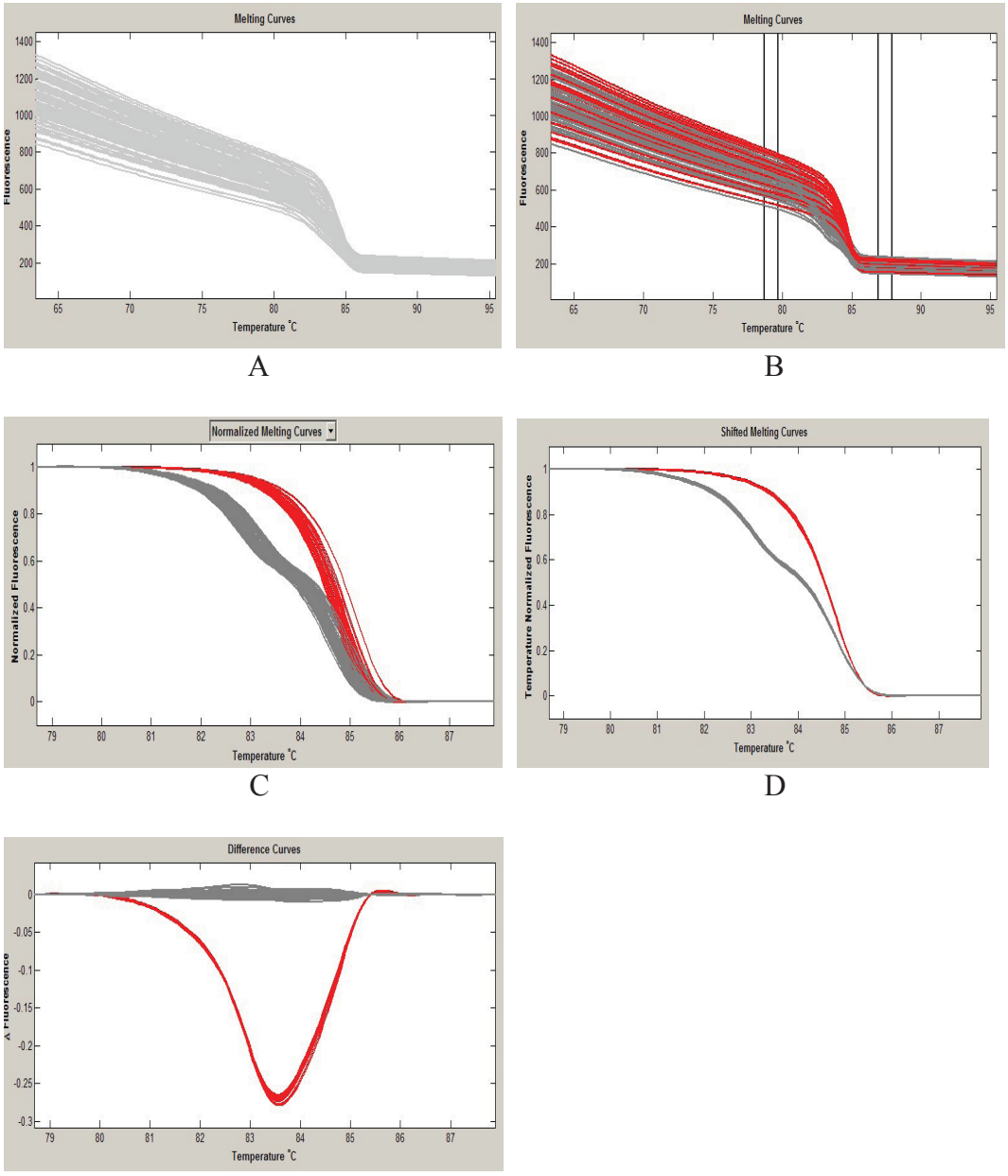


Figure 8. Amplicon genotyping in the diploid population 14946. Plot A represents the results of the negative filter on the raw fluorescence data; plot B represents the normalization of the data; plot C represents the normalized melting data; plot D represents the 5% curve-shifted plot; and plot E represents a difference plot using one group of samples (grey) as a reference.

4.5 SNP Identification by Amplicon Sequencing

Every sample was sequenced twice using one potato clone corresponding to the marker genotype. In order to reduce the labour of sequencing and generate longer reads, M13-tailed sequence (forward: 5'-GTAAAACGACGGCCAGT-3', reverse: 5'-GGAAACAGCTATGACCATG-3') were added to each primer for each marker (Table 3), except for the LS112 and LS079 markers. The amplicon PCR reaction included 1 µL of 10 ng/µL template genomic DNA, 2.5 µL of 10 x PCR buffer, 0.5 µL of 10 mM dNTPs, 0.5 µL for each 10 µM specific forward and reverse M13-tailed primers, 0.1 µL of 5U/µL Taq DNA polymerase, 2.4 µL of 25 mM $MgCl_2$ and the total volume was made to 25 µL. A thermal cycling was conducted with 40 cycles. The conditions of each PCR included incubation at 94^{o}C for 2 min, 40 cycles and each cycle contained 60 sec at 94^{o}C, 60 sec at 60^{o}C, 60 sec at 72^{o}C, and final extension 5 min at 72^{o}C. The PCR samples were stored at 4^{o}C at the end. After the PCR was finished, 20 µL of each PCR product was transferred to a new plate and added with 20 µL of sterile water, mixed well and centrifuged at 1,000 rpm for 1 min. Then it was divided into two sequencing plates with 20 µL of each PCR product. They were prepared for forward direction sequencing and reversed direction sequencing, respectively. DNA sequencing for the 12 markers was done at McGill University and Genome Quebec Innovation Centre in Montreal. The remaining two markers, LS112 and LS079, were amplified with regular primers without M13-tailed sequence and sent to the Agriculture Genomic Centre at NSAC for sequencing.

For SNP identification, the reference sequence which represented the original marker sequence was downloaded from the CPGP website for each marker. They were created to have the M13-tailed sequence as well. This was done by aligning the M13-tailed primers

and the original sequences. The CAP3 contig assembly program in the BioEdit software (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html) was used to perform this step. The reference sequence of each marker was used to compare the M13-tailed DNA sequence of each marker genotype. SNPs corresponding to different genotypes were identified by comparing the DNA sequences of different clones and the reference sequence. Mutation Surveyor software, network version 3.20 (SoftGenetics, LLC) was used to identify SNPs.

Table 3. The DNA sequences of 12 pairs of M13-tailed primers used for sequencing. Bold nucleotides indicate the sequences of the M13 tails.

Name	M13-tailed forward primer Sequences (From 5' to 3')	M13-tailed reversed primer Sequences (From 5' to 3')
LS057	**GTAAAACGACGGCCAGT**CTCTATGTGATCAGGCGAGG	**GGAAACAGCTATGACCATG**CAACATAATCGCCCATGAAG
LS030	**GTAAAACGACGGCCAGT**AGGAGTATTTCATGTGGCTAC	**GGAAACAGCTATGACCATG**GCTTTAGCACATGATTCTATGATAC
LS124	**GTAAAACGACGGCCAGT**CTCCTGATCCTAGGGTACAAA	**GGAAACAGCTATGACCATG**AGGTGAGAAAGTGGCTCG
LS163	**GTAAAACGACGGCCAGT**CACCATTCCTCGAATCATCA	**GGAAACAGCTATGACCATG**AAGCCTGTTAAACAAACGTC
BCH100	**GTAAAACGACGGCCAGT**TCTGTTCTTCTCTCCGTTA	**GGAAACAGCTATGACCATG**AGTAAACCTCTCCGATTTCTT
GP180	**GTAAAACGACGGCCAGT**TCCACCCCTGGATTCAAGAA	**GGAAACAGCTATGACCATG**TCAAACTCCTTCACAAAGCA
LS037	**GTAAAACGACGGCCAGT**CATAAGCTGCAATCTTTTCCTC	**GGAAACAGCTATGACCATG**TTTTACAGGTAAGGGTCAAGA
LS097	**GTAAAACGACGGCCAGT**TCTCAAAGAAGAACCGGGAT	**GGAAACAGCTATGACCATG**GGTTTACCGGTATCTTTTCTCTC
LS092	**GTAAAACGACGGCCAGT**TATGCATCTGCTCCTCTAGGG	**GGAAACAGCTATGACCATG**CTGGATTTTGAGTCACACCATAATT
LS201	**GTAAAACGACGGCCAGT**CGGTCGTGAATCCATTGGTA	**GGAAACAGCTATGACCATG**GCAAAACCTCCAACTCATTC
LS161	**GTAAAACGACGGCCAGT**GTTGAACATGATGATCTAAAGAGGT	**GGAAACAGCTATGACCATG**TCAGCTAGTTTCTTCCGGC
LS173	**GTAAAACGACGGCCAGT**ATGTTCGTCTACAGTCAAGATAG	**GGAAACAGCTATGACCATG**GAACATTGAGTTGAAGAGAAGTTC

4.6 Unlabelled Probe Assay for Tetraploid Germplasm

4.6.1 Unlabelled Probe Design

The design of unlabelled probes for genotyping the tetraploid population was based on the identified SNPs markers. The unlabelled probe was designed to match the SNP sequence. The location of the SNP was identified and all the SNPs present in the amplicon region were determined in order to make sure that the primers do not overlap any polymorphisms. The probe was designed with the SNP site located towards the middle of the sequence to maximize the instability with the mismatched variant. These were generated using Light Scanner Primer Design software (Idaho Technology, Salt Lake City, USA) to have a Tm of 65^{o}C and a sequence length up to 35 bp. To prevent extension of the probe, a C3 spacer was added to the 3' end of the probe (Santalucia, 2007). Usually, the most destabilizing mismatch is chosen for determining if the probe should be complement the sense or antisense strand. In this project, the unlabelled probes were designed to complement the target SNPs in both sense and antisense strands of each marker. Thus, each SNP marker had two unlabelled probes.

4.6.2 PCR Amplification with Unlabelled Probe

The asymmetric PCR amplification was used for the unlabelled probe genotyping in the tetraploid population. All PCR reactions were carried out in 96-well plates using the same equipment as that of the diploid samples. The positions of each sample in the 96-well plate are shown in Table 4. Each PCR reaction included 1 μL of 50 ng/μL template genomic DNA, 1 μL of 10 x Taq polymerase buffer, 0.2 μL of 10 mM dNTPs, 0.2 μL for each 10 μM specific forward primers, 0.04 μL for each 10 μM specific reverse primers (the amount

ratio of forward and reverse primers was 5:1), 0.2 μL for each 10 μM of probe, 0.1 μL of 5U/μL Taq DNA polymerase, 1 μL of 10 x LCGreen Plus (Bio-Chem), 0.95 μL of 25 mM $MgCl_2$ and the total volume was made to 10 μL. If the ratio of forward and reverse primers was changed to 10:1, the PCR solution should include 0.2 μL for each 10 μM specific forward primers, 0.02 μL for each 10 μM specific reversed primers. One drop of mineral oil was added on to the top of the solution. The plate was covered with adhesive film. A thermal cycling for tetraploid samples was conducted with 55 cycles. The conditions of the cycling were as follows: incubation at 94^{o}C for 2 min, 55 cycles and each cycle contained 60 sec at 94^{o}C, 60 sec at 60^{o}C, 60 sec at 72^{o}C, and final extension for 5 min at 72^{o}C. The PCR samples were stored at 4^{o}C at the end.

4.6.3 Unlabelled Probe Genotyping

The above PCR plates were briefly centrifuged at 1,000 rpm for 1 min after PCR amplification. They were consequently transferred to a 96-well Lightscanner (Idaho Technology Inc.) for HRM scanning. The steps for analyzing unlabelled probe genotyping data also involved negative filtering, normalization, and grouping (Figure 9). In Figure 9A, raw fluorescence was captured by the Lightscanner. Failed PCR reactions were excluded before the analysis. All positive samples had fluorescence higher than 600 nm. The Figure 9B displayed the full range of fluorescent data with the melting peaks in the negative differential plot. In the step of normalization, the melting temperature region of the probe was automatically placed in a suitable region for normalization by the software (Figure 9C). Since the probe was designed to perfectly match the target allele, the mismatches would have lower melting temperature than the perfect match. The next step was to assign the

samples into groups based on the melting profile of the probe. In Figure 9D, the samples collectively had two melting peaks, each representing a different sequence composition in the probe region. The relative height ratio of the perfect match peak (high melting temperature) to the mismatch peak (low melting temperature) represented the allele dosage of each genotype. For example, the peak at 72°C (the perfect match) reflects the presence of the perfect match of the target allele. In red curves, the relative peak heights at the perfect match (72°C) and the mismatch (67°C) are 0 and 0.4. The ratio is 0:4. This means that the genotype of red clone contains zero copies of the perfect match allele and four copies of the mismatch allele. Two other genotypes are 1:3 (grey) and 2:2 (blue) (Figure 9D).

Table 4. The arrangement of samples in PCR amplification with unlabelled probes.

	A	B	C	D	E	F	G	H	I	J	K	L
1	09401-03[a]	CH70-03[b]	2[e]	3	4	5	6	7	8	9	10908-06[c]	12115-07[d]
2	09401-03	CH70-03	2	3	4	5	6	7	8	9	10908-06	12115-07
3	09401-03	CH70-03	2	3	4	5	6	7	8	9	10908-06	12115-07
4	09401-03	CH70-03	2	3	4	5	6	7	8	9	10908-06	12115-07
5	09401-03	CH70-03	2	3	4	5	6	7	8	9	10908-06	12115-07
6	09401-03	CH70-03	2	3	4	5	6	7	8	9	10908-06	12115-07
7	09401-03	CH70-03	2	3	4	5	6	7	8	9	10908-06	12115-07
8	09401-03	CH70-03	2	3	4	5	6	7	8	9	10908-06	12115-07

a. 09401-03 indicates the parental female clone of the population 14945;
b. CH70-03 indicates the parental male clone of the population 14945;
c. 10908-06 indicates the parental female clone of the population 14946;
d. 12115-07 indicates the parental male clone of the population 14946;
e. Numbers 2 to 9 indicate progeny clones of the tetraploid population 14945.

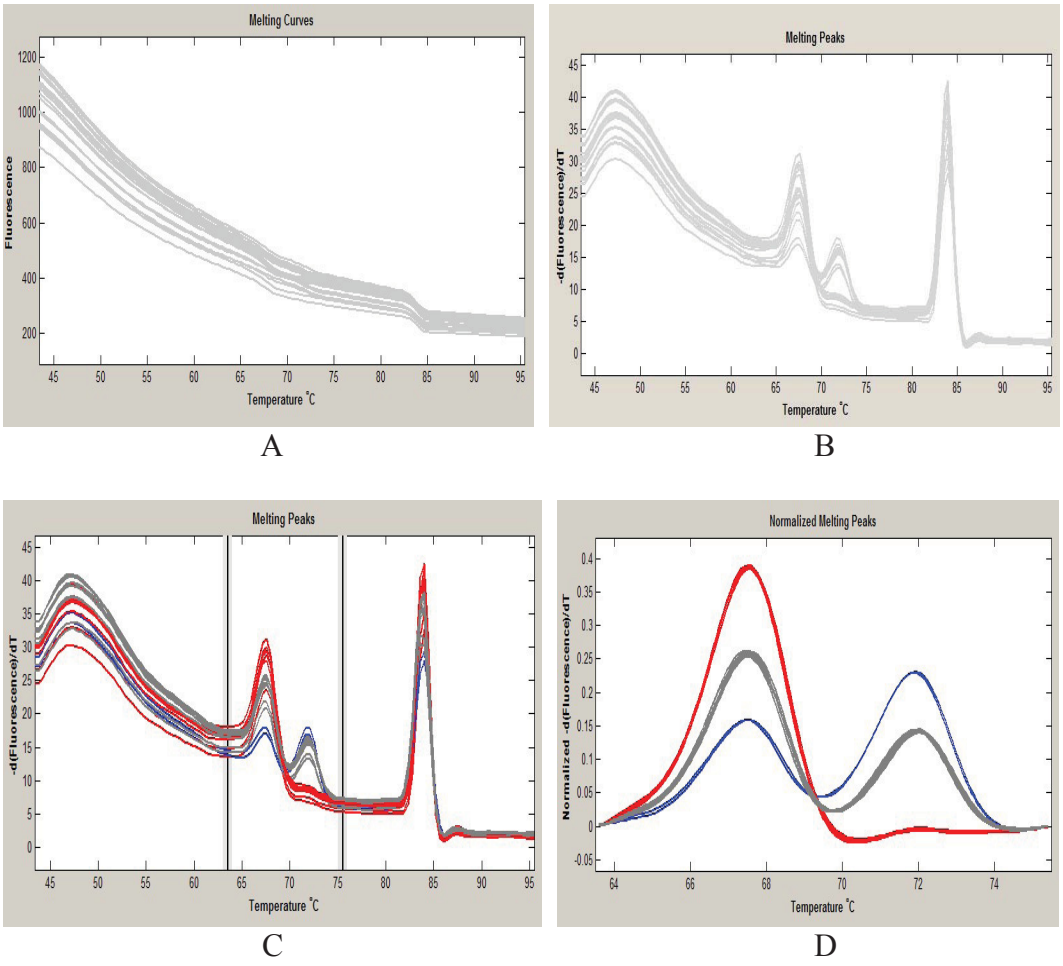


Figure 9. Unlabelled probe genotyping in the tetraploid population 14945. Plot A represents raw fluorescence data captured by the Lightscanner; B represents the negative differential plot; C represents the normalization of the data; D represents the normalized melting peaks using one group (grey) of samples as a reference.

4.7 Marker Association Analysis

4.7.1 Phenotypic Analysis of the Diploid Population

The variance of the degree of ACD in the diploid population 14946 was analyzed using Minitab 15 Statistical Software (v15.1.0.0). All the ACD data of each clone in the six trials were collected and subjected to a normal distribution test (Montgomery, 2001).

