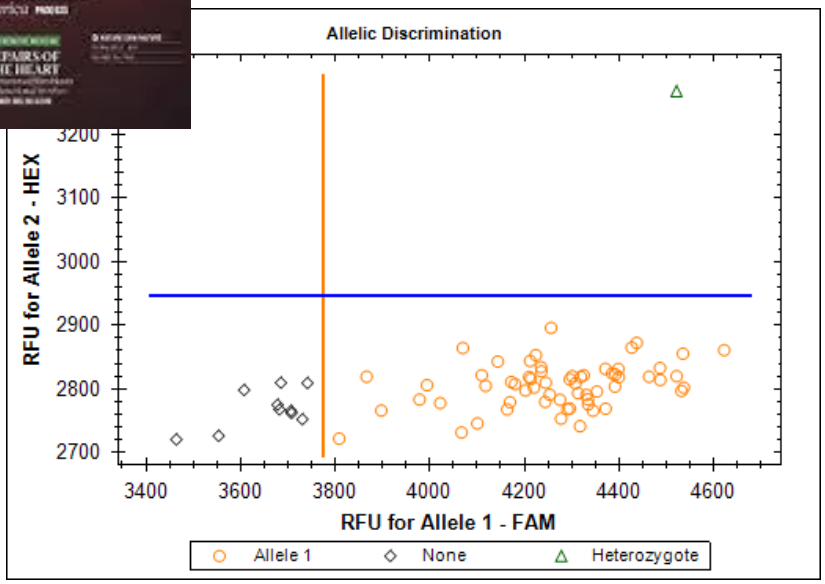
Mutation Detection with the use of KASP in an EMS derived *Solanum lycopersicum* TILLING Population

Increasing variation in tomato fruit quality related genes



Frans Ketelaars

Department of Plant Breeding,

Wageningen University

Mutation Detection with the use of KASP in an EMS derived *Solanum lycopersicum* TILLING Population

Increasing variation in tomato fruit quality related genes

Frans Ketelaars

920320430120

MSc Thesis in Plant Breeding

Supervisors:

Yury Tikunov PhD

Dr. Arnaud Bovy

March 2015

Department of Plant Breeding, Wageningen University, Wageningen

Wageningen, The Netherlands

Abstract

As a result of selecting for high yield, the tomato fruit lost its flavor. Last two decades, big efforts have been put in increasing the tomato fruit quality. In this MSc thesis report we assessed the association between existing and tomato fruit quality related traits. Also, *de novo* genetic variation in an EMS derived TILLING population has been assessed.

After two sequence experiments, we could confirm 8 mutations in genes that are involved in the tomato fruit quality. Protein prediction programs have predicted that these point mutations will alter the translated protein. The mutations were found in pools, which contain DNA of 64 individual plants. The SNP genotyping technique Kompetitive Allele Specific PCR (KASP™) was domesticated and used to detect the one mutated plant out of the pool of 64 plants.

Due to chimerism, the mutant SNP was present at low quantities in the mutant plant. KASP showed to not work when the mutant SNP is present at low quantities. Therefore we were not able to find the one mutant plant in the pool of 64 plants. It is discussed that other SNP genotyping techniques like castPCR™ or ddPCR™ could give a more satisfying outcome.

In silico experiments have revealed statistically significant clues that some existing variation in the promoter region of glycosyl transferase genes of *Solanum lycopersicum* accessions show an association with several tomato fruit quality related compounds. However, *in vivo* confirmations are still needed.

Key words: Solanum lycopersicum, tomato fruit quality, EMS, TILLING, metabolomics

Contents

1 Introduction

1.1 Background of the project

1.1.1 General introduction

Tomato (*Solanum lycopersicum*) is an important food crop. It is an important ingredient for daily nutrition all over the world. To fulfill the worldwide demands, the worldwide production increased enormously in the last 40 years and is still increasing. It increased from 0.27 Megatons in 1961 to 162 Megatons in 2012 (FAOSTAT 2012). Due to new production techniques and the understanding of the genetical content of the crop, people were able to increase the yield in The Netherlands from 80715.4 kg/Ha in 1961 to 476049.7 in kg/Ha (Figure 1). However, not everywhere around the world did the yield increase till such high quantities. For example the Yield in the USA did increase, but more steadily. The differences in yield are remarkable as is shown in Figure 1. The Figure also shows that the yield has reached a maximum, the Yield in The Netherlands stabilized.

