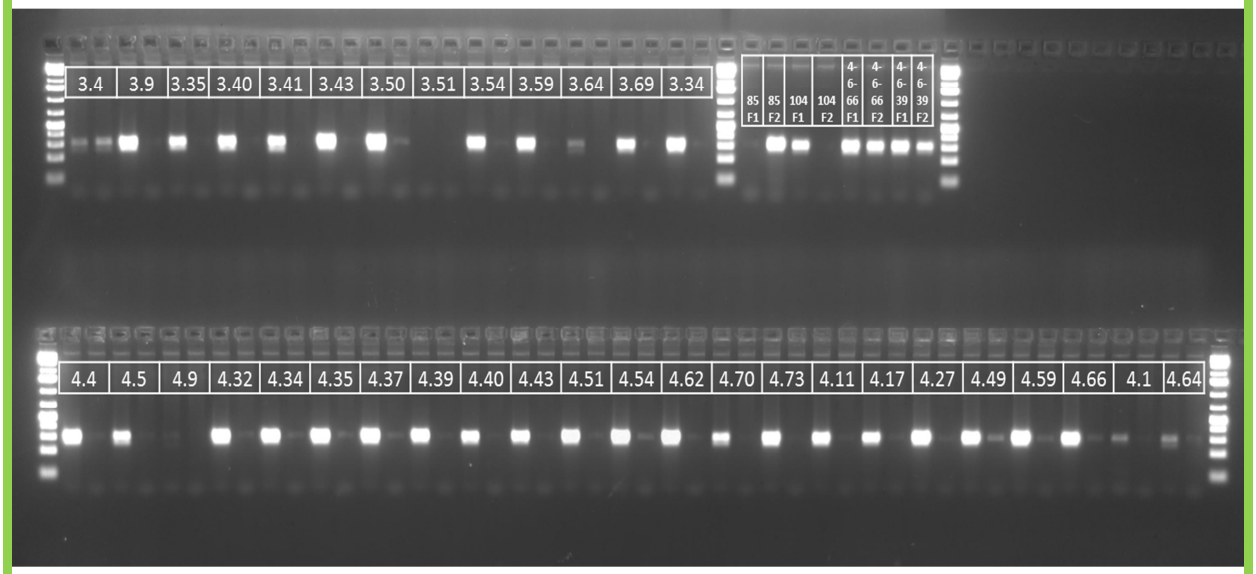
WAGENINGEN UR
For quality of life

Zygosity Diagnosis for Specific Genomic Regions of Recombinant Inbred Line Population and Development of Simple PCR-based SNP Marker Assay in Tomato



Trang Tran

MSc Plant Biotechnology

August, 2012

Zygosity Diagnosis for Specific Genomic Regions of Recombinant Inbred Line Population and Development of Simple PCR-based SNP Marker Assay in Tomato

Student: Trang Tran

Registration number: 850926841050

Supervisor: Dr. Sjaak van Heusden

Examiners: Dr. Arnaud Bovy
Dr. Sjaak van Heusden

August 2012

Acknowledgment

I would like to express my special gratefulness to my supervisor, Sjaak van Heusden, for his supports and guidance. Thanks to a great direction that he orients at the beginning of my study, I experienced a worthwhile time to learn and apply knowledge to practical approach. Moreover, I deeply appreciate his supports to complete my thesis. His valuable suggestions help me solve many obstacles during my study and improve my knowledge.

I am happy to present this thesis to Arnaud Bovy. Thanks to a good chance that he gave me for applying the method I have studied, I got a very nice result. This helps me improve not only my knowledge but also my confidence for future career.

Through this acknowledgement, I would like to thank Jos Molthoff and Fien Meijer-Dekens for guiding me laboratorial works and encouraging me during time I did my experiments.

I wish to thank my family and my boyfriend for their love and supports which inspire me to pursue my study goals. Last but not least, I like to give many thanks to my friends who encourage and help me complete my thesis.

Thank you!

Trang Tran

Summary

The main purpose of this study is making a selection to have a Recombinant Inbred Line population with the same genetic background for evaluating the effect of Non Smoky Glycosyl Transferase gene 1 (NSGT1) on smokiness trait. Following the previous studies, some regions on genome of F6 RILs still heterozygous, so that we continue screening and selecting for those specific regions. Applying Marker-Assisted Selection, two main methods was utilized for screening with SNP markers. Firstly, 116 F7 RILs were genotyped with 19 SNP markers by KASP technique at Van Haeringen Lab. Secondly, with the small scale of screening, we intend to have a simple and rapid method for selection, so that the simple PCR based SNP markers assay was developed for screening two SNP makers on 285 F8 RILs. The screening with 19 SNP markers resulted in 4 F7 RILs selected. In addition, the simple PCR based SNP marker assay was developed successfully, so we applied this method to screen F8 RILs and received 52 desired homozygous plants for the taste trial on smoky flavor.

Besides the simple PCR-based SNP marker method, we made a small trial to evaluate the possibility of using High- Resolution Melting analysis for selecting the desired homozygous genotype. The result of this trial illustrated that homozygotes can be discriminated from heterozygotes due to the difference in melting curves. However, two different homozygous genotypes did not show any difference in melting curves.

Table of contents

List of tables and figures

I. Introduction

General information about tomato history, Marker-Assisted Selection and tomato breeding

Tomato belongs to the *Solanaceae* family and it is one of the most important cultivated crops over the world for both fresh market and processing industry. In addition, tomato is also a main source of carotenoid lycopene, which has a great capacity in protecting human beings against cancer and cardiovascular disease (Giovannucci, 1999). According to Sims (1980), all related wild species of tomato originated from Andean region now encompassed by parts of Chile, Bolivia, Ecuador, Colombia and Peru. However, related to the place of domestication of the cultivated tomato, there are two competing hypotheses, one from Peru, and another from Mexico (Peralta and Spooner, 2002). Compared to wild species, the genetic diversity of cultivated tomato is generally poor. In fact, the estimation of genomes of tomato cultivars contains less than 5% of the genetic variation of their wild relatives (Miller and Tanksley, 1990). One of the reasons for lack of genetic diversity in cultivated tomato may be because the selection of horticultural crop like tomato is usually done on a single plant**,** so that the genetic variation tends to decrease (Bai and Lindhout, 2007). In order to improve the gene pool of tomato, the collection and description of genetic materials are necessary. The tomato gene bank has been established in USA and other countries where more than 7500 accessions are preserved (Peralta and Spooner, 2002). The breeding programs usually interested in characteristics which can reduce production costs, increase yield or improve quality. One of the important issues in tomato breeding is breeding for resistance to pests and pathogens because there are more than 200 hosts tomato species which can lead to severe economic losses (Bai and Lindhout, 2007). Besides, breeders are also interested in developing F1 hybrids because of heterosis which can improve the characteristics of F1 for adaptability and productivity. In addition, F1 hybrids also can help protect against illegal reproduction and provide an uniformity for production (Bai and Lindhout, 2007).

Nevertheless, when many breeding programs focus on improving yield or resistance, the flavor quality has declined. Because of the selection for certain traits, the other characteristics may not receive enough attention from breeders. Moreover, flavor is a complex quality trait since it involves a large amount of primary and secondary metabolic pathways (Klee, 2010). Therefore, studying about the relationship between the pathways and genes controlling synthesis of the component of flavor will play an important role for breeding.

The genome sequence of inbred tomato cultivar Heinz 1706 was available from May, 2012, thanks to the tomato sequencing project (SOL) with the participating of 10 countries: Korea, China, The United Kingdom, India, The Netherlands, France, Japan, Spain, Italy and the United States. The high- quality genome sequence of Heinz 1706 is approximately 900 megabase (Mb) and a 739 Mb draft sequence of its wild relative (*Solanum pimpinellifolium*) is also available. These genome sequences and the availability of millions of SNPs provide powerful tools for breeding (Zamir and Giuliano, 2012)

Nowadays, marker-assisted selection plays a crucial role in every breeding program. Molecular markers are usually used for constructing linkage map and accelerating the breeding program by assisting selection of desired traits at the early stages (Collard et al., 2005). Molecular markers represent the genetic differences between individuals of the same or different species. They can be a difference in target genes themselves or the difference of a nearby DNA sequence (Collard et al., 2005). Based on the presence or absence of a maker which is linked to a certain trait, the selection can be made in an efficient and reliable way without the need of phenotyping. There are a number of different DNA makers such as restriction fragment length polymorphism (RFLPs), random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphism (AFLPs) and simple sequence repeat (SSRs) (Mohan et al., 1997). Among these, Single Nucleotide Polymorphism (SNP) is one of the most common markers used in many breeding programs. The most straight forward way to detect SNP is sequencing. However, sequencing techniques usually require a high initial cost for investment on facilities, so those techniques are often performed at some specialized institutes or companies. Sanger technology is widely used for decoding genome sequences of many species including tomato (Zamir and Giuliano, 2012). Next generation sequencing technologies, especially Illumina system, is utilized by many companies for high throughput SNPs discovery. In order to screen for SNPs marker, there are various assays. For examples, the KBioscience Competitive Allele-Specific PCR genotyping system (KASP) patented by KBioscience company is an efficient and reliable technique for SNP screening. The principle of this technique is based on the combination of three unlabelled primers, the universal florescent reporting system and specially-developed Taq polymerase (Robinson and Holme, 2011). Besides, there are many other different ways for SNP genotyping based on PCR technique. First of all, Cleave Amplified polymorphic sequences (CAPS) assay which utilize a restriction enzyme to digest amplified fragments can be applied to screen for SNP markers. Particularly, the restriction enzyme is selected according to SNP locus so that the

polymorphic nucleotide can differentiate the number of enzyme's recognition sites between two homozygous genotypes. For example, the amplified fragments from one homozygous genotype have two recognition sites while the amplified fragments from other homozygous genotype have three recognition sites. Hence, when utilizing this restriction enzyme for digesting PCR products, the number of digested fragment types will be three and four for these two genotypes correspondingly. Heterozygous genotypes consist of both types of amplified fragments so they will produce five different types of digested fragments (Konieczny and Ausubel, 1993). However, this method is not preferred because it required an additional step for treating PCR products with restriction enzymes. Secondly, the method called a modified allele specific PCR assay is easier for operating. In principle, SNP markers can be detected by utilizing allele specific primer which is designed regarding to SNP locus. One allele specific primer has perfectly matched at 3' terminal nucleotide with specific allele and has a mismatch with non-specific allele. Because of the mismatch with non- specific allele, Taq polymerase is not efficiently in amplifying the target sequence. Meanwhile, the allele specific primer can fully bind to the end of desired target sequence, so it is much more efficient to amplify the target fragment. Therefore, SNP can be detected by the presence or absence of PCR products on standard agarose gel (Hayashi et al., 2004). Nevertheless, in practice, only one base mismatch at 3' terminus is not enough to discriminate between the two alleles by standard PCR. In order to overcome this problem, additional mismatch has been incorporated to both allele specific and non-allele specific primers within last four bases from the 3' terminus. This modified allele specific PCR based SNP technique is able to detect SNP markers in the Arabidopsis and Rice genome (Hayashi et al., 2004 and Drenkard et al., 2000).