4.7.2 The Correlation Analysis of ACD in the Six Trials

Pearson correlation was used to analyze the ACD performance in individual trials. The correlation was tested between two individual trials using the degree of ACD of each clone in the diploid population 14946. All the ACD data of each clone in the six trials were collected and subjected to correlation in SigmaPlot (v11.0). The significant level of 5% was chosen.

4.7.3 Single-Marker Analysis in Diploid and Tetraploid Populations

The relationship between a marker and the degree of ACD was analyzed by using the factor factorial design method (Montgomery, 2001). All the ACD data were subjected to analysis of variance (ANOVA) using the general linear model (GLM) procedure in SAS version 9.1 (Institute Inc., NC, USA, 2002-2003) (V.15). The significant level of 5% was set ($P< 0.05$). The statistical model for the six trials analysis with individual markers was:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

$$i = 1, 2, \ldots, 6;\ j = 1, 2;\ k = 1,2 \text{ or } 1, 2, 3, 4;\ l = 1.$$

Where μ is the overall mean of ACD; i represents the level of trial; j represents the level of month; k represents the level of marker genotype; l represents the replication; Y_{ijkl}

represents the degree of ACD contributed by the treatments and replications; α_i represents the main effect of the i^{th} level of trial; β_j represents the main effect of the j^{th} level of month; γ_k represents the main effect of the k^{th} level of marker genotype; $(\alpha\beta)_{ij}$ represents the interaction between trial and month; $(\alpha\gamma)_{ik}$ represents the interaction between trial and marker genotype; $(\beta\gamma)_{jk}$ represents the interaction between month and marker genotype; $(\alpha\beta\gamma)_{ijk}$ represents the interaction among trial, month and marker genotype; ε_{ijkl} represents the experimental error terms.

4.7.4 Identification of QTLs

According to the previous study, the QTLs of ACD in diploid population 14946 were identified with a significant level of 1%. The QTLs were determined by the most significant markers ($P<0.01$). Therefore, the chromosome regions of the significant markers were considered as QTLs related to ACD.

4.7.5 Marker Analysis in the Individual Trials

After the overall data analysis, only the significant markers in the validated QTL regions were selected and analyzed in the individual trials. All the ACD data were subjected to ANOVA under the GLM procedure in SAS. The significant level of 5% was set ($P< 0.05$). The statistical model for the individual trial analysis of significant markers was:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

$$i = 1, 2; j = 1, 2 \text{ or } 1, 2, 3, 4; k = 1.$$

Where μ is the overall mean of ACD; i represents the level of month; j represents the

level of marker genotype; k represents the replication; Y_{ijk} represents the degree of ACD contributed from the treatments and replications; α_i represents the main effect of the i^{th} level of month; β_j represents the main effect of the j^{th} level of marker genotype; $(\alpha\ \beta)_{ij}$ represents the interaction between month and marker genotype; ε_{ijl} represents the experimental error terms.

4.7.6 Marker Genotype Corresponding to the Degree of ACD

The Least Significant Difference test was used for comparison of the difference among the clones with different marker genotypes at a significant level of 5% ($P<0.05$) (Montgomery, 2001). In addition, the relationship between allele copy number and ACD was analyzed using the Least Significant Difference test.

4.7.7 Marker Contribution to the ACD

A multiple regression analysis was used for determining the percentage of the degree of ACD explained by significant markers. The most significant marker in each QTL was chosen. To use regression with the non-linear marker genotypes, these data needed to be converted to allele variables representing the number of copies of a particular allele. For each of these markers, the marker genotypes were split into 2-4 variables representing each allele in the marker genotype. These allele variables reflected the number of copies of the particular allele in the marker genotype and had values of 0, 1, or 2. For example, a marker with *lm* x *ll* segregation would have an *l* variable and an *m* variable. The values for *l* would be 1 or 2 and the values for *m* would be 1 or 0, corresponding to the *lm* and *ll* genotypes. Since some alleles provided exactly the same information as another allele of the same

marker, the final model only included independent allele variables. For both overall trials and individual trial analysis, the coefficient of determination (R^2) represents the percentages of the degree of ACD explained by the significant markers (Collard et al., 2005). The degree of ACD in each clone was calculated based on the average value of two months in each trial and the average of the six trials. All the ACD data of each clone in the six trials were collected and subjected to regression analysis using Minitab 15 Statistical Software (v15.1.1.0). The statistical model for the overall percentage of the degree of ACD explained by significant markers was:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_{10} x_{i10} + \varepsilon_i \quad i=1, 2,..., 82.$$

Where Y_i represents the observed combined average ACD value for a progeny clone; β_0, β_1,..., β_{10} are parameters and β_0 represents the regression intercept; x_1, x_2, ...x_{10} are the predictors and represent the independent alleles at the six significant loci. Specifically, x_1...x_{10} represent the following alleles: LS030-*l*, LS163-*a*, LS079-*e*, LS079-*f*, LS037-*a*, LS037-*c*, LS112-*a*, LS112-*c*, LS161-*e*, and LS161-*f*. ε represents the statistical error.

The statistical model for the percentage of degree of ACD explained by all significant markers in an individual trial was:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_{10} x_{i10} + \varepsilon_i \quad i=1, 2,..., 82.$$

Where Y_i represents the observed average ACD value of an individual trial; β_0, β_1,..., β_{10} are parameters and β_0 represents the regression intercept; x_1, x_2, ...x_{10} are the predictors and represent the independent alleles at the six significant loci. Specifically, alleles LS030-*l*, LS163-*a*, LS079-*e*, LS079-*f*, LS037-*a*, LS037-*c*, LS112-*a*, LS112-*c*, LS161-*e*, and LS161-*f* were included in the analyses. ε represents the statistical error.

The statistical models for the percentage of degree of ACD explained by a single

significant marker were:

$$1): Y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \quad i=1, 2,..., 82; \text{ or}$$

$$2): Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad i=1, 2,..., 82.$$

Where Y_i represents the observed individual trial or combined average ACD value; β_0, β_1, and β_2 are parameters and β_0 represents the regression intercept; x_1 and x_2 are the predictors and represent the independent alleles at each significant loci. Specifically, model 1 was used for analyzing markers LS030 and LS163. x_1 represents LS030-*l* or LS163-*a*. Model 2 was used for analyzing markers LS079, LS037, LS112, and LS161. x_1 and x_2 represent alleles LS079-*e* and LS079-*f*; LS037-*a* and LS037-*c*; LS112-*a* and LS112-*c*; or LS161-*e* and LS161-*f*. ε represents the statistical error.

4.8 Functional Analysis of Different Marker Genotypes

Since all the SNP markers are coding genes for production of proteins (Table 1), the difference in their genotypes may affect the codons, therefore the amino acid sequences. The change of amino acids may affect the function of the protein by altering; the biochemical activities in potato tubers. These activities could closely relate to the phenotypic difference, which is the degree of ACD. The individual nucleotide sequences of different clones were analyzed for open reading frames in the NCBI GenBank using the OPF finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The translated amino acid sequences of different clones representing different marker genotype were obtained in the program. The translated amino acid sequences of the individual clones were aligned using ClustalW program (http://www.ebi.ac.uk/clustalw).

Chapter 5. Results

Fourteen previously identified SNP markers were used to determine if they were related to ACD in a second population. The analyses used both diploid population 14946 and tetraploid population 14945. Marker association analysis was performed and 11 of the 14 SNP markers that appeared to be significantly linked to ACD in 13610 were validated in 14946. Based on the ANOVA analysis at a significance level of 1%, six QTLs on chromosomes 2, 3, 4, 6, 8, and 10 were identified in the diploid population 14946, two more from the QTLs identified in 13610. In addition, 22 unlabelled probes were designed and five of the 14 markers were tested in eight clones of the tetraploid population 14945. Three of these markers were significantly related to ACD that can be used for breeding programs developing lower ACD varieties.

5.1 Phenotypic Variance in the Diploid Population

The variance of ACD in 14946 was analyzed in order to understand its distribution in the diploid population. The degree of ACD in all six trials of the diploid population 14946 showed a continuous probability distribution (Figure 10). The chart in Figure 10 illustrated that the degree of ACD of each clone tended to cluster at a single mean value, which was 98.6 MRD. The highest degree of ACD in this population was 115.6 MRD, whereas the lowest degree of ACD was 80.5 MRD. The continuous values of ACD range from 80.5 MRD to 115.6 MRD in the population (Figure 10). The variance of the degree of ACD in the population was 68.2 and was significant at the level of 5% (Table 5). This profile revealed that the distribution of ACD in 14946 was a normal distribution.

5.2 The Correlation of the Degree of ACD in the Diploid Population

The correlations among the ACD values of each clone in the six trials were tested in order to investigate the stability of the trait performance in different trials. The correlation results are provided in Table 6. Although some clones were relatively lower in one trial and higher in another trial, the correlations among the degree of ACD of the entire population in the 6 trials were relatively high. All the correlation coefficient (r) values were higher than 0.5 (Table 6). The P-value of each correlation was lower than 0.0001, which indicated that the r was significant. It revealed a high correlation between ACD values of the same clone from different trials. In the other words, the population had a similar performance in the different environments. With regards to the correlation between individual trials, the highest correlation was between 2008 B and 2008 T with r of 0.74, whereas the lowest correlation was between 2005 B and 2009 B with r of 0.52 (Table 6). Although 2008 B and 2008 T were two trials at different locations, the population was produced in the same year. The close seasonal conditions made the ACD performance more similar between those groups; therefore, the correlation was higher than that compared with other trials. The lowest correlation between 2005 B and 2009 B was possibly because of these two trials were far apart from in the range of the five years. During the five year period, weather, soil, as well as clone quality could affect the degree of ACD in the population. The chance that these factors would change over a long period time was higher than over a short period time. In conclusion, the ACD data from the diploid population 14946 grown in six trials were valuable for the marker association analysis.

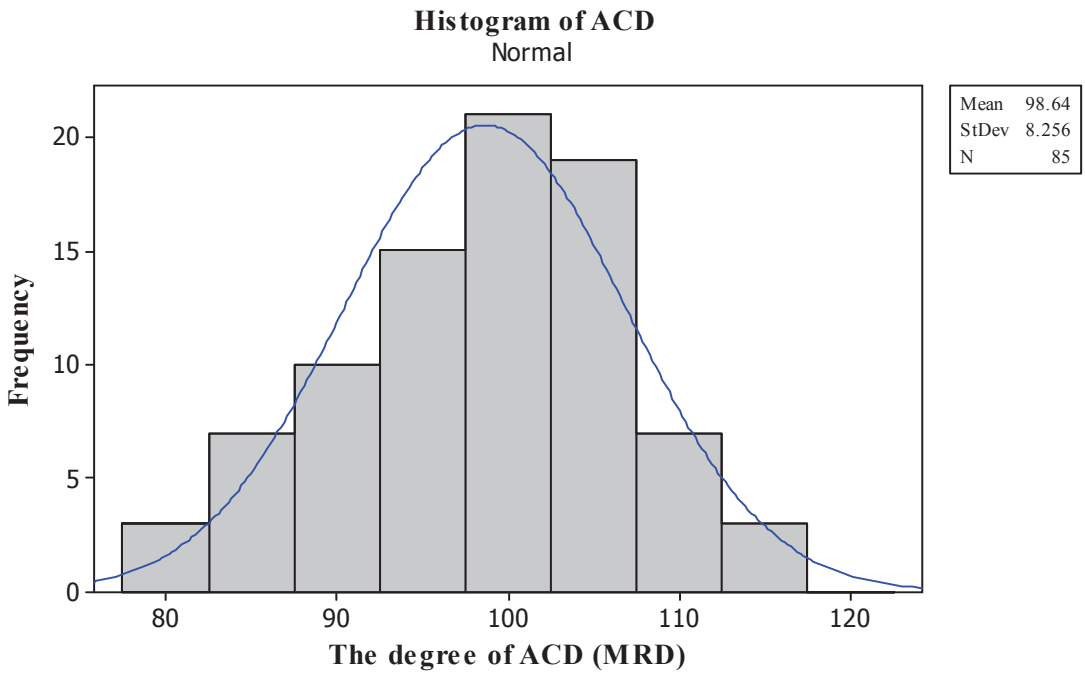


Figure 10. The distribution of ACD in the diploid population 14946. MRD: mean raw density.

Table 5. The variance of ACD in the diploid population 14946.

Parameters	14946
Mean	98.6
Standard Deviation	8.3
Sample variance	68.2
Median	100.4
Minimum	80.5
Maximum	115.6

Table 6. The correlation coefficient (r) between ACD of each clone in the six trials.

Trials	2005 B	2006 B	2007 B	2008 B	2008 T
2006 B	0.589				
2007 B	0.603	0.594			
2008 B	0.625	0.669	0.741		
2008 T	0.561	0.608	0.707	0.742	
2009 B	0.519	0.634	0.667	0.723	0.672

5.3 The Significant SNP Markers in the Diploid Population

Fourteen SNP markers previously associated with ACD in the population 13610 were tested in the diploid population 14946. The results showed that all the markers were segregating in the population. Eleven of them had a significant relationship to ACD ($P<0.05$). Six of the 11 markers with the lowest P-value ($P<0.01$) were located on the four major QTLs of chromosomes 2, 4, 6 and 10 identified in population 13610. Besides the four major QTLs, two other QTLs were identified on chromosomes 3 and 8 in the diploid population 14946. The variance of ACD contributed by these six QTLs (2, 3, 4, 6, 8, and 10) accounted for 28.5%. This was lower than the previous findings in the diploid population 13610, which was 39.9%.

5.3.1 Marker Segregation in the Diploid Population

Based on the amplicon genotyping results, five different genetic segregation types were observed from the parents of the 14 SNP markers. The five different marker genotype combinations are named as *nn* x *np*, *lm* x *ll*, *ab* x *cc*, *ef* x *eg*, and *ab* x *cd*. The combinations are written as the female parental clone crossed to the male parental clone. For example, "*nn* x *np*" means that the female parental clone *nn* was crossed to the male parental clone *np*.

The combination of the marker genotype was deduced from the segregation of progeny in the population. When the parent genotype combination was *nn* x *np*, the progeny of the population showed a 1:1 segregation for the two alleles of *n* and *p*. In this case, two different groups of melting curves were obtained in this population (Figure 11). One was homozygous *nn* and the other was heterozygous *np*. This combination of parents

produced the progeny of *nn* or *np*. Figure 11A represents the marker LS057 with the genotype *nn* x *np*. The red curves are homozygous genotype *nn*, whereas the grey curves are the heterozygous genotype *np*. The genotypes were differentiated due to the fact that fluorescence of homozygous genotypes is higher than that of heterozygous genotypes. In Figure 11B, the parental genotypes of marker LS097 also were *nn* x *np*. However, this marker had a different melting temperature compared to LS057. This is due to the different DNA sequence of each marker. It is expected that although the marker genotypes are the same, the melting curves could be different. Oddly, this plot also has two clones different from the two major groupings (Figure 11B). One was clone #17 (blue) and the other was clone #51 (orange). They did not belong to either of the groups based on their melting curves. This difference could be due to mutations in the population, or because they were mistakenly selected from another population. In fact, this observation was found for more than one marker, which could not be explained by mutation. Therefore, they were not considered for further analysis in the diploid population 14946.

When the parent genotype was *lm* x *ll*, the progeny population showed a 1:1 segregation as well. This type of parent produces *lm* or *ll* progeny (Figure 12). In Figure 12, plots A, B, and C represent the marker genotypes of LS030, LS124, and LS173, respectively. These three markers had the same genotype *lm* x *ll*. The red curves represent genotype *lm* because the fluorescence of the heterozygous was lower than that of the homozygous *ll* (grey).

When the parent genotype combination was *ab* x *cc*, the progeny population showed a 1: 1 segregation for two different alleles, *a* and *b*. All progeny have one *c* allele. However, none of the progeny had the same genotype as the parents. In this case, progeny groups are

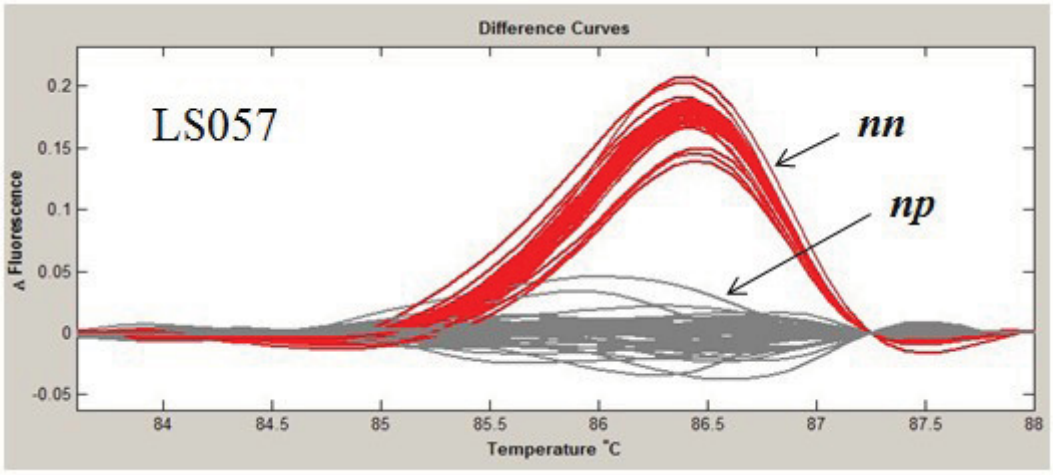
heterozygous with one common allele *c* shown as *ac* or *bc* (Figure 13). Figure 13 represents the marker genotype grouping of LS163. Based on the principle of the different melting temperatures between homozygous and heterozygous, the blue curves represent heterozygous *ab*, whereas the green curves represent homozygous *cc*. The red curves and grey curves represent progeny groups, *ac* and *bc*. However, it is impossible to distinguish which one is which in this situation. The only way of distinguishing them was sequencing of the alleles. Clone #17 is represented as orange colour and had a different melting curve than any other samples.