Pushing the yield of your crop seemed important, since the profits depends on weight of fruits sold. However a downside of breeding for high fruit yield resulted in diminishing of fruit quality, in particular its flavor and nutritional quality

Fruit flavor is a complex trait, which is determined by multiple factors such as sweetness, sourness, firmness and aroma. Quantitative expression of these factors is a product of complex biosynthetic mechanisms, which are regulated by internal genetic or external environmental factors. Besides perception of the same fruit, can be different in different parts of the world.

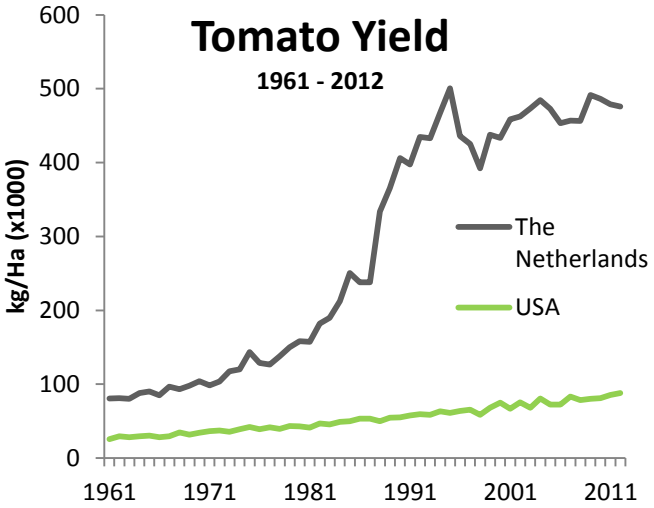


Figure 1. Tomato yield (kg/Ha) in The Netherlands and the USA since 1961. *Source: FAOSTAT (www.faostat.fao.org)*

1.1.2 Tomato fruit flavour (Aroma & Taste).

The olfaction is together with the gustatory perception the most important sense for flavour recognition (Noble 1996). This perception is determined by a combination of volatile organic compounds (VOC) and non-volatile primary (sugars, organic acids) and secondary (e.g. alkaloids) metabolites. There are over 400 aroma volatiles known in the tomato fruit, however only about 30 volatiles are present in amounts that can be perceived by the human olfactory system (Baldwin et al. 2000). Many tomato aroma compounds are stored as glycosides(Marlatt et al. 1992). Taste of tomato fruit is mostly demined by a few primary metabolites: fructose, glucose and sucrose provide sweetness, whereas citric and malic acid provide sourness. Also some amino acids, such as glutamic acid and aspartic acid are important components of tomato flavour.

Variation in the flavour related metabolites among different tomato varieties has been observed (Tikunov et al. 2005; Ortiz-Serrano and Gil 2007; Birtic et al. 2009). Thus, fortunately desirable alleles for tomato flavour seem to exist because there is genetic variation available(Tieman et al. 2012). These alleles correspond to many genes, which are either directly involved in the biosynthesis of flavour related metabolites. Sequestration and release of metabolites is also an important factor affecting flavour. Glycosyltransferases (GT) affect storage and release of secondary metabolites (Tikunov et al. 2013), as well as another enzyme family – Glycosylhydrolases (GH). On top of it, regulatory genes, such as transcriton factors can affect expression of genes involved in biosynthetic, storage and release of flavour molecules.

1.1.3 Health promoting compounds in the tomato fruit

Last decade, society started to demand products with a higher nutritional value. This is a direct result of the increasing knowledge about the positive effects of a healthy diet (Rozanski et al. 1999). Therefore not only taste related traits became of importance, but also health-promoting traits. One important health-promoting group of compounds are the flavonoids. Flavonoids, secondary plant metabolites, are potential health-protecting dietary components because of their high anti-oxidative capacity (Ness and Powles 1997; Duthie and Crozier 2000). In vitro, flavonoids have shown to be important in the synthesis of human protective enzyme systems(Choi et al. 1999; Janssen et al. 1998). Based on these two findings, has been suggested that flavonoids are important compounds to protect against coronary heart diseases and cancer (Steinmetz and Potter 1996; Hertog and Hollman 1996).

