Thesis report

Validation of candidate gene for resistance to Botrytis cinerea in Gerbera

Meng Chen
March 19, 2013
WAGENINGEN UR

Validation of candidate gene for resistance to *Botrytis cinerea* in Gerbera

Name: Meng Chen

Plant breeding and Genetic Resources

Student Number: 881108-157-020

Supervisor: Paul Arens

Ornamental Breeding Department

Wageningen University and Research Centre

19th of March 2013

Acknowledgement

I would like to express my greatest gratitude to the people who have helped and supported me throughout my project. I am grateful to my supervisor Paul Arens for his continuous support for the project, from initial advice and contacts through the whole project. I learned a lot from him not only from the research aspect but also the way of thinking. I am also grateful to Wendy van't Westende who helped me a lot about my lab work. I learned many tricks for doing experiment which make me avoid mistakes and be more skilled.

Abstract

Gerbera is one of the most important ornamental flowers, however, it is very easy to be infected by a necrotrophic fungus, *Botrytis cinerea*. Thus, it becomes valuable to find resistance for this disease and validate them by QTL mapping. In a project between Wageningen UR Plant Breeding and two gerbera breeding companies, population development and disease tests were processed at companies based on flower petal and flower heart. SNP markers were obtained from the sequences of parents by transcriptome sequencing. By combining the disease test with the genetic maps of two segregating population, QTL mapping could be performed. Within my MSce thesis project I looked for Candidate Genes for disease resistance in literature and gene databases, one candidate gene for resistance to *botrytis cinerea* called 2-ps was processed to be validated in this study. This gene was chosen as candidate gene to *botrytis cinerea* because when knocking out 2-ps gene, gerbera shows susceptibility to this disease. A full length sequence of the 2-ps gene was available from the study of Koskela et al (Koskela *et al.*, 2011). Because the parents of the gerbera crosses are heterozygous outbreeding plants and the mapping populations are F1 populations, the SNPs in the 2-ps gene of a parent can be used as an marker for mapping. At the beginning, sequences of four parents in two crosses were achieved from transcriptome data. Unfortunately, these sequences were rather scarce. Therefore, re-sequencing the parents became a necessary step. Alignment of all available sequences with 2-ps gene was applied to design primer for re-sequencing and also to avoid that a highly similar gene (chalcone synthase) being extracted. All the sequences got from re-sequencing were assembled with candidate gene to detect putative single nucleotide polymorphisms (SNP) within cultivar. Primers were designed according to the positions of these putative SNPs and were used for screening in two offspring populations using Lightscanner technique. It is an accurate method for confirming SNPs and the real markers can be added to the others from SNP data in the generic project to perform QTL mapping. Co-localisation of the 2-ps SNP marker with QTL regions for botrytis resistance would suggest that different alleles of the gene have different effect on the level of resistance and beneficial alleles can be selected for in breeding.

Contents

1. Introduction

1.1 Marker-assisted selection

Plant breeding for new crop varieties can take around 10-25 years. Traditionally, breeders select plants using phenotype which can be easily influenced by environment and can be very demanding for traits that are not easy to determine such as disease resistance. However, nowadays one tool called marker-assisted selection (MAS) can speed up breeding process and help scientists to select plant that are difficult to evaluate phenotypically much easily. Nowadays, plant breeders using marker-assisted selection with the help of molecular or genetic markers can easily identify genes of interest for traits over generations (ISAAA, 2013). Genetic markers that tightly linked to the trait are used for genetic diversity analysis, genotype identification and identification of loci which control phenotypic traits. Single nucleotide polymorphisms (SNP), as sequence-based markers, are the most abundant type of DNA sequence polymorphisms. In practice, SNPs are bi-allelic because the probability of two independent bases occurring at a single position is low. However, in principal, any of the four nucleotides can be present at any position (Vignal *et al*., 2002). In order to develop markers, identifying and genotyping sequence variants in plant is necessary. Next generation sequencing (NGS), with its high throughput is an effective alternative to Sanger sequencing which is time-consuming and labour intensive (Metzker, 2005) and considered as first generation sequencing In principal, NGS extends the Sanger sequencing process across millions of reactions in a massively parallel fashion instead of being limited to a single or a few DNA fragments. It means DNA is first divided into numerous fragments which can be sequenced in millions of parallel reactions accurately. These newly produced small pieces of sequence, called reads, are then assembled into contigs by alignment (Metzker, 2010). In brief, three steps are involved in, which are template preparation, imaging and sequencing and assembly (Figure 1) (Robinson, 2012). Illumina (Solexa) sequencing used for transcriptome is the most widely-adopted next generation sequencing method worldwide. This method makes it possible to detect single base when single bases are incorporated into the extending DNA strands. When each dNTP is added, a fluorescently-labelled terminator is imaged and then cleaved to allow the next base to be incorporated. Non-incorporated nucleotides are washed away. Due to the fact that it is a true and accurate base-by-base sequencing, it enables a broad range of applications. Transcriptome can be obtained by using this method and after that SNP markers can be identified and genotyped. In order to check if these markers are linked to a trait of importance, a linkage group followed by a QTL mapping have to be performed. To achieve this goal, several steps have to be fulfilled before. First is to check SNP scores and assessing the segregation type in population. Second is to recode the SNP score into JoinMap based on the segregation type to perform the actual mapping of markers. Third is to achieve the data from a disease test. After combining the data of the markers and disease test in MapQTL, a QTL mapping can be performed.
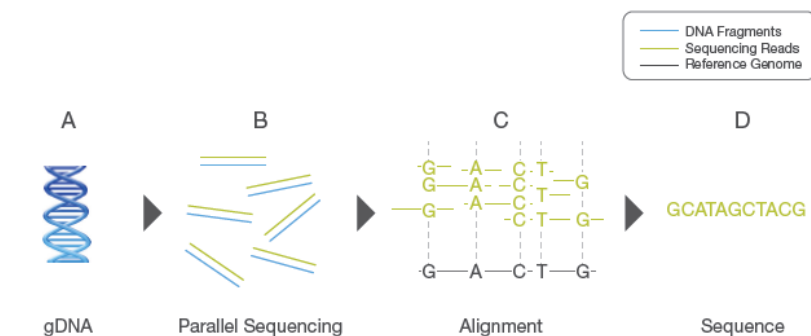
A. Extracted gDNA.
B. gDNA is fragmented into a library of small segments that are each sequenced in parallel.
C. Individual sequence reads are reassembled by aligning to a reference genome.
D. The whole-genome sequence is derived from the consensus of aligned reads.

Figure 1 Three basic steps for Next generation Sequencing

1.2 Problem analysis

The genus Gerbera, consists of approximately 30 species and it is indicated that the Gerbera-complex is originating from South Africa (Panero & Funk, 2008). Gerbera is a highly heterozygous and out-crossing species. As an outcrossing plant, Gerbera shows strong inbreeding depression when selfed (Teeri *et al*., 2006). One South African species *Gerbera jamesonii*, also known as Transvaal daisy or Barberton daisy, was described by J. D. Hooker in 1889 and published in Curtis Botanical Magazine according to the first official description (Tourjee *et al*., 1994). *G. jamesonii* has a large captitulum with prominent yellow, orange, white, pink or various red coloured ray florets. At the end of the 19th century in Cambridge, England, the breeding of gerbera started when two South African species *G.jamesonii* and *G.viridifolia* were crossed by R. I. Lynch with the goal of extending the range of colour in *G.jamesonii*. He named the hybrid Gerbera × cantebrigiensis, known today also as Gerbera hybrida. The majority of the present commercially cultivated cultivars originate from the crossing progenies of these two species. Natural hybrids of the two species have not been found. An early goal of horticulturists and breeders was cold hardiness of the crop for temperate climate established by commercial growers. Gerberas are propagated vegetatively and nowadays it is ranking at the fifth place of sold cut-flowers in the world (after rose, carnation, chrysanthemum, tulip) . Gerberas are also sold as pot plants and garden plants (SGP, 2002).

Gerbera hybrida is a typical member in the sunflower family, Asteraceae, which is the biggest flowering plants family and characterized by composite inflorescences which consist of morphologically different types of flowers (Teeri *et al.*, 2006). It belongs to the tribe Mutisieae, the fourth-most diverse subfamily. Genetically, Gerbera hybrida is a diploid plant with a somatic chromosome number of 2n=50 (Anop Kumari, 2011).

Unfortunately, Gerbera is one of the most susceptible ornamental flower to *botrytis cinerea* especially in very humid condition. *Botryits cinerea* is a necrotrophic fungus which can affects over 200 plant species (Elad, 1997). The disease is commonly known as Grey Mold which can cause rottenness on all the leaves and the common symptom is grey mold on the leaves. Infection takes place through wounds, decaying or dead plant tissue and by direct penetration of the undamaged host (Stehmann, 1995). Botrytis spores (conidia) germinate extremely well when there is free water with sugar in a thin layer. When the spore-bearing plant material is disturbed, spores are released and will attack other important organs in active development. Necrotic lesions occur on flower buds

and petals and are caused by early infections. These symptoms are increased when the relative humidity is above 93%, conditions that occur during packing flowers into boxes, transferring from cold storage into trucks for transportation, and therefore the risk of infection could be high during this process. Flowers are rejected at the auction when the botrytis causes 'pocks' on the petals during transportation. It is difficult to avoid damage caused by *B.cinerea* in the post-harvest phase especially in winter time (higher humidity), when the problem is more severe and significant losses can occur. Even though there are fungicides against this disease, many types of fungicides failed due to its genetic plasticity (Brian Williamson, 2007) and it may also increase the risk of fungicide resistance developing by the fungus. Gerbera shows large variation between cultivars for botrytis resistance. Therefore, looking for resistance to botrytis is of economic interest to breeders although the resistance is quantitative and thus not easy to handle. Few researches have been performed to look for putative quantitative trait loci (QTL) for resistance to *botrytis cinerea* in different crops such as tomato, arabidopsis, watermelon (Cho *et al.*, 2012; Denby *et al.*, 2004; Finkers *et al.*, 2007) because disease testing is quite difficult.

1.3 Generic project

My thesis study is part of a larger generic study. The generic project is QTL mapping of *Botrytis cinerea* resistance in Gerbera hybrida. Within this generic project, two F1 mapping populations have been produced and tested for botrytis disease resistance by two breeding companies. Disease testing is done by rating lesion sizes from infected ray flower petals, overall whole composite and checking the amount of lesions on the heart of the flowers as well. Because sometimes even though there is no lesion visible on the surface of the flower, the bottom part of the flower tissue is affected and will cause petal drop or other causes for loss of value.

SNP markers identified from transcriptome data are to be used for QTL mapping. Transcriptome data was obtained by Next generation sequencing using Illumina sequencing technology. QTL mapping is used to identify regions that correlate to disease resistance. The principal of determine whether the QTL is linked to one marker is to apply correlative statistics to obtain information whether there is significant correlation between a certain allele of a marker and resistance.

1.4 MSc Thesis research

Use of a candidate gene approach for Botrytis resistance. Some genes may be identified as putative candidate genes for botrytis resistance from literature. As a first validation of a candidate gene, they may be mapped in a QTL mapping population to check whether the candidate gene and a QTL for resistance co-localize. Presence of co-localization will warrant future research into expression levels or functional differences in the gene between genotypes differing in disease resistance.

1.5 Thesis Objective

Looking for candidate genes that are genes which show potential ability of controlling a trait, in this research is to look for Botrytis resistance candidate genes. According to this, a number of main steps can be identified to define the research objective for this MSc thesis research:

- Can Candidate Genes for botrytis resistance be identified from literature review or sequence databases?
- Can gerbera specific sequence information of parents for the two population used in the generic project be obtained for SNP identification?

- Can putative SNPs be used as markers to study the segregation in these two populations?
- On which position these Candidate Gene markers can be localized?

2. Materials and Methods

2.1 Identification and selection of Candidate Genes for *botrytis* resistance

To find candidate genes for botrytis, keywords combinations of resistant, *Botrytis cinerea* and Gerbera were searched for in the different databases, Scopus, Web of Science, National Centre for Biotechnology Information (NCBI) and Google Scholar were all used for searching candidate genes for botrytis. After that NCBI was browsed for looking up to the sequences of these candidate genes. Which Candidate genes were used for further validation in this study was based on BLAST results. If from the BLAST result and the literature review, the gene was found either in gerbera ESTs or the gene has homologue genes in (an)other model plant, the gene will be chosen as Candidate Gene for studying in this MSc research.

2.2 Sequence retrieval for Candidate Genes and SNP identification

In order to map Candidate Genes in the genetic maps for the generic project, SNP markers need to be developed in these Candidate Genes. Therefore sequence information of the parents of the two populations was needed and the first step to obtain these sequences was to screen the transcriptome data generated for the SNP identification in the generic project. Alignment of all the parents with the Genebank accession sequence of the Candidate Gene was needed to avoid highly similar genes being extracted which means that to make sure the gene being extract is specific the expected one. In bioinformatics, alignment is used as a way to identify similar regions of different sequences by arranging the sequence of DNA, RNA, or protein because these consensus parts may be of important function. During this procedure, gaps were kept between the residues so that identical or similar characters were aligned to achieve successive columns. The sequences were aligned in Lasergene Megalign programme which is special for DNA and protein sequences alignment and analysis. A range of pairwise and multiple alignment methods can be used in this programme, while, in this research, the W method was used for alignment to compare the sequences at the amino acid level.

Due to the fact that the parents of gerbera are heterozygous outbreeding plants and the mapping population is an F1 population, the SNP markers suitable for mapping have to be detected within cultivars. As a result, assembling sequences from the result of re-sequencing the parents with Candidate Gene for detecting SNP markers became a necessary step. Sequence assembly is to align and merge fragments of DNA to reconstruct the original sequence. Lasergene's SeqMan Pro offers accurate sequence assembly with the reference sequences into contigs and analysis of Sanger data, Illumina data and it can also offer visualization. In this research, it was performed to assemble the sequences obtained by re-sequencing from GATC service in order to discover SNPs. Putative SNPs can be viewed in the context of the alignment view. Graphic views can summarize the results by four colours which represent four nucleotide acid which are G, A, T, C, respectively. Each peak with one colour refers to one nucleotide acid while sometimes it appears two similar distribution with different colours at one position which can be considered as one putative SNP. These SNPs can be used to assess the segregation type in the offspring population in order to follow the inheritance patterns of chromosomal regions from generation to generation.

2.3 Primer design and PCR conditions

In order to re-sequence, primers were designed to nearly cover the whole Candidate Gene. Several characters have to be considered for designing primers. The generally accepted optimal length of PCR primer is between 18 and 22 base pair. The length should facilitate the primer to bind easily and achieve the specify product at the annealing temperature. The melting temperatures between 52 oC and 58 oC for primers produce the best results. The annealing temperature can be neither too high nor too low, otherwise the PCR product yield will be low or will lead to non-specific products. The GC content normally is in between 30% to 70%. Primer secondary structures such as hairpins and self-dimers have to be avoided when designing primers because these will cause a bad quality product. The primers were designed by using Primer Select programme in this research and the primer conditions were set as mentioned.

Before the primers designed to be used for re-sequencing, they were tested to confirm that they could work well. PCR reactions (Appendix 1) were used for testing the function of primers and the products were checking by comparing the length of the fragment to the 1 Kb ladder by gel electrophoresis. The band should be visible at the expected position. One special PCR method called Touch Down PCR was used (Appendix 1-3). This method can make primers work better when primers cannot work well under the setting temperature. Touch down PCR is to optimize PCRs even if the degree of primer-template complementarity is not fully known (Hecker & Roux, 1996). The protocol relies on incremental annealing temperature decrease every cycle. Touch down PCR can minimize the need of optimize annealing temperature and produce single strong target amplicons. Due to the primer initiates above the optimum annealing temperature and decreases below the melting temperature, the product yield can be significantly increased and be more specific. Because starting at a high temperature will make sure only very specific fragments can be produced after which at lower temperatures these already made fragments are more efficiently amplified.

2.4 Re-sequencing

Primer combinations designed were used for re-sequencing procedure. GATC service was chosen to do the re-sequencing. The GATC sequence service uses premixed DNA /primer samples for the analysis, so no sequence reaction is needed to be performed by our own. More details about the protocol of GATC sequencing service can be found in Appendix 2. After re-sequencing, all the sequences got were assembled with candidate gene to detect SNPs within cultivars, as described previously.

2.5 Lightscanner

In order to score the SNP markers in two populations, primers were designed according to the positions of putative SNP detected from the result of assembling. They were also tested on a small number of individuals of the population before using for the whole two populations. These SNP markers need to be scored in the two populations and therefore, an in-house SNP test with the Lightscanner was set up. Protocol is shown in Appendix 3. Lightscanner is a system that uses High Resolution Melting (HRM) analysis for mutation scanning and genotyping in an 96 well plate system. Hi-Res Melting of nucleic acid relies on the ability to collect high-density information of fluorescence which refer to the temperature changing. Images of DNA melting are recorded by a camera and reveal tiny details in DNA melting profiles. Special saturation dyes such as LC green only present in the double stranded DNA. When the sample is heated to high temperatures, the DNA denatures and

the fluorescent colour disappears as the double stranded DNA separates, generating a melting curve. Similar melting temperatures are represented as the same colour of similar curve shape and significant different temperatures will be indicated in different colours in the lightscanner software which can be interpreted as a SNP. In order to facilitate easy analysis, the fragments amplified by primers were small. In this study, they were around 200bp to avoid an additional SNP which would influence the melting curve and possibly lead to a second peak. The scanning was utilized for genotyping amplicons and samples that did not amplified were removed. For data normalization, approximately 1 ° C temperature intervals above and below the melting region was selected. The samples were grouped at a sensitivity level that best discriminated the samples. This technique was used in this research for confirming SNP markers developed from parent sequences in off-spring populations. These produced markers were used for genotyping in these two populations.

2.6 Plant population

All the leaf materials were supplied by two companies, Schreurs and Florist. Fresh leaves of four parent cultivars of Gerbera belonging to two crosses (population F and population S) called Sp1, Sp2, Fp1 and Fp2 respectively were used for re-sequencing and detecting SNPs. Leaf punches of around 250 off-spring individuals for each population were used for scoring SNPS.