When the parental genotype combination was *ef* x *eg*, it produced four different genotypes in the progeny population. Two of them had genotypes the same as parents (*ef* and *eg*), whereas the other two (*ee* and *fg*) were different from parents (Figure 14). Therefore, the progeny contained four groups in the population and two progeny groups were the same as the parental groups. In Figure 14, plots A, B, C, and D represent marker genotypes of BCH100, LS092, LS079, and LS161, respectively. Regarding the grouping results, it was easy to distinguish homozygous *ee* and heterozygous *fg* in the progeny population. The green curves represent *ee* with a higher melting temperature, whereas the blue curves represent *fg* with a lower melting temperature (Figure 14). However, it was impossible to distinguish which one was *ef* and which one was *eg* (Figure 14). These genotypes were assigned based on their similarity to one of the parents. In plot C of Figure 14**,** the light green curve indicates the clone #17, which has a different genotype compared to the other groups. Therefore, it was excluded from the population.

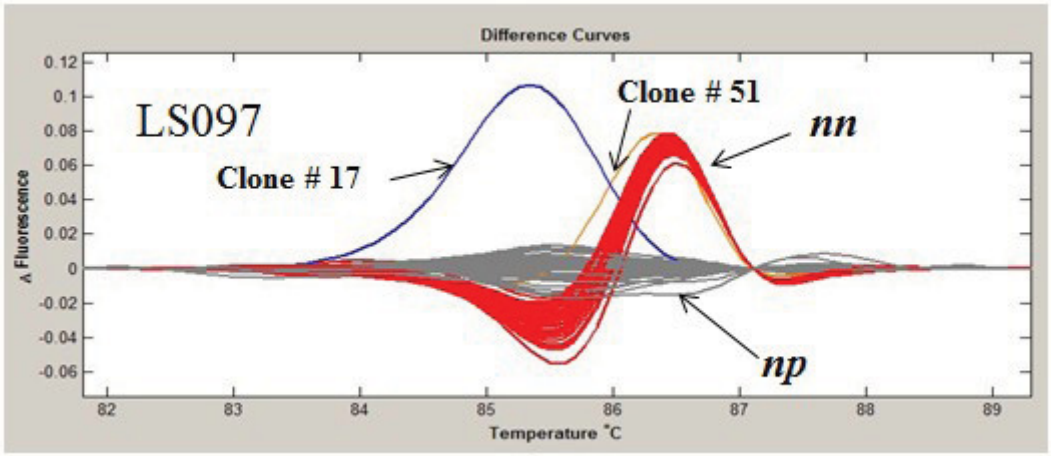
When the parent genotype combination was *ab* x *cd*, the population showed a 1:1:1:1 segregation with four different groups in the progeny (*ac*, *ad*, *bc*, and *bd*). None of them

was the same as the parents (*ab* or *cd*). Therefore, there were four different groups in the progeny (*ac*, *ad*, *bc*, and *bd*) (Figure 15). In Figure 15, plots A, B, C, and D represent marker genotypes of GP180, LS037, LS201, and LS112. They were all considered as genotype combinations *ab* x *cd*. Different melting temperatures of alleles contributed the different curve shapes of each marker. In plot C and D of Figure 15, clone #17 shown by the light green colour, grouped separately from the other genotypes.

In summary, all 14 SNP markers segregated in the diploid population 14946. The segregating genotypes of each marker are given in Table 7. Two markers (LS057 and LS097) were classified as segregating type *nn* x *np* with two groups of curves; three markers (LS030, LS124, and LS173) were designated *lm* x *ll* with two groups of curves; marker LS163 was *ab* x *cc* with four groups of curves; four markers (BCH100, LS092, LS79, and LS161) were *ef* x *eg* with four groups of curves; and the last four markers (GP180, LS037, LS201, and LS112) were *ab* x *cd* with six groups of curves in the population.

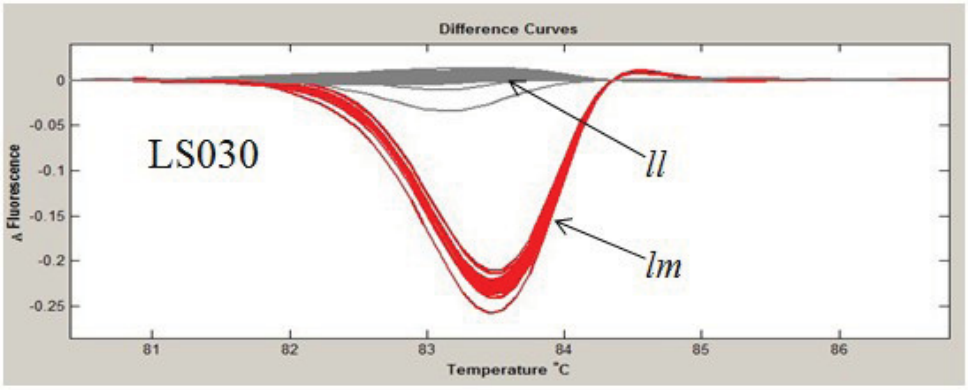


A

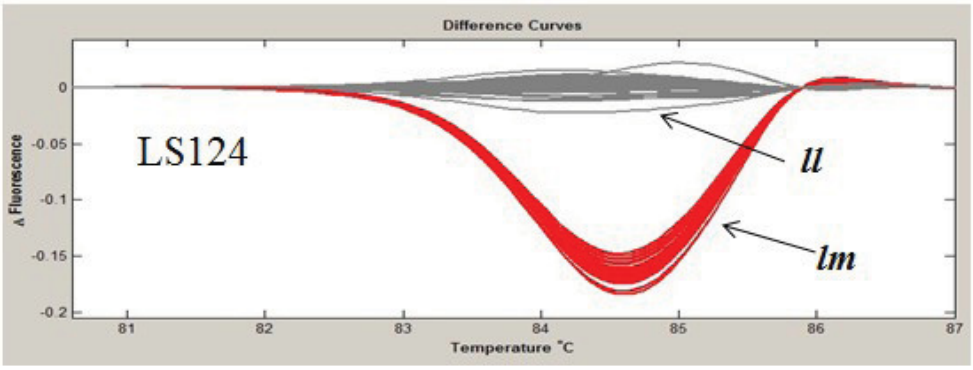


B

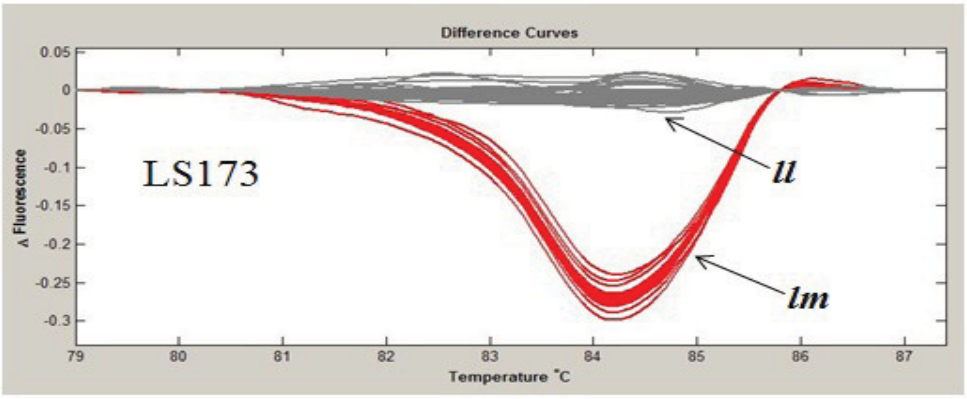
Figure 11. Grouping results of markers LS057 and LS097. Plot A (LS057) and B (LS097) represent marker genotype *nn* x *np*. In both plots, the grey group is heterozygous *np* as the reference and the red group is homozygous *nn*. In plot B, the blue curve represents clone #17 and the orange curve represents clone # 51, both did not belong to the main progeny groups and were considered outliers and therefore excluded from analysis.



A



B



C

Figure 12. Grouping results of markers LS030, LS124, and LS173. Plot A (LS030), B (LS124) and C (LS173) represent marker genotype *lm* x *ll*. The red group is heterozygous *lm*, whereas the grey group is homozygous *ll*.

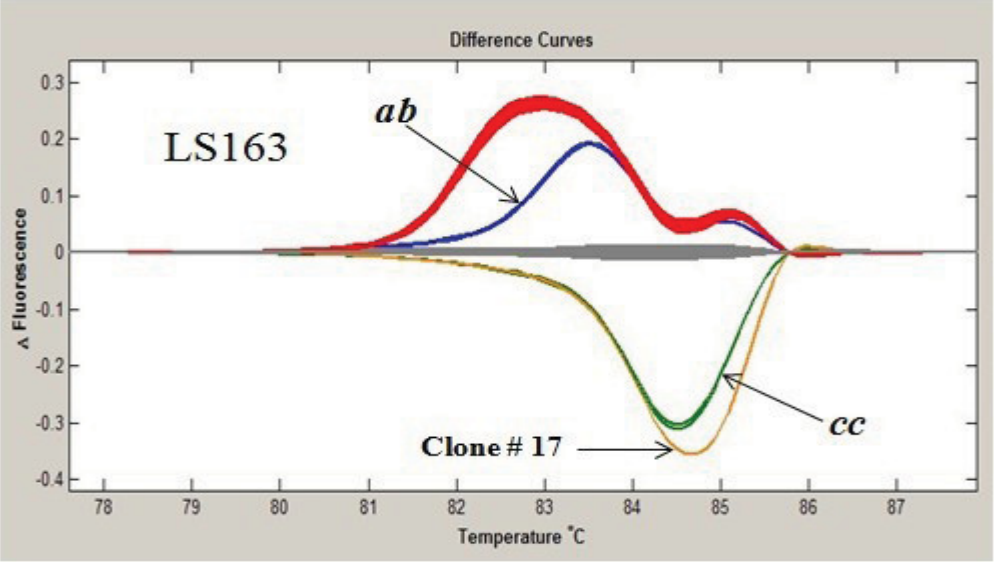


Figure 13. Grouping results of marker LS163. It represents marker genotype *ab* x *cc*. The green group is parental genotype homozygous *cc*, whereas the blue group is the other parental genotype heterozygous *ab*. The red curves and grey curves represent progeny groups, *ac* and *bc*, but they cannot be differentiated based on the melting curves. Clone # 17 is represented by an orange line.

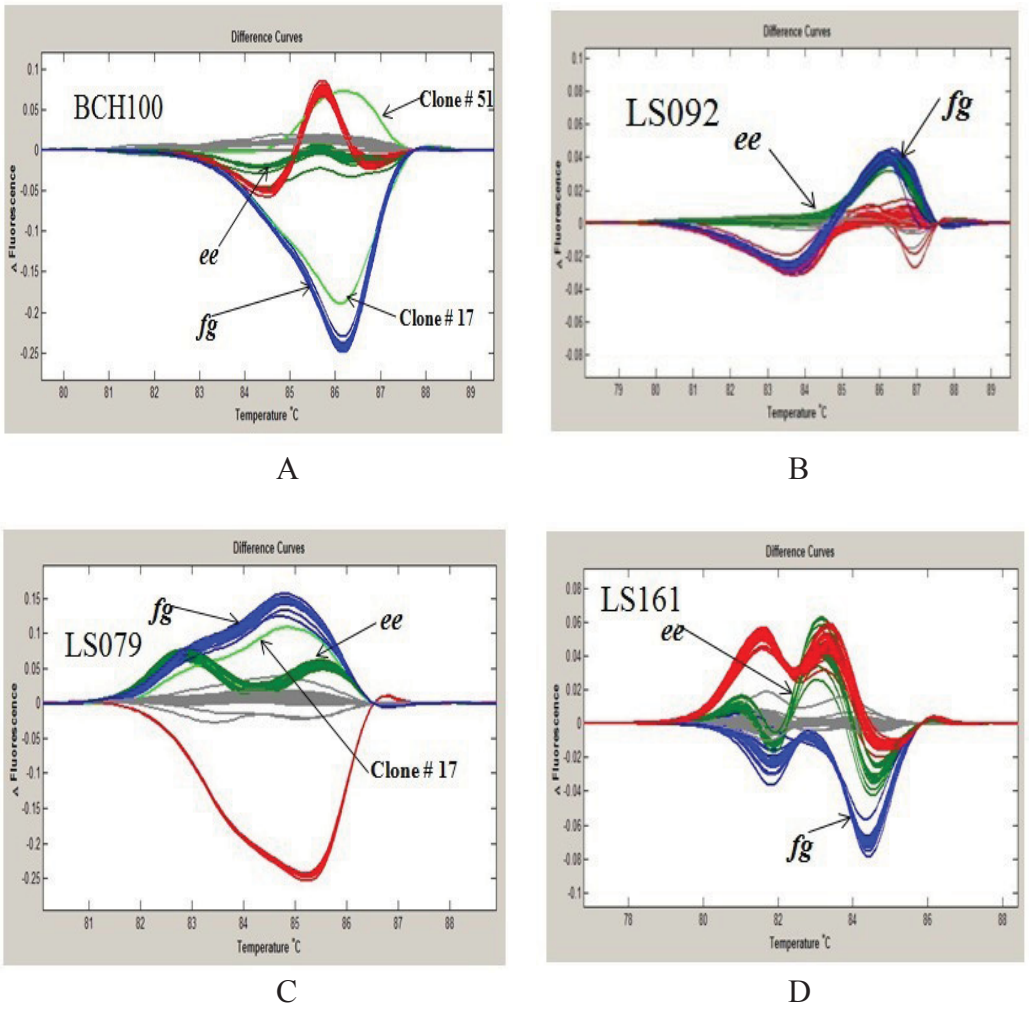


Figure 14. Grouping results of markers BCH100, LS092, LS079, and LS161. Plots A (BCH100), B (LS092), C (LS079), and D (LS161) represent maker genotype *ef* x *eg*. Green curves represent genotype *ee* of each marker; blue curves represent genotype *fg* of each marker. The light green curve in plot C represents clone #17.

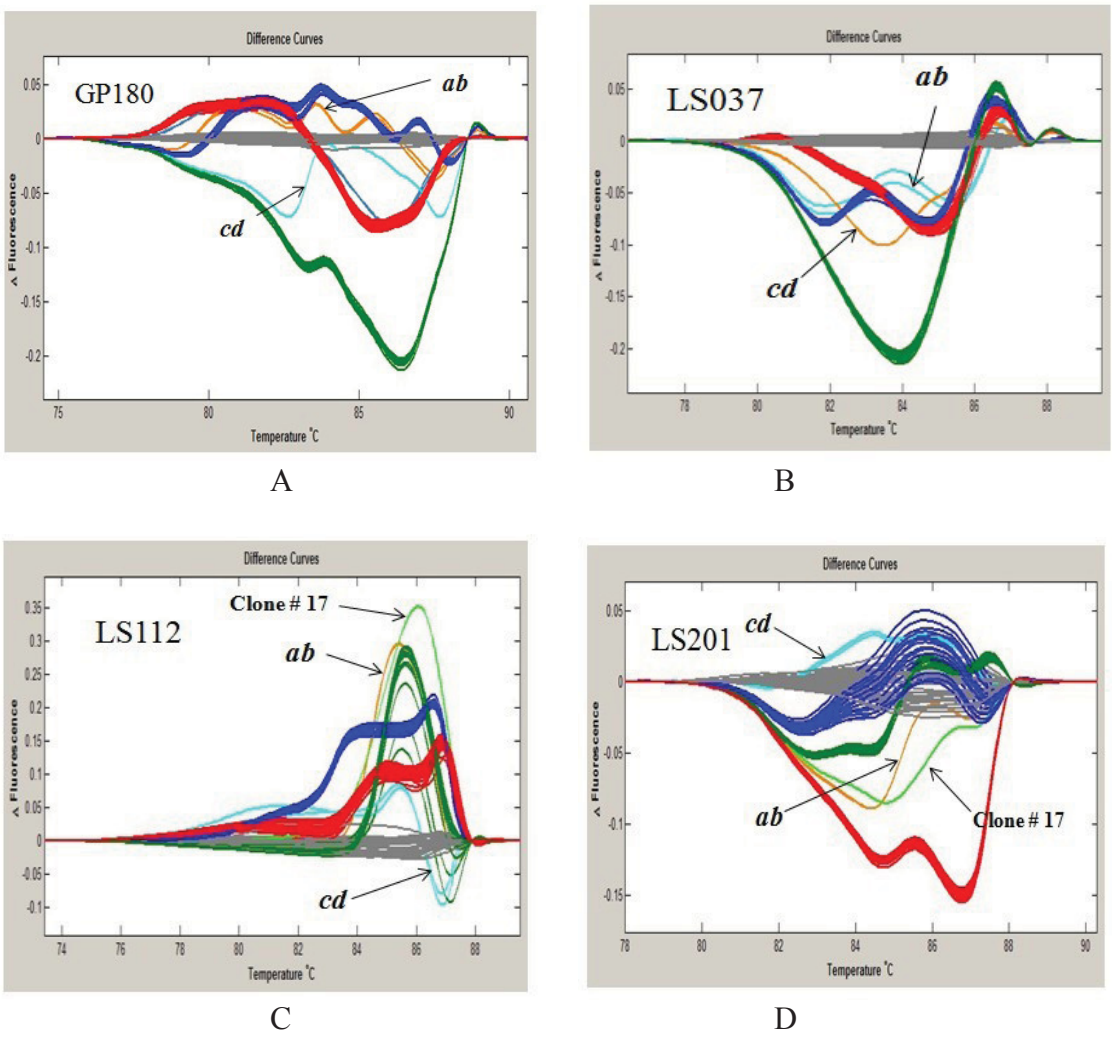


Figure 15. Grouping results of markers GP180, LS037, LS112, and LS201. Plots A (GP180) (*ab*: orange; *cd*: light blue), B (LS037) (*ab*: light blue; *cd*: orange), C (LS112) (*ab*: orange; *cd*: light blue), and D (LS201) (*ab*: orange; *cd*: light blue) represent marker genotype *ab* x *cd*. In plot C and D, light green curves represent clone #17.