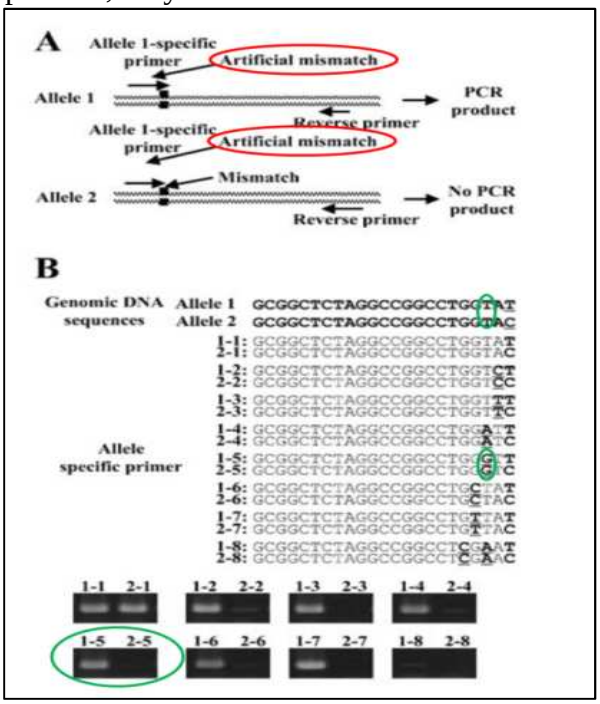


Figure1. Optimal artificial mismatch for specific and non- specific primers (source: Hayashi et al, 2004)

According to the study of Hayashi et al, 2004, the additional mismatches were incorporated to different positions within last four bases at 3' terminus of allele specific and non- allele specific primers, then PCR were performed to test for the optimal

incorporated mismatch position. The result of this study presented that the optimal mismatch should be incorporated at the third base from 3' terminus and the substitution should be either T-G or C-A (Hayashi et al., 2004) (Figure 1)

Besides above methods based on PCR, some other methods can be used for detection SNP markers due to the difference of melting temperature of PCR products. There are several High-Resolution Melting (HRM) analysis systems which used different dyes for SNP genotyping. However, in this report, we focus on lightscanner system which can detect the variation in DNA sequence by utilizing double strand DNA dye LCGreen Plus. The reason of choosing this method is because of its simplicity. Comparing to other system, lightscanner can perform without probe or unnecessary to have additional steps to separate PCR products on gel electrophoresis (Reed and Wittwer, 2004). In order to implement Hi-Res Melting analysis in the lightscanner, the target sequence needs to be amplified in the presence of LCGreen plus and then the melting profile of all samples will be analyzed by specialized software developed for lightscanner. Although lightscanner has been proven with many advantages such as highly sensitivity for mutation scanning and SNP genotyping and providing fast results, this technique often cannot discriminate between two homozygous sequence variants (Liew et al., 2004). However, this limitation can be overcome by combining DNA of unknown homozygous genotype with known homozygous parental genotype and then detect heterozygous sequence by analysing melting curve (Reed and Wittwer, 2004 and Wu et al., 2008)

Based on study of Tikunov et al. (2010) a collection of 94 tomato cultivars representing the current diversity of commercial germplasm were screened for variation of their volatiles components. As a result of that study, the emission of three phenylpropanoid (PhP) volatiles: methyl salicylate, guaiacol and eugenol were found different within the germplasm collection. Fruits from one group tomato cultivars had capacity to emit considerable amount of these PhP volatiles while other group hardly emit those volatiles. Furthermore, the study of PhP volatiles between two group of tomato cultivars has indicated that the emission of those volatiles took place upon disruption of fruit tissue through cleavage of glycoconjugates (Tikunov et al., 2010). However, at the latter stage of fruit ripening of one group of tomato fruits, those volatiles were arrested by conversion of hexose-pentoside precursors and into glycoconjugate species with a higher complexity. Therefore, this group of tomato cultivars were unable to emit PhP volatiles when fruit tissues were disrupted (Tikunov et al., 2010). In fact, it was supposed that due to modification of glycosylation transferase genes, PhP volatile emission in low-PhP volatile fruits was arrested (Tikunov et al., 2010). Then, based on

various study strategies with the combination of metabolomics, genetic, gene expression and functional gene analyses, sets of four cultivars with the maximum of diversity of characteristics have been selected for further studies.

One of parents of each of four selected lines, namely C074, C085, R075 and R104 was used for intercrossing in a half-diallel scheme (Figure 2). Six hybrid genotypes: 1, 4,6,7,9 and 20 were selfed to obtain RILs. Nevertheless, only three F6 RILs: 4, 6 and 20 were chosen. Among these three RIL populations, population 4 from the cross between cherry tomato C085 and round tomato R104 showed the segregation of PhP level. Then this F6 RI population containing 100 individuals were selected to analyze PhP level. The glycosyl transferase gene called Non Smoky Glycosyl Transferase gene 1 (NSGT1) located in chromosome 9 has been identified as a candidate gene which was responsible for glycosylation. Two alleles of NSTG1 genes are PhP^{+} and PhP^{-} and the tomato cultivars carry PhP^{+} allele has shown volatile emission and give tomato fruits smoky flavor while PhP^{-} allele plays a role in glycosylation of volatile conjugate then inhibit the emission of PhP volatile and express a non-smoky flavor. The C085 parental line is homozygous for the PhP- allele and the R104 parental line is homozygous for the PhP^{+} allele (Figure 2).

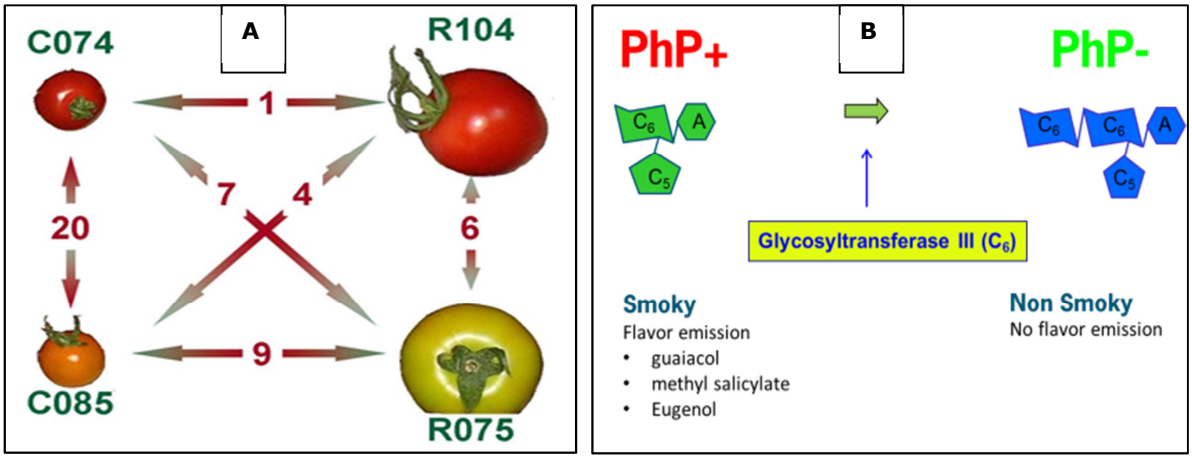


Figure2. A half-diallel scheme for intercrossing 4 parental lines (A) and two alleles of smokiness gene (B)

Purpose of this study

Following the results of previous studies on smokiness trait in tomato, F7 RILs are still heterozygous in some regions, so they may affect to result of taste trial. Hence, our goal is genotyping these heterozygous regions and making a selection on F8 RILs to have lines with the same genetic background. Particularly, the target is screening for two regions on chromosome I and IV based on two SNP makers.

Therefore, we will focus on developing simple PCR based SNP marker assay for diagnosis of these specific regions in F8 RIL population. These assays will facilitate selection for the taste trial in order to evaluate the effects of NSTG1 gene on smokiness trait. Not only in our study program but also in many other breeding programs, it is not always a lot of markers or samples need to be genotyped at the same time, so that some simple techniques will be very useful and more cost- effective for screening in a small scale.

II. Materials and methods

2.1. Plant material

In order to develop simple PCR based SNP marker assays, we used genomic DNA from two parental lines C085 and R104. Many RILs from F6 to F8 generations were also involved in this study. Particularly, two F6 RILs: 4- 6 and 4- 21, 116 F7 RILs and 285 F8 RILs which were screened with selected SNP markers by KASP and PCR based SNP marker assays.