2.7 DNA isolation

Six punches of leaf material of each individual were collected for sampling and half of them were used for DNA isolation. Leaf punches were collected in small paper envelops at the companies' greenhouse and were put into zip-lock plastic bags with silica gel. This method was used to avoid danger by traveling with nitrogen in the car. The feasibility of the method was first checked using a small number of individuals including the parents of the crosses before sampling the whole population, conformation of DNA quality was done by PCR. Punches were dried using silica gel at room temperature over 48 hours whereas fresh leaves of the parents were kept in a -80 degree fridge before DNA isolation. Fresh material of the four parents were grinded to powder in the mortar with nitrogen. All the leaf punch samples of the two off-spring populations were grinded in the shaking machine, TissueLyser II. For the powder of fresh materials, the similar amount of each sample as the dried materials with silica gel was used for DNA isolation. Fulton DNA isolation protocol was used for isolating DNA (Appendix 4). During the drying procedure with silica gel, DNA may get oxidised, some of the punches were brown which may affect the quality of DNA. As a result, additional one step for washing with 70% ethanol was added to the protocol. DNA concentration of all off-springs were checked on the 1% TBE gel comparing to lambda DNA of 10ng/ul concentration, while the DNA concentration of four parents were checked by Nano drop.

2.8 SNP marker analysis

For the generic project, I did the marker analysis for the whole SNP set of 477 SNPs for one population with the help of SNP viewer programme. These SNP markers were genotyped by KBioscience. For assess the segregation type, viewing the SNPs were done and to be able to perform the actual mapping of markers, the SNP scores were recoded into JoinMap. Finally, the scores of markers were combined with the result of disease tests. The disease test data for these two populations were supplied by two companies, Schreurs and Florist. The disease test was done on petals and flower heart. They were scored between 0 to 5 according to the level of botrytis symptom.

3. Results

According to the literature research, 2ps, WRKY33, ClWRKY70 and IVR gene might show resistance to this *Botrytis cinerea* (Koskela *et al*., 2011; Loebenstein *et al.*, 2010; Rushton, 2010). Therefore, these three genes were considered as candidate genes to *botrytis cinerea* resistance.

3.1 Candidate genes for botrytis resistance

3.1.1 2-PS

2-pyrone synthase, a new Chalcone synthase-like enzyme, is one of the secondary metabolites classified based on their synthetic origins. Plants produce diverse secondary metabolites even they are non-essential for the basic metabolic processes, however, when there is interaction with external factors they become the important factors for processing. They can inhibit the growth of microbes either by synthesising compounds in the plant during normal growth and development or inducing from the beginning in response to attack by plant pathogens (VanEtten *et al.*, 1994).

2-PS has a changed substrate specificity but remains nearly 70% similar amino acid level to Chalcone synthase which is the key enzyme of the flavonoid branch. This similarity may result some cross-reaction with related transcripts (Elomaa *et al*., 1996). 2-PS has the ability to sythesis 4-hydroxy-6-methyl-2-pyrone (triacetolactone) which is a putative precursor for gerberin and parasorboside, two abundant glucosides in gerbera. Silencing 2-PS in gerbera cannot synthesis gerberin and parasorboside subsequently lead to susceptibility to Botrytis cinerea infection (Koskela *et al.*, 2011). Some induced substances by 2-PS shows efficiently inhibiting fungal growth both in vivo and in vitro. When lacking 2-PS, Gerbera plants consequently strongly decrease the amount of producing gerberin and parasorbiside and plants show susceptibility to *botrytis cinerea*.

3.1.2 WRKY

The WRKY gene family is one of the ten largest families of transcriptional regulators in higher plants which can modulate many plant processes and it is found throughout the green lineage (green algae and land plants) (Eulgem & Somssich, 2007). They are key regulatory components of plant responses to microbial infection. A single WRKY gene might be involved in regulating separately processes such as regulate cross-talk between jasmonate and salicylate-regulated disease response pathway. In general, Salicylic acid regulate the responses against biotrophic pathogens while jasmonic acid regulate the responses to necrotrophs. Numerous reports shows the antagonistic interaction between these two pathways which mediate resistance against different types of microbial pathogens (Zheng *et al*., 2006). However, synergistic effects can happen when low concentration of salicylic acid and jasmonic acid are applied. The defining feature of WRKY transcription factors is the 60 residues length constant WRKY amino acid sequence at the N-terminus (Eulgem *et al.*, 2000; Rushton *et al.*, 1996). The WRKY amino acid sequences can be replaced by WRRY, WSKY, WKRY, WVKY or WKKY in a few WRKY proteins (Xie *et al.*, 2005). The conservation of the WRKY domain require a minimal consensus sequence (TTGACC/T) which is called W box in the promoter region of the target genes. By binding to the W-box, WRKY transcription factors can activate or repress their expression (Rushton, 2010).

WRKY33 is a member of the WRKY transcription factors and Birkenbihl showed that WRKY33 is responsible for resistance toward the necrotrophic fungus Botrytis cinerea (Birkenbihl *et al*., 2012).

Some other WRKY family members such as WRKY70 induced by salicylic acid and pathogens is also an important regulator of responses for pathogen attack. WRKY70 was found in leaves of watermelon Citrullus Lanatus (Li *et al*., 2004). Comparing to wild-type plants, overexpressing the ClWRKY70 gene in transgenic Arabidopsis in response to *B. cinerea* significantly reduced the lesion area. Its homologue sequence (AtWRKY70) in Arabidopsis has been described playing an important role in plant defence response as well (Cho *et al*., 2012).

3.1.3 IVR

Tomato plants transformed with a cDNA clone encoding the inhibitor-of-virus-replication (IVR) gene were partially resistant to Botrytis cinerea. Compared with the non-transgenic control plants, the size of lesions induced by the fungus in transgenic plants shows a significant reduction. Resistance related to presence of IVR transcripts was detected by reverse transcription-polymerase chain reaction (Loebenstein *et al*., 2010).

3.2 Bioinformatics

Sequences of candidate genes were found in National Centre for Biotechnology Information (NCBI). Gerbera hybrida mRNA for 2-pyrone synthase sequence [GenBank: Z38097.2] is 1555 base pair long and the codons is from 163 till 1371 base pair. The sequences of WRKY33 gene [GenBank: AEC09541.1] from Arabidopsis, WRKY70 gene [GenBank: ACV29874.1] from watermelon and other genes in the WRKY gene family and IVR gene [GenBank :AJ009684.1] were found in NCBI.

In order to find homologue genes for WRKY gene and IVR gene because they are not found in Gerbera, they were screened by Basic Local Alignment Search Tool (BLAST). BLAST is especially suitable for discovery of distant homologues with a conserved sequence motif. The BLAST result of WRKY33 shows that most of the homologue genes are from Arabidopsis thaliana, some are unknown protein mRNA, some are full-length cDNA sequences from clone genes. The BLAST result of IVR gene shows that this gene is only found in Tabaco and tomato and no homologues could be found in gerbera or other species outside the Solanaceae. In order to further study the diversity of WRKY gene, alignment was performed to try to figure out the consistent region between the two genes mentioned here and other members of the family .

3.2.1 Alignment

AtWRKY33 and ClWRKY70 were aligned to check the similarity because they both show potential resistant to *botrytis cinerea* . From Figure 2, high level of diversity is shown and the conserve part of these two WRKY genes is only the WRKY region. The full alignment can be found in Appendix 6. Between the two genes there is hardly any similarity and diversity between different members over different classes identified gives no clues how to identify the corresponding genes in gerbera. The WRKY region is the only conserved part according to the alignment of the whole family. That is the reason why this study only focused on 2-PS gene and did not proceed with WRKY genes.

Figure 2 Part of the alignment of AtWRKY33 with ClWRKY70

Alignment of all available sequences with 2-ps gene was applied to design primer for re-sequencing and also to avoid that a highly similar gene (chalcone synthase) being extracted. The sequence information of parents were obtained from transcriptome data. Parent 1 in Florist population (Fp1) obtained two contigs while parent 1 in Schreurs population (Sp1), parent 2 in Schreurs population (Sp2) and parent 2 in Florist population (Fp2) all obtained one contig for each and there was also one contig for all the parents. In total, six sequences achieved from four parents with the sequence of 2-PS were aligned together (Figure 3). The full alignment of 2-PS with six contigs from the four parents were shown in Appendix 7.

Figure 3 One part of alignment of 2-PS gene with six contigs from the four parents

On the left side of figure 3 are the names of contigs. The numbers on the top of figure 3 are the positions of single nucleotide based on the 2-PS gene [GenBank: Z38097.2]. The shaded part of this figure shows the conserve sequence for all the contigs and the white part, shown the single nucleotide, are the putative SNPs between cultivars. The orientation of first six contigs is from 3' UTRs to 5' UTRs while the last sequence of 2-ps is from 5' UTRs to 3' UTRs.

3.2.2 Re-sequencing by GATC service and SNP identification

Because SNP markers have to be detected within cultivars and they can be found in assembled sequences from re-sequencing result of parents, primers were designed to nearly cover the whole

2–PS gene first. Table 1 illustrates the information of the four designed primer pairs designed according to the SNP positions between cultivars. Information of primer position, primer sequence and expected product length are included in the table. The numbers after the 2-PS represent the primer position on 2-PS gene while F represents forward primer and R represents the reverse primer. The detailed characteristics of these primers are shown in Appendix 8.

Table 1 Information for four primer pairs

Primer	Primer sequence	Product length
2-PS_32F	TTCTTTTCTTTTCATTCCCTTCCCTCAA	352bp
2-PS_383R	TATCGTTTCTTTATCGCTGTTTTCTCAC	
2-PS_448F	TTAAACGCTCGACAAGACCTA	412bp
2-PS_859R	GGCCTGAACCCACAATGA	
2-PS_919F	CCGGACACTGAGAAGGCAATGAAG	427bp
2-PS_1345R	TAACGCGGACGCTACGAAGAACAA	
2-PS_1235F	CTATGGCGGAAGGGAAGA	249bp
2-PS_1483R	AAATAAAATCATAAACAAGAAAAATACA	

These four primer pairs were used for re-sequencing and one part of the re-sequencing result of Sp1 by 1345R primer are shown in Figure 4. Each colour stands for one nucleotide. Green, blue, black and red curve represent Adenine, Cytosine, Guanine and Thymine respectively. After getting the result of re-sequencing, these sequences were assembled in SeqMan Pro to detect SNPs within cultivars by comparing two sequences from the same primer pair (Figure 5).
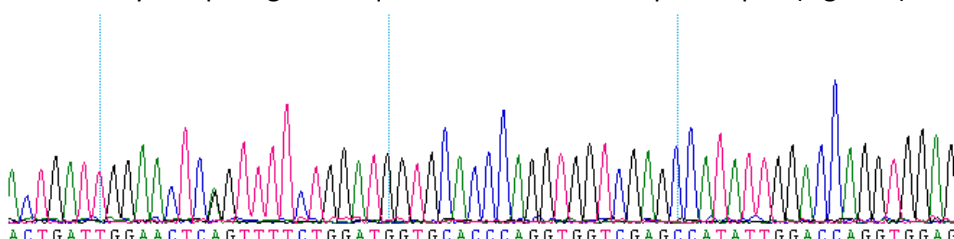
Figure 4 One part re-sequencing result of Sp1
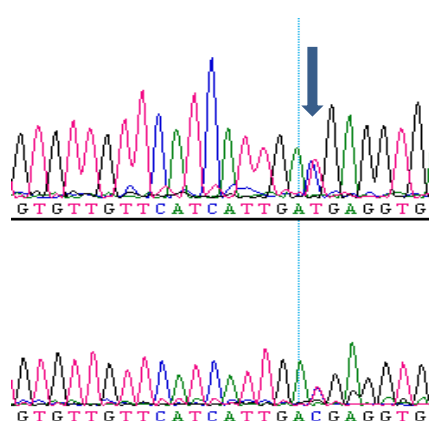
Figure 5 One part of assembled sequences of cultivar Sp1

Figure 5 illustrates one part of the assemble result of Sp1 using F919-1345R primer pair. Putative SNPs were detected by having two curves with similar distribution at one position and this situation happen in both sequences got from one primer pair. From this figure, a C/T SNP is shown.

3.2.3 Primer design based on the putative SNP positions

In order to score the SNP markers in two populations, eight primer pairs were designed according to these putative SNP positions within cultivars (Table 2). These primer pairs were designed by narrowing the length of the products to a maximum of around 200bp to avoid a second SNP involved in the product. Even so, there were some primers contained an additional SNP in the product. As a result, three best primer pairs were chosen for confirming the SNP markers in Lightscanner (Table 3).

Table 2 Eight primer pairs designed based on putative SNP positions within cultivar

SNP position	Primer combination	Cultivar
802	775F/859R	Fp1
992	831F/1012R	Sp1
1071	1021F/1240R	Sp1
1221	1104F/1290R	Sp1
1345	1235F/1433R	Sp1
1407	1235F/1433R	Sp1
1321	1107F/1345R	Fp2
1343	1235F/1433R	Fp2

Table 3 Information for three primer pairs chosen based on putative SNP positions within cultivar

primer	SNP position	Primer sequence	Product length	SNP pattern
2PS-775F-Fp1	802	GGACCCAATGAGAACCACCTT	85bp	A/G
2PS-859R-Fp1	802	GGCCTGAACCCACAATGAGT		
2PS-831F-Sp1	992	AGCTGCAGCACTCATTGT	182bp	G/A
2PS-1012R-Sp1	992	TGTTCTTTGCGACCATCA		
2PS-1107F-Fp2	1321	GGACCAGGTGGAGCGAAAACTAAA	239bp	C/T
2PS-1345R-Fp2	1321	TAACGCGGACGCTACGAAGAAC		

3.2.4 Lightscanner, an in-house SNP test for gerbera

For checking SNP identification process , an in-house SNP test called Lightscanner (HRM) was set up. It was used to check if this method can be suitable used in Gerbera and also to confirm putative SNPs in parent.

Due to the transcriptome sequences, ten primer pairs were obtained according to the contigs of Sp1 and eleven primer pairs were obtained according to the contigs of Sp2. These primer pairs were designed by QualitySNP programme according to the SNP positions. For testing if the putative SNPs are real in the fragment amplified by the primer pairs, these primer pairs were used for running Lightscanner. Gel was run for checking PCR products before running the Lightscanner to make sure the products amplified were the right ones.

For each primer, DNA samples of four parents were loaded on the Lightscanner plate and each parent sample had a duplication next to each other. By this way, it is more efficient to compare the result when scanned in the lightscanner machine. The amplification products of the primers were next to each other and the product length was compared to the 1 KB ladder. The products amplified showed the expected length. For Lightscanner result, some of the SNPs were real ones because two different curves with two colours and good duplication were shown in the lightscanner machine. Some primers did not get polymorphic allele and for some primers, parents did not have good

duplication which cannot conclude from the result if there is polymorphic. One example is shown below (Figure 6) and other results can be found in Appendix 9.
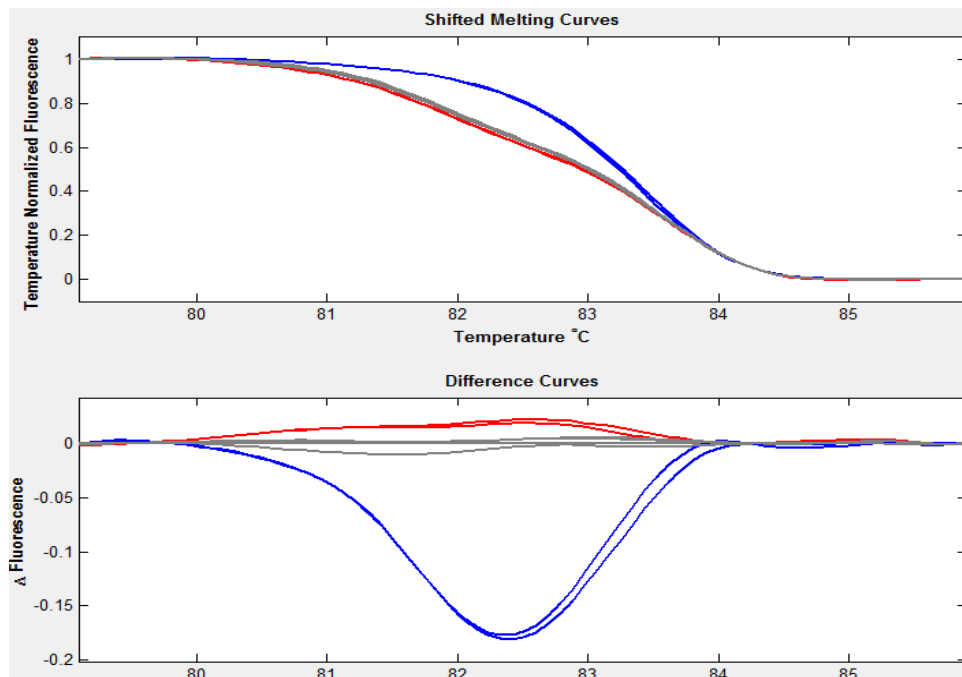

Figure 6 Lightscanner result of primer 16382-30F/116R designed by Sp1

Polymorphic alleles are shown within one cross of the parents by two different colour and pattern of curves. Figure 6 shows that all the samples have good duplication and each cross of two parents shows a polymorphic allele even though for the first cross, the difference is a small. Red, grey, blue and another two grey curves stand for Sp1, Sp2, Fp1 and Fp2 respectively.

