The identification of new genes through the use of molecular marker technology in barley (Hordeum vulgare L.)
by Deven Robert See

A thesis submitted in partial fulfilment of the requirements for the degree of Masters of Science in Agronomy
Montana State University
© Copyright by Deven Robert See (2000)

Abstract:
Molecular marker technology has permitted many advancements in crop breeding. Three applications of this marker technology are discussed within this paper.

Single nucleotide polymorphisms (SNPs) represent the most prevalent class of mutation within a genome. In this project a new technique for characterizing SNPs is presented. This technique is applicable to any genetic analysis which needs to identify allele states within a segregating population. The SNP technique permits quick identification of both homozygous and heterozygous alleles, and is adaptable to PCR multiplexing.

Quantitative Trait Loci (QTL) have been identified for many agronomically important traits. The identification and location of new grain protein content QTL inherent in Karl is reported. The characterization of these QTL and tightly linked markers should lead to the use of marker assisted selection to decrease grain protein content in malting varieties.

The application of SNP markers in mapping led to the discovery of a multigene family in barley (Hordeum vulgare L). The puroindoline pinA and pinB genes are important factors in the end use milling and halting quality in wheat (Triticum aestivum). Wheat has a single copy of the pinB gene on each of its three chromosome, however only the pinB gene on 5D is expressed. This report demonstrates that barley has multiple, expressed copies of the pinB gene on barley chromosome 7 (5H).

THE IDENTIFICATION OF NEW GENES THROUGH THE USE OF MOLECULAR MARKER TECHNOLOGY IN BARLEY (*HORDEUM VULGARE L.*)

by

Deven Robert See

A thesis submitted in partial fulfilment
of the requirements for the degree

of

Masters of Science

in

Agronomy

MONTANA STATE UNIVERSITY
Bozeman, Montana

September 2000

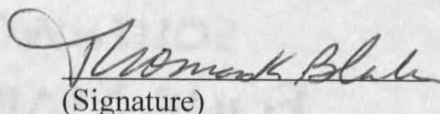
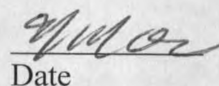
N378
Se31

APPROVAL

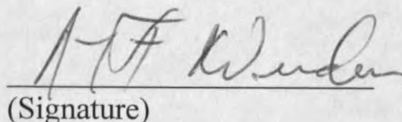
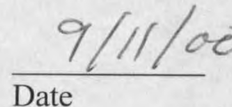
of a thesis submitted by

Deven Robert See

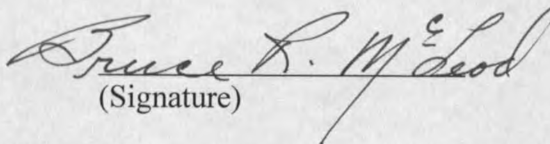
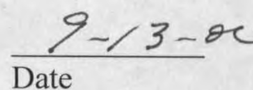
This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Thomas K. Blake  
(Signature) Date

Approved for Department of Plant Sciences

Norman F. Weeden  
(Signature) Date

Approved for the College of Graduate Studies

Bruce R. McLeod  
(Signature) Date

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirement for a masters of science degree at Montana State University, I agree that the library shall make it available to borrowers under rules of the library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in part may be granted only by the copyright holder.

Signature

Date 9-12-00

ACKNOWLEDGMENT

I sincerely appreciate the assistance and guidance of my major advisor, Dr. Tom Blake. I would also like to thank the members of my committee, Dr. Michael Giroux, Dr. Luther Talbert, and Dr. Janice Bowman, whose assistance, friendship and knowledge has been appreciated.

Lastly I would like to thank Dr. Vladimir Kanazin and Hope Talbert whose friendship and technical expertise aided me greatly in my pursuit of this degree.