Table 7. Summary of marker genotype segregation in population 14946 for the 14 SNP markers associated with ACD in population 13610.

Marker genotypes classification	# of curve groups in the population	Markers
nn x *np*	2	LS057, LS097
lm x *ll*	2	LS030, LS124, LS173
ab x *cc*	4	LS163
ef x *eg*	4	BCH100, LS092, LS79, LS161
ab x *cd*	6	GP180, LS037, LS201, LS112

5.3.2 SNP Positions of PCR Amplicons

The purpose of sequencing the PCR amplicons of the 14 SNP markers previously associated with ACD in population 13610 was to design unlabelled probes based on the SNPs presented in the amplions. With the sequence data of each genotype, allele sequences of each marker were determined. It was, therefore, verification of amplicon genotyping results for each marker. Genotypes of each marker for each SNP position in the amplicon are presented in Table 8. SNPs were detected in all the markers, except for marker LS173. In this case, the segregation in the population could be because one deletion found in the beginning of the sequenced region. The haplotype (which is essentially the allele) of each marker was identified based on the genotype at the SNP positions of each marker. For example, two SNPs were detected at amplicon position 31 and 68 in the marker LS057. The direct sequencing of the PCR products demonstrated that there were two haplotypes for marker LS057 based on these two SNP positions in clones 10908-06 and 12115-07 (Table 8).

Besides marker LS173, the other three markers (LS079, GP180, and LS201) were excluded from Table 8. Marker LS079 only had one SNP (A/T) at position 81 of the sequence. The three haplotype sequences could not be generated based on the single SNP. The sequence quality of this marker was low, particularly at the beginning of the amplicon. Therefore, there could have other SNPs in that region so that the haplotypes *e* and *g* could be differentiated. The relative qualities of the SNPs shown in markers GP180 and LS201 were much lower than the other markers. Some of the samples had small secondary peaks at the SNP positions, so it was difficult to be 100% certain of the true genotype. Therefore, they were excluded from the sequence analysis.

Table 8. Genotypes at each SNP position and derived haplotypes of 10 amplicons based on sequencing results.

LS057	Base position								
Clone	31	68							Genotype
10908-06	TT	GG							*nn*
12115-07	TC	GT							*np*
Haplotype									
n	T	G							
p	C	T							
LS030	Base position								
Clone	29	95							Genotype
10908-06	AG	TC							*lm*
12115-07	AA	TT							*ll*
Haplotype									
l	A	T							
m	G	C							
LS124	Base position								
Clone	58								Genotype
10908-06	AC								*lm*
12115-07	AA								*ll*
Haplotype									
l	A								
m	C								
LS163	Base position								
Clone	54	57	132	155	164	179	194	217	Genotype
10908-06	CT	TC	GG	TA	GG	TA	AA	GA	*ab*
12115-07	CC	TT	AA	AA	CC	TT	GG	GG	*cc*
14946-27	CC	TT	GA	TA	GC	TT	AG	GG	*ac*
14946-73	TC	CT	GA	AA	GC	AT	AG	AG	*bc*
Haplotype									
a	C	T	G	T	G	T	A	G	
b	T	C	G	A	G	A	A	A	
c	C	T	A	A	C	T	G	G	

Table 8. (Cont')

BCH100	Base position							
Clone	61	73	89	94	97	124		Genotype
10908-06	GG	CG	GA	TT	AG	TA		*ef*
12115-07	GC	CG	GG	TA	AG	TT		*eg*
14946-30	GC	GG	AG	TA	GG	AT		*fg*
14946-40	GG	CC	GG	TT	AA	TT		*ee*
Haplotype								
e	G	C	G	T	A	T		
f	G	G	A	T	G	A		
g	C	G	G	A	G	T		
LS037	Base position							
Clone	40	55	67	73	127	133	164	Genotype
10908-06	TA	CT	CT	CC	AA	AA	TC	*ab*
12115-07	AT	TT	CC	TC	TA	GA	CC	*cd*
14946-07	AT	TT	TC	CC	AA	AA	CC	*bd*
14946-10	TT	CT	CC	CC	AA	AA	TC	*ad*
14946-19	TA	CT	CC	CT	AT	AG	TC	*ac*
14946-39	AA	TT	TC	CT	AT	AG	CC	*bc*
Haplotype								
a	T	C	C	C	A	A	T	
b	A	T	T	C	A	A	C	
c	A	T	C	T	T	G	C	
d	T	T	C	C	A	A	C	
LS097	Base position							
Clone	226							Genotype
10908-06	GG							*nn*
12115-07	GA							*np*
Haplotype								
n	G							
p	A							
LS092	Base position							
Clone	267	360						Genotype
10908-06	TT	CG						*ef*
12115-07	TA	CC						*eg*
14946-30	TA	GC						*fg*
14946-40	TT	CC						*ee*
Haplotype								
e	T	C						
f	T	G						
g	A	C						

Table 8. (Cont')

LS112	Base position			
Clone	75	79	107	Genotype
10908-06	GG	GG	GA	*ab*
12115-07	CG	GA	AA	*cd*
14946-14	GC	GG	GA	*ac*
14946-19	GG	GA	AA	*bd*
14946-30	GG	GA	GA	*ad*
14946-40	GC	GG	AA	*bc*
Haplotype				
a	G	G	G	
b	G	G	A	
c	C	G	A	
d	G	A	A	
LS161	Base position			
Clone	30	36	72	Genotype
10908-06	TC	AA	CC	*ef*
12115-07	TC	AG	CT	*eg*
14946-30	CC	AG	CT	*fg*
14946-40	TT	AA	CC	*ee*
Haplotype				
e	T	A	C	
f	C	A	C	
g	C	G	T	

5.3.3 Markers Significantly Associated with ACD in the Diploid Population 14946

The results of single-marker analysis using ANOVA in all six trials showed that 11 of the 14 SNP markers were significantly associated with ACD in the diploid population 14946 (Table 9). The lower the P-value, the closer the marker is linked to a QTL. Six of the 11 significant markers (LS030, LS124, LS079, LS037, LS201, and LS112) had the P-values less than 0.0001; three of them (LS163, GP180, and LS161) had P-values of 0.0071, 0.0018, and 0.0030, respectively; and the other two markers (LS097 and LS092) had P-values of 0.0394 and 0.0311, respectively. Markers LS057, BCH100, and LS173 were considered insignificant because the P-values were higher than 0.05 (Table 9). In addition, among all factors tested, month showed an insignificant relationship to ACD. Because the two evaluation months (January and February) were close together, it is not expected that month would have a significant effect on ACD. For all the significant markers, both trial and marker genotype had a significant relationship to ACD ($P<0.05$) (Table 9). This means that ACD was influenced by both environmental and genetic factors.

Table 9. The effects of the significant markers on ACD in the six trials

Chr. #	Marker	P-value of trial	P-value of genotype
1	LS057	<0.0001	0.1080
2	LS030	<0.0001	<0.0001
2	LS124	<0.0001	<0.0001
3	LS163	<0.0001	0.0071
3	BCH100	<0.0001	0.0930
4	GP180	<0.0001	0.0018
4	LS079	<0.0001	<0.0001
6	LS037	<0.0001	<0.0001
7	LS097	<0.0001	0.0394
8	LS092	<0.0001	0.0311
8	LS201	<0.0001	<0.0001
8	LS112	<0.0001	<0.0001
10	LS161	<0.0001	0.0030
10	LS173	<0.0001	0.1750

5.3.4 ACD QTLs in the Diploid Population 14946

The detection of QTLs in 14946 was performed by using ANOVA at a significance level of 1%. The result showed that nine markers, located on chromosomes 2, 3, 4, 6, 8, and 10, had the strongest association with ACD in the diploid population 14946. These nine markers were LS030 (chr. 2), LS124 (chr. 2), LS163 (chr. 3), GP180 (chr. 4), LS079 (chr. 4), LS037 (chr. 6), LS201 (chr. 8), LS112 (chr. 8), and LS161 (chr. 10). In the previous study using the diploid population 13610, four major QTLs were identified on chromosomes 2, 4, 6 and 10. Six of the validated SNP markers in this study (LS030, LS124, GP180, LS079, LS037, and LS161) were located in the four major QTL regions previously identified. This means that the chromosomes 2, 4, 6, and 10 were verified as important QTLs in the diploid population 14946. The other three significant markers LS163, LS201, and LS112 were located on two other QTLs of chromosomes 3 and 8. The P-values of markers associated with ACD in populations 13610 and 14946 are listed in Table 10.

Table10. Markers significantly associated with ACD in populations 13610 and 14946.

QTLs (Chr. #)	13610 Marker and (P-value)	14946 Marker and (P-value)
2	LS030 (<0.05) LS124 (<0.0001)	LS030 (<0.0001) LS124 (<0.0001)
3	LS163 (<0.01)	LS163 (<0.005)
4	LS079 (<0.05) GP180 (<0.0001)	LS079 (<0.0001) GP180 (<0.005)
6	LS037 (<0.0005)	LS037 (<0.0001)
8	LS201 (<0.005) LS112 (<0.05)	LS201 (<0.0001) LS112 (<0.0001)
10	LS161 (<0.01) LS173 (<0.0005)	LS161 (<0.001)

5.3.5 Marker Analysis in the Individual Trials

Based on the nine significant markers located in the six QTLs, the relationship between each marker and ACD was analyzed in individual trials. Individual trials were not replicated in the field, so this analysis will only identify the most significant markers not influenced by field location. The results indicated that markers LS030, LS124, and LS037 had significant relationship to ACD in all the six trials ($P<0.05$) (Table 11). This means that genetic factors near these three markers have more effects on ACD than other factors in this population. The markers LS163 and LS161 had no significant relationship to ACD in any trial ($P>0.05$) (Table 11). However, they showed significant relationship to ACD with the overall average of six trials. It is possible that these markers were not very close to the underlying QTL. However, it is more likely that the phenotypic estimates of ACD from combined trials are closer to the true phenotypic values of the clones than the individual trial data. This combined data allows for detection of smaller effect QTLs. The other four significant markers (GP180, LS079, LS201, and LS112) had significant effects on some of the trials but not all the trials. This means that environmental factors or within field variation may have more effects on ACD than these genetic factors in some trials. In summary, markers LS030, LS124, and LS037 were the most significant markers for ACD in the diploid population 14946. LS030 is in the gene encoding dihydroflavonol 4-reductase; LS124 is in the gene encoding soluble starch synthase; LS037 is in the gene encoding cinnamic acid 4-hydroxylase.

Table 11. The effects of the significant markers on ACD in the individual trials. Three markers in bold have significant relationship to ACD in all six individual trials.

Marker	2005 B	2006 B	2007 B	2008 B	2008 T	2009 B
LS030[a]	0.0119	0.0057	0.0119	0.0029	0.0245	0.0002
LS124[a]	0.0119	0.0057	0.0119	0.0029	0.0245	0.0002
LS163[b]	0.0956	0.9397	0.2296	0.2671	0.2702	0.1323
GP180[c]	0.4200	0.0063	0.1301	0.0340	0.2780	0.0067
LS079[c]	0.0009	0.0007	0.1450	0.3437	0.2780	0.0458
LS037[d]	0.0044	0.0057	0.0051	0.0017	0.0076	0.0360
LS201[e]	0.3369	0.0075	0.0646	0.3299	0.1578	0.3205
LS112[e]	0.6698	0.0681	0.0001	0.1250	0.0326	0.0120
LS161[f]	0.3013	0.1850	0.1151	0.6550	0.1092	0.4154

a. Markers LS030 and LS124 are located on chromosome 2;
b. Marker LS163 is located on chromosome 3;
c. Markers GP180 and LS079 are located on chromosome 4;
d. Marker LS037 is located on chromosome 6;
e. Markers LS201 and LS112 are located on chromosome 8;
f. Marker LS161 is located on chromosome 10.

5.4 Marker Genotypes Corresponding to the Degree of ACD

The mean comparisons revealed that different marker genotypes were associated with one group of ACD value, either relatively high ACD or relatively low ACD. The differences of mean ACD in the different genotypes of each significant marker are shown in Table 12. The results showed that the most significant genotype corresponding to the relative high degree of ACD was genotype *bc* of the marker LS037 (93.7 MRD). Oppositely, the genotype *ad* of the marker LS037 was the most significant genotype corresponding to a relatively low degree of ACD (101.6 MRD). Other genotypes corresponding to the relative low degree of ACD were *lm* of the marker LS030 (100.5 MRD), *lm* of the marker LS124 (100.5 MRD), *ee* of the marker LS079 (101.4 MRD), and *ee* of the marker LS112 (100.7 MRD), respectively (Table 12).

In order to determine if allele copy number contributed to ACD in each marker, the analysis was also performed based on the data presented in Table 12. The results showed that the highest degree of ACD were contributed by the clones with two copies of *l* allele of the marker LS030 (ACD=95.7 MRD) and the clones with two copies of the *l* allele of marker LS124 (ACD=95.7 MRD), whereas the lowest degree of ACD was contributed by the clones with two copies of the *e* allele of marker LS079 (ACD=101.4 MRD) (Table 13).

Table 12. Mean ACD value of each marker genotype. Grouping was done based on the type of segregation. Letters a and b indicated the significance of the relationship between each groups. Groups with the same letters are not significantly different.

Marker genotype	Mean ACD (MRD)
LS030_*lm*	100.5a
LS030_*ll*	95.7b
LS124_ *lm*	100.5a
LS124_ *ll*	95.7b
LS163_ *ab*	99.1a
LS163_ *cc*	97.3b
LS079_*ee*	101.4a
LS079_*eg*	97.5b
LS079_*ef*	94.9c
LS079_*fg*	98.1b
GP180_ *ac*	99.3a
GP180_ *ad*	96.9b
GP180_ *bc*	99.8a
GP180_*bd*	96.1b
LS037_ *ac*	97.8b
LS037_ *ad*	101.6a
LS037_ *bc*	93.7c
LS037_*bd*	100.1a
LS201_ *ac*	99.4a
LS201_ *ad*	99.6a
LS201_ *bc*	96.8b
LS201_*bd*	96.0b
LS112_ *ac*	100.7a
LS112_ *ad*	95.66c
LS112_ *bc*	98.9a
LS112_*bd*	97.1b
LS161_*ee*	96.3b
LS161_*eg*	99.3a
LS161_*ef*	99.6a
LS161_*fg*	98.1a

Table 13. The relationship between allele copy number and ACD value for each marker. The letter grouping for individual markers indicates significant differences among the different groups of ACD value corresponding to the allele copy number in each marker genotype.

Allele	0 copy	1 copy	2 copies
LS030-*l* allele	—	100.5a	95.7b
LS030-*m* allele	95.7b	100.5a	—
LS124-*l* allele	—	100.5a	95.7b
LS124-*m* allele	95.7b	100.5a	—
LS163-*a* allele	97.3b	99.1a	—
LS163-*b* allele	99.1a	97.3b	—
LS163-*c* allele	—	98.2a	—
GP180-*a* allele	98.0a	98.1a	—
GP180-*b* allele	98.1a	98a	—
GP180-*c* allele	96.5b	99.6a	—
GP180-*d* allele	99.6a	96.5b	—
LS079-*e* allele	98.1b	96.2c	101.4a
LS079-*f* allele	99.5a	96.5c	—
LS079-*g* allele	98.1a	97.8a	—
LS037-*a* allele	96.9b	99.7a	—
LS037-*b* allele	99.7a	96.9b	—
LS037-*c* allele	100.9a	95.8b	—
LS037-*d* allele	95.8b	100.9a	—
LS201-*a* allele	96.4b	99.5a	—
LS201-*b* allele	99.5a	96.4b	—
LS201-*c* allele	97.8a	98.1a	—
LS201-*d* allele	98.1a	97.8a	—
LS112-*a* allele	96.4b	99.8a	—
LS112-*b* allele	99.8a	96.4b	—
LS112-*c* allele	97.3b	98.9a	—
LS112-*d* allele	98.9a	97.3b	—
LS161-*e* allele	98.1a	99.4a	96.3b
LS161-*f* allele	97.8a	98.8a	—
LS161-*g* allele	97.9a	98.7a	—

5.5 Percentage of ACD Variance Explained by Markers

The most significant marker in each QTL was selected to examine individual or combined contribution to ACD. The percentage of ACD explained by each marker is shown in Table 14. Based on the single-marker analysis, the selected markers were LS030 (Chr. 2), LS163 (Chr. 3), LS079 (Chr. 4), LS037 (Chr. 6), LS112 (Chr. 8) and LS161 (Chr.10), respectively. The percentages of ACD explained by individual markers in overall trials were 9.2%, 1.3%, 2.5%, 10.2%, 5.8%, and 0.8%, respectively. This result illustrated that marker LS037 on chromosome 6 contributed the highest percentage of ACD. This means that the marker was linked to the most important QTL. In addition, the six markers contributed the highest percentage of ACD in trial 2009 B (28.1%). The six validated QTL regions together contributed a total of 28.5% of the ACD (Table 14).

Table 14. Percentage of ACD (R^2) explained by the six markers in the diploid population 14946.