2.2. Source of SNP markers

In order to select for SNP markers which are polymorphic between the two parental lines C085 and R104, around 6000 SNP markers of tomato were used to screen on these two parental lines by *Illumina Infinium* Genotyping Assay. This screening resulted in about 1500 good markers spread over 12 chromosomes. Together with those selected SNP markers, the flanking sequences (400 base pairs) are also available for designing primers.

2.3. DNA isolation

For large scale of DNA isolation, tomato seeds were sown for two weeks, then leaf material from each plantlet was grinded according the RETSCH Mixer Mills method. After that, the KingFisher® Flex 96 genomic DNA purification system protocol was applied. DNA samples isolated from this method were used for identifying PhP alleles, screening SNP markers by *Illumina Infinium* Genotyping Assay and screening with KASP technique. The KingFisher method hash some advantages such as fully automated, high speed purification, high throughput 96 sample processing and easy operation.

However, there is an alternative method which also provide good quality DNA and suitable for small amount of samples. For F8 RILs, we applied this method for isolation DNA. Young leaf material from each planet, which was cultivated at Hollandplant Company for about two weeks, was sampled and stored in dry ice before doing DNA isolation. These leaf materials were grinded by Retsch Mixer Mills system, then added 200 µl extraction buffer, 250 µl nuclei lysis buffer and 25 µl sarkosyl, vortexed very well and incubated in a water bath at 65^{o}C for at least one hour. The protocol was followed by adding 200 µl chloroform/isoamylalcohol (24:1) and mixed by inversing the tubes then centrifuge 4 min at 8000 rpm and rT. The supernatant (about 400 µl) was pipetted to a new 2 ml tube and added 400 ul isopropanol (mix carefully) and centrifuged 1 min at 8000 rpm. The DNA is a pellet which was rinsed for 20 minutes in 125 µl 76% EtOH with 10 mM NH4Ac (MW 77.08). In

the final step, DNA pellet was dried and dissolved in 50 µl TE (1 µl stock RNase(2 mg/ml) for 100 µl TE)

2.4. Development PCR based SNP markers assay

In order to assay SNP markers, we applied the principle of allele specific PCR method (Kwok et al., 1990; Hayashi et al., 2004). The SNP genotypes can be determined based on the presence or absence of PCR products. This method required three primers in which two primers need to be designed according SNP position. Particularly, these two primers have their sequences with only one base difference according to SNP locus and each of them will be combined to the same primer pair in two different PCR reactions. Nevertheless, this allele specific PCR method has been shown their difficulties in discriminating between two alleles based only on one base mismatch.

Therefore, in this report, we increase the specificity of these allele specific primers by applying the modified allele specific PCR method which were described by Kwok et al. (1990); Drenkard et al. (2000) and Hayashi et al. (2004). We designed primers by primer 3 program (http://frodo.wi.mit.edu) in which two forward primers had 3' end corresponding to two polymorphic alleles at SNP locus. In addition, we incorporated an additional mismatch to the third base from 3' end of forward primers (Figure 3), so that due to two mismatches on non- specific allele primer, the target sequence could not be amplified through standard PCR. Meanwhile, with only on mismatch on the specific allele primer, the PCR product of target sequence was amplified successfully. Then the presence or absence of PCR products can be visualized on 2% agarose gel.

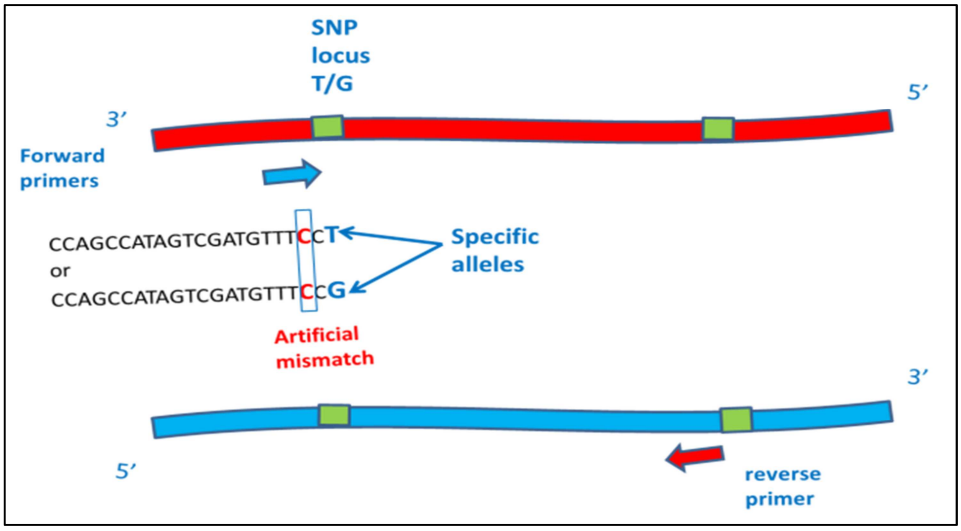


Figure3. Two allele specific primers of SNP seq-rs 207 for simple PCR based SNP marker assay.

Before applying this simple PCR based method for screening four F8 RILs with two SNP markers called seq-rs207 and seq-rs9033, we have made a small test based on two SNP markers on chromosome I, namely seq-rs7782 and seq-rs7854 in order to investigate the effect of an additional mismatch at third base from 3'end of forward primers and optimize the PCR protocol. Particularly, we designed allele specific primers which only have one mismatch at SNP locus, then doing PCR to see whether they can discriminate between two alleles.

After that, we incorporated an artificial mismatch to the third base from 3' end of forward primers to investigate the effect on the amplification products. Besides, we check for the most suitable annealing temperature for each primer pair and optimize the number of PCR cycles which can help distinguish two alleles at SNP locus. Each PCR reaction(volume 15µl) contains 10.76 µl mili-Q water, 1.5 µl Dream taq buffer (10x), 0.6 µl dNTP (5mM), 0.24 µl Dream taq polymerase, 1 µl template DNA (300ng measured by NanoDrop) and 0.45 µl for each of alleles specific forward primer and reverse primer (10mM). The amplification protocol consists of 94 ^{o}C for 2 minutes, followed by 25 to 32 cycles of 94 ^{o}C for 30 seconds, and annealing temperature which was depended on each primer pair for 30 seconds, 72 ^{o}C for 30 seconds and final elongation step for 5 minutes at 72 ^{o}C.

2.5. Primer design and PCR optimization for High Resolution Melting Analysis

Besides applying the simple PCR based SNP marker assay for screening four F8 RILs, we also make a small trial for utilizing lightscanner to screen one F8 RIL population (line 4-21-3). The reason for this is because if we succeed in using lightscanner for discriminating desired homozygous individuals based on melting curves, we need to do only one PCR reaction for each individual and can skip the step for visualizing PCR products on agarose gel. However, as stated in introduction, lightscanner usually can only distinguish between heterozygous and homozygous individuals but not between two different homozygous lines. Therefore, thanks to this trial, we can make a conclusion on the reliability of utilizing Hi-Res Melting analysis to select for desired homozygous plants based on SNP markers.

In order to apply Hi-Re Melting Analysis, we designed primers to amplify the target sequence which contains SNP locus. The Primer 3 program was also used to design primers and then those primers were synthesized at Biolegio BV company. For this trial, a primer pair was designed to amplify the target sequence which has 164 base pairs and contains the SNP marker seq rs 207. Because one of the most crucial elements to obtain the good data from Hi-Res Melting analysis is having robust single PCR product, the best annealing temperature was selected by doing gradient PCR. The gradient PCR was performed with annealing temperature ranging from 55 to 70 ^{o}C. When the optimal annealing temperature was identified, PCR amplifications was performed in total volume of 10 µl reaction mixture which consists of 5 µL Milli-Q water, 2 µL of reaction buffer (5X), 0.1 µL dNTPs, 1 µL LC green plus, 0.25 µL for each of forward and reverse primer and 1 µL DNA samples (about 150ng). Then each PCR mixture was overlaid by 30 µL mineral oil in order to prevent evaporation of samples during scanning. PCR amplification procedure start at 98 ^{o}C for 30 seconds and followed by 40 cycles of 98 ^{o}C for 5 seconds, 63 ^{o}C for 5 seconds and 72 ^{o}C for 20 seconds. After that, heteroduplex formation was performed by heating samples to 94 ^{o}C for 30 seconds then reduced temperature to 25 ^{o}C for annealing for 30 seconds.

Hi-Res Melting analysis was performed by lightscanner in which samples were analysed between 77 ^{o}C and 95 ^{o}C and hold at 74 ^{o}C. Data obtained was analysed by the lightscanner software to see the difference in melting temperature of samples.

III. Results

3.1. Development simple PCR-based SNP markers assay

As stated in materials and methods, before applying the simple PCR based SNP marker for screening, a small trial was conducted to see the effect of allele specific primers on amplification target sequence. The results of the test on two SNP markers on chromosome I present three main aspects. First of all, as our prediction, due to only one nucleotide mismatch on non-specific primer at 3'end of forward primer, we cannot discriminate between two alleles. This result is also consistent with result of Hayashi et al (2004) on rice genome and result of Drenkard et al. (2000) on Arabidopsis genome. Secondly, the annealing temperature plays a pivotal role in reducing the efficiency of non-specific primers in amplifying specific alleles. Thirdly, adjustment on number of PCR cycles is also a helpful tool which helps inhibit non-specific primers from amplifying a significant amount of PCR product from specific alleles. In fact, for some cases, non-specific primer is able to amplify specific allele but less efficiency than specific primer, so that if we reduce the number of PCR cycles, we can inhibit the visualization of unwanted PCR products on agarose gel.