TABLE OF CONTENTS

Page

LIST OF TABLES....vii
LIST OF FIGURES....viii
ABSTRACT....ix

1. INTRODUCTION....1
   Nitrogen and Malting....1
   Marker Technology....4
   Mapping and Marker Assisted Selection....6

2. SINGLE NUCLEOTIDE POLYMORPHISM DETECTION....9
   Introduction....9
   Materials and Methods....11
      SNP Identification....11
      Production of Fluorescent Product....11
      Detection Platforms....12
   Results....14
   Discussion....18

3. MAPPING THE GENES UNDERLYING PROTEIN CONTENT IN A LEWIS BY KARL CROSS....22
   Introduction....22
   Materials and Methods....24
      Population Development....24
      Protein and Morphological Markers....25
      STS, SNP, and SSR Markers....25
      AFLP Marker Detection....26
      Map Construction....27
      Statistical Analysis....27
   Results....28
      STS, SNP, and SSR Markers....28
      Protein and Morphological Markers....30
      AFLP Markers....30
      Map Construction and QTL Analysis....31
      Statistical Analysis....38
      Protein Quantitative Traits....38
      Heading Date QTL....40

TABLE OF CONTENTS-continued

Plant Height QTL....................40
Yield QTL....................41
Discussion....................43
Markers and Mapping....................43
Quantitative Trait Loci....................44
Marker Assisted Selection....................47

4. *PINB* A MULTIPLE GENE FAMILY IN BARLEY (*HORDEUM VULGARE* L.)....................50
Introduction....................50
Materials and Methods....................51
Results....................53
Sequence Analysis....................53
RT-PCR Analysis of *PinB* Genes....................58
Mapping of *PinB*....................58
Discussion....................59

LITERATURE CITED....................60

LIST OF TABLES

table Page

1. SNP Sequence and Primers.....21
2. Chromosome Alignment.....32
3. Statistical Analysis of Quantitative Traits.....38
4. Correlations Between Quantitative Traits.....38
5. Marker Means.....39
6. Marker Interaction.....39
7. Heritability.....42
8. AFLP Markers.....49
9. PinB Sequences.....55
10. Protein Sequences.....56

LIST OF FIGURES

Figure Page

1. *Fnu*4HI digestion of ABC255 in Steptoe and Morex....................................15
2. Flourescent detection of ABC255 in Steptoe and Morex................................15
3. Fluorescent detection of multiplexed SNP primers..16
4. Fluorescent detection of heterozygous allele at ABC255................................16
5. Segregation of TB33,34 in Lewis and Karl mapping population......................28
6. SNP detection of ABC305 in Lewis and Karl mapping population.................29
7. Protein segregation in Lewis and Karl mapping population.............................30
8. QTL placement on chromosomes...34
9. Genomic digestion of pinB..54
10. Digestion of RT-PCR clones in pinB...57

ABSTRACT

Molecular marker technology has permitted many advancements in crop breeding. Three applications of this marker technology are discussed within this paper.

Single nucleotide polymorphisms (SNPs) represent the most prevalent class of mutation within a genome. In this project a new technique for characterizing SNPs is presented. This technique is applicable to any genetic analysis which needs to identify allele states within a segregating population. The SNP technique permits quick identification of both homozygous and heterozygous alleles, and is adaptable to PCR multiplexing.

Quantitative Trait Loci (QTL) have been identified for many agronomically important traits. The identification and location of new grain protein content QTL inherent in Karl is reported. The characterization of these QTL and tightly linked markers should lead to the use of marker assisted selection to decrease grain protein content in malting varieties.

The application of SNP markers in mapping led to the discovery of a multigene family in barley (*Hordeum vulgare L.*). The puroindoline *pinA* and *pinB* genes are important factors in the end use milling and baking quality in wheat (*Triticum aestivum*). Wheat has a single copy of the pinB gene on each of its three chromosome, however only the *pinB* gene on 5D is expressed. This report demonstrates that barley has multiple, expressed copies of the *pinB* gene on barley chromosome 7 (5H).

CHAPTER 1

INTRODUCTION

Nitrogen and Malting

Barley (*Hordeum vulgare* L.) Is primarily utilized for livestock feed and malting. In beer production one of the most important factors affecting malt quality is grain protein content (Bezant et al., 1996). The American Malting Barley Association AMBA requires six-rowed and two-rowed malt barleys to maintain grain protein lower than 135 and 130g kg^{-1} respectively (Weston et al., 1993). Many studies have shown that while high total grain protein content is good for monogastric feed, it is negatively correlated with malt extract yield (Garcia del Moral et al.,1998, Ulmer et al., 1985).