Trial / Chr.	2005 B	2006 B	2007 B	2008 B	2008 T	2009 B	All
2	3.1	5.2	6.4	5.6	3.7	9.9	9.2
3	1.2	0	0.9	0.5	0.8	1.6	1.3
4	3.2	3.1	1.2	1.4	1.8	2.5	2.5
6	7.6	6.2	8.6	9.5	8.2	6.3	10.2
8	0.2	5.0	10.9	2.5	1.2	7.6	5.8
10	1.7	1.5	0.2	0	3.7	1.6	0.8
All	17.6	18.4	27.7	21.0	19.4	28.1	28.5

5.6 Amino Acid Sequences of the Three Most Significant Markers

The *c4h* gene encoding cinnamic acid 4-hydroxylase is one of the essential enzymes related to ACD (Babin et al., 2011). In this study, it found that the marker LS037 located in the *c4h* gene was one of the most significant markers related to ACD in the diploid population 14946 and 13610 in the previous study. The marker representing a QTL region on chromosome 6 had the highest effect (10.2%) on ACD. Moreover, this marker contributed 10.8% of ACD variance in the diploid population 13610. Based on the DNA sequencing data, clones 10908-06, 12115-07, 07, 10, 19, and 39 of the diploid population 14946 represented *c4h* marker genotypes of *ab*, *cd*, *bd*, *ad*, *ac*, and *bc*. Seven SNPs were present in the DNA sequences of this marker. As a result, seven DNA sequence changes among these clones were indicated by an *underline* of the DNA sequences in Figure 16. However, six of them did not cause the amino acid changes. Only one SNP at the sequence position 32 caused the amino acid differences among the clones. The codon of the DNA was AAG instead of AGG (Figure 16). Codon of AAG represents lysine (K), whereas AGG represents arginine (R). At position 32 of the amino acid sequences, clones 12115-07, 07, and 39 showed R in the sequences. The rest of clones (10908-06, 10, and 19) showed K in the sequences (Figure 16).

In addition, the DNA sequences of the other two most significant markers, LS124 and LS030, were analyzed to determine if the amino acid sequences are changed in the different genotypes. There was only one SNP present in the marker LS124. The genotype *lm* (10908-06) had a nucleotide C instead of A (*ll*) at the position 58. However, the DNA sequences of CAC in *lm* and CAU in *ll* are both corresponding to the same amino acid histidine (H). In marker LS030, the result of DNA sequence change did not cause the

amino acid change either.

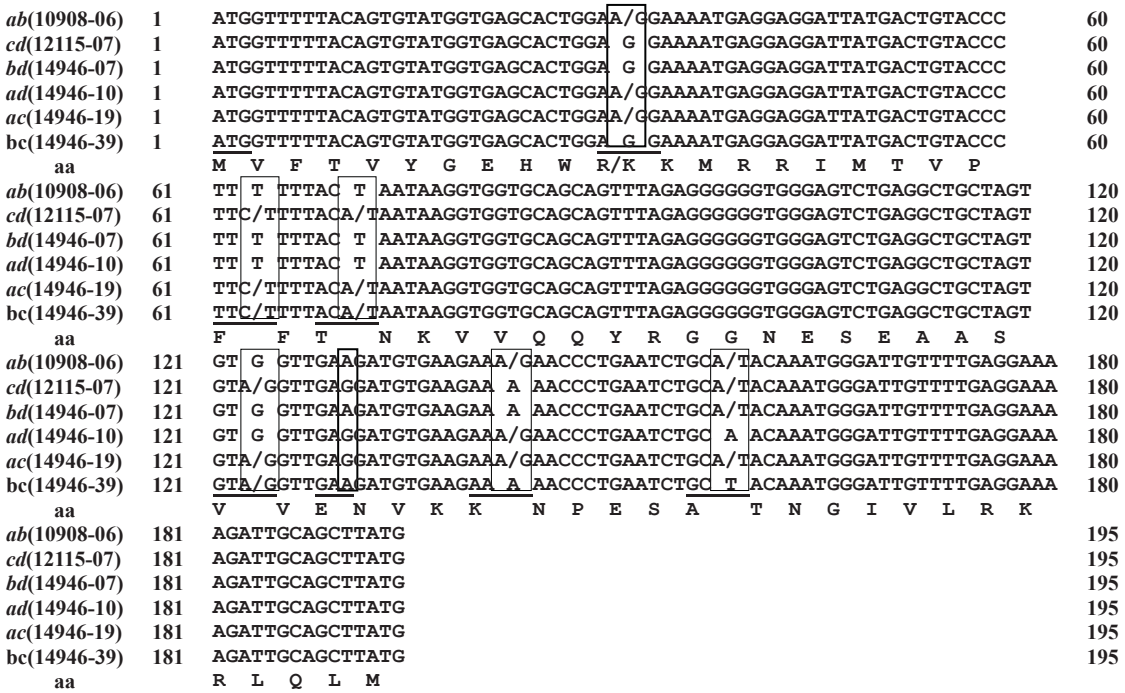


Figure 16. Alignment of DNA sequences of the *c4h* marker in the six marker genotypes (*ab*, *cd*, *bd*, *ad*, *ac*, and *bc*) and corresponding amino acid sequences. SNPs are indicated by *rectangles* in the DNA sequences. Underlined DNA sequences are codons for the seven amino acids. Codon AA/GG=arginine (R) or lysine (K); codon TTC/T=phenylalanine (F); codon ACA/T=threonine (T); codon GTA/G=valine (V); codon GAA/G=glutamic acid (E); codon AAA/G=lysine (K); and codon GCA/T=alanine (A).

5.7 Markers Significantly Associated with ACD in the Tetraploid Population 14945

Based on the SNPs identified within each marker in the diploid population 14946, unlabelled probes were designed and used for tetraploid genotyping by using an HRM unlabelled probe assay. The probes were designed to complement the target SNP, but they were not designed for all the SNPs in the marker region. Based on this design, three out of 11 tested SNP markers were demonstrated to be significantly related to ACD in the eight tested tetraploid samples. Although the significant markers in the tetraploid population 14945 were less than the diploid population 14946 tested, the results confirmed that HRM can be used as a technique for tetraploid genotyping.

5.7.1 Design of Probes

Based on the SNP positions in each marker shown in Table 8, the unlabelled probes were designed to complement the sense strand and antisense strand of the amplicon sequence. The probe sequence and target SNP information for each marker are given in Table 15. Based on the sequencing result, marker LS173 did not have any SNPs in the amplicon region, and therefore was not evaluated. In addition, probes were not designed for markers LS079 and LS112 because of time constraints of this project. Therefore, only 11 markers had unlabelled probes designed for the tetraploid genotyping.

Table 15. Probe sequences designed for the 11 HRM markers. The locations of SNPs are shown in red bold of each probe.

Name	Sequences/3' modification (From 5' to 3')
LS057_68[c]T+[a]	GAAATGTCCGGTT[d]ACTGTTTGTGGT
LS057_68G-[b]	CCACAAACAGTCACCGGACATTT
LS030_29A+	ATTCTATGATACTTAACATTCCCCTGACTGT
LS030_95C+	ATGTCATGAGAGTTATGCTAGTATTAATCATCCAG
LS124_58C+	CAATTGCACTCCACATATAATGACCGC
LS124_58A-	GCGGTCATTATATTTGGAGTGCAATTGAT
LS163_57C+	AAGATCGATATTCATCGGCGATGAAGTC
LS163_155A+	CATAACAAACATACAGTGGCTAAGACAAGGT
BCH100_73C+	GAGGTTGGCCCGTTTGCTTTGT
BCH100_94T+	GCTTTGTGCTGAAGAATGAAAAATTGAATAGTACT
GP180_112A+	CTTTCTGGAGCAACCATAGAAAGAGTTTGT
GP180_176C+	TTTTGCAGGTATGCTGGTGTGCAG
LS037_55C+	CAGATTCAGGGTTCTTCTTCACATCCT
LS037_55T-	AGGATGTGAAGAAAAACCCTGAATCTG
LS097_226A+	ATAAGCATGTCAAGAGTATAACTGGGCTT
LS097_226G-	GCCCAGTTATACCCTTGACATGCTT
LS092_267T-	ATGAACAAGACCGAGCTATGAAATTCCC
LS092_360G-	ACACCATAATTATGTCGAAACGGTTGAAAC
LS201_78G-	CGATTGACTTCCGAACTCAGAGAGATCATAG
LS201_160T-	GACCTCGATCCAATCATACTCAACTTTCTAG
LS161_36G-	ACTGATATAGATAATTGGAGTGAGCCTAATAATGT
LS161_72C+	CTTAAAACCAAGTGTCTTTGTTGGTGTACTA

a. + indicates that the probe was designed to match the sense strand;
b. – indicates that probe was designed to match the antisense strand;
c. The number indicates the SNP position in the amplicon region;
d. The red highlights indicate the target SNPs.

5.7.2 Tetraploid Scanning and Genotyping

Twenty-two different probes for 11 HRM amplicon markers were tested and scanned for tetraploid genotyping (Table 15). All the probes were successfully used for tetraploid genotyping. However, only five markers corresponding with seven probes showed segregation in the tetraploid population 14945. They were LS030_95C+, LS092_267T-, LS163_57C+, LS163_155A+, LS161_36G-, LS201_78G-, and LS201-160T-. The other markers did not show segregation in the progeny. Therefore, they were excluded from the further analysis. The corresponding probe melting results of these segregating markers are shown in Figure 17.

In Figure 17A, the curves represented are the melting result of the marker LS030 with the unlabelled probe LS030_95C+. The relative heights of the perfect match melting peak (72^{o}C) to the mismatch melting peak (67.5^{o}C) showed that clones within red group (parental clone 09401-03, progeny clones # 21, 27, and 32) had four copies of the mismatch allele and zero copies of the perfect match allele. Clones presented in grey group (progeny clones # 02, 38, 85, 93, and 102) had three copies of the mismatch allele and one copy of the perfect match allele. The blue curves corresponded to the parental clone CH72-03 that had two copies of the perfect match allele and two copies of the mismatch allele. In the other words, the genotypes of progeny clones in the tetraploid population 14945 are 0:4 and 1:3 for the marker LS030. Technically, the parental clones 09401-03 (0:4) and CH72-03 (2:2) can produce three different progeny with genotypes of 0:4, 1:3, and 2:2. However, only genotypes of 0:4 and 1:3 were present in the eight progeny.

Figure 17B shows the melting curves of the marker LS092 with the unlabelled probe LS092_267T-. The relative heights of the perfect match melting peak (72^{o}C) to the

mismatch melting peak (69^{o}C) showed that clones within red group (parental clone 09401-03 and progeny clone # 21) had two copies of the mismatch allele and two copies of the perfect match allele (genotype is 2:2). Clones within the grey group (clones # 02, 27, 32, 38, 85, and 102) had one copy of the mismatch allele and three copies of the perfect match allele (genotype is 3:1). Clones within the blue group (parental clone CH72-03 and progeny clone # 93) have four copies of the perfect match allele (genotype is 4:0). This means that the genotypes of 2:2 and 4:0 of the parental clones can produce three different genotypes (4:0, 3:1, or 2:2) in the progeny.

In figure 17C are the melting curves of the marker LS163 with the unlabelled probe LS163_57C+. The relative heights of the perfect match peak (70^{o}C) to the mismatch peak (65^{o}C) showed that clones within red group (parental clone 09401-03) had four copies of the perfect match allele (genotype is 4:0). The green group (parental clone CH70-03 and progeny clone # 21) had two copies of the mismatch allele and two copies of the perfect match allele. The rest of progeny clones in the grey group had one copy of the mismatch allele and three copies of the perfect match allele. This means that the parental clones, 09401-03 (4:0) and CH72-03 (2:2), only produced two different genotypes of progeny in the population. These are genotypes of 3:1 and 2:2 in the marker LS163.

Figure 17D shows the melting curves of the marker LS163 with the other unlabelled probe LS163_155A+. The relative heights of the perfect match melting peak (70^{o}C) to the mismatch melting peak (66^{o}C) showed that clones within orange group (parental clone CH70-03 and progeny clone # 21) had two copies of the mismatch allele and two copies of the perfect match allele. Parental clone 09401-03, presented in red, had four copies of the mismatch allele. The other progeny clones within the grey group had three copies of the

mismatch allele and one copy of the perfect match allele. This means that the parental clones (2:2 and 0:4) produced two kinds of progeny (1:3 and 2:2) in the population. This result was identical with the genotypes of LS163_57C- (Figure 17C). Because the target SNPs in the same marker were different, the melting temperatures were different in Figure 17C and Figure 17D. However, the two different probes of the same marker had the same genotyping result in the population.

Figure 17E are the melting curves of the marker LS161 with the unlabelled probe LS161_36G-. The relative heights of the perfect match peak (68^{o}C) to the mismatch peak (63.5^{o}C) showed that clones within the grey group (parent clone CH72-03, progeny clones # 21, 32, 93, and 102) had four copies of the mismatch allele (0:4). Clones within the red group (progeny clone # 02, 27, 38, and 85) had three copies of the mismatch allele and one copy of the perfect match allele (1:3). The orange curves indicated that the parental clone (09401-03) had four copies of the perfect match allele (4:0). However, the parent clones with genotypes of 4:0 and 0:4 cannot produce progeny clones with a genotype of 1:3. In this case, two possible reasons of getting 1:3 progeny clone would be: 1) the genotyping was wrong; or 2) this parental clone (09401-03) was not the parent of the population as recorded.

Figure 17F shows the melting curves of the marker LS201 with the unlabelled probe LS201_78G-. The relative heights of the perfect match peak (73.5^{o}C) to the mismatch peak (68.5^{o}C) showed that clones within red group (parental clone 09401-03, progeny clones # 02, and 21) have four copies of the mismatch allele (0:4). Progeny clones within the grey group (clones # 27, 32, 85, and 93) have three copies of the mismatch allele and one copy of the perfect match allele (1:3). Progeny clone #38 and parental clone CH72-02 within the

blue group have two copies of the mismatch allele and two copies of the perfect match allele (2:2). The progeny clone # 102 presented in a green curve was different from other genotypes (3:1). Based on the cross of 0:4 and 2:2, the potential genotypes of the progeny were 0:4, 1:3, or 2:2. Therefore, this clone with a 3:1 genotype was considered unexpected. The reasons for this result could either be a sample mix-up during the breeding, DNA extraction or an incorrect genotyping result.

In figure 17G are the melting curves of the marker LS201 with the unlabelled probe LS201_160T-. The blue group indicated that the parental clone 09401-03 had four copies of the mismatch allele at the melting peak of 67.5^{o}C (0:4). The red group (parental clone CH70-03 and progeny clone # 21) indicated that they had four copies of the perfect match allele at the melting peak of 69.5^{o}C (4:0). Typically, this cross of two parents only can produce one kind of progeny, which is having two copies of the mismatch allele and two copies of the perfect match allele (2:2). All of the progeny clones, except # 21, were this kind of genotype. Therefore, progeny clone # 21 should be excluded from the population.

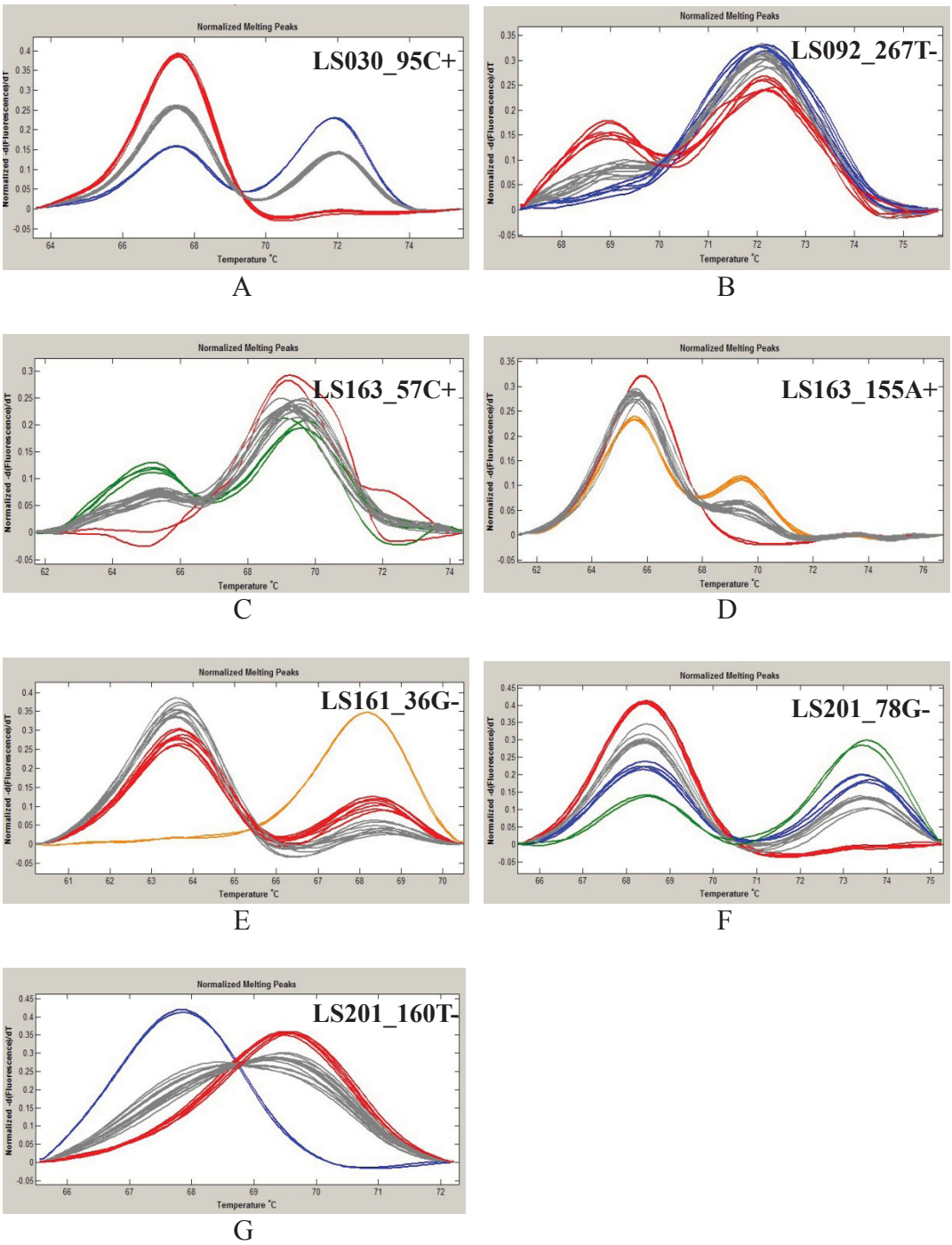


Figure 17. Probe melting profiles for the two parents and eight clones of the tetraploid population 14945. Plot A represents the result of LS030_95C+; B represents the result of LS092_267T-; C represents the result of LS163_57C+; D represents the result of LS163_155A+; E represents the result of LS161_36G-; F represents the result of LS201_78G-; and G represents the result of LS201_160T-.