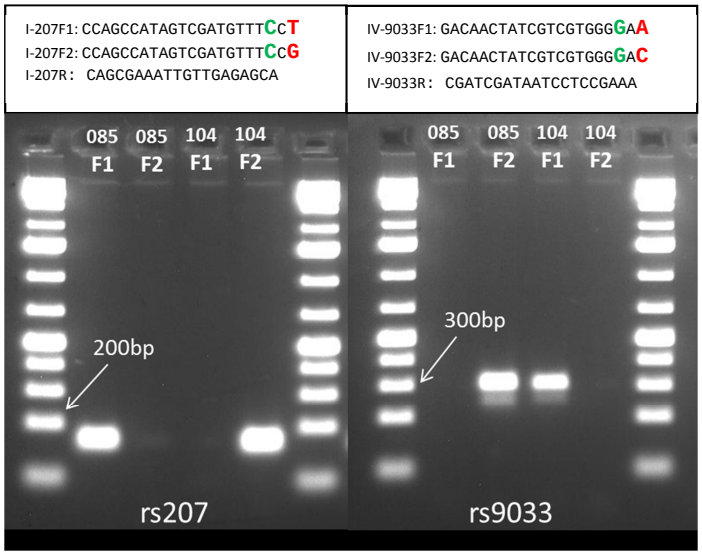


Figure4. Gel picture of PCR based SNP marker assay on two parental lines C085 and R104

Thanks to useful information from the first test, we succeeded in designing primers for assay two SNP markers: seq-rs207 and seq-rs9033 located in chromosome I and chromosome IV respectively. For SNP maker seq-rs207, the information from previous study indicated that parent C085 and R104 are homozygous for T and G alleles correspondingly. To assay this SNP marker, two forward primers were designed with 3' end matching with SNP position and substituted C for A at third base from 3' end (Figure 3). Each of these two forward primers was combined with the same reverse primer in two different PCR reactions to amplify a 148 base pairs target sequence. The size of PCR product which was visualized on 2% agarose gel (Figure 4) confirmed its size of 148bp.

Similarly, with SNP marker seq-rs 9033, parent C085 and R104 were indicated homozygous for C and A alleles respectively. Two specific forward primers aimed to amplify specific alleles at SNP locus was incorporated an artificial mismatch at third base from 3' end. When used with a reverse primer to amplify 287bp target sequence, each specific primer demonstrated their ability to distinguish the specific allele (Figure 4).

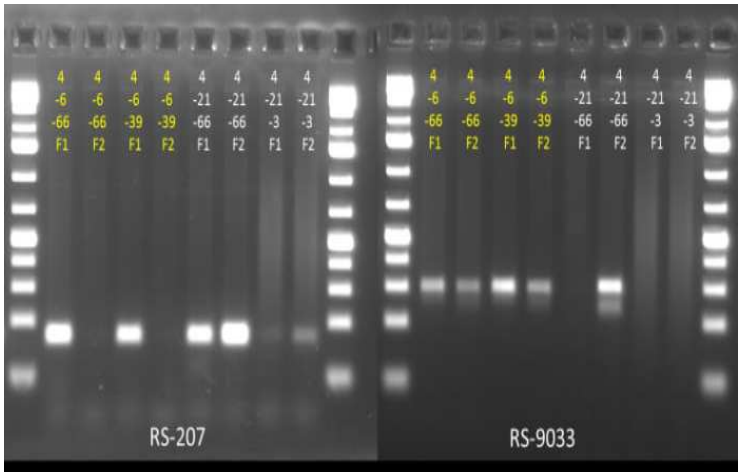


Figure5. PCR based two SNP markers on four F7 lines: 4-6-66, 4-6-39 and 4-21-66, 4-21-3

After that, we have evaluated for the effectiveness of these specific primers on four F7 RILs, namely 4-6-66, 4-6—39 and 4-21-66, 4-21-3 which were selected for study on smokiness trait. The results which revealed in figure 5 strengthen the reliability of these specific primers. With SNP marker seq-rs207, line 4-21-66 showed the heterozygous genotype based on the presence of amplification products from both allele specific primers while line 4-6-66 and 4-6-39 illustrated the homozygous pattern due to the absence of PCR products from non-specific primers (Figure 5). These results are consistent with results of screening with KASP method which will be presented in next section. Likewise, for SNP marker seq-rs9033, our results confirmed heterozygous pattern of two lines 4-6-66 and 4-6-39 and homozygous genotype of line 4-21-66. Nevertheless, we did not get the result of both SNP markers for line 4-21-3. The reason is probably because DNA concentration of this line is much higher compared to other three lines, the efficiency of primers was inhibited in amplifying target fragments. Fortunately, this obstacle was overcome after that when DNA concentration of line 4-21-3 was diluted into a lower level.

3.2. Screening F7 RI population with selected SNP markers by KASP method

Table1. 19 selected SNP markers for screening F7 RIL population by KASP method

DNA/Asssay	Chromosome	Position	4-6 inf	4-21 inf
seq-rs207	1	86094056	AA	AB
seq-rs7794	1	89449632	AB	AA
seq-rs7476	2	40330288	AB	BB
seq-rs7516	2	41799208	AB	BB
seq-rs8522	4	1896090	AB	AA
seq-rs9033	4	58340636	AB	BB
seq-rs2012	5	4313620	AB	BB
seq-rs2122	5	16092483	AB	BB
seq-rs2289	5	31363376	AB	BB
seq-rs4572	5	59633772	AB	AA
seq-rs3536	6	34100828	BB	AB
seq-rs7941	6	34819080	AA	AB
seq-rs4564	7	64363264	AB	AA
seq-rs9061	9	64794424	AB	AB
seq-rs9073	9	65152920	AB	AB
seq-rs9078	9	65295836	AA	AB
seq-rs5612	10	6380151	AA	AB
seq-rs6141	10	18031984	BB	AB
seq-rs4761	10	56224968	BB	AB

After six to eight selfings of RILs, it can be presumed that plants are homozygous for a majority of loci. Only about 1/32 of all loci that segregated in F2 were estimated heterozygous in F6. This estimation is consistent with our screening results on genetic backgrounds of two F6 lines which were selected for study on smokiness trait. Unfortunately, those heterozygous regions which segregated in F7 lines had a considerable effect on the taste trial for smoky flavor. Therefore, in order to investigate the precise effect of Non Smoky Glycosyl Transferase gene 1, we need to have lines with the same genetic background. We selected for 19 good SNP markers which represented those heterozygous regions on 8 chromosomes (table 1) for screening F7 population by KASP technique at Van Haeringen Lab. The result in figure 6 illustrates the segregation pattern of these 19 SNP markers on 57 and 59 offspring of line 4-21 and 4-6 correspondingly. Among them, we are interested in four lines, namely 4-21-66; 4-21-3 and 4-6-66; 4-6-39 which are only heterozygous for one SNP marker. In particular, line 4-21-66 and 4-21-3 was identified homozygous for PhP^- and PhP^+ alleles respectively and they share the same homozygous genetic background for many chromosomes. In fact, among 19 SNP markers, only the SNP marker seq rs-207, which located on chromosome I, presented the heterozygous result. Likewise, two lines 4-6-66 and 4-6-39 also revealed the homozygous results for two alleles of NSTG1 gene and share the same homozygous pattern for 18 SNP markers. Only one SNP marker, which located in chromosome IV, showed the heterozygous result. In summary, these four F7 lines belong to two groups in which each group exists one heterozygous region that require a selection for homozygosity in F8 population. Followed this study objective, we will present the results of DNA isolation and screening for those four lines by a simple PCR based SNP marker method in next sections.

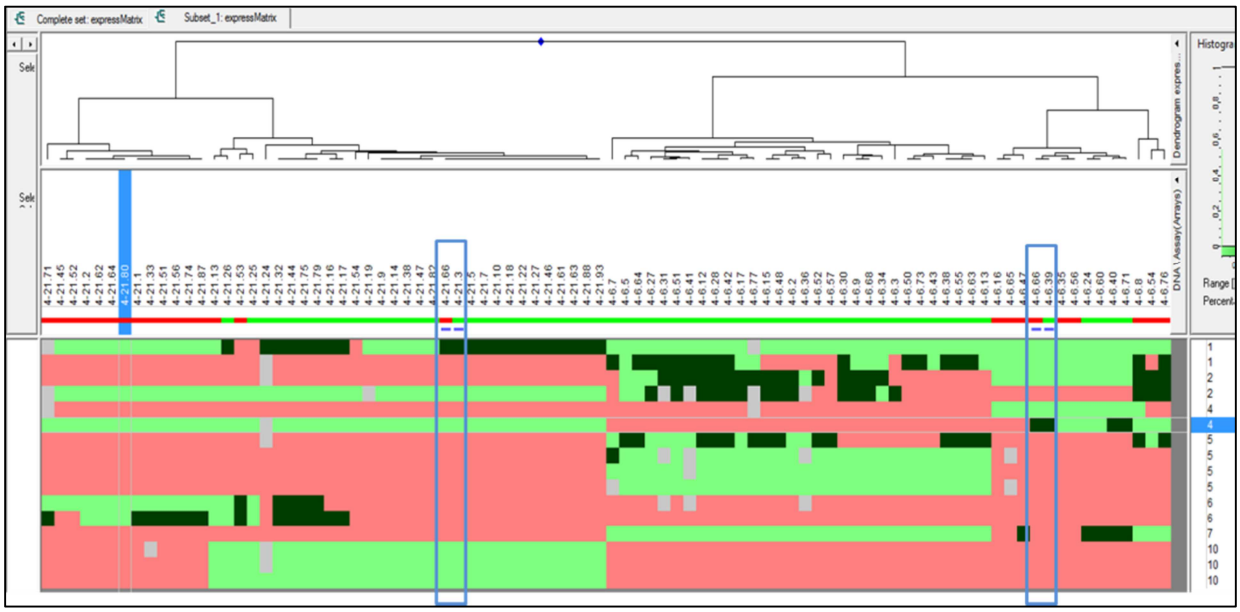


Figure6. Results of screening F7 RILs with 19 SNP markers by KASP method: Black illustrate for heterozygous genotype; Pink: homozygous with R104 allele; Green: homozygous with C085 allele (Source: Arnaud Bovy)

3.3. DNA isolation and screening F8 population by PCR based SNP makers assay

DNA isolation

Figure7. Four F8 RI populations: 4-21-66; 44-21-3 and 4-6-66; 4-6-39 grown at Hollandplant Company for screening and selecting.