Trait improvement depends upon the availability of heritable variance, and upon effective selection. While most barley varieties are highly responsive to both fertilizer treatment and moisture stress with respect to grain protein content, one variety, 'Karl' has been shown to be relatively non-responsive. 'Karl' produces grain with low grain protein content under a wide array of environments (Weston et al., 1993). Karl has been used as a parent in many barley improvement programs, and to date no accepted 6-rowed barley variety has been developed which combines low, stable grain protein percentage with acceptable agronomic performance and malting quality. Since variation for grain protein percentage is both reasonably heritable and easily measured, this is either the result of poor application of selection or of undesirable repulsion linkages. In this project we

mapped the chromosomal locations of genes responsible for variation in grain protein content, plant height, flowering date and grain yield. This project was undertaken with the expectation that a better understanding of the genetics underlying variation in grain protein content would help us understand how best to select for desired protein content.

Marker Technology

Genetic markers can be made into useful tools for plant breeding. In plants like barley, morphological markers like rachilla length, awn texture or head type, are easy to identify making them quick determinants of the source of inheritance of specific traits. The limited number of good morphological markers emphasizes a need for more common and informative classes of markers . The characterization of DNA sequence diversity has brought about new genetic markers. The first class of genetic markers based upon sequence differences were restriction fragment length polymorphisms (RFLP) (Botstein et al., 1980). RFLPs rely upon digestion of genomic DNA using restriction endonucleases, blotting (Southern et al., 1975), and hybridization with a radioactively labeled probe (Rigby et al., 1975). Restriction endonucleases cleave DNA at specific recognition sequences producing polymorphic products between alleles of different genotypes.

With the invention of the polymerase chain reaction (PCR) by Mullis and Faloona, (1987), more rapid methods for sequence polymorphism detection became available. PCR permits the exponential amplification of specific DNA sequences. PCR uses primers, short lengths of DNA, in conjunction with buffers and DNA Polymerase II derived from bacteria living in extreme thermal environments. PCR provided the machinery to be able to produce billions of identical copies of DNA from a single copy of sequence present in the genomic DNA. The discovery of PCR technologies by Mullis

opened up molecular biology to new and faster methods for research. These new methods include the use of sequence tagged site (STS) markers (Tragoonrung et al., 1992), simple sequence repeat (SSR) markers (Liu 1996), and amplified fragment length polymorphisms (AFLP) (Vos et al., 1995). STS markers are usually derived from known sequences from RFLP probes. STS markers are amplified using short primer pairs ~ 20bp in length that are used to prime the amplification of DNA. This DNA can derive from two sources, genomic DNA or cDNA which is complementary DNA produced by reverse transcriptase from a mRNA template. SSRs, are simple sequence repeat motifs which mutate rapidly due to DNA polymerase slippage during synthesis and repair. Primer pairs which direct amplification of SSRs often identify multiple alleles within a population. SSRs are excellent markers in mapping. (Liu et al., 1996). The AFLP technique (Vos et al., 1995) is a combination of RFLP and PCR. The AFLP technique uses four and six base restriction enzymes to digest genomic DNA. PCR amplification of the digested fragments using selectable primer combinations produces multiple restriction fragments per gel. While RFLPs were able to identify one polymorphism at a time, AFLPs are able to identify multiple polymorphisms at once thus increasing the efficiency of the technique and research productivity. In a comparison between various marker types and their ability to detect polymorphisms, Russell et al., (1997) determined that AFLPs produce more polymorphisms than RFLPs. This high rate of polymorphism detection makes AFLPs an ideal tool for detecting differences between individuals within a population (Powell et al., 1997). They are also ideal for constructing a genetic map

(Becker et al., 1995, Waugh et al., 1997) since they produce multiple scored markers per primer combination.

A recent type of molecular markers are single nucleotide polymorphisms (SNPs). SNPs are the most abundant type of discriminating DNA sequence differences that can be found in any genome. Marker technologies based upon point mutations take advantage of single base differences between alleles of individuals. These single base differences permit determination of allele states. Recently Cho et al., (1999) developed a genetic map in (*Arabidopsis thaliana*) using SNPs. Since SNPs are the most abundant type of mutations within populations this technology will become more prevalent as more cost effective and accurate development and detection approaches are developed.