5.7.3 Single-Marker Analysis in the Tetraploid Population

Based on the single-marker analysis, of five markers, three are related to ACD in the tetraploid clones in 14945 (Table 16). ACD data were collected from 2005 to 2009 trials. Mean comparisons of the three significant markers are shown in Table 17. In Table 16, markers LS030, LS163, and LS201 on chromosomes 2, 3, and 8 showed a significant relationship to ACD with unlabelled probe genotyping. The P-values were less than 0.0001. Other segregating markers did not have a significant relationship to ACD. In Table 17, marker genotypes of LS030_95C+ were 1:3 and 0:4. The average ACD value of genotype 1:3 was 122.8 MRD that is corresponded to light ACD performance in clones # 02, 38, 85, 93, and 102, whereas the average ACD value of genotype 0:4 was 106.9 MRD that corresponds to the dark ACD performance in clones # 21, 27, and 32. Marker genotypes of LS163_57C+ were 2:2 and 1:3. The ACD value of genotype 2:2 was 88.5 MRD that corresponded to dark ACD clone # 21, whereas the average ACD value of genotype 1:3 was 120.8 MRD that corresponded to light ACD. In addition, marker genotypes of LS201_78G- were 2:2 (clone # 38), 1:3 (clones # 27, 32, 85, and 93), and 0:4 (clones # 02 and 21). There was no significant difference between genotypes 2:2 and 1:3. The average ACD values of 2:2 and 1:3 were 129.5 MRD and 118.1 MRD, respectively. The other group of 0:4 was significant different from these two groups. This genotype had an average ACD value of 99.3 MRD. In addition, marker LS030 (chr. 2) is a significant marker tested in the diploid populations 13610 and 14946; markers LS163 (chr. 3) and LS201 (chr. 8) are significant markers tested in populations 14946 and 14945.

In summary, five markers segregated in the related probe regions of the eight progeny of tetraploid population 14945. Only three probes (LS030_95C+, LS163_57C+, and

LS201_78G-) showed a significant association with ACD and are candidates for genotyping other tetraploid potato populations. However, the number of segregating markers and significant associations could increase if the population size was increased.

Table 16. The effects of the significant markers on ACD in the tetraploid clones of 14945.

Chr. #	Marker with probe	Significant factor	P-value
2	LS030_95C+	marker genotype	<0.0001
3	LS163_57C+	marker genotype	<0.0001
8	LS201_78G-	marker genotype	<0.0001

Table 17. Mean ACD value of each marker genotype in the progeny clones of 14945. Letters a and b indicate the significance of the relationship between each group. Groups with the same letters are not significantly different.

Marker genotype	Mean ACD (MRD)
LS030_95C+ (2:2)	—
LS030_95C+ (1:3)	122.8a
LS030_95C+ (0:4)	106.9b
LS163_57C+ (2:2)	88.5b
LS163_57C+ (1:3)	120.8a
LS163_57C+ (0:4)	—
LS201_78G- (2:2)	129.5a
LS201_78G- (1:3)	118.1a
LS201_78G- (0:4)	99.3b

Chapter 6. Discussion

6.1 Validation of Genetic Markers

In this study, a diploid population 14946 was chosen for validating the identified SNP markers using the HRM technique. The use of appropriate plant materials and the application of phenotyping methods with high reliability can help to validate markers more accurately (Yamamoto et al., 2009). In the previous study, a cross of two contrasting diploid parental clones derived the diploid mapping population 13610. Compared to the previous population, the diploid mapping population 14946 used in this study shared one of the parental clones 10908-06 with the population 13610. Both plant materials in the two projects were F_1 populations. Therefore, the diploid populations 14946 and 13610 possessed similar genetic backgrounds. On one hand, using two related populations would increase the probability that the identified markers would demonstrate a significant relationship to ACD. On the other hand, the similar genetic background could be a shortcoming when attempting to identify the usefulness of these markers in other populations. If the previously identified QTLs are real, it is likely that the same markers would show a significant relationship to ACD in the two populations. However, they may not necessarily have a significant relationship to ACD in another population.

The previous study generated the linkage groups for ACD by using 28 SSR markers and 115 SNP markers in a diploid population 13610 with 138 clones in the progeny. The number of the markers mapped and the size of the population used led to the reliance on the data produced by the diploid population 13610. Therefore, the results from this study indicated that the hypothesis can be accepted. Most markers identified in the diploid population 13610 showed a similar relationship to ACD in the other diploid population

14946.

The validated markers were used to detect QTLs in the second population. Based on the ANOVA approach used in this study, QTL location was indicated only by looking at which markers give the greatest differences between the means of the degree of ACD in each genotype groups. Therefore, if the number of tested marker increases, the number of QTL could increase as well. In this study, only 14 identified markers were tested in the diploid population. Six QTLs were localized from the nine significantly validated markers. Four major QTLs on chromosomes 2, 4, 6 and 10 identified in the diploid population 13610 were validated in 14946. In addition, two smaller QTLs on chromosomes 3 and 8 were confirmed in the diploid population 14946. The apparent QTL effect from a marker could be smaller than the true QTL effect as a result of the recombination between the marker and the QTL.

Another factor influencing the study was that there was no replication in the field experiment design. Although the ACD value was evaluated from four tubers of each clone, the lack of replication in the field trials may have influenced the marker effect between the ACD means of different genotype groups. In order to enhance the strength of this experiment, further studies should include replication in field trials. This means that the experiments have to be replicated in different plots planted under the same environmental conditions.

6.1.1 Population Size

The population size of the diploid population 14946 is a potential shortcoming for the marker analysis. Population sizes used in preliminary genetic mapping studies generally

range from 50 to 250 individuals (Mohan et al., 1997). There were 138 individuals in the progeny of the previous diploid population 13610 for ACD QTL mapping, whereas there were only 83 clones (excluding clones # 17 and 51) in the diploid population of 14946 for the validation. The smaller size of the population could affect the phenotypic variance of the trait; hence, affect the detection of markers. This means that the number of significant markers in the diploid population 14946 could be different with increasing population size. Using markers with less statistical significance could also change the validated QTLs. Therefore, using a larger size of the population could improve the accuracy of the result when validating QTLs.

6.1.2 Phenotypic Variance

The digital imaging method for ACD evaluation indicated that it was accurate and precise which helped greatly for the trait correlation analysis. This analysis showed high reliability of the degree of ACD in six trials in the population 14946. The ACD distribution in the diploid population 14946 displayed from low ACD to high ACD as a continuous variation (Figure 10). Moreover, based on the correlation results showed in the Table 6, the performance between ACD values of clones in each trial were highly correlated with each other. This means that the ACD performance of the population of 14946 was quite stable from 2005 to 2009 trials. Both normal distribution and stable ACD performance indicate that the diploid population 14946 is suitable for QTL analysis.

6.2 Efficiency of HRM Genotyping

Besides the high reliability of phenotyping, an efficient and accurate genotyping is

very important for detection and validation of markers. A good method for genotyping should have several characteristics including ease to use, robust, and low cost. In addition to these considerations, cultivated potatoes are autotetraploid crop where allele dosage and multi-allelism are thought to play more important roles in determining phenotypic variation compared to diploid species (Gallais, 2003; De Koeyer et al., 2010). Therefore, the method should be sensitive for detecting multiple alleles in potatoes.

6.2.1 Utilization of HRM for Diploid Potato Genotyping

Amplicon genotyping is one of the approaches for genotyping SNPs using HRM. It was very effective in detecting polymorphisms in the diploid potato clones. In this study, all the SNP markers were successfully used for determining segregation of individuals within the population. Although three markers (LS057, BCH100, and LS173) were insignificantly related to the degree of ACD ($P>0.05$), they were all segregating in the population (Figure 11, 12 and 14).

Before the application of HRM in potatoes, SNPs were genotyped by using other methods, such as real-time PCR (Lay and Wittwer, 1997; Song et al., 2002), pyrosquencing (Rickert et al., 2002), and allele-specific oligonucleotide microarrays (Rickert et al., 2005). However, the disadvantages of these methods are that they are either time-consuming or dependent on a probe. For example, a fluorescent probe is designed to detect and monitor DNA amplification in the real-time PCR. The presence of the mutation can be evaluated by monitoring fluorescence while the sample is being heated through the melting temperature of the probe. A single base change will cause the probe to melt at a lower temperature than if the probe is completely complementary (Lay and Wittwer, 1997). The principle is similar

to the unlabelled probe assay of HRM. However, HRM does not require a probe for diploid genotyping. In the other words, if the SNP exists in the marker, amplicon genotyping can detect the polymorphism without a probe. This also means that the diploid clones can be genotyped without knowing the sequences and efficiently reduce the time of genotyping. Therefore, HRM is recommended for potato marker discovery and QTL analysis.

The success of HRM is largely due to two factors: instrumentation with highly controlled temperature transitions and data acquisition, and fluorescent DNA-binding dyes with improved saturation properties (De Koeyer et al., 2010). Heterozygous genotypes have a characteristic melting curve shape and a broader width than homozygous genotypes, which are assigned based on small (1^{o}C) melting temperature differences. Also, the three replications of each marker used in this study provided consistent results (Figures 11, 12, 13, 14, and 15). Thus, HRM genotyping is reliable. The whole assay from PCR to generating the melting profiles is performed in a single tube and no additional post PCR processing of the DNA amplicons is required. The disadvantage is that, although the HRM genotyping is fast, it could only distinguish different groups within the population. There is no way to know how many SNPs are contributing to the melting curve or to know the identity of the nucleotides causing these SNPs. Further DNA sequencing is needed to find out the SNP information.

The amplicons were also sequenced in this study to prove that the HRM is suitable for genotyping. The genotyping results (Figures 11, 12, 13, 14, and 15) indicated that all the markers were segregating and that polymorphisms existed in the markers. Later, SNPs within each marker were detected by DNA sequencing (Table 8). These results showed that the melting curves could reflect a few SNPs rather than only one SNP. No matter how

many SNPs were in the markers, it was possible to detect the polymorphisms as long as SNPs existed.

6.2.2 Application of HRM on Tetraploid Potatoes

Compared to other dosage sensitive genotyping methods, such as pyrosequencing (Ronaghi et al., 1996), HRM is relatively inexpensive and easy to use. The HRM, in addition to genotyping, can be used in quantifying haplotype dosage and variant scanning (De Koeyer et al., 2010). It was, therefore, applied in the project for tetraploid genotyping.

In the project, 23 unlabelled probes assays were designed for tetraploid potato genotyping. All the probes worked for genotyping, except GP180. Hofinger et al. (2009) reported that it was difficult to identify SNPs located very close to the amplicon's primers and this method has low accuracy in detecting single SNPs when using amplicons larger than 300 bp or 400 bp. Therefore, these two limitations of the HRM assay may have been the main reason that the GP180 marker was failed. Experiments for this marker failed in that none of the characteristic curves were achieved based on the probe. It could have been incorrect probe design, or wrong SNP information. As well, for marker GP180, the amplicon size was higher than 400 bp. The probe design might have failed if the sequencing results for SNP had low accuracy. These problems likely caused the genotyping of GP180 probes to be unsuccessful.

Except for the two mentioned limitations of the HRM, the unlabelled probe assay was very effective in detecting polymorphisms in the 8 tetraploid clones and parents. The unlabelled probe assay allowed a better classification of the allelic composition of the genotypes. In this study, the probe can easily be used to distinguish as many as five

different genotypes. For example, the genotypes reflected the ratio of the perfect match allele to the mismatch allele. This means that the genotype of a tetraploid could be 4:0, 0:4, 2:2, 1:3, and 3:1 (Figure 17). The results in Figure 17 indicated that HRM was a dosage sensitive method for tetraploid genotyping. Also, in some situations, although the progeny did not segregate, the parents showed different genotypes. As De Koeyer et al. (2010) mentioned, HRM unlabelled probe genotyping could be used to select parents or progeny with multiple copies of a desirable marker. This advantage is also reflected in this study.

6.3 Phenotypic Variance Explained by QTLs

Estimates of the phenotypic variance explained by the QTLs in the diploid population 14946 were less than those of the previous population 13610. Regarding the previous population, 39.9% of the variance for ACD was explained by the four QTLs in the diploid population 13610. However, only 28.5% of ACD was explained by a total of six QTLs in the diploid population 14946. This means that the four major QTL regions in population 13610 have larger effects on ACD than that the six QTL regions in the population of 14946. Regarding the percentage of ACD explained by the single marker in the diploid population 14946, LS030, LS163, LS079, LS037, LS112, and LS161 contributed 9.2%, 1.3%, 2.5%, 10.2%, 5.8%, and 0.8% effects to the degree of ACD, respectively (Table 14).

Of the six QTLs, chromosomes 3 (LS163) and 10 (LS161) have the lowest contribution to ACD compared with the other QTLs. Not only from the overall data but also from the individual trial's data, both QTLs have the lowest percentage value. However, when the QTLs on chromosomes 3 and 10 were removed from the calculation, the total percentage of ACD explained by markers decreased slightly from 28.5% to 25.9%. This

also illustrated that the QTLs on chromosome 3 and 10 had less effects on the degree of ACD than the other four QTLs.

The phenotypic variance explained by six QTLs in the diploid population 14946 was lower than that in the study reported by Bradshaw et al. (2008). QTLs on chromosomes of 2, 4, and 6 were common in both studies. Bradshaw et al. (2008) identified six QTLs of ACD that accounted for 49% of the phenotypic variance in a tetraploid potato population. They generated two genetic maps based on both parents (12601 and Stirling) by using 514 AFLP and SSR markers. Since there were no common markers used in the 13610 map and the study from Bradshaw et al., it is not possible to compare the QTL locations within the common chromosomes. The higher number of markers used (514), larger population size (227), and replicated field trials over 3 years likely contributed to the identification of the QTLs explaining more of the phenotypic variance in Bradshaw's study than observed in our studies.

Chapter 7. Direction of Future Research

The four major QTLs on chromosomes 2, 4, 6, and 10 in both diploid populations 13610 and 14946 now allow future research to be conducted in areas of MAS and gene functional analysis. The QTLs could be used for MAS in the breeding programs. The markers related to the QTLs could be used for parent selection in MAS. For instance, the marker LS030 with *m* allele is associated with low ACD, whereas the *l* allele indicates high ACD. The desirable clones with *m* allele of LS030 could be used as a parental clone to produce new cultivars. Subsequently, the same marker allele would be used for selection in the next breeding cycle. This means that this marker could be employed in each generation to select desirable clones. Also, the HRM proved to be a fast and accurate method for genotyping diploid and tetraploid potatoes. The successful unlabelled probes designed in this study could be employed in tetraploid genotyping, saving a lot of time in the selection process.

The identified QTLs provide additional information in the search of more genes related to ACD. For example, the marker LS037 in coding *c4h* gene was considered as the most significant marker close to the QTL on chromosome 6. The marker itself was not shown to be the reason for causing ACD changes. Therefore, the future work could focus on searching markers or genes between the marker LS037 and the ACD QTL on chromosome 6.

The genes related to ACD could be further analyzed to learn more about their functions. Also, gene expression studies could be used for verifying the relationship between candidate genes and ACD. If the gene has a negative function by enhancing ACD, RNA interference (RNAi) experiments could be used for inhibiting the expression of the

gene to lower ACD production. If the gene has a positive effect, being correlated to a low ACD, over-expression of this gene in a transgenic plant could lead to new cultivars with low levels of ACD.

Recently, a homozygous doubled-monoploid potato clone was used to sequence and assemble 84% of the 844-megabase potato genome. A total of 39,031 protein-coding genes were found in this study (Xu et al., 2011). Furthermore, the Solanaceae Coordinated Agricultural Project (SolCAP) has released 69,011 SNPs identified from Altantic, Premier Russet, and Snowden (http://solcap.msu.edu/). These SNPs would greatly help for developing more SNP markers useful for QTL identification. Quantified SNP markers and accurate methods of QTL analyses could generate new QTL maps. Since the unlabelled probe assay can be used in tetraploid genotyping, new QTL mapping could be conducted in tetraploid populations.

Chapter 8. Conclusion

This project successfully achieved its two objectives. First of all, the markers from the 13610 population, which were found to be informative of ACD levels, worked equally well for the second breeding population 14946. Fourteen SNP markers located in eight chromosomes of the genome were tested in the diploid population 14946 by using HRM method. Single marker analysis indicated that 11 identified SNP markers had a significant relationship to the degree of ACD in the diploid population 14946. The results of marker genotyping illustrated that amplicon genotyping by HRM is a very useful and a convenient method for diploid clones in potatoes. Based on the 11 validated markers, six QTL regions on chromosomes 2, 3, 4, 6, 8 and 10 were identified in this study. The six QTL regions explained 28.5% of the variation in ACD in the population 14946. Compared to the previous study of the diploid population 13610, the findings in this study were similar for 14946. It suggests that the significant markers in both populations could be applied in the potato breeding program in the future. These markers could be used for selecting parents to breed low ACD potato cultivars.

Second, the HRM unlabelled probe assay was demonstrated to be useful for genotyping tetraploid potatoes. This approach can distinguish five different dosages of alleles in the tetraploids. Although the number of the segregating markers was low, three significant markers showing a relationship to ACD in the tetraploid population 14945 were identified. These results suggest that the unlabelled probe assay can be successfully used in genotyping tetraploid potatoes. Therefore, using these validated markers with the HRM method would be another option for selecting low ACD lines in MAS.

References

Andre, C. M., Ghislain, M., Bertin, P., Oufir, M. Del Rosario Herrera, M., Hoffmann, L., Hausman, J. F., Larondelle, Y., and Evers, D. 2007. Andean potato cultivars (Solanum tuberosum L.) as a source of antioxidant and mineral micronutrients. *J. Agric. Food Chem.* 55:366-378.