From the result, Lightscanner can be used to confirm putative SNPs in gerbera. Thus, two primers designed from Florist cross and one primer designed from Schreurs cross were used to proceed. All the samples of two populations were screened in the Lightscanner programme using the designed primers. One example of Schreurs population is shown below (Figure 7). The Lightscanner result of Florist population can be found in Appendix 9.
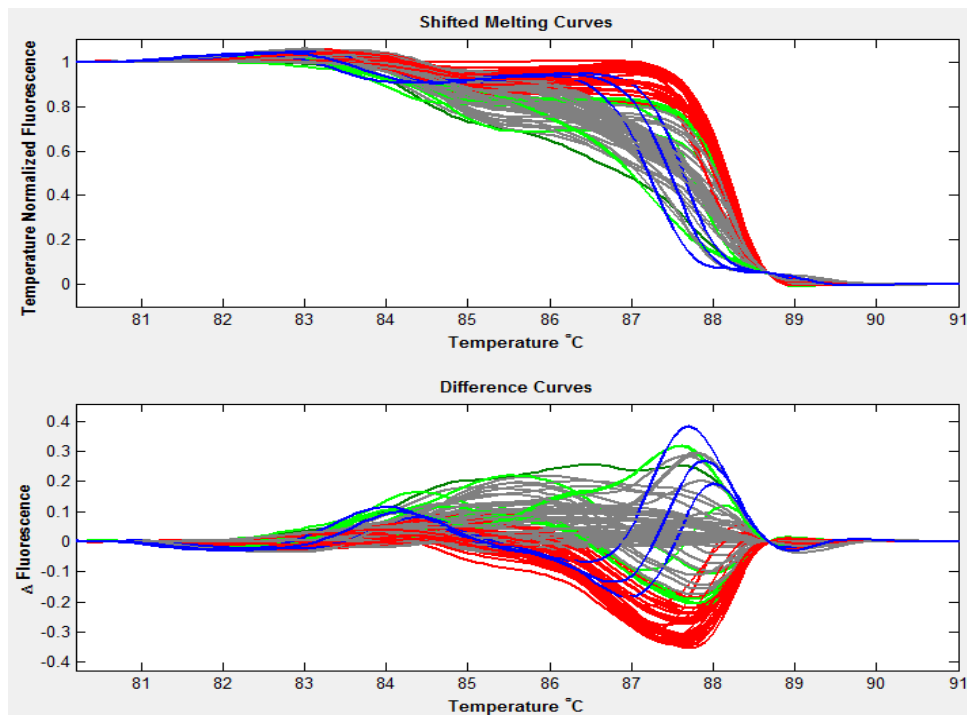
Figure 7 Lightscanner result of Schreurs population using primer pair 831F/1012R

In Figure 7, two highest green curves stand for Sp1 while two higher blue curves stand for Sp2. They both showed good duplications. Two clearly groups which were distinguish by two colours, grey and red. It seems that the population showed a 1:1 segregation ration. According to the SNP information from assembling result, putative SNP (G/A) was in Sp1 and no SNP (G) was in Sp2.

Before all the primers designed were used, they were tested in small population to be sure that they could work well. Primer pair 32F/383R did not work no matter using Touch down PCR reaction or changing DreamTaq to PhireTaq.

3.3 DNA isolation

DNA samples were isolated first for the four parents followed by 12 individuals for each population, after that DNA isolation was done for two whole populations. Four parents material were fresh leaves and 12 individuals materials were punches dried by silica gels. The concentration of the DNA for the four parents were checked by nanodrop. The concentration of Sp1, Sp2, Fp1, Fp2 were 908ng/ul, 445ng/ul, 469 ng/ul and 540ng/ul respectively. They were diluted to get the similar concentration which was around 10 ng/ul. Because 10ng/ul is the proper concentration for re-sequencing. All the off-spring samples in two populations were diluted to 30 times. All the DNA were loaded on the gel to check the quality and concentration by comparing to the 2ul, 4ul, 6ul, 8ul lambda DNA before dilution to around 10 ng/ul.

3.4 SNP marker analysis

In the generic project, SNP markers were identified in transcriptome data from the four parents and SNPs were genotyped by KBioscience. The result files with SNP scores were checked in the programme SNP viewer for assessing the segregation type. To be able to perform the actual mapping of markers, the SNP scores were recoded into JoinMap. I did the marker analysis for a whole SNP set for one population.

In SNP viewer software, homologous and heterozygous scores were distinguished by different positions in a plot and given a different colour. The heterozygous individuals were given green colour and the homologous were given either blue or red colour. If one of the parents was polymorphic and the other was not, the off-spring population would show a 1:1 segregation and two groups of dots were seen in the screen (Figure 8-1). On the other hand, if both of the parents are polymorphic, the off-spring population would show a 1:2:1 segregation, thus three groups of dots will be present (Figure 8-2). Moreover, if both the parents are monomorphic, there would be no segregation in the population (Figure 8-3). One special situation is that polymorphic allele is in only one parent but the off-spring population shows a 1:2:1 segregation (Figure 8-4). This is because one null allele is in the parent which shows homologous phenotype. For instance, GA genotype crosses to GØ genotype will lead to GG, GØ, AG and AØ genotype. GG and GØ cannot be distinguished and were just presented as the GG phenotype which subsequently should be of the most individuals. Additionally, AØ genotype shows the AA phenotype. As a result, three different phenotypes result in a 1:2:1 segregation.

Pink colour dots were uncertain samples, but if they were very clearly near one group, they were changed to be the same genotype as this group in the excel data sheet (Table 3). This leaded to a different amount of individuals belonging to the group.

In total, there were three plates for each population and the parents were located in C1, F1 for plate 1, B12, G1 for plate 2 and B11, H1 in plate 3. For each marker consistency over these scores were checked before further analysis started. In all cases markers were consistent over these multiple parent samples.
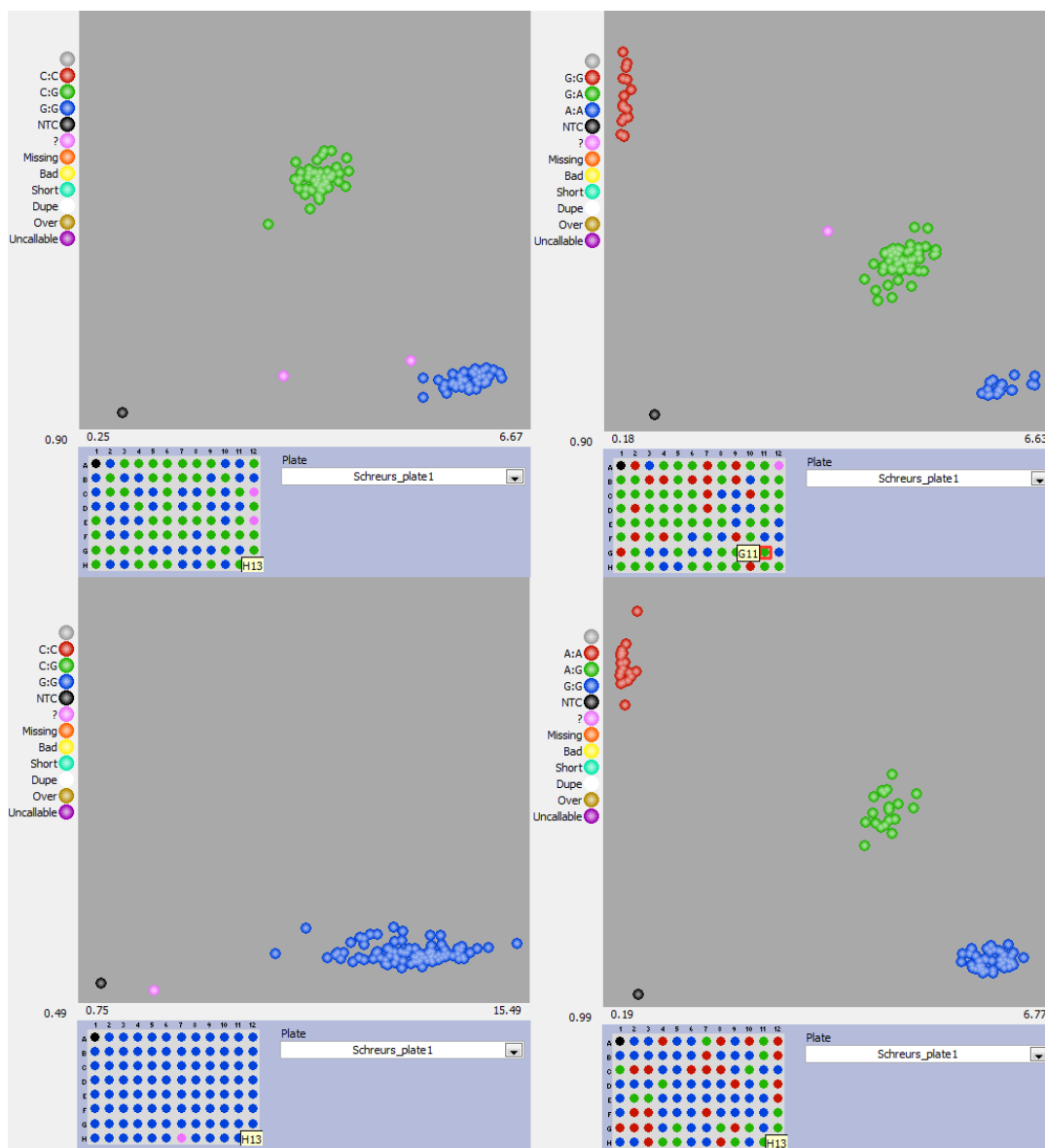
Figure 8 Four different situations visualized in SNP viewer

Table 4 Example of excel sheet for SNP data analysis

SNP	Segregation type	Check parents	Adaptation to data	Polymorphic nr.	S001	S002	S003	S004	...
Contig1	nnxnp	X	Y	141	C:C	C:C	T:C	T:C	...
Contig2	lmxll	AG GØ	Y	61	G:G	G:G	A:A	G:G	...
Contig3	hkxhk	X	Y	152	G:G	A:G	A:A	A:G	...
Contig4	mono(GGXGG)	X	Y	0	G:G	G:G	G:G	G:G	...
...	...	...	...	...	...	...	...	...	...

The first column of table 4 indicates marker names, the second column shows the segregation type by 'hkxhk' for heterozygous in both parents, 'lmxll' for heterozygous in first parent and 'nnxnp' for heterozygous in second parent, 'mono' stands for monomorphic or no clear segregation was shown. The third column was filled with an X when parents of all three plates had the same result and fitted the segregation type. 'Y' was filled in the 'Adaptation to data' column if data were changed while 'N' was filled in for no data was changed. The changed data were highlighted as shown in the brown colour. The number of heterozygous scores should be around half of the individuals and this was

used for a quick check on normal segregation. All the other columns were individual numbers in the population. The special situation, the null allele situation, was indicated with genotypes for both parents in the 'check parents' column.

Due to the time limited and because the SNP data came later than expected during the thesis period, data analysis was not fully finished.

4. Discussion

Gerbera hybrida is the fifth most important cut and pot flower marketed in the world, after rose, carnation, chrysanthemum and tulip, in the global cut flower trade. The genome sequencing of Gerbera is not finished at the moment, however, 16,998 gerbera expressed sequence tags (ESTs) are in the GenBank (Gong & Deng, 2010). Lack of information of genome sequencing makes it difficult to ensure the accurate position of the candidate gene in gerbera genome. But when the linkage groups are available from the generic project , it can be known which linkage group the candidate gene is located on.

From this study, eight SNPs were found in the 2-PS gene according to the re-sequencing result. Eight primers were designed based on these SNPs and three of them were chosen for confirming three SNP markers in Lightscanner. Because the other primers contained an additional SNP in the product according to the primer positions and SNP positions. These different alleles of the gene lead to a number of gene variant, and this may lead to different effect on the level of resistance. This difference can be checked preliminary according to co-localisation of the 2-ps SNP marker with QTL regions for botrytis resistance. Furthermore, statistics method can be used to check how much percentage the markers contribute to this QTL. From the result, beneficial alleles can be selected for in breeding.

In the generic project, some markers did not show Mendel segregation ratio and difficult to explain. This is because they are a number of SNP markers with null-alleles, for instance GAxGØ and GAxØØ. These markers were changed manually. As a result, markers with null alleles are not as powerful as other good markers and it is doubtful to use these markers for linkage map.

4.1 Molecular markers and their impact on plant breeding

The use of molecular markers accelerate the plant genetic analysis and facilitate the molecular breeding of crops. From the eighties, molecular markers such as RFLPs became popular used in plant molecular genetics, SSR markers came into use in late 90s and during the last five years SNP markers broke the dominant position of SSRs (JafarMammadov, 2012). Specially for gerbera, there is project running to develop and use SSR markers in order to facilitate breeding Gerbera. These markers will be used to identify polymorphic DNA fragments linked to powdery mildew resistance in gerbera in order to develop powdery mildew resistant cultivars. Structural or regulatory genes for anthocyanidin biosynthesis in gerbera will be amplified and sequenced to detect SNPs associated with leaf or flower colour phenotypes (Deng, 2010). Another paper published was to develop SSR markers in gerbera using Roche 454 GS-FLX for investigating genetic diversity and differentiation in gerbera (Seo *et al.*, 2012). According to Li (2010), SNPs were used to improve primer specificity in order to distinguish between very closely related fungal species. This research was done on *Fusarium oxysporum* f. sp. *chrysanthemi* which is an economically important pathogen of ornamentals including *Gerbera jamesonii* (Li, *et al*. 2010). No SNPs associated to plant disease were found in gerbera at the moment and there are only a limited number of examples of applications of SNP markers in other plant breeding (Ganal *et al.*, 2009). In this generic project, transcriptome re-sequencing using next-generation sequencing was used to rapidly discover SNPs and avoid highly

repetitive regions of genome. However, this technique is lack the power to recognize and eliminate duplicated sequences which cause the detection of false SNPs. This makes the validation of a marker a necessary step. It is the process of designing an assay according to the discovered polymorphism and genotyping the segregating population. That is why the detected SNPs in the candidate gene related to *botrytis cinerea* were detected by re-sequencing the parents and then they were validated in the population with the help of the high-resolution melting analysis in this study. Nowadays, developed genome complexity reduction technologies such as Complexity Reduction of Polymorphic Sequences (CRoPS) and Restriction Site Associated DNA (RAD) are capable of removing the duplicated SNPs (JafarMammadov, 2012). KBioscience genotyping platform was used for SNP validation in this research and some other genotyping platforms are currently popular such as Illumina's Bead Array technology-based Golden Gate (GG) and Infinium assays, Life Technologies' TaqMan assay coupled with Open Array platform. The application of SNP markers in MAS is becoming a tendency in plant breeding due to the genome-wide abundance and high-throughput detection platforms. Additionally, the commercial organizations are the main drivers for this breeding method according to the paper published by Monsanto and Syngenta (Ribaut, 2007; Rosso, 2011; Eathington, 2007).

4.2 Application of high-resolution melting

High-resolution melting analysis became widely used since 2003 and it is a high-throughput and cost-effective method for genotyping in MAS (Tan *et al*., 2013). Nowadays, this technique was used in many plant species including barley, grapevine and olive, perennial ryegrass, potato, wheat, rice, *brassica rapa*, almond, alfalfa, capsicum and bean (Botticella *et al*., 2011; Costa *et al*., 2012; de Koeyer *et al*., 2009; Ganopoulos *et al*., 2012; Han *et al*., 2012; Hofinger *et al*., 2009; Jeong *et al*., 2012; Lehmensiek, 2008; Li *et al*., 2011; Lochlainn *et al*., 2011; MacKay *et al*., 2008; Studer *et al*., 2009), but not yet used in gerbera. As a result, the utility of high-resolution melting (HRM) was checked for polymorphism detection in diploid gerbera in this study. One candidate gene, 2-PS, was studied using marker-assisted selection in gerbera breeding programs. Part of the study illustrates the use of HRM in gerbera genetics. It is considered as a useful and promising tool for basic and applied studies to discover superior alleles controlling important traits in model and other plant species. The obvious benefit for MAS of this technique is the short analysis time.

4.3 Candidate Gene approach in plant genetics

Candidate gene approach started to be used in plant genetics since the 1990s (Pflieger *et al*., 2001). This approach focuses on associations between variation of the gene of interest and phenotypes or disease states. This method is often used for study according to the known knowledge of gene's biological function correspond to the trait or disease. The candidate gene can be structural genes or genes act as a regulator in a metabolic pathway (Pflieger *et al*., 2001). In this case, 2-PS is a gene coding for a gene for one of the secondary metabolites. Comparing to genome-wide association studies and quantitative trait locus mapping, which scan the entire genome for genetic variation, candidate gene approach is time-saving. It is usually used in a population-based sample of individuals, so it has the advantage of both the statistical efficieny of complex diseases and the biological understanding of the phenotype for a certain disease (Tabor *et al*., 2002). However, this approach is also limited by the reliance on known knowledge about the physiological, biochemical or functional aspects of candidate genes (Zhu & Zhao, 2007). Paying attention to allelic variation in candidate gene because the only one SNP difference may directly impact the function of the gene and lead to subtle

differences in phenotype or disease level. In this study, map co-segregation between candidate gene and QTL was used to identify the candidate gene. One SNP in the 2-PS leads to two groups divided from one offspring population. One group is relevant resistant than the other group, which is the same situation as parents, one parent is relevant resistant than the other parent in one cross. The disease test data could be combined with the genotype result to present more accurate and comparable result.

Generally, candidate gene approach consists of three steps, selection of candidate genes, screening candidate genes, validation of candidate genes (Figure 9). Firstly, selection of candidate genes can be according to known gene functions in the biochemical or physiological pathways related to a trait of interest. If neither gene nor cDNA sequences are available for the crop of interest, sequence information from model plant can be used to detect conserved motifs and to design degenerate primers to try to get sequence information in this crop of interest. In this study, the sequence of 2-PS was found in the database for another gerbera cultivar. Secondly, co-localization of a functional candidate gene and a QTL. The QTL are quite imprecise because it may cover several hundred to thousand genes. The co-localization of a candidate gene with a QTL can be confirmed by statistical association studies or other validation experiments. There are three parts based on the traits studied for candidate gene-QTL co-localization, physiological, biochemical and disease resistance traits. In this case, 2-PS is a potential resistant gene to *botrytis cinerea*. Lastly, validation of candidate genes can be performed by physiological analysis of candidate gene expression or activity and genetic transformation. In sunflower, transformation of non-deficient plants with an antisense construction to approach validation was used. However, very few candidate genes have been validated in plants by transformation methods.
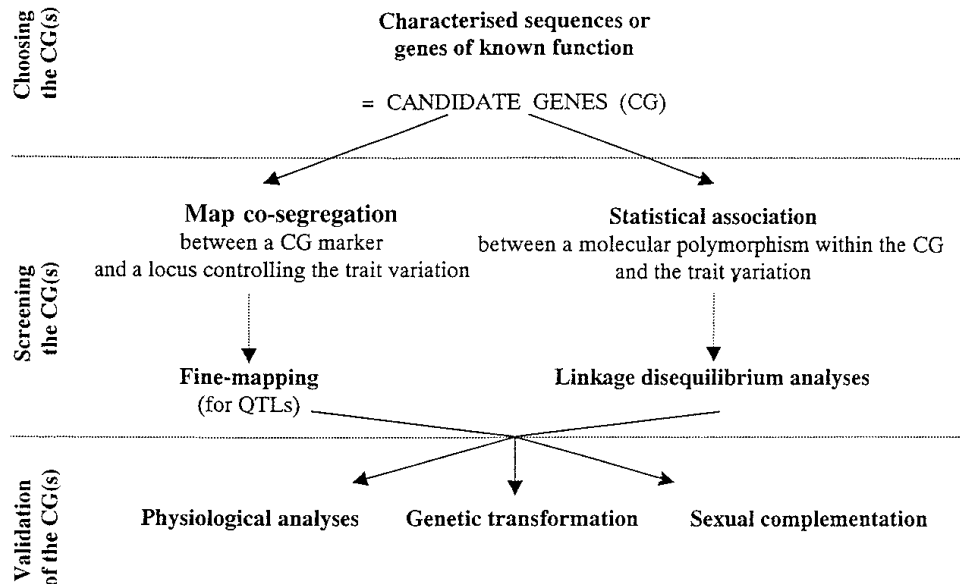
Figure 9 A synopsis of the candidate gene methodology (Pflieger *et al.*, 2001)

Candidate gene approach has been widely applied for disease research, genetic association studies and drug target selection from animals to humans (Zhu & Zhao, 2007). It is a powerful and promising method for genetic study.