Mapping & Marker Assisted Selection

Linkage maps and quantitative trait analysis can help breeders determine whether or not a trait with marginal heritability may be suitable for improvement through selection. Genetic maps locate genes contributing to variation for a trait of interest, and as a consequence identify genetic markers tightly linked to valuable alleles. While most traits are determined by the complex interactions of many gene products with the environment, a few gene differences generally control most of the heritable variation for many traits of interest, rendering marker-assisted selection often practically useful.

Quantitative traits by definition are traits that are controlled by more than one gene. Genetic maps are ideal for isolation of the numerous loci regulating quantitative traits especially in crop plants where genotype by environment interactions are high. Genetic maps also assist in the cloning of genes of interest. Wing et al., (1994) used map based cloning to identify the position and sequence of the jointless gene in tomato.

In barley, as in many crop plants genetic maps have been developed to not only identify genes regulating traits, but to use these maps to identify markers that can be used in future breeding programs.

Marker assisted selection depends upon the development of genetic maps to identify linked markers to valuable traits. Marker assisted selection applies genetic markers to population development to increase the likelihood of favorable alleles in the outcome of crosses. Disease resistance is a good example of simple marker assisted

selection, where marker assisted selection is used to report the successful incorporation of favorable alleles (Toojinda et al., 1998). 'Valier' (PI 610264) a feed barley released in Montana in 1999, was developed with the aid of molecular markers. Marker assisted selection should be effective for tightly linked markers to favorable alleles under the conditions that there is little or no epistatic interactions between genes. One concern in marker assisted selection is the resolution of the QTL with respect to linked markers. Selection using markers that are weakly associated with QTL may actually reduce trait heritability in selected populations if the alleles of interest are inadvertently lost through recombination (Lande and Thompson, 1990). Sufficiently high resolution mapping is the key to successful identification of informative markers that are tightly linked to QTLs of interest. Good genotype by environment statistical analysis is essential in determining the importance of gene by environmental effects, and therefore the reliability of heritability estimates for specific QTL. Applying stringent criteria to mapping and analysis of QTLs helps alleviate problems before they arise.

Technology development, technique application and serendipity are each important components of a research project. In this project, we developed a practically useful approach to SNP detection. We then used SNP and other marker technologies to develop a genome-wide linkage map in a population of recombinant inbred lines derived from the cross 'Lewis' x 'Karl'. Using three years' replicated yield trial data, we identified the major genes contributing to variation for grain protein content which were observed to segregate in this population. We demonstrated that two of these genes were

associated with genes modifying flowering date, a result which perhaps explains the slow progress made to date in grain protein content improvement. A serendipitous observation in SNP analysis led to a series of experiments in which we discovered that a gene *(PinB)* which is single-copy in wheat is multi-copy in barley.

CHAPTER 2

SINGLE NUCLEOTIDE POLYMORPHISM DETECTION

Introduction

Several approaches have been described which permit the assay of DNA sequence polymorphisms within individual targeted loci (Chen et al., 1997, Haff et al., 1997, Landegren et al., 1998, Liu et al., 1997, Shumaker et al., 1996). The value of massively multiplexed linkage analysis has been reported for the yeast genome utilizing microarrays (Winzeler et al., 1998). A study using human volunteers utilized a similar detection platform and reported discrimination among alleles at over 100 loci (Wang et al., 1998). The human genome (3×10^9 base pairs per haploid genome) is roughly 300 times larger than the yeast genome, resulting in a 300-fold dilution of target sequence per nanogram of DNA. Preamplified specific genomic sequences containing characterized allelic variants prior to allele detection are needed in order to make allele detection successful (Wang et al., 1998). Products of amplification were then annealed to arrays carrying short oligonucleotides which spanned regions containing previously characterized mutations. Differential duplex stability due to internal sequence mismatch helped to differentiate among alleles within samples. Recently, a chip-based platform was described which provides the possibility of electrophoretic concentration of amplification products at hybridization sites (Gilles at al.,1999). Like that of Wang (Wang et al.,1998),

this detection approach depends upon preamplification, then utilizing two differently-labeled allele-specific primers per locus, each of which matches one of two possible allele states at the 3' nucleotide. Following initial amplification labeled internal primers were added to the preamplification mix along with fresh Taq polymerase, unlabeled opposite strand primer, dNTPs and reaction buffer. Fluorescent products were amplified then resolved using either polyacrylamide gels for detection on the FMBIO II (Hitachi) gel scanning system, or denaturing polyacrylamide gels for detection on the ABI377 automated DNA sequencer. These approaches permit characterization of products both by size and by fluorescent label, which permits efficient multilocus analysis.