Genomic DNA of 285 plants of F8 population which were grown at greenhouse of Hollandplant Company (Figure 7) was isolated by normal Retsch method. This method yielded approximately 2-5 µg/µl (50 µl in total) genomic DNA per sample. The concentration and quality of DNA samples were measured by NanoDrop Spectrophotometer V3.3. The ratio absorbance at 260nm to 280nm ranged from 1.8 to 2.0 which indicated the good quality. After

diluting genomic DNA to a lower concentration about 300ng/ µl for doing PCR, we confirmed the quality and quantity of DNA again by loading 1 µl each sample on 0.8% agarose gel. The results in figure 8 illustrated a very good quality and confirmed the DNA quantity of all samples that we need for screening.

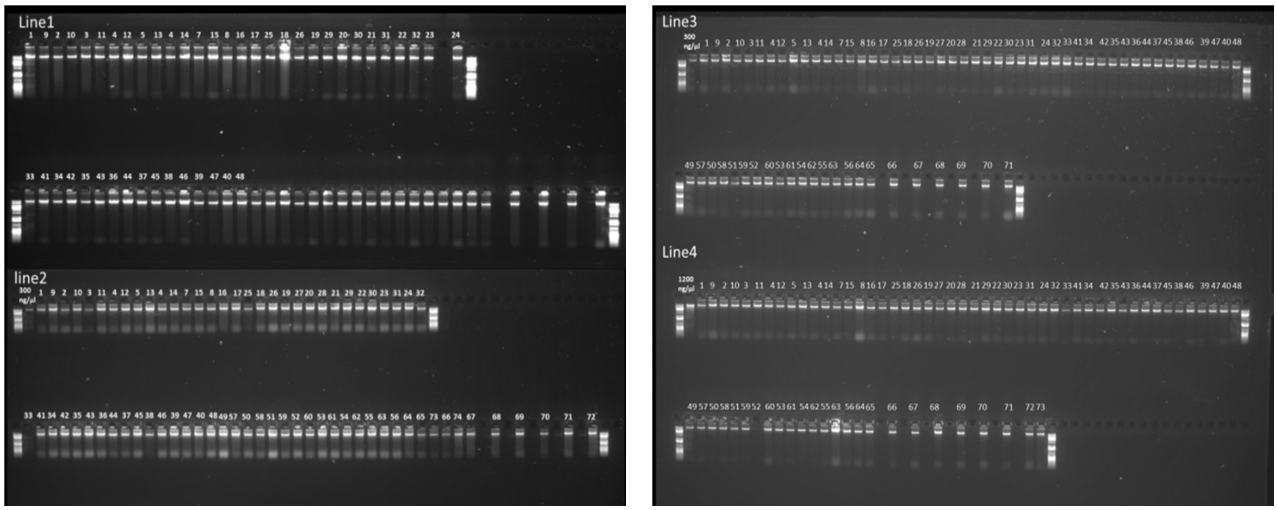


Figure8. Checking for DNA quality and quantity of four F8 RI populations on agarose gel (0,8%)

Screening F8 population by PCR based SNP maker assay

Thanks to the success in development PCR based SNP makers assay and following the results of screening on F7 RILs, we applied this simple PCR technique to make a selection for homozygous plants in F8 population based on two SNP markers. The purpose of this selection is preparing for the taste trial on smokiness trait. As stated in 3.1, these two SNP markers are seq-rs207 and seq-rs9033 which linked to terpenes level and sweetness QTL accordingly. PCR based SNP marker seq-rs207 were applied to screen two F8 populations which are offspring of lines 4-21-66 and 44-21-3. These two lines have segregated for smokiness trait. In fact, line 4-21-66 is homozygous for PhP^{-} allele and line 4-21-3 is homozygous for PhP^{+} allele, so that we need to have the same genetic background for these two lines to clearly see the expression of smokiness gene. With this requirement, our target is selecting for those plants which are homozygous for C085 alleles in these two F8 populations. Regarding to the SNP marker seq-rs9033, we utilized PCR-based assay to select for R104 allele in two F8 lines: 4-6-66 and 4-6-39. The reason for this selection is because we would like to see if sweetness influence smoky evaluation perception. In fact, line 4-21 has high sweetness background and we need to select for low sweetness background from line 4-6 (allele R104).

Although PCR-based SNP marker assays worked very well on two parents C085 and R104 and four F7 selected lines, we confronted a trouble with SNP maker seq-rs207. The primers which were designed for distinguishing two alleles of SNP marker seq-rs207 did not succeed in amplifying target sequence for a majority number of DNA samples from two populations: 4-21-66 and 4-21-3. In order to overcome this problem, we looked for the possible causes that made PCR failed and then we found a similar situation that happened to F7 line 4-21-3 which was mentioned in 3.1. It is possible that the whole genome sequence of tomato is approximately 900Mb while the target sequence for amplifying is only 148bp, so that within a high concentration of genomic DNA, the primers cannot bind to the target sequence then lead to fail in amplifying the PCR product. With this hypothesis, DNA concentration was diluted to around 150ng/ µl and doing PCR again. Fortunately, our problem was overcome and we got clear results which were presented in table 2. In theory, the segregation ratio for a heterozygous locus should be 1:2:1. However, for the F8 population 4-21-66, we got only 6 plants which were homozygous for C085 allele, the remaining 67 plants consisted 33 heterozygotes and 28 homozygotes for allele R104. Based on this result, 6 homozygous plants with C085 allele were selected for taste trial. A better situation for F8 population 4-21-3, we received 11 desired plants which were homozygous with C085 allele.

Regarding to SNP marker seq-rs9033, PCR-based assays for two populations: 4-6-66 and 4-6-39 were succeeded without any problems and we received 12 and 23 desired homozygous plants with R104 allele respectively. Before selecting those plants together with 17 plants from PCR-based SNP seq-rs 207 assays, we confirmed the result by repeating PCR. The reason is that we are unable to have the negative control with this simple PCR-based method, so repeating PCR is a good way to avoid any possible mistakes may lead to the wrong selection. The results of this confirmation were presented in figure 9. With the high level of reliability thanks to two times doing PCR, those selected plants will be grown until harvest to carry out the taste trial.

Table2. Results of screening four F8 RI populations: 4-21-66; 44-21-3 and 4-6-66; 4-6-39 by simple PCR based SNP marker assays

Sample names		SNP seq- rs 207	Sample names		SNP seq- rs 207
Parental lines	C085	a	Parental lines	C085	a
	R104	b		R104	b
Line 1: 4-21-66	1	h	Line 2: 4-21-3	1	h
Non Smoky	2	b	Smoky	2	b
	3	h		3	b
	4	b		4	b
	5	b		5	b
	6	h		6	h
	7	h		7	h
	8	b		8	b
	9	h		9	h
	10	h		10	h
	11	b		11	h
	12	h		12	h
	13	h		13	b
	14	b		14	a
	15	a		15	b
	16	h		16	b
	17	b		17	h
	18	h		18	b
	19	h		19	b
	20	b		20	b
	21	h		21	h
	22	h		22	no data
	23	h		23	a
	24	a		24	h
	25	b		25	h
	26	h		26	a
	29	h		27	h
	30	h		28	a
	31	b		29	h
	32	h		30	h
	33	b		31	h
	34	b		32	h
	35	b		33	no data
	36	b		34	a
	37	b		35	b
	38	b		36	h
	39	h		37	b
	40	h		38	h
	41	b		39	h
	42	b		40	b
	43	h		41	b
	44	h		42	h
	45	h		43	no data
	46	b		44	a
	47	h		45	a
	48	b		46	h
	49	h		47	h
	50	b		48	h
	51	h		49	h
	52	b		50	b
	53	h		51	b
	54	h		52	b
	55	b		53	h
	56	h		54	b
	57	a		55	h
	58	a		56	h
	59	h		57	h
	60	h		58	a
	61	b		59	b
	62	b		60	h
	63	h		61	h
	64	a		62	h
	65	a		63	h
	66	h		64	b
	67	b		65	a
	68	b		66	h
	69	b		67	h
				68	a
				69	h
				70	h
				71	h
				72	h
				73	b
				74	a

Sample names		SNP seq- rs 9033	Sample names		SNP seq- rs 9033
Parental lines	C085	c	Parental lines	C085	c
	R104	d		R104	d
Line3: 4-6-66	1	no data	Line4: 4-6-39	1	d
Non Smoky	2	c	Smoky	2	h
	3	h		3	c
	4	h		4	d
	5	c		5	d
	6	c		6	c
	7	c		7	h
	8	c		8	no data
	9	d		9	d
	10	h		10	h
	11	h		11	d
	12	c		12	h
	13	h		13	c
	14	c		14	c
	15	c		15	h
	16	no data		16	c
	17	no data		17	d
	18	h		18	h
	19	h		19	h
	20	c		20	h
	21	h		21	h
	22	h		22	h
	23	c		23	h
	24	no data		24	h
	25	no data		25	c
	26	h		26	h
	27	h		27	d
	28	h		28	h
	29	h		29	h
	30	c		30	h
	31	h		31	h
	32	c		32	d
	33	no data		33	h
	34	d		34	d
	35	d		35	d
	36	h		36	h
	37	h		37	d
	38	c		38	c
	39	h		39	d
	40	d		40	d
	41	d		41	c
	42	h		42	h
	43	d		43	d
	44	c		44	h
	45	h		45	h
	46	h		46	h
	47	c		47	c
	48	c		48	h
	49	c		49	d
	50	d		50	h
	51	d		51	d
	52	h		52	no data
	53	c		53	h
	54	d		54	d
	55	c		55	h
	56	h		56	h
	57	no data		57	h
	58	c		58	c
	59	d		59	d
	60	h		60	h
	61	h		61	h
	62	h		62	d
	63	h		63	c
	64	d		64	d
	65	no data		65	h
	66	h		66	d
	67	c		67	c
	68	h		68	h
	69	d		69	h
	70	h		70	d
	71	no data		71	h
				72	no data
				73	d