Recommendations

During the thesis period, two other candidate genes involved in resistance to *botrytis cinerea* was coming out which are EARLI and PgD gene. These two candidate genes can be further studied for their anti-fungal activity.

EARLI gene, found in Arabidopsis consists of 168 amino acids, shows auxiliary functions in protecting plants against freezing-induced cellular damage and could inhibit the growth of the necrotrophic fungi (Li *et al.*, 2012). EARLI1 can be induced by *Botrytis cinerea* and it may be involved in plant defense system. Overexpression of EARLI1 could inhibit the growth of *Botrytis cinerea* (Du *et al.*, 2012).

Defensins, small cationic antimicrobial peptides, were found in vertebrates and invertebrates and have been reported in plants. These defensins inhibit growth of a broad spectrum of fungal plant pathogens such as *Fusarium oxysporum, Verticillium dahliae* and B*otrytis cinerea* (Picart *et al.*, 2012). PgD 1, Picea glauca defensin 1, is a plant defensin with 50 amino acids causing extensive growth inhibition of three fungal pathogens which are *cylindrocladium floridanum, Fusarium oxysporum* and *Nectria galligena.* It can be expressed during seed germination and also up-regulated by wounding and jasmonic acid treatment indicating the function in both the constitutive and induced defense mechanisms (Pervieux *et al.*, 2004). PgD 2-4 only show 7-9 amino acids differ from PgD 1 while PgD 5 has 64% sequence identity. They display non-morphogenic antifungal activity, possibly by altering membrane permeability (Picart *et al.*, 2012).

According to the Lightscanner result, it is obvious to see one group or two different groups. It could be further checked the ratio of the segregation and reliability using statistics method to support the Lightscanner result. After the ratio is confirmed, the result of the disease test can be combined with genotyping result. Each group can has its disease resistant level and can be compared to the resistant level of parents. Then the different alleles of the gene have different effect on the level of resistance can be confirmed. After perform the QTL mapping, statistics method can be used to check how much percentage the markers contribute to this QTL, in order to obtain accurate statistics proof.

References

Anop Kumari, K. S. P. a. M. C. (2011). Genetic variability studies in Gerbera. *Plant Biology, 1* (5), 01-04.

Birkenbihl, R. P., Diezel, C., & Somssich, I. E. (2012). Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. *Plant Physiology, 159*(1), 266-285.

Botticella, E., Sestili, F., Hernandez-Lopez, A., Phillips, A., & Lafiandra, D. (2011). High resolution melting analysis for the detection of EMS induced mutations in wheat SbeIIa genes. *BMC Plant Biology, 11*.

Brian Williamson, B. T., Paul, Tudzynski, Jan A. L. Van Kan (2007). Botrytis cinerea: the cause of grey mould disease. *Molecular Plant Pathology, 8*(5), 561–580.

Cho, S. M., Kang, E. Y., Min, K. H., Lee, Y. K., Kim, Y. C., Yang, K. Y., . . . Cho, B. H. (2012). A positive regulatory role of the watermelon ClWRKY70 gene for disease resistance in transgenic Arabidopsis thaliana. *Biologia Plantarum, 56*(3), 560-565. doi: 10.1007/s10535-012-0070-x

Costa, J., Mafra, I., & Oliveira, M. B. P. P. (2012). High resolution melting analysis as a new approach to detect almond DNA encoding for Pru du 5 allergen in foods. *Food Chemistry, 133*(3), 1062-1069.

de Koeyer, D., Douglass, K., Murphy, A., Whitney, S., Nolan, L., Song, Y., & de Jong, W. (2009). Application of high-resolution DNA melting for genotyping and variant scanning of diploid and autotetraploid potato. *Molecular Breeding, 25*(1), 67-90.

Denby, K. J., Kumar, P., & Kliebenstein, D. J. (2004). Identification of Botrytis cinerea susceptibility loci in Arabidopsis thaliana. *Plant Journal, 38*(3), 473-486.

Deng, Z. (2010). Oranmental and landscape pant cultivar development Retrieved 11th of March, 2013, from http://www.reeis.usda.gov/web/crisprojectpages/0223712-ornamental-and-landscape-plant-cultivar-development.html

Du, Z., Xu, D., Li, L., Shi, Y., Schläppi, M., & Xu, Z. Q. (2012). Inhibitory effects of Arabidopsis EARLI1 against Botrytis cinerea and Bradysia difformis. *Plant Cell, Tissue and Organ Culture, 110*(3), 435-443.

Elad, Y. (1997). Responses of plants to infection by Botrytis cinerea and novel means involved in reducing their susceptibility to infection. *Biological Reviews of the Cambridge Philosophical Society, 72*(3), 381-422.

Elomaa, P., Helariutta, Y., Kotilainen, M., & Teeri, T. H. (1996). Transformation of antisense constructs of the chalcone synthase gene superfamily into Gerbera hybrida: Differential effect on the expression of family members. *Molecular Breeding, 2*(1), 41-50.

Eulgem, T., Rushton, P. J., Robatzek, S., & Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends in Plant Science, 5*(5), 199-206.

Eulgem, T., & Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Current Opinion in Plant Biology, 10*(4), 366-371.

Finkers, R., van den Berg, P., van Berloo, R., ten Have, A., van Heusden, A. W., van Kan, J. A. L., & Lindhout, P. (2007). Three QTLs for Botrytis cinerea resistance in tomato. [Article]. *Theoretical and Applied Genetics, 114*(4), 585-593. doi: 10.1007/s00122-006-0458-0

Ganal, M. W., Altmann, T., & Röder, M. S. (2009). SNP identification in crop plants. *Current Opinion in Plant Biology, 12*(2), 211-217.

Ganopoulos, I., Bosmali, I., Madesis, P., & Tsaftaris, A. (2012). Microsatellite genotyping with HRM (High Resolution Melting) analysis for identification of the PGI common bean variety Plake Megalosperma Prespon. *European Food Research and Technology, 234*(3), 501-508.

Gong, L., & Deng, Z. (2010). EST-SSR markers for gerbera (Gerbera hybrida). *Molecular Breeding, 26*(1), 125-132.

Han, Y., Khu, D. M., & Monteros, M. J. (2012). High-resolution melting analysis for SNP genotyping and mapping in tetraploid alfalfa (Medicago sativa L.). *Molecular Breeding, 29*(2), 489-501.

Hecker, K. H., & Roux, K. H. (1996). High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. *Biotechniques, 20*(3), 478-485.

Hofinger, B. J., Jing, H. C., Hammond-Kosack, K. E., & Kanyuka, K. (2009). High-resolution melting analysis of cDNA-derived PCR amplicons for rapid and cost-effective identification of novel alleles in barley. *Theoretical and Applied Genetics, 119*(5), 851-865.

ISAAA. (2013). Pocket K No. 19: Molecular Breeding and Marker-Assisted Selection Retrieved 13,January, 2013, from http://www.isaaa.org/resources/publications/pocketk/19/default.asp

J. M. Ribaut, M. R. (2007). Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *Journal of Experimental Botany, 58*(2), 351-360.

JafarMammadov, R. A., Ramesh Buyyarapu, and Siva Kumpatla. (2012). SNPMarkers and Their Impact on Plant Breeding. *International Journal of Plant Genomics, 2012*. doi: 10.1155/2012/728398

Jeong, H. J., Kwon, J. K., Pandeya, D., Hwang, J., Hoang, N. H., Bae, J. H., & Kang, B. C. (2012). A survey of natural and ethyl methane sulfonate-induced variations of eIF4E using high-resolution melting analysis in Capsicum. *Molecular Breeding, 29*(2), 349-360.

Koskela, S., Söderholm, P. P., Ainasoja, M., Wennberg, T., Klika, K. D., Ovcharenko, V. V., . . . Teeri, T. H. (2011). Polyketide derivatives active against Botrytis cinerea in Gerbera hybrida. *Planta, 233*(1), 37-48.

Lehmensiek, A., Sutherland, Mark W, McNamara, Raymond B. (2008). The use of high resolution melting (HRM) to map single nucleotide polymorphism markers linked to a covered smut resistance gene in barley. *Theoretical & Applied Genetics, 117*(5), 721-728.

Li, J., Brader, G., & Palva, E. T. (2004). The WRKY70 Transcription Factor: A Node of Convergence for Jasmonate-Mediated and Salicylate-Mediated Signals in Plant Defense. *Plant Cell, 16*(2), 319-331.

Li, J., Wang, X., Dong, R., Yang, Y., Zhou, J., Yu, C., . . . Chen, J. (2011). Evaluation of High-Resolution Melting for Gene Mapping in Rice. *Plant Molecular Biology Reporter, 29*(4), 979-985.

Li, L., Zhang, C., Xu, D., Schläppi, M., & Xu, Z. Q. (2012). Expression of recombinant EARLI1, a hybrid proline-rich protein of Arabidopsis, in Escherichia coli and its inhibition effect to the growth of fungal pathogens and Saccharomyces cerevisiae. *Gene, 506*(1), 50-61.

Li, Y., Garibaldi, A., & Gullino, M. L. (2010). Molecular detection of Fusarium oxysporum f. sp. chrysanthemi on three host plants: Gerbera jamesonii, Osteospermum sp. and Argyranthemum frutescens. *Journal of Plant Pathology, 92*(2), 525-530.

Lochlainn, S. T., Amoah, S., Graham, N. S., Alamer, K., Rios, J. J., Kurup, S., . . . Broadley, M. R. (2011). High Resolution Melt (HRM) analysis is an efficient tool to genotype EMS mutants in complex crop genomes. *Plant Methods, 7*(1).

Loebenstein, G., David, D. R., Leibman, D., Gal-On, A., Vunsh, R., Czosnek, H., & Elad, Y. (2010). Tomato plants transformed with the inhibitor-of-virus-replication gene are partially resistant to botrytis cinerea. *Phytopathology, 100*(3), 225-229.

M. L. Rosso, S. A. B., L. M.Maupin, K.M. Rainey. (2011). Development of breeder-friendly markers for selection of MIPS1 mutations in soybean. *Molecular Breeding, 28*(1), 127-132.

MacKay, J. F., Wright, C. D., & 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 Methods, 4*(1).

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research, 15*(12), 1767-1776.

Metzker, M. L. (2010). Sequencing technologies [mdash] the next generation. *11*(1), 31-46.

Panero, J. L., & Funk, V. A. (2008). The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution, 47*(2), 757-782.

Paul J. Rushton, I. E. S., Patricia Ringler and Qingxi J. Shen. (2010). WRKY transcription factors. *Trends in Plant Science, 15*(15), 247-258.

Pervieux, I., Bourassa, M., Laurans, F., Hamelin, R., & Séguin, A. (2004). A spruce defensin showing strong antifungal activity and increased transcript accumulation after wounding and jasmonate treatments. *Physiological and Molecular Plant Pathology, 64*(6), 331-341.

Pflieger, S., Lefebvre, V., & Causse, M. (2001). The candidate gene approach in plant genetics: A review. *Molecular Breeding, 7*(4), 275-291.

Picart, P., Pirttilä, A. M., Raventos, D., Kristensen, H. H., & Sahl, H. G. (2012). Identification of defensin-encoding genes of Picea glauca: characterization of PgD5, a conserved spruce defensin with strong antifungal activity. *BMC Plant Biology, 12*.

Robinson. (2012). An Introduction to Next-Generation Sequencing Technology Retrieved 13,January, 2013, from http://www.illumina.com/Documents/products/Illumina_Sequencing_Introduction.pdf

Rushton, P. J., Torres, J. T., Parniske, M., Wernert, P., Hahlbrock, K., & Somssich, I. E. (1996). Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *EMBO Journal, 15*(20), 5690-5700.

S. R. Eathington, T. M. C., M. D. Edwards, R. S. Reiter, J. K. Bull. (2007). Molecular markers in a commercial breeding program. *Crop Science, 47*(3), 154-163.

Seo, K., Lee, G., Park, S., Yoon, M., Ma, K., Lee, J., . . . Lee, M. (2012). Genome shotgun sequencing and development of microsatellite markers for gerbera (Gerbera hybrida H.) by 454 GS-FLX. *African Journal of Biotechnology, 11*(29), 7388-7396.

SGP. (2002). The compositae Genome project Retrieved 14, January, 2013, from http://compgenomics.ucdavis.edu/compositae_data.php?name=Gerbera+hybrida

stehmann, C. (1995). Biological activity of triazole fungicides towards Botrytis cinerea.

Studer, B., Jensen, L. B., Fiil, A., & Asp, T. (2009). "Blind" mapping of genic DNA sequence polymorphisms in Lolium perenne L. by high resolution melting curve analysis. *Molecular Breeding, 24*(2), 191-199.

Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nature Reviews Genetics, 3*(5), 391-397.

Tan, Y. Y., Fu, H. W., Zhao, H. J., Lu, S., Fu, J. J., Li, Y. F., . . . Shu, Q. Y. (2013). Functional molecular markers and high-resolution melting curve analysis of low phytic acid mutations for marker-assisted selection in rice. *Molecular Breeding, 31*(3), 517-528.

Teeri, T. H., Elomaa, P., Kotilainen, M., & Albert, V. A. (2006). Mining plant diversity: Gerbera as a model system for plant developmental and biosynthetic research. *BioEssays, 28*(7), 756-767.

Teeri, T. H., Uimari, A., Kotilainen, M., Laitinen, R., Help, H., Elomaa, P., & Albert, V. A. (2006). Reproductive meristem fates in Gerbera. *Journal of experimental botany, 57*(13), 3445-3455.

Tourjee, K. R., Harding, J., & Byrne, T. G. (1994). Early development of gerbera as a floricultural crop. *HortTechnology, 4*(1), 34-40.

VanEtten, H., Mansfield, J. W., Bailey, J. A., & Farmer, E. E. (1994). Two classes of plant antibiotics: phytoalxins versus phytoanticipins. *Plant Cell 6*(9), 1191-1192.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. [Review]. *Genetics Selection Evolution, 34*(3), 275-305. doi: 10.1051/gse:2002009

Xie, Z., Zhang, Z. L., Zou, X., Huang, J., Ruas, P., Thompson, D., & Shen, Q. J. (2005). Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiology, 137*(1), 176-189.

Zheng, Z., Qamar, S. A., Chen, Z., & Mengiste, T. (2006). Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant Journal, 48*(4), 592-605.

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: Progress and challenges. *International Journal of Biological Sciences, 3*(7), 420-427.

Appendix

1. Protocol for PCR reaction

PCR Reaction 1

Template DNA 2ng/μl	5	(Normally is 10-25 ng/reaction)
DreamTaq buffer 10 ×	2	
dNTP's (5mM)	0.4	
DreamTaq (5u/μl)	0.08	
MiliQ water	12.12	
Primers F+R 10 μM each	0.4	
End volume	20	

Programme setting for PCR machine

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	3 min	1
Denaturation	94	30sec	
Annealing	58	30 sec	
Extension	72	2 min	30
Final Extension	72	10 min	1
On hold	15		

PCR reaction 2

Template DNA 2ng/µl   2   (Normally is 10-25 ng/reaction)
Phire buffer × 5   4
dNTP's (5mM)   0.4
phire enzyme   0.2
Forward primer 10 pmol   1
Reverse primer 10 pmol   1
MiliQ water   11.4
                    End volume   20
Programme setting for PCR machine

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	98	30sec	1
Denaturation	98	5sec	
Annealing	55	5sec	
Extension	72	15sec	30
Final Extension	72	1min	1
On hold	15	Forever	

PCR reaction 3-Touch Down PCR reaction

Temperature (°C)	Time	Number of cycles
95	3 min	1
94	30sec	
60	30 sec	
60-55	Decrease by 0.5°C every cycle	
72	2 min	10
94	30sec	
55	30sec	
72	2 min	20
72	10 min	1
15	forever	

2. Protocol GATC sequencing service

General

The GATC sequence service uses premixed dna/primer samples for their analyses, so we do not have to perform a sequence reaction anymore!!

A barcode sticker on every tube or plate takes care of the Sequence → Sample → Person logistics.

Barcode stickers for single samples and full (96) plates are available

Tubes/plates are bi-weekly collected by GATC

Results available 24 hours after arrival of the samples at GATC

Sequences can be collected from the N/common/GATC sequence service folder

Step by step

Add your sample and primer (according to the sample requirements) to a normal 1.5 ml safe lock eppendorf tube or in case of 96 samples to a special 96 plate (available at the GATC sequence service fridge in E1.123).

Take one barcode for every eppendorf tube or 96 plate from the GATC sequence service order (next to the fridge E1.123). Remark: a single sample or full plate needs a different type of Barcode!!!!!

Register your full name to every barcode that you use (in the same GATC sequence service order).

For every barcode there is one sticker for your sample and a corresponding one for your lab record. Use these for your sequence→sample logistics.

Place your samples, before 1500 hours, every Tuesday and Thursday in the GATC sequence service fridge (E1.123) . The samples will be collected only on those two days.

The sequences will be available in the N:/common/GATC sequence service folder optimally 24 hours after arrival of the samples at GATC.