Materials and Methods

SNP Identification

Total barley DNA was isolated from foliar tissue of the barley cultivars Steptoe (CI15229), Morex (CI15773), Lewis (CI15856), Baronesse (PI568246) and Karl (CI15487) using the technique of Dellaporta et al.,(1983). Amplification was done according to Tragoonrung et al., (1992), PCR products were cleaned up to remove excess salts and primers on Quiagen columns. Sequencing was done on an ABI377 automated DNA sequencer and sequence was read in both the forward and reverse direction. Data analysis was done with DNASIS. A total of 20.86kb of DNA was analyzed per cultivar, of that 13.12kb was from genomic DNA and 7.74kb was from cDNA. From the DNA analyzed a few representative SNP's were chosen for fluorescent analysis.

Production of Fluorescent Product

Following amplification, the alternative alleles of ABC255, ABG65, and ABC305 for these five barley cultivars were sequenced using an ABI377 automated DNA sequencer and the Perkin Elmer Big Dye sequencing reaction kit. Sequencing was performed in both directions, and final sequences analyzed using DNASIS (Hitachi corp.) see (Table 1) for specific sequence and primers. Allele-specific fluorescently labeled primers were synthesized by IDT using either 6-carboxyfluorescein,(Fam) or 4,7,2',4',5',7' -Hexachloro-6-carboxyfluorescein (Hex). ABC255 and ABC305 map to barley chromosome 1, while ABG65 maps to barley chromosome 6 (Kleinhofs et al., 1993). Since the ABC255 polymorphism lies within an *Fnu4H1* restriction site, we were

able to validate it using restriction analysis. Primary amplification products from ABC 255 were combined with 1 unit of *Fnu4H1* and one microliter of 10x reaction buffer (New England Biolabs), then incubated for 1 hour at 37°C prior to electrophoretic analysis. In the parallel fluorescent SNP detection reactions one microliter of primary PCR product was mixed with 1 unit Taq polymerase and the solution brought to a final concentration of 1mM of each dNTP, 1.5mM $MgCl_2$, 10mM Tris-Cl pH 8.0 and 50 ng each of both fluorescent internal primers and unlabeled reverse strand primers.
Following one minute of denaturation at 92°C, samples were amplified through thirteen cycles (30 sec 92°C, 30 sec 45°C, 45 sec 72°C), with five minutes at 72°C prior to reaction termination at 4°C when using the FMBIOII. Only seven cycles of amplification were required for detection using the ABI377. A simulated heterozygote was produced by mixing equal quantities of DNA from both Lewis and Karl. This sample was treated as previously described.

Detection Platforms

The Hitachi FMBIO II is a flatbed fluorescence detection platform which can be utilized as a gel scanner. One was generously loaned to our project for the purposes of conducting this experiment by Hitachi-USA (San Jose, CA).
Ten microliters of secondary reaction mix or restriction digest were mixed with 2 microliters of 50% glycerol containing a small amount of bromphenol blue, then loaded onto a 7% acrylamide gel running in a continuous 0.5X TBE (Maniatis 1989) buffer system at 200V (20V/cm). HaeIII-digested PhiX174 was utilized as molecular weight

marker in the restriction digestion. Gels were run until the bromphenol blue migrated to the edge of the gel. Following electrophoresis, the restriction digest gel was stained with ethidium bromide. Both gels were transferred onto a borosilicate glass plate, placed in the FMBIO II and scanned. The restriction digest gel used a 605 nm cutoff filter which permits detection of ethidium bromide fluorescence (figure 1). The SNP detection gel was scanned twice, using filter sets which differentiate among FAM at 505nm and HEX at 585 nm. In figure 2 the FAM-labeled products are green and the HEX-labeled products are red. The ABI 377 is primarily used in our laboratory for sequencing and amplified fragment length polymorphism (AFLP) detection (Vos 1995). This platform also works well to characterize SNPs . One microliter of secondary amplification product was diluted 1:30 in 24ul ddH_20 and 5ul Promega loading dye. Samples were denatured at 94^0C for 3 min. One microliter of diluted product was loaded on a 48 well membrane comb CAM48-750 obtained from The Gel Company. The samples were run in a 6% polyacrylamide denaturing gel in a 1X TBE buffer. The gel was prerun for 15 min. then loaded and run for 3 hrs. at 750V and 51^0C.