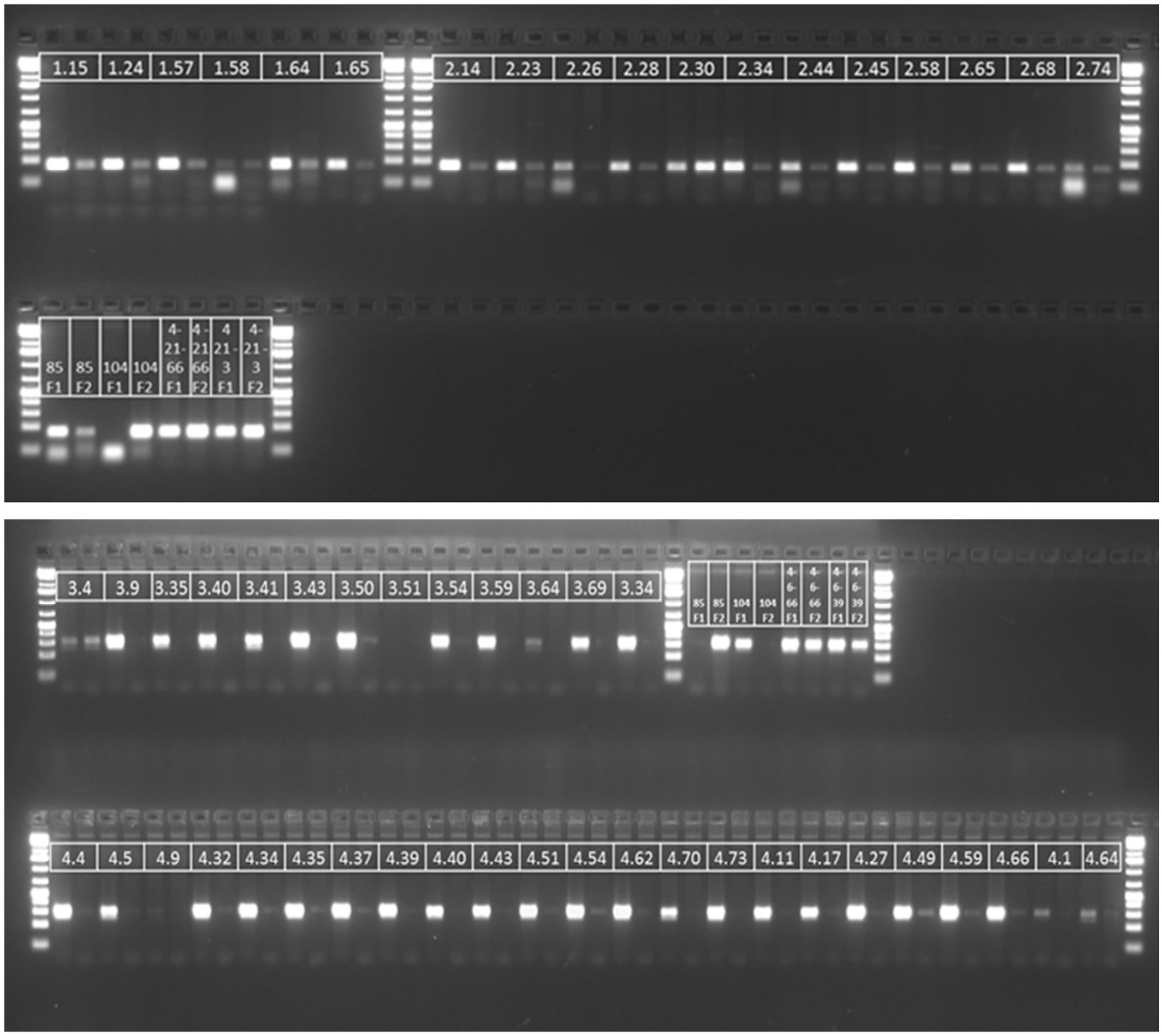


Figure9. Confirming the results of selected homozygous plants from four F8 RI populations: 4-21-66 (1); 44-21-3(2) and 4-6-66 (3); 4-6-39(4) by simple PCR based SNP marker assays

3.4. Screening F8 RIL population - offspring of line 4-21-3 with lightscanner

Although we got the desired results based on the simple PCR-based SNP markers assays, we wonder whether lightscanner can discriminate desired homozygous plants for taste trial. This is the reason for us to implement this small trial on F8 population 4-21-3 and based on the result of this experiment, we can compare to the PCR-based method to identify the most efficient technique for this kind of screening.

TTT**TCGAAACCCTAACGAACCAG**CCATAGTCGATGTTTAC**{T/G}**CTGCTTGCGCTTACGGTGA

>>>>>>>>>>>>>>>>>>>>

Forward primer (5' ->3')

TTTTGACAAACTCCGGAAATTCGTTGAAGTTGACGGCGTTTCCGTTTACCAGCCGGACGG

CAACGGATATTACGCGCTTCAATGGGC**TGCTCTCAACAATTTCGCTG**ATATTGTTCAATA

<<<<<<<<<<<<<<<<<<<

(3'<-5') **ACGAGAGTTGTTAAAGCGAC**

Reverse primer

Figure10. Designing primers for amplifying the target sequence containing SNP locus for Hi-Res Melting analysis

In order to apply Hi- Res Melting analysis, a primer pair was designed to amplify 164bp target sequence including SNP locus (Figure 10), then the gradient PCR was performed to identify the most suitable annealing temperature for this primer pair. The result which presented in figure 11 illustrates high specificity of this primer pair. In fact, only one robust PCR product was amplified although the annealing temperature ranged from 55 to 70^{o}C. This result was stable with both parents C085 and R104, F7 line 4-21-3 and two F8 lines 4-21-3-13; 4-21-3-24 (Figure 11).

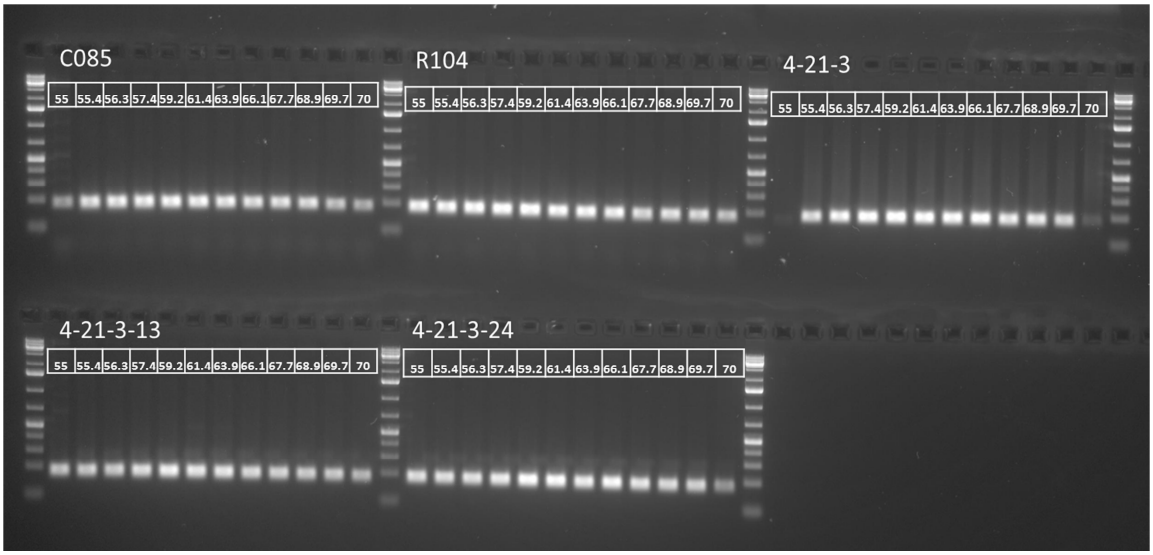
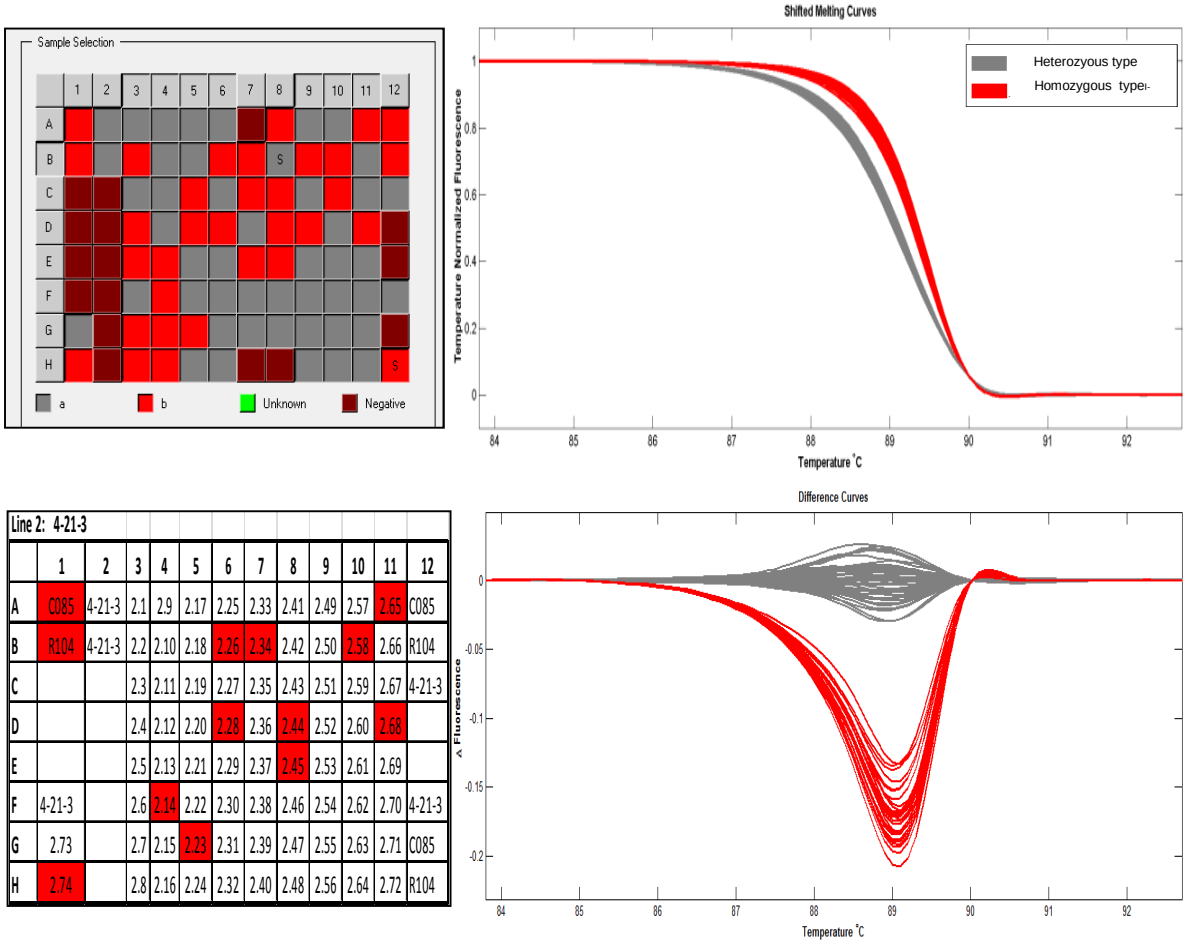