Sample Requirements

5 µl purified plasmid DNA or purified PCR fragment

Purified plasmid DNA: 80 - 100 ng/µl

Purified PCR product: 20 – 80 ng/µl

5 µl of primer (5 µM (5 pmol/µl))

Please send the total amount of 10 µl in a 1.5 safe lock tube

Primer characteristics:

The melting temperature of the primer should be between 52° and 58°C and the length should be between 17 – 19bp. Ideally, the GC content of an 17mer should be 10 G+C; for an 18mer 8 – 9 G+C and for a 19mer 7 – 9 G+C.

G or C should be at the 3' end, but not more than 3 Gs or Cs.

The primer sequence should be a good mix of all 4 nucleotides with no more than 4 identical bases in a row (AAAA or GGGG).

3. HRM using Phire hotstart (Finnzymes)

2 μl 5x phire buffer

0.4 μl dNTP's 5 mM

1 μl F primer (10 pmol/μl)

1 μl R primer (10 pmol/μl)

1 μl 10x LC green, genescan reagent (BioChem)

1 μl DNA 10 ng/μl

0.1 μl Phire enzyme (10 u/μl)

3.5 μl MQ

10 μl End volume

10 μl mineral oil

PCR profile:

Start with 98 degree initial denaturation for 30 seconds, followed by 40 cycles of denaturation step of 5 seconds for 98 degree, 5 seconds for 60 degree and 10 seconds for 72 degree. After that, 72 degree for one minute, 94 degree for 30 seconds and 25 degree for 30 seconds. Finally, keep the temperature to 10 degree.

4. Fulton DNA isolation

50 - 200 mg is sufficient

Starting material is lyophillized, grinded leaf (small leaves or leaf disc, grinded with glass pearls)

All centrifugations are at maximum speed.

Switch waterbath on at 60°C

Prepare fresh microprep buffer 2,5 parts extraction buffer (EB) (+ RNAse 100µg/ml), 2.5 parts lysis buffer and 1.0 part 5% sarcosyl (w/v). Add 0.38 g sodium bisulfite/100 ml buffer immediately before use.

Add 750 µl micro-prep buffer to leaf powder, mix well

Incubate in 60°C waterbath for 30-60 min.

Extract with 800 µl chloroform (mix well)

Spin 5 min.

Pipet off 400-600 µl aqueous phase

Add equal volume cold isopropanol and invert tubes repeatedly untill DNA precipitates.

Spin 5 min.

Wash with 500 µl 70% ethanol (pellet often very loose), spin 5 min. pour off.

Dry pellet

Resuspend DNA in TE^{-4}

Caution: chloroform is carcinogenic: wear gloves. Waste choroform in cat. cans

DNA extraction buffer:

      Per liter:

0.35 M Sorbitol	- 63.7 g	
0.1 M Tris-HCl	- 12.1 g	pH 7.5
5 mM EDTA	- 1.7 g	

Lysis buffer:

0.2 M Tris	
0.05 M EDTA	pH 7.5
2 M NaCl	
2% CTAB	

Sarcosyl ⟶ Do not autoclave!   TE^{-4} : 10 mM Tris/0,1 mM EDTA pH 8.0

5. BLAST result of IVR

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage
Transcripts				
AJ009684.1	Nicotiana tabacum mRNA for inhibitor of virus replication (IVR)-like pr	1877	1877	100%
AK328373.1	Solanum lycopersicum cDNA, clone: LEFL2053H23, HTC in fruit	920	920	73%
AK378439.1	Bombyx mori mRNA, clone: fmqV42J01	52.8	52.8	2%
XM_002164852.1	PREDICTED: Hydra magnipapillata similar to collagen, type II, alpha 1	52.8	52.8	2%
EU950870.1	Zea mays clone 645197 mRNA sequence	52.8	52.8	2%
Genomic sequences[show first]				
BX649611.8	Mouse DNA sequence from clone RP23-445E3 on chromosome 4 Cont	54.7	54.7	2%
AC188444.2	Gallus gallus BAC clone CH261-78K2 from chromosome z, complete se	52.8	52.8	2%
CR848830.7	Zebrafish DNA sequence from clone DKEY-226F21 in linkage group 11	52.8	52.8	2%

From the BLAST result of IVR, it can be seen that IVR gene only exits in Nicotiana tabacum and Solanum lycopersicum.

6. Alignment of WRKY33 with ClWRKY70

7. Alignment of 2-ps with the contigs from four parents* in two crosses

```
----T--C-----TT------ GCCCTCAGATTAGAGAACAAAAGGCCTACTCAAGCCTTGAAATTCTCTTTTCTTTTCTT
         10        20        30        40        50        60        70        80
```

Row end positions: 54, 79, 80, 0, 79, 80, 41

```
TTCTXTTXATTCCCTTCCCTCAAATTATAAACTTACCTTTCTGTTTXTCTTTCAAAGAATTTAGCTGXCTCAAACGAAGA
         90       100       110       120       130       140       150       160
```

Row end positions: 134, 152, 160, 0, 159, 160, 114

```
TCTTCATATCTCATTTGTTAGGATATACAAACATCAATCTCGAGTAAAATGGGATCTTACTCATCCGATGATGTGGAGGT
        170       180       190       200       210       220       230       240
```

Row end positions: 214, 232, 240, 0, 239, 240, 194

```
GATTXGTGAGGCCGGACGGGCACAAGGTTTAGCCACGATTCTTGCCATTGGCACTGCTACTCCTCCCAATTGCGTCGCTC
        250       260       270       280       290       300       310       320
```

Row end positions: 294, 312, 320, 66, 319, 320, 274

```
AAGCTGATTATGCAGACTATTATTTTCGTGTCACTAAGAGCGAACATATGGTTGATCTTAAAGAGAAATTTAAACGCATT
        330       340       350       360       370       380       390       400
```

Row end positions: 374, 392, 400, 146, 399, 400, 354

*The order of the sequences are Sp1,Sp2,Fp1-1,Fp1-2,Fp2,All four,2-PS.

```
TGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGACTATCTGCAAGAGAACCCAACAATGTGTGAGTT
          410       420       430       440       450       460       470       480
TGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGACTATCTGCAAGAGAACCCAACAATGTGTGAGTT  454
TGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGACTATCTGCAAGAGAACCCAACAATGTGTGAGTT  472
TGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGACTATCTGCAAGAGAACCCAACAATGTGTGAGTT  480
TGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGACTATCTGCAAGAGAACCCAACAATGTGTGAGTT  226
TGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGACTATCTGCAAGAGAACCCAACAATGTGTGAGTT  479
TGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGACTATCTGCAAGAGAACCCAACAATGTGTGAGTT  480
TGTGAGAAAACAGCGATAAAGAAACGATACCTAGCCCTCACCGAAGACTATCTGCAAGAGAACCCAACAATGTGTGAGTT  434

CATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAGAAGCCGCAGTCAAGG
          490       500       510       520       530       540       550       560
CATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAGAAGCCGCAGTCAAGG  534
CATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAGAAGCCGCAGTCAAGG  552
CATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAGAAGCCGCAGTCAAGG  560
CATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAGAAGCCGCAGTCAAGG  306
CATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGTCCCGATGCTTGGCAAAGAAGCCGCAGTCAAGG  559
CATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGTCCCGATGCTTGGCAAAGAAGCCGCAGTCAAGG  560
CATGGCTCCATCCTTAAACGCTCGACAAGACCTAGTGGTCACCGGCGTCCCAATGCTTGGCAAAGAAGCCGCAGTCAAGG  514

CCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCACCGCTGGCGTTGACATGCCCGGT
          570       580       590       600       610       620       630       640
CCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCACCGCTGGCGTTGACATGCCCGGT  614
CCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCACCGCTGGCGTTGACATGCCCGGT  632
CCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCATCTCATCTTCTGCACCACCGCTGGCGTTGACATGCCCGGT  640
CCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCATCTCATCTTCTGCACCACCGCTGGCGTTGACATGCCCGGT  386
CCATTGATGAATGGGGGCTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCACCGCTGGCGTTGACATGCCCGGT  639
CCATTGATGAATGGGGGCTACCAAAATCCAAGATCACCCATCTCATCTTCTGCACCACCGCTGGCGTTGACATGCCCGGT  640
CCATTGATGAATGGGGACTACCAAAATCCAAGATCACCCACCTCATCTTCTGCACCACCGCTGGCGTTGACATGCCCGGT  594

GCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAACGCTATATGTTGTACCAACAGGGATGTGCCGC
          650       660       670       680       690       700       710       720
GCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAACGCTATATGTTGTACCAACAGGGATGTGCCGC  694
GCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAACGCTATATGTTGTACCAACAGGGATGTGCCGC  712
GCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAACGCTATATGTTGTACCAACAGGGATGTGCCGC  720
GCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAACGCTATATGTTGTACCAACAGGGATGTGCCGC  466
GCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAACGCTATATGTTGTACCAACAGGGATGTGCCGC  719
GCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAACGCTATATGTTGTACCAACAGGGATGTGCCGC  720
GCTGACTATCAACTCGTCAAACTCCTTGGTCTCTCCCCTTCAGTCAAACGCTATATGTTGTACCAACAGGGATGTGCCGC  674

CGGCGGCACAGTCCTCCGGCTAGCCAAGGACCTTGCTGAAAACAACAAGGGCTCACGAGTCCTTATCGTCTGCTCCGAGA
          730       740       750       760       770       780       790       800
CGGCGGCACAGTCCTCCGGCTAGCCAAGGACCTTGCTGAAAACAACAAGGGCTCACGAGTCCTTATCGTCTGCTCCGAGA  774
CGGCGGCACAGTCCTCCGGCTAGCCAAGGACCTTGCTGAAAACAACAAGGGCTCACGAGTCCTTATCGTCTGCTCCGAGA  792
CGGCGGCACAGTCCTCCGGCTAGCCAAGGACCTTGCTGAAAACAACAAGGGCTCACGAGTCCTTATCGTCTGCTCCGAGA  800
CGGCGGCACAGTCCTCCGGCTAGCCAAGGACCTTGCTGAAAACAACAAGGGCTCACGAGTCCTTATCGTCTGCTCCGAGA  546
CGGCGGCACAGTCCTCCGGCTAGCCAAGGACCTTGCTGAAAACAACAAGGGCTCACGAGTCCTTATCGTCTGCTCCGAGA  799
CGGCGGCACAGTCCTCCGGCTAGCCAAGGACCTTGCTGAAAACAACAAGGGCTCACGAGTCCTTATCGTCTGCTCCGAGA  800
CGGCGGCACAGTCCTCCGGCTAGCCAAGGACCTTGCTGAAAACAACAAGGGCTCACGAGTCCTTATCGTCTGCTCCGAGA  754
```

TCACTGCTATCTTATTCCATGGACCCAATGAGAACCACCTTGACTCACTCGTCGCTCAAGCTTTATTCGGAGACGGAGCT

```
        810       820       830       840       850       860       870       880
TCACTGCTATCTTATTCCATGGACCCAATGAGAACCACCTTGACTCACTCGTCGCTCAAGCTTTATTCGGAGACGGAGCT  854
TCACTGCTATCTTATTCCATGGACCCAATGAGAACCACCTTGACTCACTCGTCGCTCAAGCTTTATTCGGAGACGGAGCT  872
TCACTGCTATCTTATTCCATGGACCCAATGAGAACCACCTTGACTCGCTCGTCGCTCAAGCTTTATTCGGAGACGGAGCT  880
TCACTGCTATCTTATTCCATGGACCCAATGAGAACCACCTTGACTCGCTCGTCGCTCAAGCTTTATTCGGAGACGGAGCT  626
TCACTGCTATCTTATTCCATGGACCCAATGAGAACCACCTTGACTCACTCGTCGCTCAAGCTTTATTCGGAGACGGAGCT  879
TCACTGCTATCTTATTCCATGGACCCAATGAGAACCACCTTGACTCGCTCGTCGCTCAAGCTTTATTCGGAGACGGAGCT  880
TCACTGCTATCTTATTCCATGGACCCAATGAGAACCACCTTGACTCACTCGTCGCTCAAGCTTTATTCGGAGACGGAGCT  834
```

GCAGCACTCATTGTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCAATATTCGAGATCGTGTCAACTGATCAAACAAT

```
        890       900       910       920       930       940       950       960
GCAGCACTCATTGTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCAATATTCGAGATCGTGTCAACTGATCAAACAAT  934
GCAGCACTCATTGTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCAATATTCGAGATCGTGTCAACTGATCAAACAAT  952
GCAGCACTCATTGTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCAATATTCGAGATCGTGTCAACTGATCAAACAAT  960
GCAGCACTCATTGTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCAATATTCGAGATCGTGTCAACTGATCAAACAAT  706
GCAGCACTCATTGTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCAATATTCGAGATCGTGTCAACTGATCAAACAAT  959
GCAGCACTCATTGTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCAATATTCGAGATCGTGTCAACTGATCAAACAAT  960
GCAGCACTCATTGTGGGTTCAGGCCCTCACTTGGCCGTAGAACGGCCAATATTCGAGATCGTGTCAACTGATCAAACAAT  914
```

CTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCT

```
        970       980       990       1000      1010      1020      1030      1040
CTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTGCCCT  1014
CTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTGCCCT  1032
CTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCT  1040
CTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCT  786
CTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCT  1039
CTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCT  1040
CTTGCCGGACACTGAGAAGGCAATGAAGTTACACTTGAGAGAGGGAGGGTTGACGTTTCAGTTGCATAGAGATGTACCCT  994
```

TGATGGTCGCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACTCAGTTTTC

```
        1050      1060      1070      1080      1090      1100      1110      1120
TGATGGTCGCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACTCAGTTTTC  1094
TGATGGTCGCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACTCAGTTTTC  1112
TGATGGTCGCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACTCAGTTTTC  1120
TGATGGTCGCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACTCAGTTTTC  866
TGATGGTCGCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACTCGGTTTTC  1119
TGATGGTCGCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACTCGGTTTTC  1120
TGATGGTCGCAAAGAACATAGAGAACGCAGCGGAGAAAGCGTTGTCTCCACTAGGGATAACTGATTGGAACTCAGTTTTC  1074
```

TGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGATAAGTTAAGGGC

```
        1130      1140      1150      1160      1170      1180      1190      1200
TGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGATAAGTTAAGGGC  1174
TGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGATAAGTTAAGGGC  1192
TGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGATAAGTTAAGGGC  1200
TGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGATAAGTTAAGGGC  946
TGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGATAAGTTAAGGGC  1199
TGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGATAAGTTAAGGGC  1200
TGGATGGTGCACCCAGGTGGTCGAGCCATATTGGACCAGGTGGAGCGAAAACTAAACCTTAAGGAAGATAAGTTAAGGGC  1154
```

```
      TAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTGTTCATCATTGACGAGGTGAGGAAGAGAT
           1210      1220      1230      1240      1250      1260      1270      1280
      TAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTGTTCATCATTGACGAGGTGAGGAAGAGAT  1254
      TAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTGTTCATCATTGACGAGGTGAGGAAGAGAT  1272
      TAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTGTTCATCATTGACGAGGTGAGGAAGAGAT  1280
      TAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTGTTCATCATTGACGAGGTGAGGAAGAGAT  1026
      TAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTGTTCATCATTGACGAGGTGAGGAAGAGAT  1279
      TAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTGTTCATCATTGACGAGGTGAGGAAGAGAT  1280
      TAGCAGGCATGTGCTTAGTGAATACGGAAACCTGATTAGCGCTTGTGTGTTGTTCATCATTGACGAGGTGAGGAAGAGAT  1234

      CTATGGAGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTGTTTTGTTTGGATTTGGACCGGGTATGACTGTT
           1290      1300      1310      1320      1330      1340      1350      1360
      CTATGGAGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTGTTTTGTTTGGATTTGGACCGGGTATGACCGTT  1334
      CTATGGAGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTGTTTTGTTTGGATTTGGACCGGGTATGACCGTT  1352
      CTATGGAGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTGTTTTGTTTGGATTTGGACCGGGTATGACTGTT  1360
      CTATGGAGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTGTTTTGTTTGGATTTGGACCGGGTATGACTGTT  1106
      CTATGGAGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTGTTTTGTTTGGATTTGGACCGGGTATGACTGTT  1359
      CTATGGAGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTGTTTTGTTTGGATTTGGACCGGGTATGACTGTT  1360
      CTATGGCGGAAGGGAAGAGTACAACCGGTGAAGGTTTGGATTGCGGTGTTTTGTTTGGATTTGGACCGGGTATGACTGTT  1314

      GAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTCCGGTTGCCAATGGAAACTGATCATTGTTGTTTGCAAAATATTA
           1370      1380      1390      1400      1410      1420      1430      1440
      GAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAATGGAAACTGATCACTGTTGTTTGCAAAATATTA  1414
      GAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAATGGAAACTGATCACTGTTGTTTGCAAAATATTA  1432
      GAGACCGTTGTTCTTCGTAGCGTCCGCGTTACTGCTCCGGTTGCCAATGGAAACTGATCATTGTTGNTTGCAAAATATTA  1440
      GAGACCGTTGTTCTTCGTAGCGTCCGCGTTACTGCTCCGGTTGCCAATGGAAACTGATCATTGTTGNTTGCAAAATATTA  1186
      GAGACTGTTGTTCTTCGTAGCGTCCGCCTTACTGCTCCGGTTGCCAATGGAAACTGATCATTGTTGTT-GCAAAATATTA  1438
      GAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTCCGGTTGCCAATGGAAACTGATCATTGTTGTT-GCAAAATATTA  1439
      GAGACTGTTGTTCTTCGTAGCGTCCGCGTTACTGCTGCGGTTGCCAATGGAAACTGATCACTGTTGTTTGCAAAATATTA  1394

      CTTTTTACTACGGTATGTTTCCTTGTTTATGAGTTTGTCATTCACCTATGATAATAGGGTTTGTATTTTTCTTGTTTATG
           1450      1460      1470      1480      1490      1500      1510      1520
      CTTTTTACTACGGTATGTTTCCTTGTTTATGAGTTTGTCATTCACCTATGATAATAGGGTTTGTATTTTTCTTGTTTATG  1494
      CTTTTTACTACGGTATGTTTCCTTGTTTATGAGTTTGTCATTCACCTATGATAATAGGGTCTGTATTTTTCTTGTTTATG  1512
      CTTTTTACTACGGTATGTTTCCTTGTTTATGAGTTTGTCATTCACCTATGATAATAGGGTTTGTATTTTTCTTGTTTATG  1520
      CTTTTTACTACGGTATGTTTCCTTGTTTATGAGTTTGTCATTCACCTATGATAATAGGGTTTGTATTTTTCTTGTTTATG  1266
      CTTTTTACTGCGGTATGTTTCCTTGTTTATGAGTTTGTCATTCACCTATGATAATAGGGTTTGTATTTTTCTTGTTTATG  1518
      CTTTTTACTGCGGTATGTTTCCTTGTTTATGAGTTTGTCATTCACCTATGATAATAGGGTTTGTATTTTTCTTGTTTATG  1519
      CTTTTTACTACGGTATGTTTCCTTGTTTATGAGTTTGTCATTCACCTATGATAATAGGGTCTGTATTTTTCTTGTTTATG  1474

      ATTATATATTCTCAAAGATGCATGATGTAAGTTGGCAATTAAATAAAGATTGTTTTTCCTATGAATAATATAAGATTACA
           1530      1540      1550      1560      1570      1580      1590      1600
      ATTATATATTCTCAAAGATGCATGATGTAAGTTGGCAATTAAATAAAGATTGTTTTTCCTATGAATAATATAAGATTACA  1574
      ATTTTATATTCTCAAAGNNNNATGATGTAAGTTGGCAATTAAATAAAGATTGTTTTTCCTATGAATAATATAAGATTACA  1592
      ATTATATATTCTCAAAGATGCATGATGTAAGTTGGCAATTAAATAAAGATTGTTTTTCCTCTGAATAATATAAGATTACA  1600
      ATTATATATTCTCAAAGATGCATGATGTAAGTTGGCAATTAAATAAAGATTGTTTTTCCTCTGAATAATATAAGATTACA  1346
      ATTATATATTCTCAAAGATGCATGATGTAAGTTGGCAATTAAATAAAGATTGTTTTTCCTATGAATAATATAAGATTACA  1598
      ATTATATATTCTCAAAGATGCATGATGTAAGTTGGCAATTAAATAAAGATTGTTTTTCCTCTGAATAATATAAGATTACA  1599
      ATTTTATTTTCTCAAAGATG----ATGTAAGTTGGCAATTAAATAAAGATTGTTTTTCCTATGAATAATATAAGATTACA  1550
```