Results

To determine the validity of identifying point mutations two experiments were performed. The first experiment tested whether the ABC 255 transition mutation at position 330 would be faithfully detected by the fluor-tagged primers. This was done by comparing a restriction digestion of the point mutation in an array of segregants using *Fnu4HI* and then evaluating the products of secondary amplification with fluor-tagged primers. The samples tested included the cultivars Steptoe and Morex, two cultivars which differ at position 330 by a C/T transition mutation, and 13 samples from a double haploid line (DHL) population derived from a Steptoe by Morex cross (figure 1). The *Fnu4H1* digestion produced a 410bp fragment in the Steptoe allele and a 316bp fragment in the Morex allele. Figure 2 shows Steptoe and Morex and the 13 DHL segregants with the transition mutation C (red, Morex) or T (green, Steptoe) alleles. The second experiment involved multiple allele and multiple cultivar analysis using ABC 255, ABG 65, and ABC305 with the cultivars Lewis, Karl, Morex, Steptoe, and Baronesse. This was a two part assay evaluating both multiplexing and heterozygosity analysis. The first part of this test involved multiplexing the primer sets ABC 255, ABG 65, and ABC305 through both the initial PCR and secondary SNP amplification. All of the samples in figure 4 were run under the same amplifying conditions. The first panel of figure 3 shows the results obtained when all three primer sets were used to detect polymorphisms simultaneously. The next three panels show the results obtained using the primer sets individually. Panels are separated by molecular weight ladders.

Figure 1

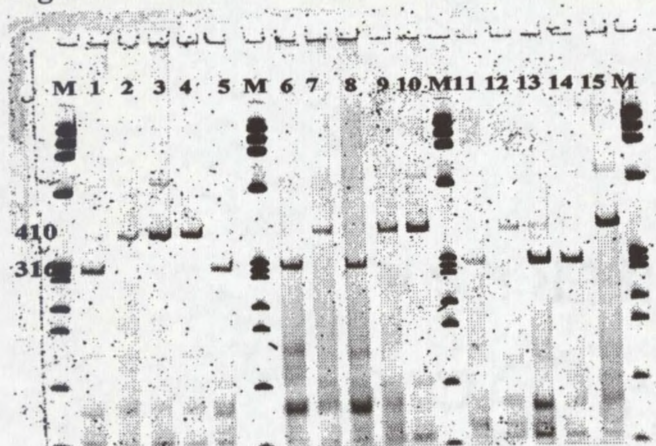


Figure 1. FMBIOII, ethidium stained *Fnu4H1* restriction of ABC255 amplification products of Morex (lane 1), Steptoe (lane 2) and a subset of doubled haploid lines from their F_1. Molecular weight standards Phi X lambda are in lanes labeled 'm'.

Figure 2

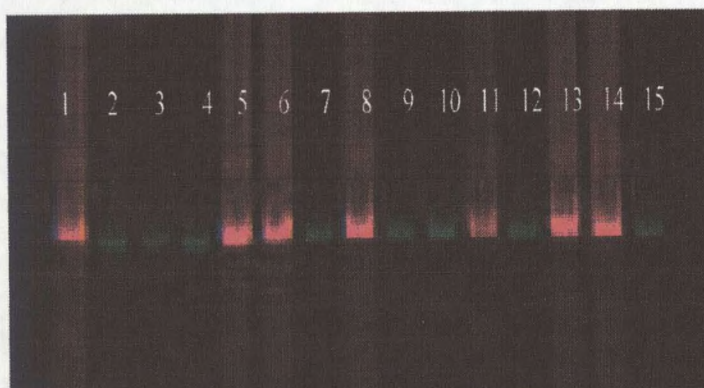


Figure 2. FMBIOII, Fluorescent primer detection of the ABC255 C/T transition mutation using the same DNA samples used in figure 2.