Augustin, J. 1975. Variations in the nutritional composition of fresh potatoes. *J. Food Sci.* 40:1295-1299.

Babin, S., Narayanan, K., Murphy, A., De Jong, H., and Wang-Pruski, G. 2011. Identification of the gene encoding cinnamic acid 4-hydroxylase and its function to after-cooking darkening in potatoes. *Am. J. Plant Sci. Biotechnol.* 5:112-121.

Bachem, C. W. B., Speckmann, G., van der Linde, P., Verheggen, F., Hunt, M., Steffens, J, and Zabeau, M. 1994. Antisense expression of polyphenol oxidase genes inhibits enzymatic browning in potato tubers. *Bio. Technol.* 12:1101-1105.

Barker, J. 1938. Changes of sugar content and respiration in potatoes stored at different temperatures. *Rep. Fd. Invest.* 1937:175-177.

Barone, A. 2004. Molecular marker-assisted selection for potato breeding. *Am. J. Potato Res.* 81:111-117.

Beckmann, J. S. and Soller, M. 1983. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping, and costs. *Theor. Appl. Genet.* 67:35-43.

Bonierbale, M., 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. *Genet.* 120:1095-1103.

Booth, A. 1963. The role of growth substances in the development of stolons. In: Ivins, J. D. and Milthorpe, F. L. (Ed.), The Growth of the potato, *Butterworths, London*. pp: 99-113.

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

Bradshaw, J. E., Hackett, C. A., Pande, B., Waugh, R., and Bryan, G. J. 2008. QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum subsp. tuberosum*). *Theor. Appl. Genet.* 116:193-211.

Brigneti, G., Garcia-Mas, J., and Baulcombe, D. C. 1997. Molecular mapping of potato virus Y resistance gene Ry_{sto} in potato. *Theor. Appl. Genet.* 94:198-203.

Brookes, A. J. 1999. The essence of SNPs. *Gene.* 234:177-186.

Cantle, S. 2004. Gene structure and expression of cinnamic acid 4-hydroxylase in potato and its relationship to after-cooking darkening. M.Sc. Thesis. Nova Scotia Agricultural College, Nova Scotia, Canada.

Cantle, S. and Wang-Pruski, G. 2005. Genbank. Accession number: bankit776390 DQ341174. Available at: http://www.ncbi.nlm.nih.gov/ [Accessed: 2010 November 2]

Ching, A., Caldwell, K. S., Jung, M., Dolan, M., Smith, O. S., Tingey, S., Morgante, M., and Rafalski, A. J. 2002. SNP frequency, haplotype structure and linkage disequibrium in elite maize inbred lines. *BMC Genet.* 3:19-33.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J.B., and Pang, E.C.K. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphyt.* 142:169–196.

Corsini, D. L., Pavek, J. J., and Dean, B. 1992. Differences in free and protein-bound tyrosine among potato genotypes and their relationship to internal blackspot resistance. *Am. Pot. J.* 69:423-434.

Cutter, E. G. 1992. Structure and development of the potato plant. In: Harris, P. M. (Ed.), The potato crop: The scientific basis for improvement. *Chapman and Hall, New York. NY*. pp: 65-161.

Dalianis, C. D., Plaisted, R. L., and Peterson, L. C. 1966. Selection for freedom from ACD in a potato breeding program. *Am. Potato J.* 43:207-215.

De Koeyer, D., Douglass, K., Murphy, A., Whitney, S., Nolan, L., Song, Y., and De Jong, W. 2010. Application of high-resolution DNA melting for genotyping and variant scanning of diploid and autotetraploid potato. *Mol. Breed.* 25:67-90.

Denny, F. E. 1929. Role of mother tuber in growth of potato plant. *Bot. Gaz.* 87:157-194.

D'hoop, B. B. 2009. Association mapping in tetraploid potato. Ph.D. Thesis. Wageningen University, Netherlands.

Doyle, J. J. and Doyle, J. L. 1990. Isolation of plant DNA from fresh tissue. *Focus* 1: 13-15.

Eagles, H. A., Bariana, H. S., Ogbonnaya, F. C., Rebetzke, G. J., Hollamby, G. L., Henry, R. J., Henschke, P. H., and Garter, M. 2001. Implementation of markers in Australian wheat breeding. *Aust. J. Agric. Res.* 52:1349-1356.

Fairbanks, D. J. and Andersen, W. R. 1999. Quantitative genetics. In: Genetics: the continuity of life. *Pacific Grove, Canada. Wadsworth*. pp: 608-625.

Fairweather-Tait, S. 1983. Studies on the availability of iron in potatoes. *Brit. J. Nutr.* 50:15-23.

Falconer, D. S. and Mackay, T. F. C. 1996. Quantitative trait loci. In: Introduction to quantitative genetics. Fourth edition. *Longman Group Ltd. England.* pp:356-361.

FAO. 2007. Available from http://faostat.fao.org/default. [Accessed: 2009 February 17]

FAO. 2010. Available from http://faostat.fao.org/default. [Accessed: 2010 October 12]

Feingold, S., Lloyd, J., Norero, N., Bonierbale, M., and Lorenzen, J. 2005. Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). *Theor. Appl. Genet.* 111:456-466.

Flinn, B., Rothwell, C. Griffiths, R., Lague, M., De Koeyer, D., Sardana, R., Audy, P., Goyer, C., Li, X. Q., and Wang-Pruski, G. 2005. Potato expressed sequence tag generation and analysis using standard and unique cDNA libraries. *Plant Mol. Biol.* 59:407-433.

Friedman, M. 1997. Chemistry, biochemistry, and dietary role of potato polyphenols. *J. Agric. Food Chem.* 45:1523-1540.

Gallais, A. 2003. Quantitative genetics and breeding methods in autopolyploid plants. *Institut National de la Recherche. Agronomique, Paris*. 515p.

Glendinning, D. R. 1983. Potato introductions and breeding up to the early 20th century. *New Phytol.* 94:479-505.

Griffiths, D. and Bain, H. 1997. Photo-induced changes in the concentrations of individual chlorogenic acid isomers in potato (*Solanum tuberosum*) tubers and their complexation with ferric ions. *Potato Res.* 40:307-315.

Hawkes, J. G. 1988. The evolution of cultivated potatoes and their tuber-bearing wild relatives. *Biomed. Life Sci.* 36:189-208.

Hawkes, J. G. 1992. Biosystematics of the potato. In: Harris, P. (Ed.), The potato crop. *Chapman & Hall, New York, USA*. pp: 13-14.

Hawkes, J. G. 1994. Origins of cultivated potatoes and species relationships. In: Bradshaw, J. E. and Mackay, G. R. (Ed.), Potato genetics. *CAB International Publishers, Wallingford, UK.* pp: 3-5.

Hegeny, M, A. and McPharlin, I. R. 2000. Response of summer-planted potatoes to level of applied nitrogen and water. *J. Plant. Nutr.* 23:197-218.

Heisler, E. G., Sicilano, J., Treadway, R. H., and Woodward, C. F. 1963. After cooking discoloration in potatoes. Iron content in relation to blackening tendency of tissues. *J. Food Sci.* 28:453-459.

Hofinger, B. J. Jing, H. C., Hammond-Kosack, K. E., and Kanyuka, K. 2009. High-resolution melting analysis of cDNA-derived PCR amplicons for rapid and cost-effective identification of novel alleles in barley. *Theor. Appl. Genet.* 119:851-865.

Howell, W. M., Jobs, M., Gyllensten, U., and Brooks, A. J. 1999. Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat. Biotechnol.* 17:87-88.

Hughes, J. C. and Evans, J. L. 1967. Studies on after-cooking blackening in potatoes. IV. Field experiments. *Eur. Potato J.* 10:16-36.

Hughes, J. C. and Swain, T. 1962a. After-cooking blackening in potatoes. II. Core experiments. *J. Sci. Food Agric.* 13:229-236.

Hughes, J. C. and Swain, T. 1962b. After-cooking blackening in potatoes. III. Examination of the interaction of factors by *in vitro* experiments. *J. Sci. Food Agric.* 13:358-363.

Jones, C. J., Edwards, K. J., Castaglione, S., Winfield, M. O., Sala, F., van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volkaert, G., Rueda. J., Linaeero, R., Vazquez, A., and Karp, A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by network of European laboratories. *Mol. Breed.* 3:381-390.

Kaldy, M. S. and Lynch, D. R. 1983. Chlorogenic acid content of Russet Burbank potato. *Am. Potato. J.* 60:375-377.

Killick, R. J. 1977. Genetic analysis of several traits in potato by means of a diallel cross. *Ann. Appl. Biol.* 86:279-289.

Klug, W. S., Cummings, M. R., Spencer, C. A., and Palladino, M. A. 2006. Concepts of genetics. Ninth Edition, *Pearson Benjamin Cummings, San Francisco, US.* pp:678-679.

Ky, C. L., Louarn, J., Guyot, B., Charrier, A., Hamon, S., and Noirot, M. 1999. Relations between and inheritance of chlorogenic acid contents in an interspecific cross between Coffea pseudozanguebarizae and Coffea liberica var ‘deweveri’. *Theor. Appl. Genet.* 98:628-637.

Lay, M. and Wittwer, C. 1997. Real-time fluorescence genotyping of factor V leiden during rapid cycle PCR. *Clin. Chem.* 43:2262-2267.

Lehmensiek, A., Sutherland, M. W., and McNamara, R. B. 2008. The use of high resolution melting (HRM) to map single nucleotide polymorphism markers linked to a covered smut resistance gene in barley. *Theor. Appl. Genet.* 117:653-669.

LeRiche E. 2007. A study of potato (Solanum tuberosum L.) tuber element distribution and its implication for after-cooking darkening. M.Sc. Thesis, Nova Scotia Agricultural College, Nova Scotia, Canada.

Lopez De Rommana, G., Graham, G. G. Mellits, E. D., and Maclean, W. C. 1980. Utilization of the protein and energy of the white potato by human infants. *J. Nutr.* 110:1849-1857.

Mackay, J. F., Wright, C. D., and Bonfiglioli, R. G. 2008. A new approach to varietal identification in plants by microsatellite high resolution melting analysis: application to the verification of grapevine and olive cultivars. *Plant Meth.* 4:8.

Mackill, D. J., Nguyen, H. T., and Zhang, J. X. 1999. Use of molecular markers in plant improvement programs for rainfed lowland rice. *Field Crops Res.* 64:177-185.

Mazza, G. and Qi, H. 1991. Control of after-cooking darkening in potatoes during growth and long-term storage. *Am. Potato J.* 60:145-159.

Meksem, K., Leister, D., Paleman, J., Zebeau, M., Salamini, F., and Gebhardt, C. 1995. A high-resolution map of chromosome V based on RFLP and AFLP markers in the vicinity of the R1 locus. *Mol. Gen. Genet.* 249:74-81.

Meyer R.C., Milbourne, D., Hackett, C.A., Bradshaw, J.E., McNichol, J.W., and Waugh, R. 1998. Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (*Phytophthora infestans*) *Mol. Gen. Genet.* 259:150-160.

Milbourne, D., Meyer, R. C., Collins, A. J., Ramsay, L. D., Gebhardt, C. and Waugh, R. 1998. Isolation, characterisation and mapping of simple sequence repeat loci in potato. *Mol. Gen. Genet.* 259:233-245.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R., and Sasaki, T. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed.* 3:87-103.

Montgomery, D. C. 2001. Design and analysis of experiments. 5th Ed. John Wiley& Sons, *Inc. New York*. pp:25-44.

Montgomery, J., Wittwer, C. T., Palais, R., and Zhou, L. 2007. Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. *Nat. Protoc.* 2:59-66

Murphy, P. J., Kong, F., Pinto, D. M., and Wang-Pruski, G. 2010. Relative quantitative proteomic analysis reveals wound response proteins correlated with after-cooking darkening. *Proteom.* 23:4258-4269.

Narayanan, K. 2005. Gene expression studies in potato and their relationship to after-cooking darkening. M.Sc. Thesis. Dalhousie University, Nova Scotia, Canada.

Oberacher, H., Parson, W., Holzl, G., Oefner, P. J., and Huber, C. G. 2004. Optimized suppression of adducts in polymerase chain reaction products for semi-quantitative SNP genotyping by liquid chromatography-mass spectrometry. *J. Am. Soc. Mass Spectrom.* 15:1897-1906.

Pika, N. A., Tarasenko, V. A., and Mitsko, V. N. 1984. Combining ability of potato varieties and hybrids for tuber flesh blackening after cooking. Selektsiia-i-Semenorodstro, USSR 7. pp: 16-17.

Razdan, M. K. and Mattoo, A. K. 2005. Molecular markers in identification of genotypic variation. In: Genetic Improvement of Solananceous crops. *Science Publishers Inc., England, NH, USA*. pp:115-130.

Reed, G. H., Kent, J. O., and Wittwer, C. T. 2007. High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogen.* 8:597-608.

Reeve, R. M., Hautala, E., and Weaver, M. L. 1969. Anatomy and compositional variation within potatoes II. Phenolics, enzymes, and other minor components. *Am. Potato J.* 46:374-386.

Rickert, A. M., Premstaller, A., Gebhardt, C., and Oefner, P. J. 2002. Genotyping of SNPs in a polyploidy genome by pyrosequencing. *Biotech.* 32:592-598.

Rickert, A. M., Ballvora, A., Matzner, U., Klemm, M., and Gebhardt, C. 2005. Quantitative genotyping of single-nucleotide polymorphisms by allele-specific oligonucleotide hybridization on DNA microarrays. *Biotechnol. Appl. Biochem.* 42:93-96.

Rieman, G. H., Tottingham, W. E., and McFarlane, J. S. 1944. Potato varieties in relation to blackening after cooking. *J. Agric. Res.* 69: 21-31.

Ronaghi, M., Karamoharmed, S., Petterson, B., Uhlen, M., and Nyren, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242:84-89.

Ronning, C. M., Stommel, J. R., Kowalski, S. P., Sanford, L. L., Kobayashi, R. S., and Pineada, O. 1999. Identification of molecular markers associated with leptine production in a population of *Solanum chacoence* Bitter. *Theor. Appl. Genet.* 98:39-46.

Rouppe van der Voort, J. N. A. M., van Eck, P., Folkertsma, H. J., Hutten, R. T., Draaistra, R. C. B., Gommers, J., Jacbsen, F. J., Helder, E., and Baker, J. 1997. Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. *Mol. Gen. Genet.* 255:438-447.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., 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., Marth, G., Sherry, S., 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., Etten, W. J. V., Reich, D., Higgins, J. Daly, M. J., Blumenstiel, D., 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.

Salaman, R. N. 1949. The history and social influence of the potato. In: Hawkes, J. G. (Ed.), *Cambridge University Press, Cambridge*. pp:26-29.

SantaLucia, J. 2007. Physical principles and visual-OMP software for optimal PCR design. *Methods Mol. Biol.* 402:3-34.

Sattarzadeh, A., Achenbach, U., Lubeck, J., Strahwald, J., Tacke, E., Hofferbert, H. R., Rothsteyn, T., and Gebhardt, C. 2006. Single nucleotide polymorphism (SNP) genotyping as basis for developing a PCR-based marker highly diagnostic for potato varieties with high resistance to *Globodera pallida* pathotype Pa2/3. *Mol. Breed.* 18: 301-312.

Shekhar, V. C., Iritani, W. M., and Arteca, R. 1978. Changes in ascorbic acid content during growth and short-term storage of potato tubers (*Solanum tuberosum* L.). *Am. Potato J.* 55:663-670.

Simko, I., Pechenick, D. A., McHale, L. K. Truco, M. J., Ochoa, O. E., Michelmore, R. W., and Scheffler, B. E. 2009. Association mapping and marker-assisted selection of lettuce dieback resistance gene *Tvrl*. *BMC Plant Biol.* 9:135.

Smith, O. and Nash, L. B. 1940. Potato quality I. Relation of fertilizers and rotation systems to specific gravity and cooking quality. *Am. Potato J.* 17:163-169.

Smith, O. 1987. Effect of cultural and environmental conditions on potatoes for processing. In: Talburt, W. F. and Smith, O. (Ed.), Potato processing. 4th ed. *Van Nostrand Reihold Company Inc., New York.* pp: 108-110.

Song, J., Ohkura, T., Sugimoto, M., Mori, Y., Inagi, R., Yamanishi, K., Yoshizaki, K., and Nishimoto, N. 2002. Human interleukin-6 induces human herpesirus-8 replication in a body cavity-based lymphpma cell line. *J. Med. Virol.* 68:404-411.

Stark, J. C. and Love, S. L. 2003. Harvest management. In: Potato production systems. *University of Idaho Agricultural Communications, USA*. pp: 345-359.

STATCAN. 2011. Available from http://www.statcan.gc.ca/start-debut-eng.html [Accessed: 2011 February 20]

Storey, R. M. J. and Davies, H. V. 1992. Tuber quality. In: Harris, P. M. (Ed.), The Potato Crop: The scientific basis for improvement. *Chapman and Hall, London.* pp: 508-530.

Stuber, C. W., Polacco, M., and Lynn, S. M. 1999. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Sci.* 39:1571-1583.

Subar, A. F., Krebs-Smith, S. M., Cook, A., and Kahle, L. L. 1998. Dietary source of nutrients among US adults, 1989 to 1991. *J. Am. Diet. Assoc.* 98:537-547.

Sun, G., Wang-Pruski, G., and Mayich, M. 2003. RAPD and pedigree-based genetic diversity estimates in cultivated diploid potato hybrids. *Theor. Appl. Genet.* 107:110-115.

Swiniarski, E. 1968. After-cooking darkening and some chemical constituents of potato tubers. *Hod. Rosl. Akl. Nas.* 12:369-384.

Talburt, W. F., Schwimmer, S., and Burr, H. 1987. Structural and chemical composition of potato tuber. In: Talburt, W. F. and Smith, O. (Ed.), Potato processing. *AVI Publishing Co., Westport, CT.* pp: 11-39.