8. Characteristics of four designed primers for re-sequencing

| Upper Primer: | 28-mer | 5' TTCTTTTCTTTTCATTCCCTTCCCTCAA | 3' |
| Lower Primer: | 28-mer | 5' TATCGTTTCTTTATCGCTGTTTTCTCAC | 3' |

DNA 250 pM, Salt 50 mM	Upper Primer	Lower Primer
Primer Tm	61.4 °C	57.5 °C
Primer Overall Stability	-52.7 kc/m	-49.2 kc/m
Primer Location	32..59	383..356

Product Tm - Primer Tm	16.5 °C
Primers Tm Difference	3.8 °C
Optimal Annealing Temperature	54.2 °C

Product Length	352 bp
Product Tm (%GC Method)	74.1 °C
Product GC Content	39.2%
Product Tm at 6xSSC	95.7 °C

| Upper Primer: | 21-mer | 5' TTAAACGCTCGACAAGACCTA | 3' |
| Lower Primer: | 18-mer | 5' GGCCTGAACCCACAATGA | 3' |

DNA 250 pM, Salt 50 mM	Upper Primer	Lower Primer
Primer Tm	49.7 °C	51.0 °C
Primer Overall Stability	-38.4 kc/m	-35.8 kc/m
Primer Location	448..468	859..842

Product Tm - Primer Tm	30.5 °C
Primers Tm Difference	1.3 °C
Optimal Annealing Temperature	56.2 °C

Product Length	412 bp
Product Tm (%GC Method)	80.3 °C
Product GC Content	53.6%
Product Tm at 6xSSC	101.9 °C

| Upper Primer: | 24-mer | 5' CCGGACACTGAGAAGGCAATGAAG | 3' |
| Lower Primer: | 24-mer | 5' TAACGCGGACGCTACGAAGAACAA | 3' |

DNA 250 pM, Salt 50 mM	Upper Primer	Lower Primer
Primer Tm	60.8 °C	61.6 °C
Primer Overall Stability	-46.2 kc/m	-48.5 kc/m
Primer Location	919..942	1345..1322

Product Tm - Primer Tm	17.2 °C
Primers Tm Difference	0.9 °C
Optimal Annealing Temperature	57.9 °C

Product Length	427 bp
Product Tm (%GC Method)	78.0 °C
Product GC Content	48.0%
Product Tm at 6xSSC	99.6 °C

| Upper Primer: | 24-mer | 5' CCGGACACTGAGAAGGCAATGAAG | 3' |
| Lower Primer: | 24-mer | 5' TAACGCGGACGCTACGAAGAACAA | 3' |

DNA 250 pM, Salt 50 mM	Upper Primer	Lower Primer
Primer Tm	60.8 °C	61.6 °C
Primer Overall Stability	-46.2 kc/m	-48.5 kc/m
Primer Location	919..942	1345..1322

Product Tm - Primer Tm	17.2 °C
Primers Tm Difference	0.9 °C
Optimal Annealing Temperature	57.9 °C

Product Length	427 bp
Product Tm (%GC Method)	78.0 °C
Product GC Content	48.0%
Product Tm at 6xSSC	99.6 °C

9. Lightscanner result
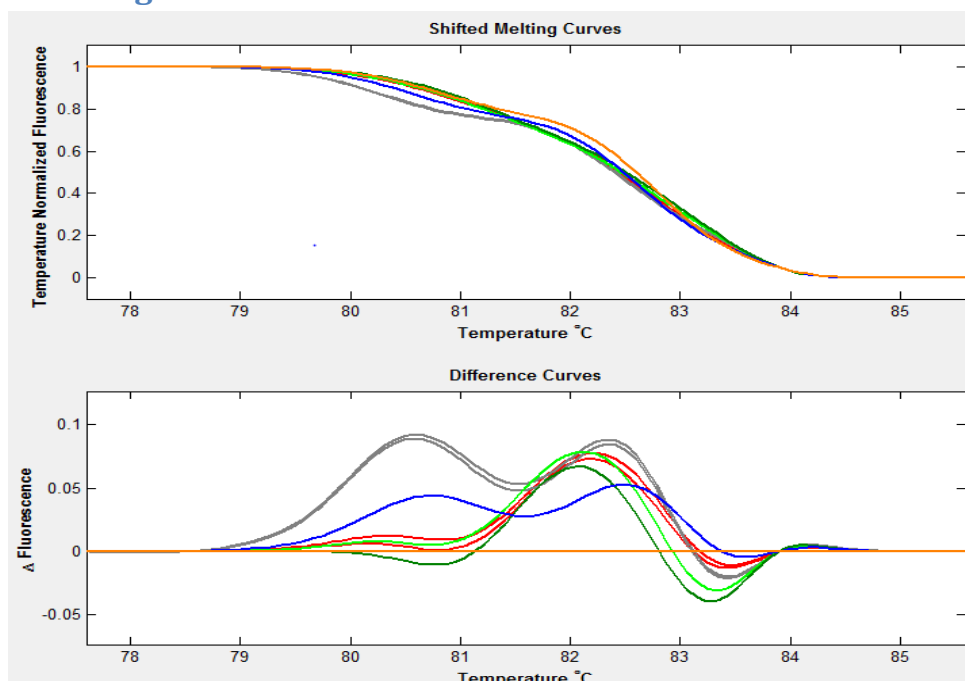
Figure 10 Lightscanner result of primer 14735-16F/124R from 10 primers designed by Sp1

Two red curves stand for Sp1 and two grey curves stand for Sp2 and this cross had good duplication and shows a polymorphic allele. The other two parents of the other cross did not have good replication which cannot conclude.
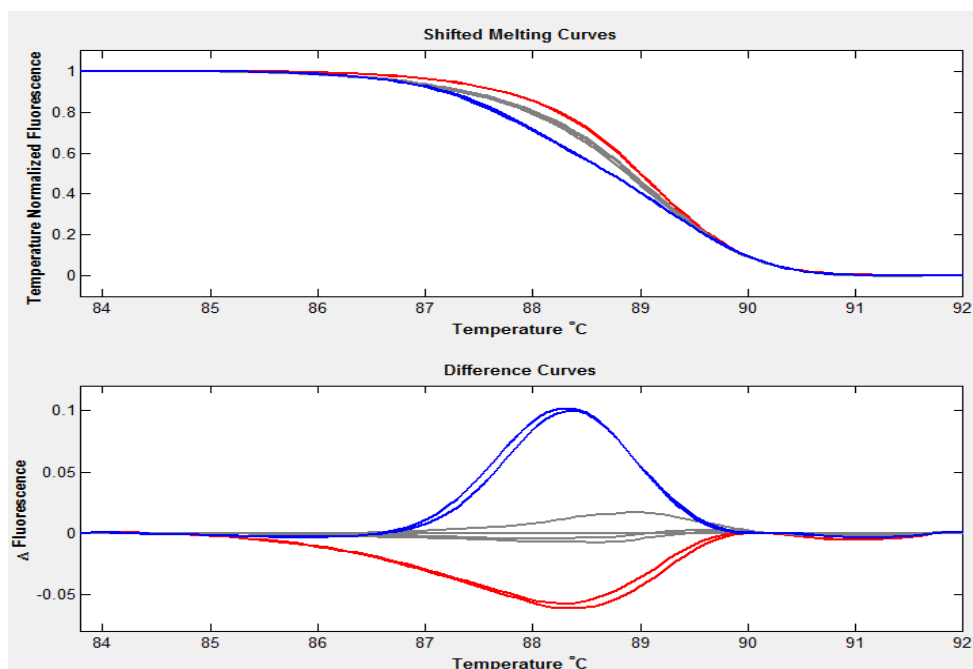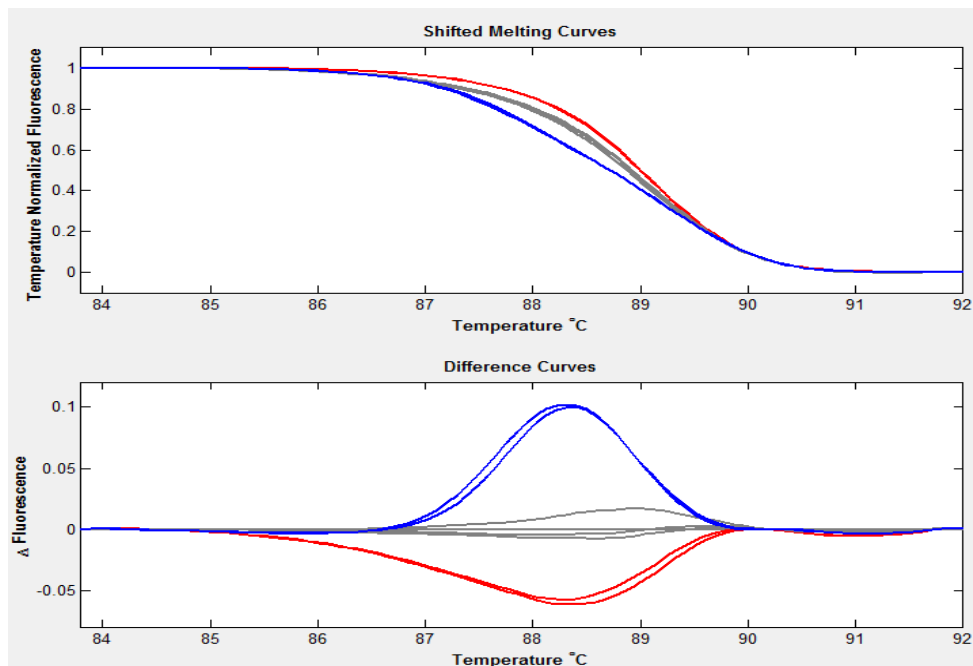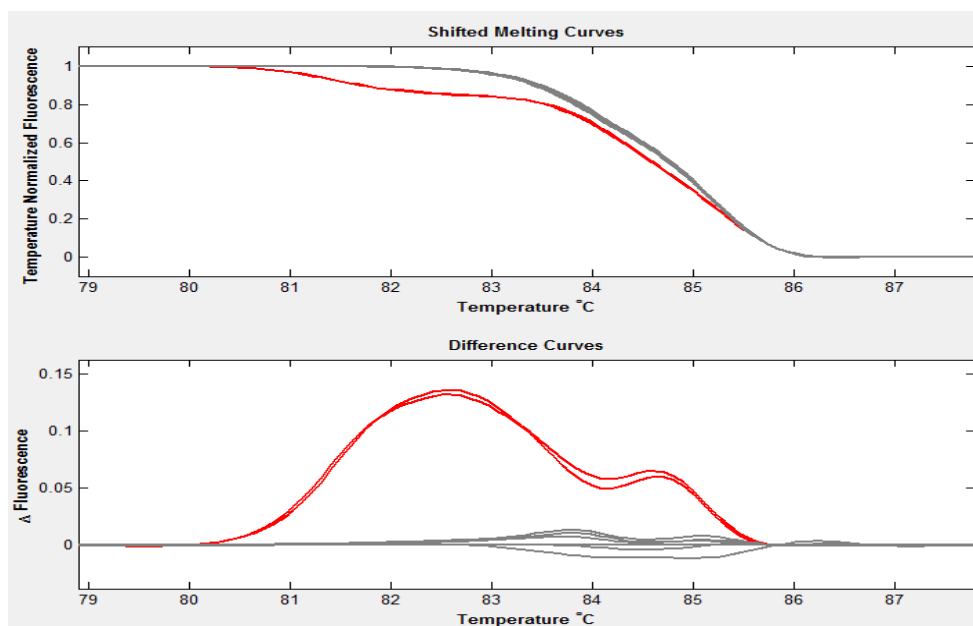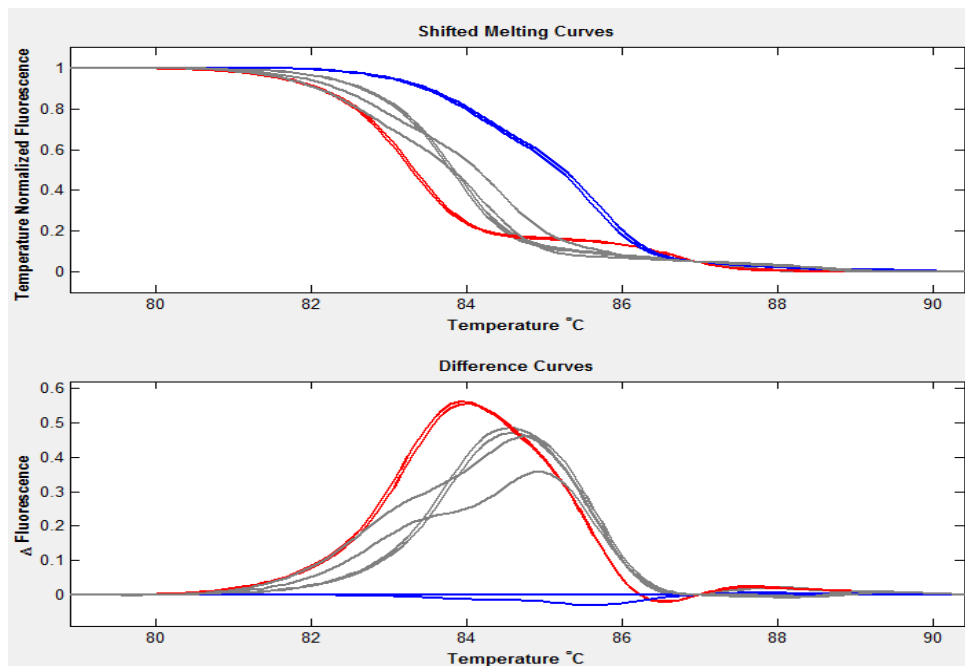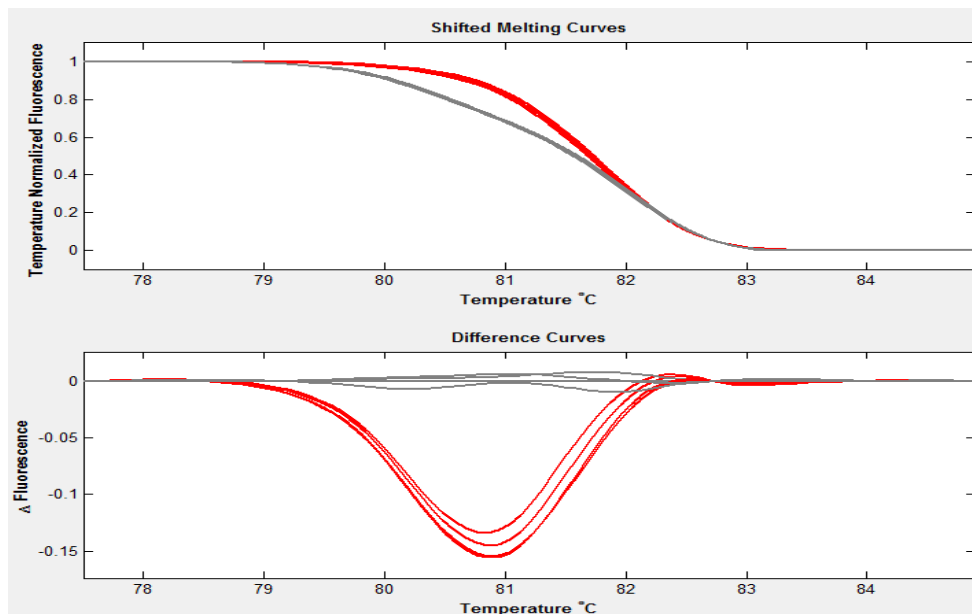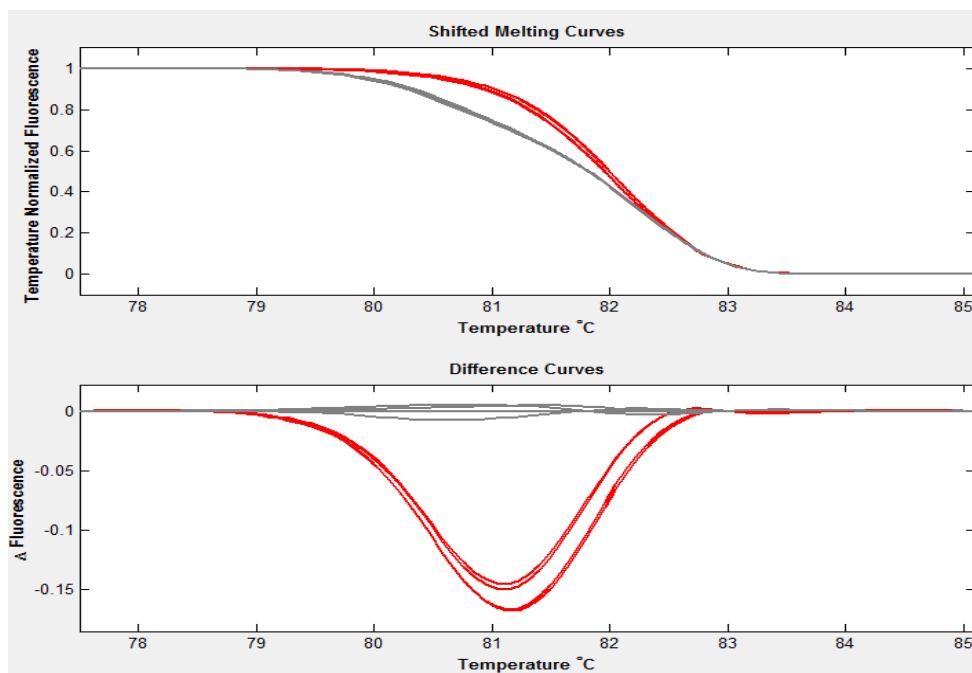

Figure 11 Lightscanner result of primer 9530-30F/134R from 10 primers designed by Sp1

Two of the grey curves stand for Sp1, two red curves stand for Sp2, two other grey curves stand for Fp1 and two blue curves stand for Fp2. They all had good duplication and show a polymorphic allele in both crosses.