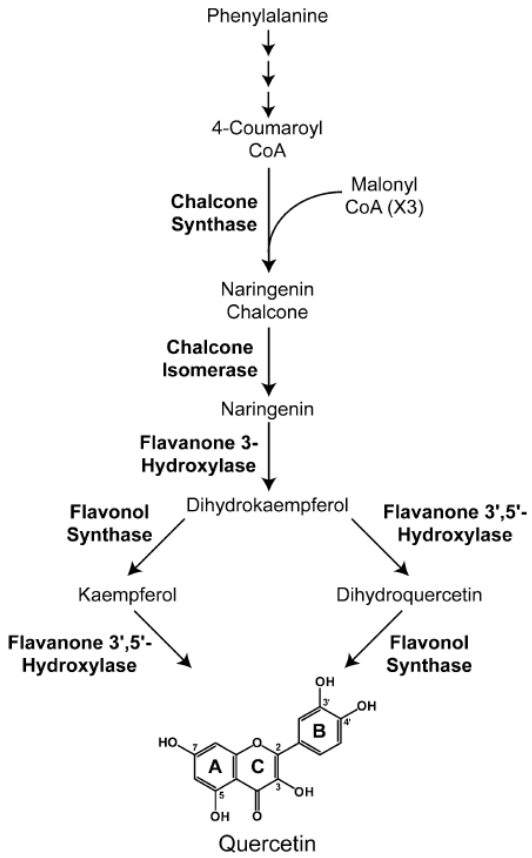


Figure 2. The biosynthetic pathway of flavonoids in Tomato. *Source: (Willits et al. 2005)*

Rutin (quercetin-3-rutinoside) and naringenin chalcone are the main flavonoids in the ripe tomato fruit. However they are present at low quantities and confined only to the peel (Stewart et al. 2000; Muir et al. 2001; Adato et al. 2009). The synthesis of rutin in the peel is limited by low expression of the chalcone isomerase (Muir et al. 2001; Willits et al. 2005; Adato et al. 2009), the pathway is shown in Figure 2. Chalcone isomerase is the enzyme that catalyzes the conversion of naringenin chalcone to naringenin quercetin, therefore there is accumulation of naringenin chalcone in the peel (Hunt and Baker 1980). Recently, a QTL was found in *Solanum habrochaites* that increases the expression of the chalcone isomerase gene (Hanson et al. 2014). It is highly probable that introduction of the QTL restores the flavonoid pathway resulting in elevated rutin content in the peel. Most genes of the flavonoid pathway are not expressed in the tomato fruit flesh, which results in the typical low flavonoid content in the cultivated tomato fruits (Willits et al. 2005).

A lot of studies have been performed to improve the flavonoid content in tomato. Different approaches have been used with success: 1) modification of the flavonoid pathway through regulatory or structural genes (Muir et al. 2001; Bovy et al. 2002; Verhoeyen et al. 2002; Le Gall et al. 2003), 2) RNAi to block degrading steps in the pathway (Schijlen et al. 2007) and 3) Interspecific crosses with *S. chilense* and *S. pennellii* (Willits et al. 2005).

1.1.4 Glycosyltransferase genes

Glycosylation is the conjugation of the aglycone (the glycosyl acceptor) to a carbohydrate (the glycosyl donor). This conjugation is catalyzed by a ubiquitous family of enzymes called glycosyltransferases. Glycosylation is involved in the transport and storage of secondary plant metabolites, such as flavonoids (Gachon et al. 2005), alkaloids and volatile compounds. Glycosylation patterns can also play an important role for tomato fruit quality. Release of certain flavor molecules upon consumption of tomato fruit by a human can depend on the structure of their glyconjugates (Tikunov et al. 2013). Human intestinal absorption of quercetin glycosides is better (52%) compared to the pure aglycone (24%) (Hollman and Katan 1999).

More than 400 glycosyltransferase genes organized into 97 different families based on their structure and function have been identified (http://www.cazy.org/GlycosylTransferases.html). They are found all over the tree of life: in archaea, bacteria, animals, plants and viruses. In the tomato genome, also a vast number of different glycosyltransferase genes have been identified.