Figure11. Gradient PCR with annealing temperature ranging from 55 to 70 ^{o}C on two parental lines C085 and R104; F7 line 4-21-3 and two F8 lines 4-21-3-13; 4-21-3-24

The results of Hi-Res Melting analysis which presented in figure 12 can discriminate between homozygous and heterozygous genotypes based on the difference in melting curves. The normalized melting curves of 74 plants formed two groups. The group homozygous plants have an entire curve with higher melting temperature compared to heterozygous group which have a different shape at a lower melting region. The principle of this Hi-Res Melting analysis is based on heteroduplex- detecting DNA dye. One heterozygous sample produces two heteroduplexes and two homoduplexes which give a skewed composite melting curve, so that it is easily to be discriminated from homozygous melting curve (Graham et al., 2005; Liew et al., 2004). Moreover, compared to the results that received from PCR based SNP marker assay, all 11 selected plants are members of homozygous group which presented in red color (Figure 12) but they could not be distinguished from other homozygous type.



Line 2: 4-21-3

	1	2	3	4	5	6	7	8	9	10	11	12
A	C085	4-21-3	2.1	2.9	2.17	2.25	2.33	2.41	2.49	2.57	2.65	C085
B	R104	4-21-3	2.2	2.10	2.18	2.26	2.34	2.42	2.50	2.58	2.66	R104
C			2.3	2.11	2.19	2.27	2.35	2.43	2.51	2.59	2.67	4-21-3
D			2.4	2.12	2.20	2.28	2.36	2.44	2.52	2.60	2.68	
E			2.5	2.13	2.21	2.29	2.37	2.45	2.53	2.61	2.69	
F	4-21-3		2.6	2.14	2.22	2.30	2.38	2.46	2.54	2.62	2.70	4-21-3
G	2.73		2.7	2.15	2.23	2.31	2.39	2.47	2.55	2.63	2.71	C085
H	2.74		2.8	2.16	2.24	2.32	2.40	2.48	2.56	2.64	2.72	R104

Figure12. Hi-Res Melting analysis in F8 RIL population 4-21-3: tables in left illustrate the position and sample's names (red and grey color present homozygous and heterozygous genotypes respectively); figures on the right shows difference in melting curves of homozygotes and heterozygotes.

IV. Discussion

Single Nucleotide Polymorphisms are one of the most common makers which are widely used for mapping and assisting selection, especially when the whole genome sequence is available. There are many ways to genotype SNPs. However, no one method meets the needs of all studies using SNPs (Kwok and Chen, 2003). Although nowadays many companies have provided high throughput services for SNP genotyping, it is not always the best choice because in many situations, only a limited number of SNP markers are used to screen for a small population.

The PCR based SNP markers assay that we developed for diagnosis the specific region in tomato genome has showed their simplicity and reliability for applying. The facilities required for this method are very basically that any molecular laboratories can perform. In fact, we can see directly the result and make a rapid selection through doing standard PCR. The method required three primers and two PCR reactions for each SNP marker. We have also assessed the feasibility of this simple method through screening 285 plants with two SNP markers. The effectiveness of this method was proven based on desired homozygous plants which were selected for taste trial.

However, every method has some drawbacks. Although this PCR-based assay illustrates many advantages such as its applicability for any molecular biology laboratory, it has a few disadvantages. For example, we are unable to have false negative control for this method, so that it may cause mistakes in scoring the results if a heterozygous genotype was success in one PCR reaction and fail in another reaction.

Besides, we also made an evaluation for the possibility of applying High Resolution Melting analysis on screening with SNP marker. With heteroduplex detecting DNA dye, it is quite common for distinguishing between homozygous and heterozygous genotypes (Graham et al., 2005; Wittwer et al., 2003 and Wu et al., 2008). The result of our trial also illustrates a similar pattern. In fact, there is no difference in melting curves between two homozygous genotypes. However, by an additional step, it is possible to see the difference of these two homozygous genotypes. Based on the fact that Hi-Res Melting analysis can discriminate between homozygous and heterozygous genotypes, DNA of homozygous samples can be mixed with DNA from one of homozygous parent, so that two complement homozygous genotypes can be detected in form of heterozygous genotype. Compared to PCR based SNP markers assay, Hi-Res Melting analysis require an additional step to distinguish two

homozygous genotypes. However, with Hi- Res Melting analysis, we need to do only one PCR reaction per sample per SNP marker and we can skip the step for visualize PCR product on agarose gel. Therefore, applying Hi-Res Melting analysis is also a good choice for genotyping SNP markers. Both methods have their advantages and disadvantage, so that depending on particular purpose and availability of laboratory facilities we can determine the most suitable method for screening.

V. Conclusions and Recommendations

Conclusions

It can be concluded that both methods can produce reliable results for SNP genotyping. Each method has their own strengths and weaknesses; therefore, depending on the availability of facilities and the amount of samples, we can decide to choose the most suitable method. For HRM analysis, it can be easily adapted to a large scale of genotyping compared to simple PCR based SNP maker assay (Han et al., 2012). However, with a small scale of genotyping and requirement of simple facilities, the PCR-based method will be the best choice.

In our case, the simple PCR-based SNP markers assay have been applied successfully for selecting 52 desired homozygous plants. These plants are growing until harvest in order to perform taste trial. Thanks to this selection, we have the homozygous RIL population which only segregated for NSTG1 gene, so that the effect of two different alleles of this smokiness gene can be evaluated in a precise way.

Recommendations

Our main target in this report is selecting for desired plants which are homozygous for genetic background to perform the taste trial. In order to reach this goal, PCR based SNP markers assay have been developed. This method is simple to implement, but some attentions may help avoid unwanted difficulties. First of all, PCR product size should be not too short because it may cause trouble in distinguishing between PCR product and primer dimer. Secondly, a high DNA concentration of samples may affect to primer activity in amplifying target sequence. We confronted a difficulty with DNA concentration around 300ng/ µl (measured by NanoDrop). The primers for SNP marker seq-rs 207 was fail in amplifying PCR product, but then the trouble was solved when DNA concentration to a lower level about 150ng/µl. Hence, a suitable DNA concentration is highly recommended for a success PCR.

Moreover, the Hi-Res Melting analysis is also applicable for genotyping SNP markers. Although this technique can only discriminate between homozygotes and heterozygote, an additional step in which the homozygous DNA samples are mixed with DNA from one of homozygous parents is recommended to see the difference in melting curves. If the sample has the same homozygous type of the parent, they will not lead any changing in melting curves, but if they have different homozygous genotypes, the melting curve can be scored as the heterozygous genotype.

References

Bai, Y., and P. Lindhout, 2007: Domestication and Breeding of Tomatoes: What have We Gained and What Can We Gain in the Future? Oxford Journals **100** 1085-1094.

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

Drenkard, E., B. Richter, S. Rozen, L. Stutius, N. Angell, M. Mindrinos, R. Cho, P. Oefner, R. Davis, and F. Ausubel, 2000 A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. Plant Physiol **124**, 1483-92.

Giovannucci, E., 1999: Nutritional factors in human cancers. Advances in Nutrition and Cancer 2 **472**, 29-42.

Graham, R., M. Liew, C. Meadows, E. Lyon, and C.T. Wittwer, 2005: Distinguishing different DNA heterozygotes by high-resolution melting. Clinical Chemistry **51**, 1295-1298.

Han, Y., D. Khu, and M. Monteros, 2012: High-resolution melting analysis for SNP genotyping and mapping in tetraploid alfalfa (Medicago sativa L.). Mol Breed **29**, 489-501.