Figure 12 Lightscanner result of primer 14735-27F/107R from 10 primers designed by Sp1

Two grey curves stand for Sp1 and two other grey curves which have similar pattern stand for Sp2. It did not show a polymorphic within this cross. Two red curves stand for Fp1 while for Fp2, they did not have a good duplication, but it seems that they are polymorphic due to the pattern differs.

Figure 13 Lightscanner result of primer 22553-32F/153R from 10 primers designed by Sp1

Two red curves stand for Sp1, two grey curves stand for Sp1. They had big difference between each other and do show polymorphic. The other four grey curves stand for Fp1 and Fp2. They did not have polymorphic obviously.

Figure 14 Lightscanner result of primer 3758-39F/130 from 10 primers designed by Sp1

Two grey curves stand for Sp1, two red curves stand for Sp2 and in this cross there is polymorphic. Two blue curves stand for Fp1 and two other grey curves stand for Fp2. There is polymorphic in this cross as well.
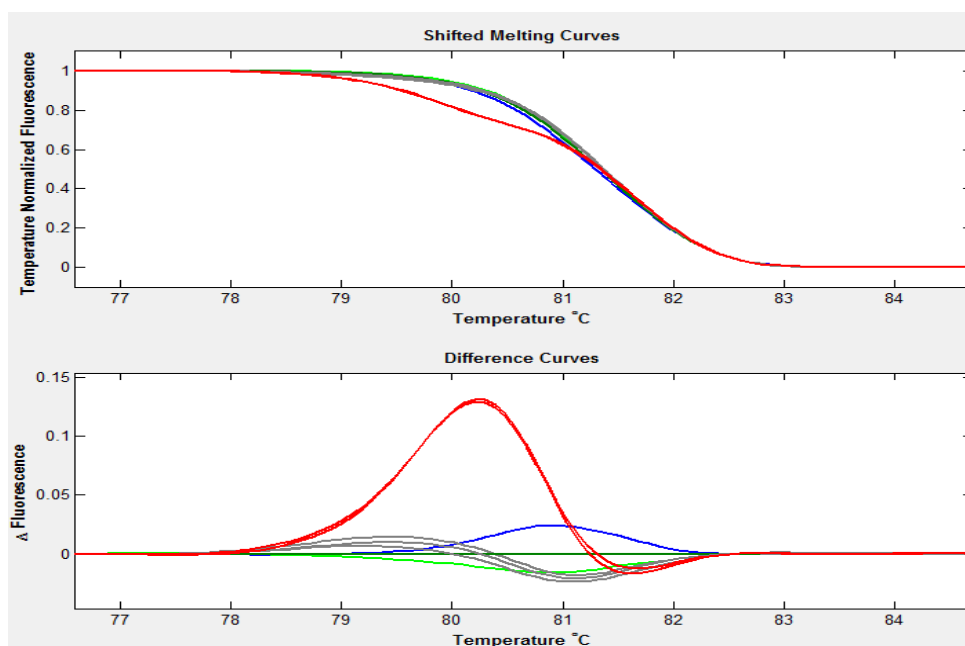
Figure 15 Lightscanner result of primer 3935-57F/150R from 10 primers designed by Sp1

In the first cross, the parents did not have similar duplication. Two grey curves stand for Fp1 and two red curves stand for Fp2. Within this cross, it shows polymorphic.
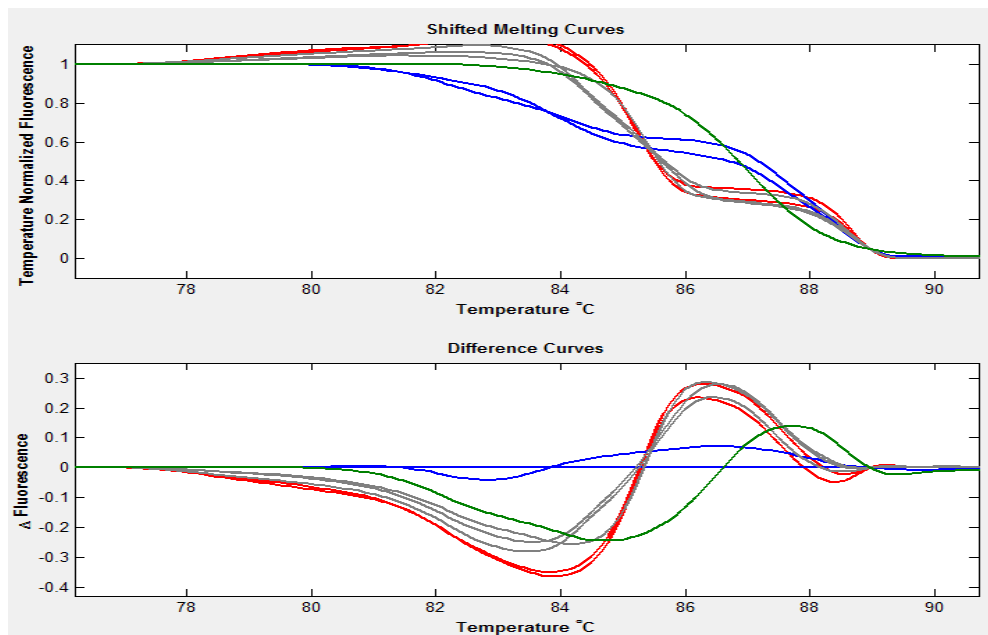
Figure 16 Lightscaner result of primer 3094-37F/148R from 10 primers designed by Sp1

Two of red curves stand for Sp1 and two of the grey curves stand for Sp2. Base on the good duplication, there is polymorphic in this cross can be conclude. The other two red curves stand for Fp1 and two grey curves stand for Fp2. There is polymorphic as well.

Figure 17 Lightscanner result of primer 8729-15F/126R from 10 primers designed by Sp1

Two of red curves stand for Sp1, two grey curves stand for Sp2. The pattern of the curves show a big difference and thus there is polymorphic in this cross. The situation is the same as the other cross. They also have polymorphic.
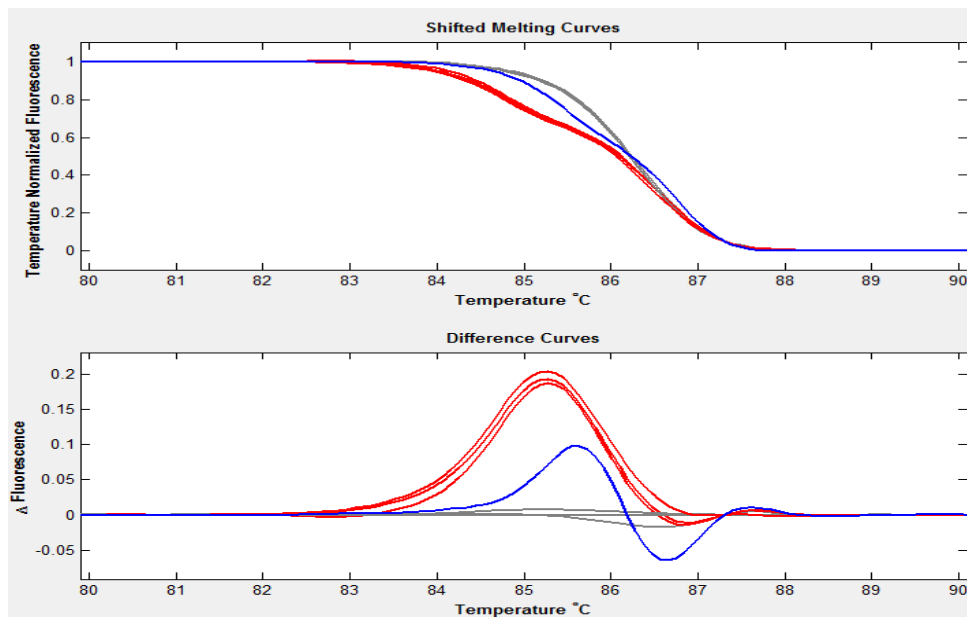
Figure 18 Lightscanner result of primer 8936-39F/142R from 10 primer designed by Sp1

One bad data of Sp1 was ticked off in the software. Two red curves stand for Sp2 and there seems show a polymorphic allele due to the big difference but lack of information of the duplication of Sp1. Two grey curves stand for Fp1 and for Fp2, there is no good duplication but it seems there is polymorphic in this cross according to the pattern difference.
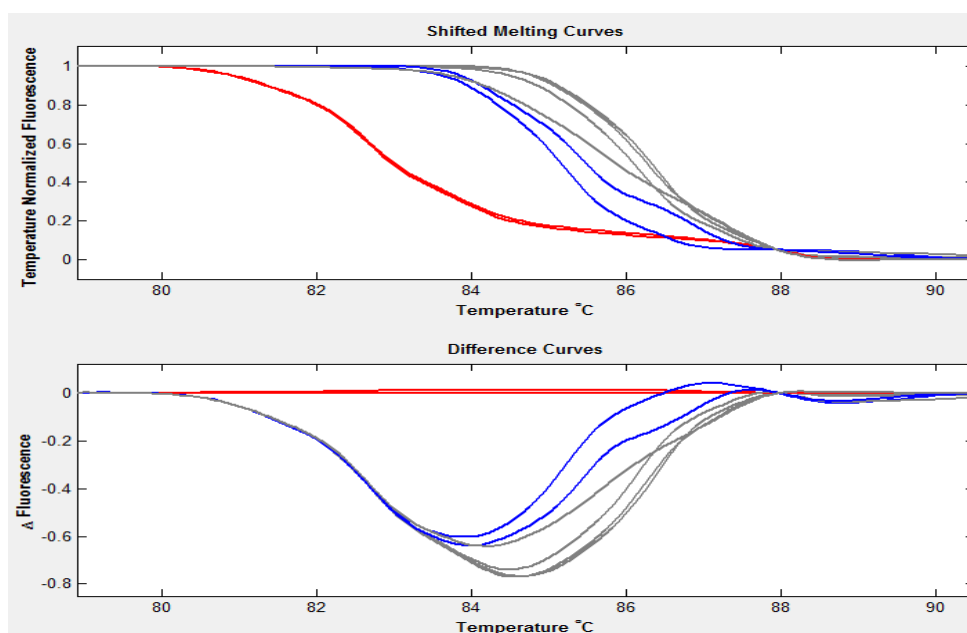
Figure 19 Lightscanner result of primer 767-13F/136R from 11 primer designed by Sp2

Sp1 were indicated by two grey curves and Sp2 were indicated by two red curves. There shows a polymorphic allele within this cross. For the second cross, they did not have a good duplication.

Figure 20 Lightscanner result of primer 16460-15F/141R from 11 primers designed by Sp2

Two red curves stand for Sp1 and two blue curves stand for Sp2. They show polymorphic in the cross. In the second cross, Fp2 did not have a good replication, so the result cannot be analysed.
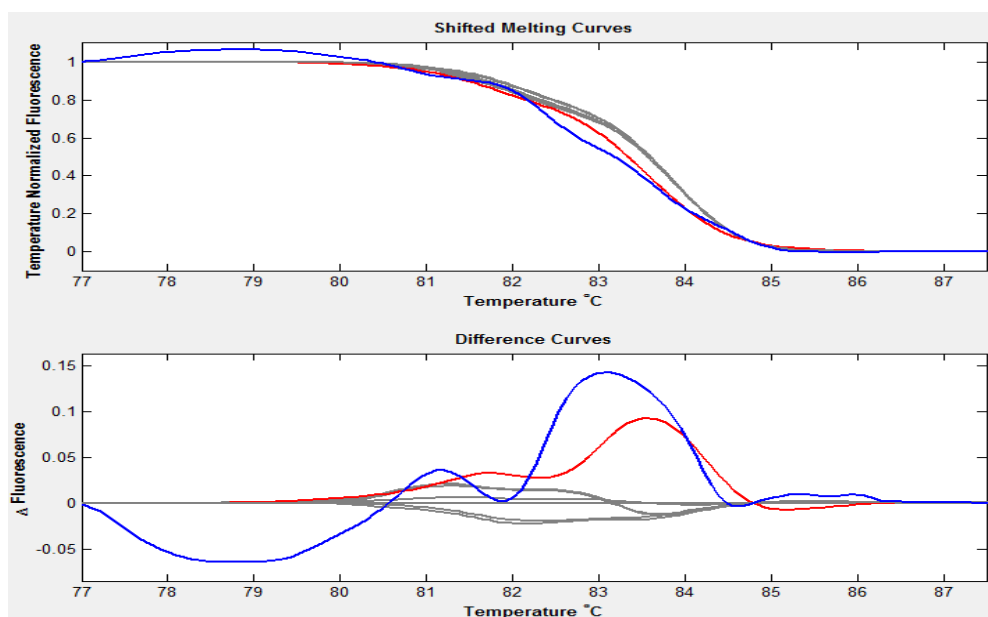
Figure 21 Lightscanner result of primer 7715-13F/106R from 11 primers designed by Sp2

Four grey curves stand for Sp1 and Sp2, thus there is no polymorphic in the cross. The second cross cannot be analysed due to the bad duplication result.
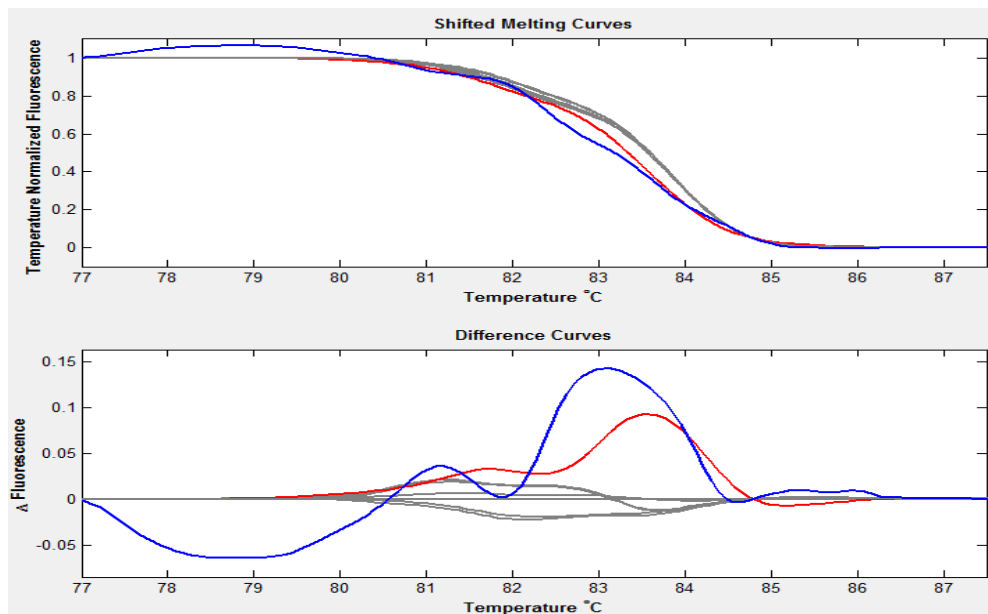
Figure 22 Lightscanner result of primer 9429-24F/151R from 11 primers designed by Sp2

For the first cross, they all show a grey curve which means there is no polymorphic. One red curve and one grey curve both stand for Fp2, but they show big difference, so it cannot be concluded from this situation.
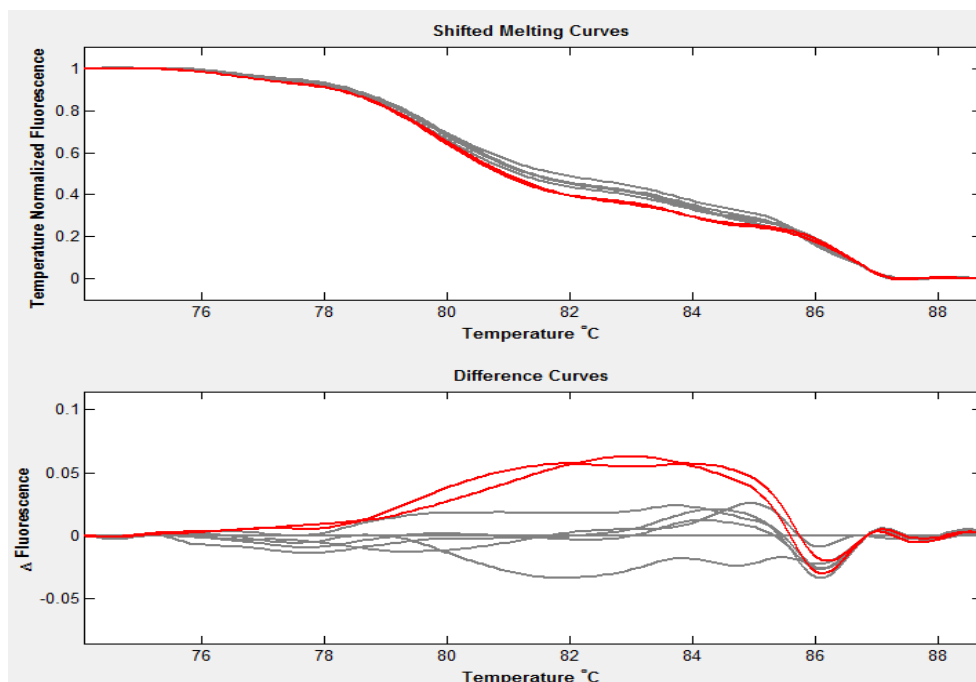
Figure 23 Lightscanner result of primer 8334-11F/129R from 11 primers designed by Sp2