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

Thomson, N. R. 1987. Potato cultivars. In: Talburt, W. F. and Smith, O. (Ed.), Potato processing. 4th ed. *Van Nostrand Reihold Company Inc., New York.* pp: 107-119.

Topley, H. 2004. Gene expression analysis of potato (*Solanum tuberosum* L. cv. Russet Burbank) tubers during long-term storage and its relationship with after-cooking darkening. M.Sc. Thesis, Nova Scotia Agricultural College, Nova Scotia, Canada.

Twyman, R. M. 2005. Single nucleotide polymorphism (SNP) genotyping techniques—An overview. *Encycl. Diagn. Genom. Proteom.* DOI: 10.1081/E-EDGP 120020761. pp: 1202-1207.

Van Eck, H. J., Rouppe van der Voort, J., Draaistra, J., van Zandvoort, P., van Enckevort, E., Segers, B., Peleman, J., Jacoben, E., Helder, J., and Bakker, J. 1995. The inheritance and chromosomal location of AFLP markers in a non-inbred potato offspring. *Mol. Breed.* 1:397-410.

Vreugdenhil, D. and Bradshaw, J. 2007. Potato biology and biotechnology. Oxford UK, *Elsevier.* pp: 55-84; 471-500.

Wang-Pruski, G**.** 2006. Digital imaging for evaluation of potato after-cooking darkening and its comparison with other methods. *Int. J. Food Sci. Technol*. 41 (8): 885-891.

Wang-Pruski, G. 2007. The Canon of Potato Science: 47. After-cooking darkening. *Potato Res.* 50:403–406.

Wang-Pruski, G., Astatkie, T., De Jong, H., and Leclerc, Y. 2003. Genetic and environmental interactions affecting potato after-cooking darkening. *Acta. Hort.* 619:45-52.

Wang-Pruski, G. and Nowak, J. 2004. Potato after-cooking darkening. *Am. J. Potato Res.* 81:7-16.

Wang-Pruski, G., Cantle, S., and Narayanan, K. 2005. Patent Filed: United States and Canada. Cinnamic acid 4-hydroxylase. No.2,510,950;11/452,342.

Wang-Pruski, G., Griffiths, R., Lagüe, M., Murphy, P., DeKoeyer, D., Pinto, D. Rothwell, C., Gustafson, V., Flinn, B., and Regan, S. 2006. Identification of genes related to potato after-cooking darkening using potato genomic and proteomic approaches. Plant Genomics European Meetings, October 11-14, Venice, Italy.

Wang-Pruski, G., Zebarth, B. J., Leclerc, Y., Arsenault, W. J., Botha, E. J., Moorehead, S., and Ronis, D. 2007. Effect of soil type and nutrient management on potato after-cooking darkening. *Am. J. Potato Res.* 84:291-299.

Wang-Pruski, G., Babin, S., Murphy, P. Kong, F., Wu Y., Douglass, K., Murphy, A., De Koeyer, D., and Pinto, D. 2011. Identification of genes and gene markers by genomics, genome mapping and comparative proteomic and their relationship to potato tuber qualities. *ActaHort in press*.

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.* 18:6531-6535.

Yamamoto, Y. Y., Yoshitsugu, T., Sakurai, T., Scki, M., Shinozaki, K. and Obokata, J. 2009. Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *Plant J.* 60:350-362.

Young, N. D. 1996. QTL mapping and quantitative disease resistance in plants. *Ann. Rev. Phytopathol.* 34:479-501.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., et al. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475:189-195.

Zane, L., Bargelloni, L. and Patarnello, T. 2002. Strategies for microsatellite isolation: a review. *Mol. Ecol.* 11:1-16.

Zhou, L. M., Wang, L., Palais, R., Pryor, R. and Wittwer, C. T. 2005. High-resolution DNA melting analysis for simultaneous mutation scanning and genotyping in solution. *Clin. Chem.* 51:1770-1777.

Appendix I

Fourteen marker sequences represented by the transcript sequences of the genes. The underlined sequences indicate the primers of each marker. Each gene sequence can be obtained from NCBI database or Potato Genome Sequencing Consortium based on the accession numbers given in the brackets following each marker name.

LS057 (AY325818)

ATGCCGTCTCATGCAGATCTAGATCGGCAGATCGAACAATTGATGGAGTGCAAGCCGTTA
TCGGAAGCTGATGTGAAAACTCTGTGTGATCAAGCGAGGGCTATACTTGTTGAGGAATGG
AATGTTCAACCGGTGAAATGTCCGGTGACTGTTTGTGGTGATATTCATGGACAGTTTTAC
GACCTGATTGAGCTGTTTCGGATTGGAGGAAATGCTCCTGATACTAATTATCTCTTCATG
GGCGATTATGTTGACCGTGGATACTATTCAGTGGAGACTGTCACACTTTTGGTTGCTCTG
AAGGTTCGTTATAGAGATAGAATCACGATTCTTAGGGGAAATCATGAAAGCCGGCAAATC
ACCCAAGTGTATGGTTTTTATGATGAATGCTTGAGGAAATATGGCAATGCAAATGTTTGG
AAGTATTTCACTGATCTTTTTGATTATCTACCACTAACAGCACTGATAGAGAGTCAGATA
TTCTGTTTGCATGGAGGACTCTCACCTTCTCTGGATACACTGGATAATATCCGAGCGCTG
GACCGTATACAAGAGGTTCCACATGAAGGGCCAATGTGTGA

LS030 (AY289923)

ATGGCAAGTGAAGTTCATGCAGTTGTTGATGCCCATTCTCCTCCGAAGACGCCAACGGTT
TGCGTCACAGGAGCAGCTGGATTTATCGGCTCTTGGCTTGTCATGAGACTCCTTGAACGC
GGTTATAATGTCCATGCTACTGTTCGTGATCCTGGAACCAGAAGAAGGTGAAACATCTGT
TGGAACTGCCAAAAGCTGATACAAACTTAACGCTGTGGAAAGCAGACTTGGCAGTGGAAG
GAAGCTTTGATGAAGCCATTCAAGGCTGTCAAGGAGTATTTCATGTGGCTACACCGATGG
ATTTCGAGTCCAAGGACCCAGAGGTACTATAAAGTAACAGAGTAACTCATCAATACTAAT
GGACTTCTTACATTTCTGGATGATTAATACTAACATAACTCTTATGACATTGATATACGT
GACAGAACGAAGTAATTAAACCAACAGTCAGGGGAGTGTTAAGTATCATAGAATCATGTG
CTAAAGCTAACACAGTGAAGAGGCTGGTTTTCACTTCATCTGCTGGAGCTCTTGATGTCC
AAGAGGACCAAAAACTCTTTTGTGACGAGACCAGCTGGAGCGATTTGGACTTCATATATG
CTAAGAAGATGACAGGATGG

LS124 (GQ337857)

ATGTTGTTGAAGGAAAAACAGCAGCCAAGGAGGCTTTGCAGCGAAAACTTGGATTGAAAC
AGGCTGACCTTCCTTTGGTAGGAATTATCACCCGCTTAACTCACCAGAAAGGAATCCACC
TCATTAAACATGCTATTTGGCGCACCTTGGAACGGAACGGACAGGTAGTCTTGCTTGGTT
CTGCTCCTGATCCTAGGGTACAAAACGATTTTGTTAATTTGGCAAATCAATTGCACTCCA
AATATAATGACCGCGCACGACTCTGTCTAACATATGACGAGCCACTTTCTCACCTGATAT
ATGCTGGTGCTGATTTTATTCTAGTTCCTTCAATATTTGAGCCATGTGGACTAACACAAC
TTACCGCTATGAGATATGGTTCAATTCCAGTCGTGCGTAAAACTGGAGGACTTTATGATA
CTGTATTTGATGTTGACCATGACAAAGAGAGAGCACAGCAGTGTGGTCTTGAACCAAATG
GATTCAGCTTTGATGGAGCAGATGCTGGCGGAGTTGATTATGCTCTGAATAGAGCTCTCT
CTGCTTGGTACGATGGTCGGGATTGGTTCAACTCTTTATGCAAGCAGGTCATGGAACAAG
ATTGGTCTTGGAACCGACCTGCTCTTGATTATTTGGAGCTTTACCATGCTGCTAGAAAGT
TAGAATAG

BCH100 (HM013963)
ATGGCGGCCGGAATTTCAGCCTCCGCTGGTTCCCGAACTATTTGCCTCCGTCACAACCCG
TTTCGCGGTCCAAAATCCGCCTCAACCGCCCCGCCAGTTCTGTTCTTCTCTCCGTTAACT
CGCAATTTTGGCGCAATTTTGCTGTCTCGAAGAAAGCCGAGGTTGGCGGTTTGCTTTGTG
CTGAAGAATGAGAAATTGAATAGTACTATCCAAAGTGAGTGTGAAGTAATACAGGATCAG
ATACAAGTAGAGATTAATGAGGAGAAGAGTTTAGCTGCCTGTTGGCTGGCGGAAAAATTG
GCGAGGAAGAAATCGGAGAGGTTTACTTATCTTGTGGCGGCTGTCATGTCTGTTTAG

LS163 (PGSC0003DMT400036743)
CACCATTCCTCGAATCATCAATTCCTCTAAAACAGCATCTGTTGCAACCTCCACACTCCC
GTCTCACAATCTCTACAGGCCTTTAGCTGGAACATTTTCGCGTTTTATTTCTTCTAATGA
GACTTCATCGCCGATGAATATCAATCTTTCTGACGAAGAGAGCAAAAGACGTTTGTTTAA
CAGGCTT

LS079 (PGSC0003DMT400040781)
AAAAAATGTCATCATCTGGTTCATTCAAGTTGCAACCTCCAATTAATGGTAAATTCATCA
CAATTCTCAGCATCGATGGAGGTGGCATTAGAGGATTAATTCCTGCAACTATTCTTGAAT
ATCTTGAATCTCAGCTCCAGGAATTAGACGGAGAGGATGCTAGGCTCGCAGATTACTTTG
ATATAATCGCAGGAACAAGCACGGGTGGCCTTGCTACCGCTATGCTAACTGGTCCAAACA
AAGATAATCGTCCTCTATTTGCTGCTAAGGATATAAAAGCCCTTTCTATCTTGAACCATG
GTCCTAACGATATTTCCACAAAATCAAAGTATGATGGGAAGTATCTTCCACAAGTCCTAA
AGGACAAATTAGGGCGACCTAATTGCGTGACGACTTTGACTACTGTCGTTATTCCAACTT
TTTGATATCCAGACATATACAAACTTAAATTTTTCTCCCCTTTTAAGGCCAAAA

GP180 (PGSC0003DMT400019304)
TCCACCCCTGGATTCAAGAAAGGCCGCCATTTGCCCTCGAATGCTCAATTTTTTTAGATG
ACACGCCTTTGCCAGCACATCTTTCTTGCAACTTTCTGGTGCAACCATAGAAAGAGTTTG
TAGACAAGAAGGTTTACCTGTTGTGACAGTCGGGGATGGCAATTTTGCAGGTATGCTGGT
GTGCAGATGCCTCAACTGTAACATGTTCCATATGTCTGCTTTTACGTCAAGAGTGGGCTC
TGAGGTACTTGTATTAAGTATAAGAGTTTGTAAATTCCAGAATTTACCAAAGGTCGGAGG
GAGGGCCTTAAAGTCACCTGAGATAGCAATATACCTCAAATGAAATAAATTGTTAAAGTC
TTTGGAGAAAAGAAATTTGAGAGATTCAACATCCAAGACTCTCATAAGTGGAAATGCTTT
GTGAAGGAGTTTGA

LS037 (DQ34174)
ATGGATCTTCTCTTACTGGAGAAGACCTTAATAGGTCTTTTCTTTGCTATTTTAATCGCT
ATTATTGTCTCTAAACTTCGTTCCAAGCGATTTAAACTACCCCCAGGTCCAATTCCAGTC
CCAGTTTTTGGAAATTGGCTTCAAGTTGGTGATGATTTGAACCATAGAAACCTTACTGAG
TATGCTAAAAAGTTTGGTGATGTGTTCTTGCTTAGAATGGGGCAAAGGAACTTAGTTGTT
GTGTCATCCCCTGAATTAGCTAAAGAAGTTTTACACACACAAGGGGTTGAATTTGGTTCA
AGAACAAGAAATGTTGTTTTTGATATTTTTACAGGGAAGGGTCAAGATATGGTTTTTACA
GTGTATGGTGAGCACTGGAGGAAAATGAGGAGGATTATGACTGTACCCTTTTTTACTAAT
AAGGTGGTGCAGCAGTATAGAGGGGGGTGGGAGTCTGAGGCTGCTAGTGTAGTTGAGGAT
GTGAAGAAAAACCCTGAATCTGCTACAAATGGGATTGTTTTGAGGAAAAGATTGCAGCTT
ATGATGTATAATAACATGTTTAGGATTATGTTTGATAGGAGATTTGAGAGTGAAGATGAT
CCCCTTTTTGTTAAGCTTAGGGCTTTGAATGGTGAGAGGAGTAGATTGGCTCAGAGCTTT
GAGTACAACTATGGTGATTTTATCCCTATTTTGAGGCCTTTCTTGAGAGGGTACTTGAAG

ATTTGTAAGGAGGTTAAGGAGAAGAGGTTGAAGCTATTCAAAGACTACTTTGTTGATGA

LS097 (PGSC0003DMT400015673)
TCTCAAAGAAGAACCGGGATCTTGCTGAACGCCTGCTAACTCGCAAGCCCAGTTATACC
CTTGACATGCTTATTCGAGAGAGGTACCCTAAATTCATTGATGCACTTAGAGATTTGGA
TGACTGCCTTAGTATGGTACACTTGTTTGCTGCATTACCTGCTGTAGAGAGAGAAAAGA
TACCGGTAAACC

LS092 (PGSC0003DMT400062230)
TATGCATCTGCTCCTCTAGGGAACATTGTACCTCCTCCTGGAAATTTGAAAACGTTTCC
CTGAAACTGTACCCAGTTCTGGACTGCCTTCTCAACTGTCAAGTGTTTGTAAGGAACAT
TAGCATAGTAGGCATAATCACGGCTTTTTGGCCATGGAAATGGAGTTGTGTATCCTTTG
GGTGCTAGAATCAGACAACGAAGCTTTTCATCATCTGGAGGGCAATGCCTTTCTCTATA
TATCATATCTTCCCGTGGGAATTTCATAGCTCGGTCTTGTTCATGGCAGGGAGTATAGT
CAGTGTATTTAGCTTCACATGATTTGAATCTTTTAGTTGTTAGTTCTGAAGTTTCAACC
GTTTCCACATAATTATGGTGTGACTCAAAATCCAG

LS112 (CX700086)
CAAGCTGTGCATAGTAGTAGTACATCTCTCAAAACTCCTTTTACTTCTTCCTCCACTGG
CTTTATCTTCCACTCCTAAGCCCTCTCAACTTTTCATCCATGGAAAACGTAACCAAATG
TTCAAAGTTTCATGCAAGGTTACCAATAATAACGGTGACCAAAACCAAAACGTTGAAAC
AAATTCTGTTGATCGAAGAAATGTTCTTCTTGGCTTAGGTGGTCTTTATGGTGTTGCTA
ATGCTATACCATTAGCTGCATCCGCTGCTCCAGCTCCACCTCCTGATCTCTCGTCTTGT
AGTATAGCCAGGATTAACGAAAATCAGGTGGTGCCGTACAGTTGTTGCGCGCCTAAGCC
TGATGATATGGAGAAAGTTCCCGTATTACAAGTTCCCTTCTATGACTAAGCTCCGTGTT
CGTCAGCCTGCCTCATGAAGCTAATGAGGAGTATATTGCCAAGTACAATCTGGCGATTA
GTCGAATGAAAGATCTTGATAAGACACAAACTTTAAACCCTATTGGTTTTAAGCAACAA
GCTAATATACATTGTGCTTATTGTAACGGTGCTTATAGAATTGGTGGCAAAGAGTTACA
AAGTTCATAATTCTTGGCTTTTCTTCCCGTTCCATAGATGGTACTTGTACTTCCACGAG
AGAATCGTGGGAAAATTCATTGATGATCCAACTTTCGCTTTACCATATTGGAATTGGGA
CCATCCAAAAGGTATGCGTTTCCTGCCATGTATGATCGTGAAGGGACTTCCTTTTCGAT
GTACACGTTACAAAGTCACCGAATGGAGCATAATC

LS201 (PGSC0003DMT400045174)
CGGTCGTGAATCCATTGGTAAAGTGTTTAACGGAGAGAAGAAGAAGAAAGGATGTTCTT
CTCCTATGATCTCTCTGAGTTCGGAAGTCAATCGTCGCGGTATACGCCGGGTTTCATCT
GAATCCGACGTGATTCGGTCGGTTTTGAAGGTTCGAGTGTAACTAGAAAGTTGAGTAAG
ATTGGATCGAGGTCATTATCGTCAAAAATACCGGAGGAGGATTGCGGTTCTGAACAGAA
TGAGTTGGAGGTTTTGC

LS161 (PGSC0003DMT400029253)
GTTGAACATGATGATCTAAAGAGGTTGTTTCATGTATCAAAACCAATTAGAGAAGCGAC
TTTGGTTGCAAAAAGATGGCATTTTGAATATAGTACACCAACAAAAACACTTGGTTTTA
AGAATGCTACTGATATAGATAATTGGAGTGAGCCTAATGATGTTGAAGTGCCTAATGCT
CCTAGGCAATCCAAGTTTCTGAGGGCTCGTTTAAGCCGGAAGAAACTAGCTGA

LS173 (PGSC0003DMT400049371)

ATGTTCGTCTACAGTCAAGATAGGGAATTGCTTCTGCAGAAAGGCAAGAAAATGCGCTC
GAGCAGCTTAGCTTCTTACTATCACTTTTGGAAAACCTACTTTGAACTTCTCTTCAACT
CAATGTTC