Hayashi, K., N. Hashimoto, M. Daigen, and I. Ashikawa, 2004: Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. Theoretical and Applied Genetics **108**, 1212-1220.

Klee, H.J., 2010: Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. New Phytologist **187**, 44-56.

Konieczny, A., and F. Ausubel, 1993: A procedure for mapping *Arabidopsis* mutations using co-dominat ecotype-specific PCR-based markers. The Plant Journal **4**, 403-410.

Kwok, P.Y., and X. Chen, 2003: Detection of Single Nucleotide Polymorphisms. Curr Issues Mol Biol **5**, 43-60.

Kwok, S., D.E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson, and J.J. Sninsky, 1990: Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic Acids Res **18**, 999-1005.

Liew, M., R. Pryor, P. R, M. C, E. M, L. E, and W. C., 2004: Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clin Chem **50**, 1156-64.

Miller, J.C., and S.D. Tanksley, 1990: RFLP analysis of phylogentic relationship and genetic variation in the genus Lycopersicon. Theor Appl Genet **80**, 437-448

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

Peralta, I., and D. Spooner, 2002: History, Origin and Early Cultivation of Tomato (Solanaceae) Genetic Improvement of Solanaceous Crops **2**, 1-27.

Reed, G., and C. Wittwer, 2004: Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. Clin Chem **50**, 1748-54.

Robinson, E., and J. Holme, 2011: KASP version 4.0 SNP Genotyping Manual. KBio science.

Rozen, S., and H. Skaletsky, 2000: Primer3 on the WWW for General Users and for Biologist Programmers Methods in Molecular Biology **132**, 365-386.

Sims, W.L., 1980: History of tomato production of industry around the world. International Society for Horticultural Science **100**, 25-26.

Sini Suomalainen, M.S., Virpi Puro, Vuokko Kytöniemi and Arja Lamberg, ?: Optimization of the DNA Purification Protocol for Thermo Scientific KingFisher Flex and the BindIt 3.1 Software. Thermo Fisher Scientific.

Tikunov, Y., d.R.H. Vos, A. González Paramás, R. Hall, and A. Bovy, 2010: A Role for Differential Glycoconjugation in the Emission of Phenylpropanoid Volatiles from Tomato Fruit Discovered Using a Metabolic Data Fusion Approach. Plant Physiology **152**, 55-70

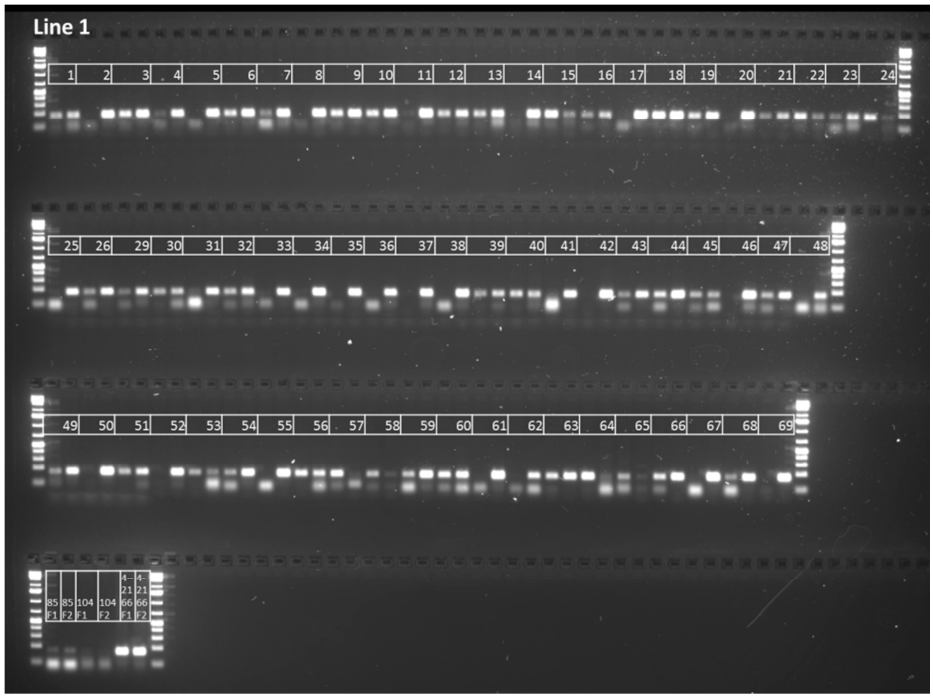
Wittwer, C.T., G.H. Reed, C.N. Gundry, J.G. Vandersteen, and R.J. Pryor, 2003: High-resolution genotyping by amplicon melting analysis using LCGreen. Clinical Chemistry **49**, 853-860.

Wu, S., M. Wirthensohn, P. Hunt, J. Gibson, and M. Sedgley, 2008 High resolution melting analysis of almond SNPs derived from ESTs Theor Appl Genet **118**, 1-14.

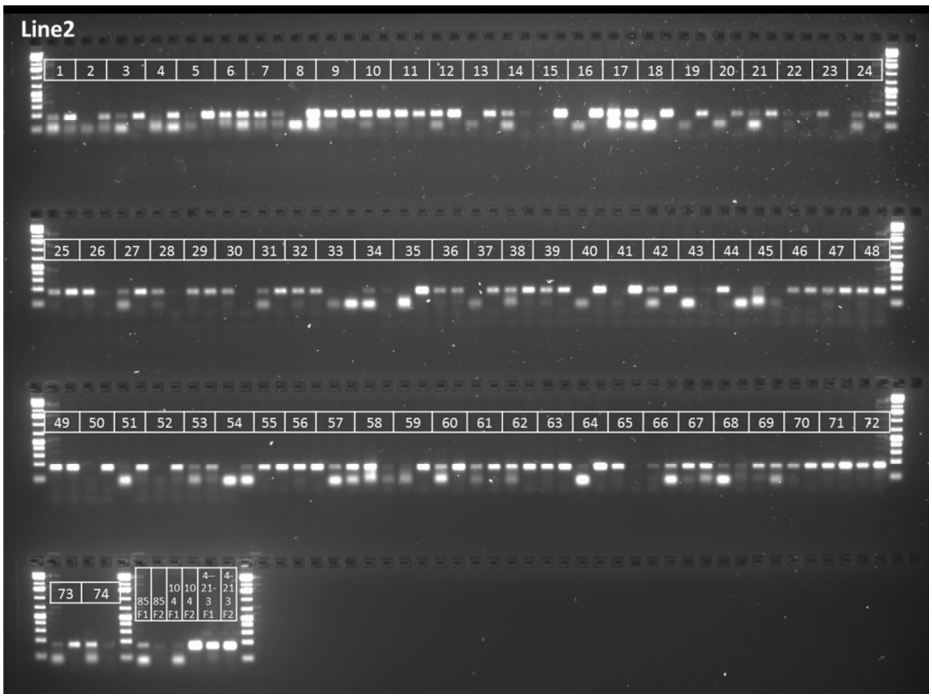
Zamir, D., and G. Giuliano, 2012: The tomato genome sequence provides insights into fleshy fruit evolution. nature **485**, 635–641.

Appendix

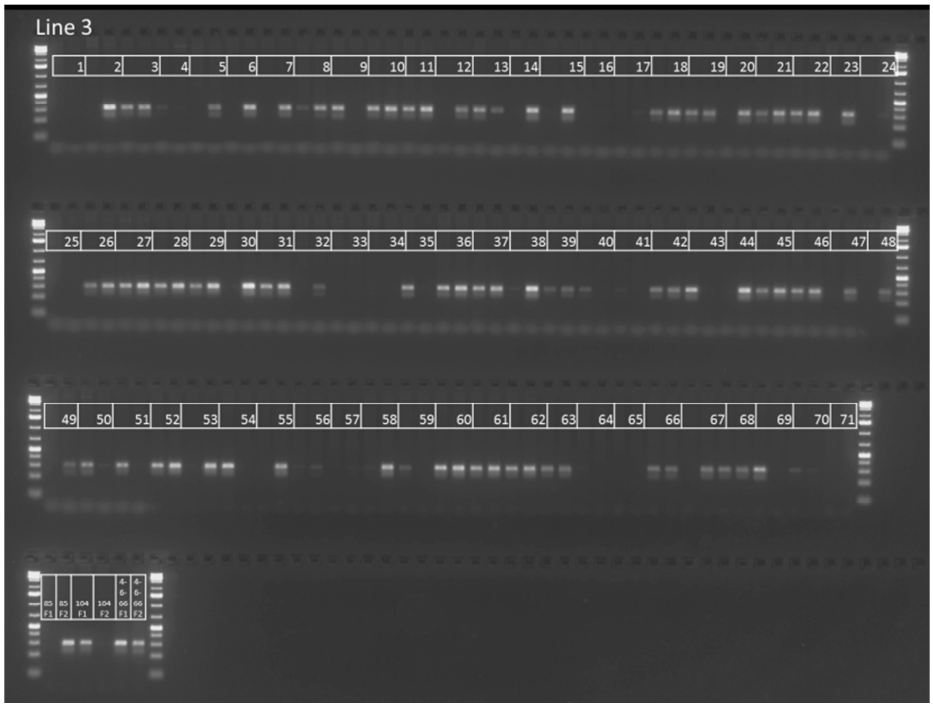
Appendix I. Screening F8 population 4-21-66 with SNP marker seq-rs207 by simple PCR based assay



Appendix II. Screening F8 population 4-21-3with SNP marker seq-rs207 by simple PCR based assay



Appendix III. Screening F8 population 4-6-66 with SNP marker seq-rs9033 by simple PCR based assay



Appendix IV. Screening F8 population 4-6-39 with SNP marker seq-rs9033 by simple PCR based assay