There is no polymorphic in the first cross due to they were all indicated by grey. The other cross shows different colour and there is polymorphic in this cross.
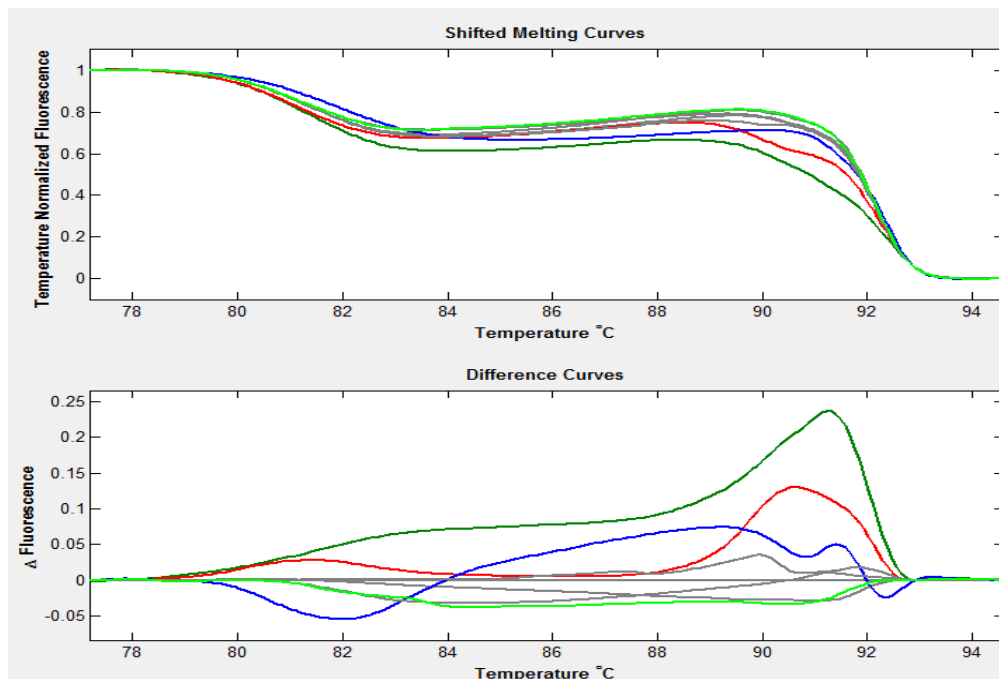
Figure 24 Lightscanner result of primer 17793-28F/127R from 11 primers designed by Sp2

This result cannot be analysed due to the fact that only Sp2 had good duplication.
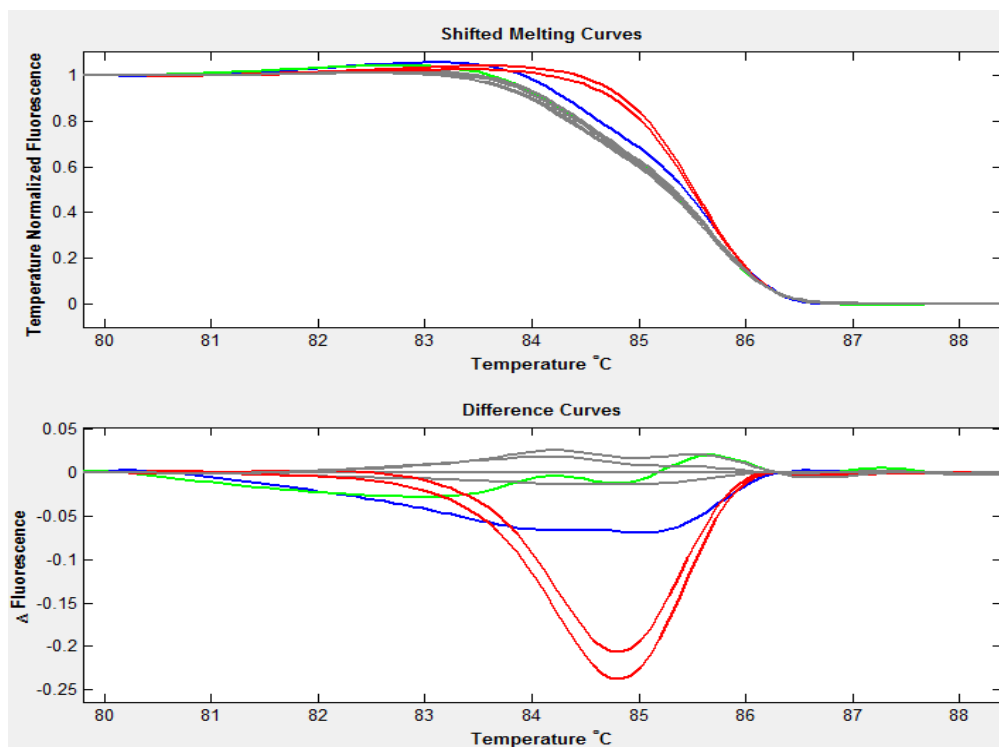
Figure 25 Lightscanner result of primer 4329-24F/138R from 11 primers designed by Sp2

For Sp2, the duplication were not good, then the first cross was no need to analysis. But due to the difference between the curves, there can be a SNP. For the second cross, they show big differences and there is a real SNP in this cross.
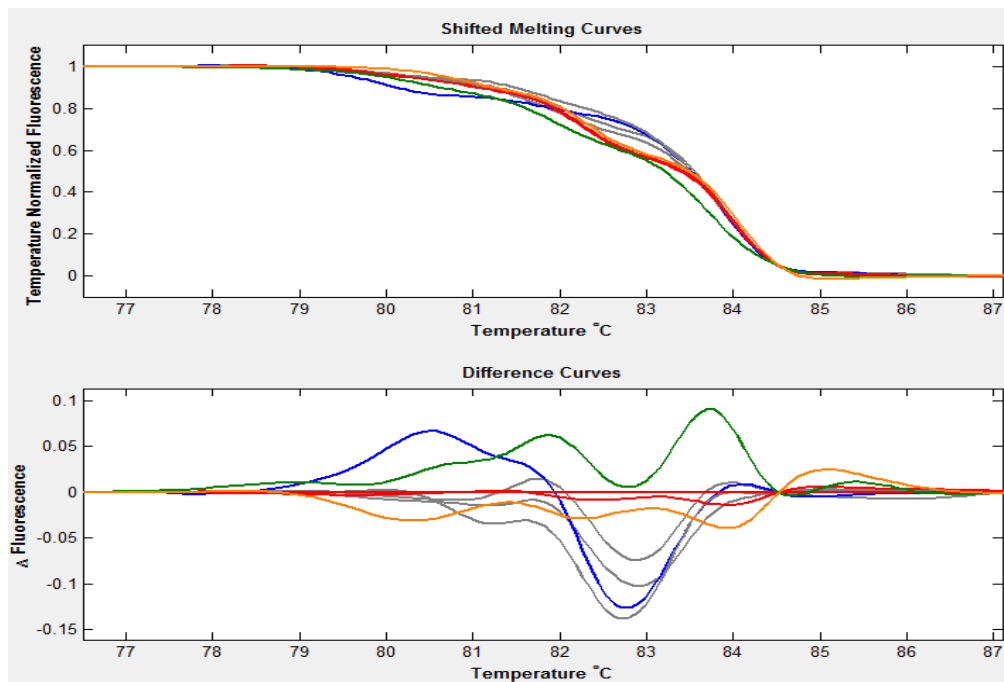
Figure 26 Lightscanner result of primer 9447-15F/153R from 11 primers designed by Sp2

Only Sp1 and Fp1 had good duplication which were shown by grey and red respectively. The other two parents did not have good duplication.
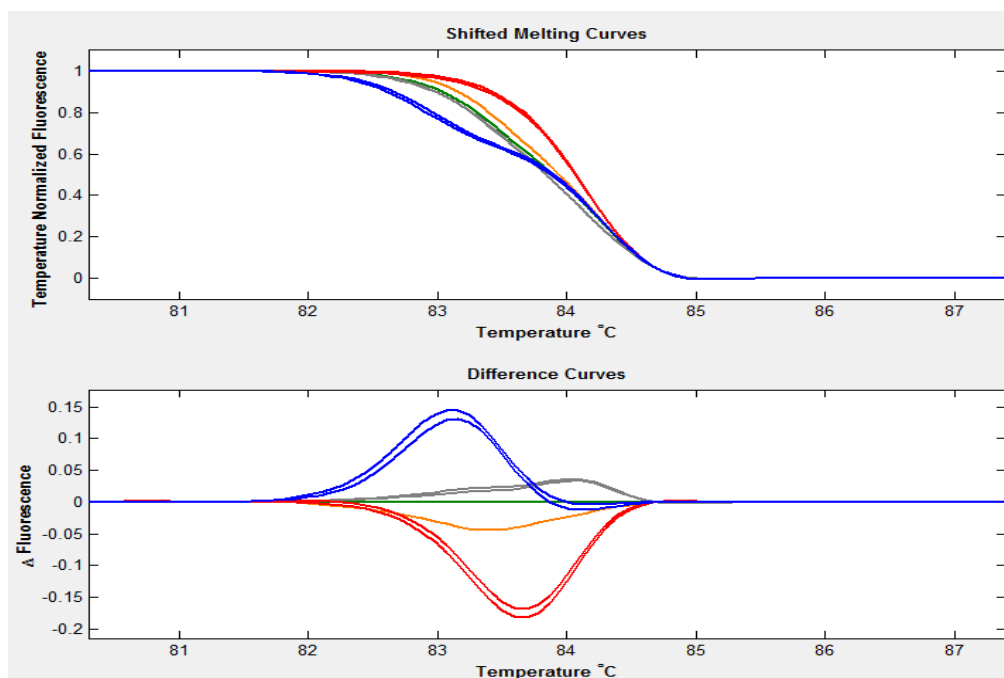

Figure 27 Lightscanner result of primer 16410-11F/136R from 11 primers designed by Sp2

Sp1 did not have good duplication which were shown by green and orange, but both show difference to grey curves which stand for Sp2. There might be a polymorphic allele in this cross. Fp1 were shown in red and F2 were shown in blue. There is polymorphic in this cross.
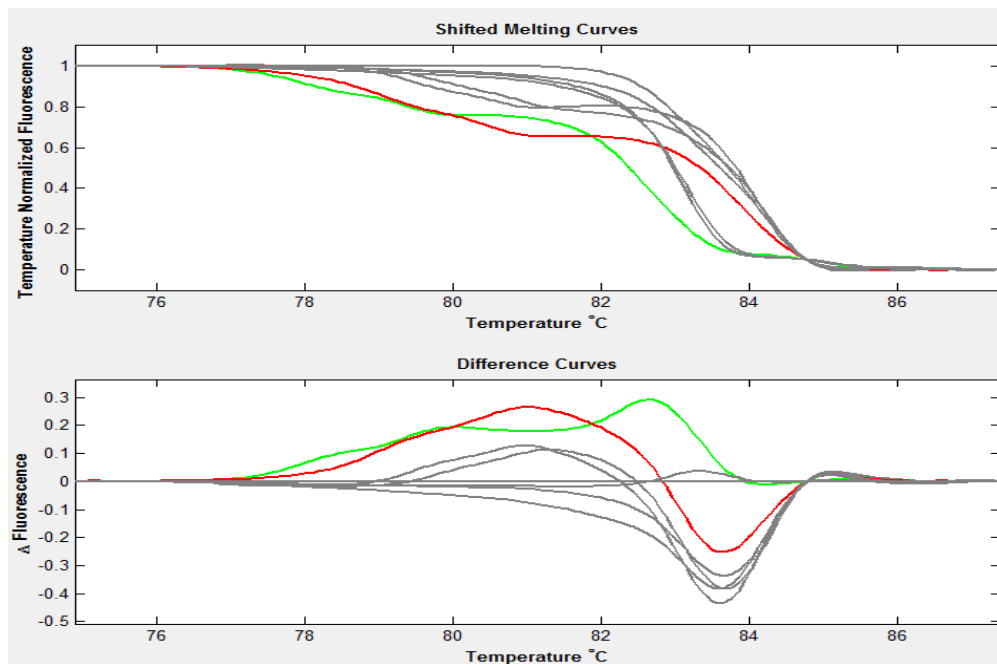
Figure 28 Lightscanner result of primer 11031-24F/126R from 11 primers designed by Sp2

Sp2 did not have good duplication which cannot be analysed. For the second cross, they were both shown in the grey colour which means there is no polymorphic allele.
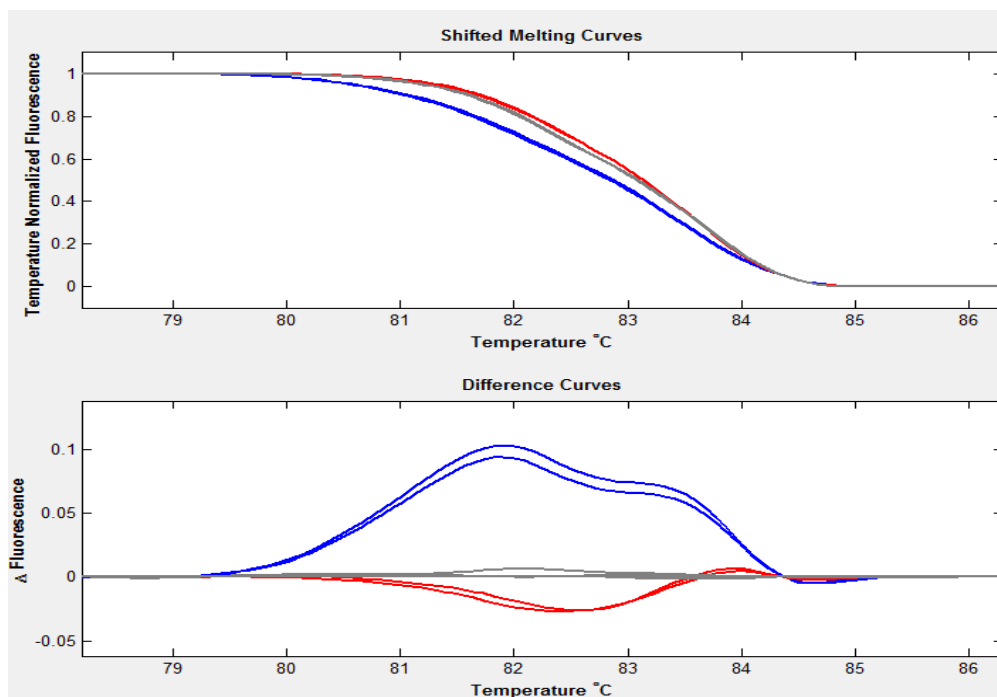

Figure 29 Lightscanner result of primer 13979-11F-120R from 11 primers designed by Sp2

Two blue curves stand for Sp1 and two red curves stand for Sp2. They show big difference so there is a real SNP due to this result. For the second cross, they were all shown in grey, so there is no polymorphic in this cross.
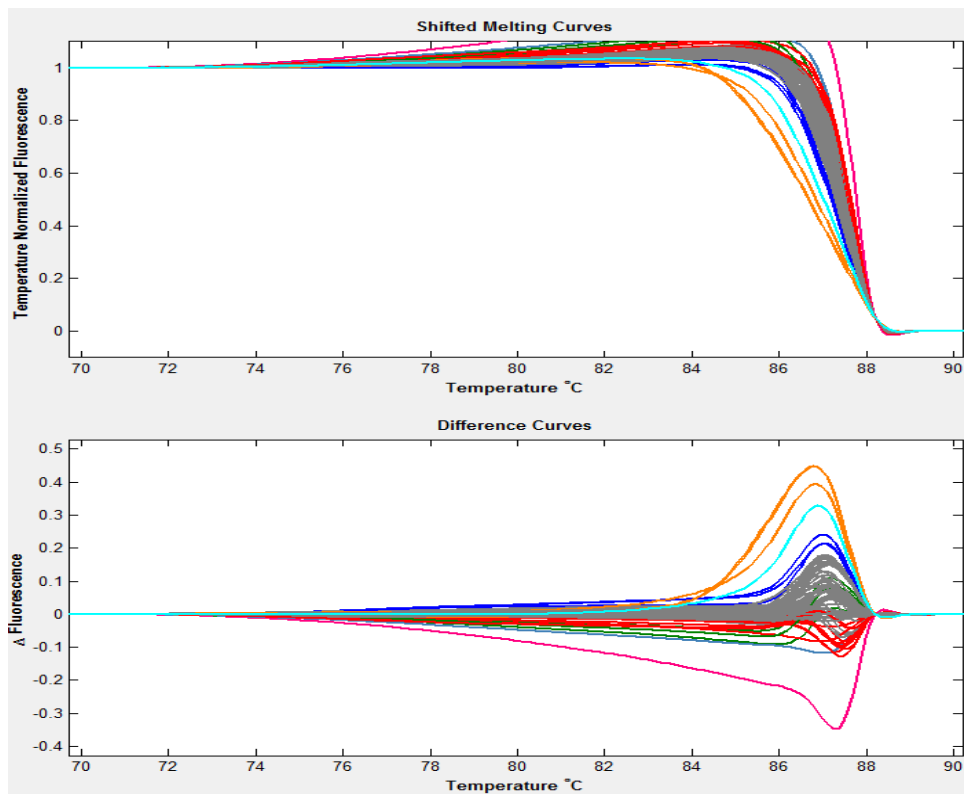
Figure 30 Lightscanner result of Florist population in one 96-well micro-plate using primer combination 1107F, 1345R

The result of Lightscanner for Florist population using this primer combination did not show a segregation. It may be because the detected SNP from the re-sequencing result was due to background noise.