1.1.5 Targeted Induced Local Lesions IN Genomes

One way to improve the consumer quality of tomato fruit would be to find naturally occurring alleles of genes, which positively affect the fruit quality and to introduce them into the crop via breeding. It seems there is some natural allelic variation in genes involved tomato fruit quality (Tieman et al. 2012). Another way to improve the consumer quality of tomato fruit would be to use artificially created variation in the form of *de novo* mutations. Targeting Induced Local Lesions IN Genomes (TILLING) combines traditional chemical mutagenesis with high-throughput genome-wide screening for point mutations in desired genes. TILLING is therefore a powerful way of creating novel mutant alleles for crop improvement (Piron et al. 2010; Chen et al. 2014). Another very important advantage of TILLING for the use in crop improvement is the non-transgenic method (Slade et al. 2005). Therefore, the *de novo* mutants are free of regulatory restrictions imposed on genetically modified organisms and the novel variation can be inherited stably (Dong et al. 2009; Till, Reynolds, et al. 2003). Induced mutations unleash the potentials of plants for food production and other agricultural uses (Mba 2013). Due to this high potential, 2250 mutated crop varieties have been released in the past 70 years either as direct mutants or as offspring from mutated plants (Ahloowalia et al. 2004). A third advantage of TILLING is the applicability to any species, no matter its genome size and ploidy level (Parry et al. 2009; Uauy et al. 2009).

Literature describes examples in TILLING in all kinds of organism. In animals it is used in zebrafish (Amsterdam and Hopkins 2006), Drosophila (Winkler et al. 2005) and *C. elegans* (Gilchrist et al. 2006). In fungi it is used in Phytophthora (Lamour et al. 2006). In plants it is used in almost all important (crop) species. Examples are *Arabidopsis thaliana* (Greene et al. 2003; Jander et al. 2003), *Pisum sativum* (Dalmais et al. 2008), wheat (Slade et al. 2012), rice (Till et al. 2007), maize (Till et al. 2004), *Sorghum* (Xin et al. 2008) and Tomato (Gady et al. 2009; Menda et al. 2004; Minoia et al. 2010; Piron et al. 2010; Saito et al. 2011)

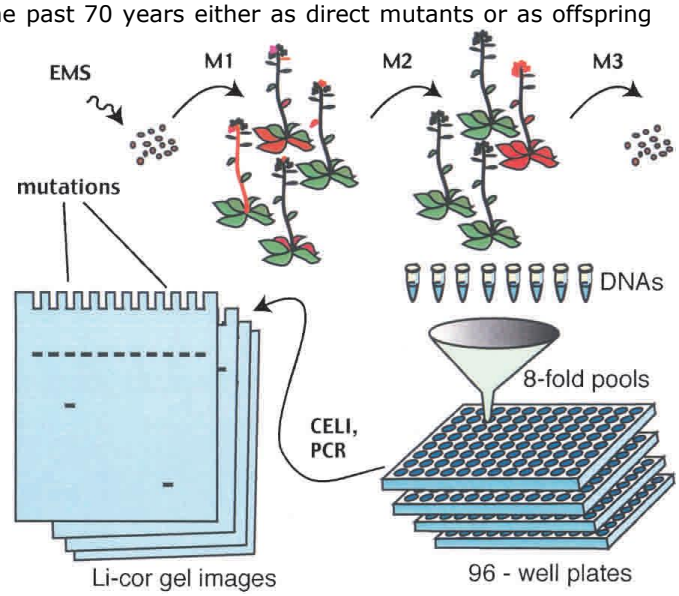


Figure 3 shows the steps that are taken in a standard TILLING experiment. Step 1 is the mutagenesis step. The seeds are treated with a mutagen, e.g. Ethyl methanesulfonate (EMSAn alternative mutagen is radiation, X- or gamma-rays (Wu et al. 2005). Next, these mutated seeds are sown and will give rise to the M1 population. To fix mutations, the chimeric plants in the M1 population are selfed, giving rise to the M2 population. In the M2 population, plants that are homogenic for the mutant alleles can now be found. DNA samples are taken from plants in the M2 population and DNA sequences of genes of interest are amplified.There are a number of ways to detect plants which carry mutant alleles (Wang et al. 2012). For example, CEL I method (Oleykowski et al. 1998) (Figure 3) has often been used. The cost efficiency of large-scale high-throughput DNA sequencing has increased, which enabled the application of Next Generation Sequencing (NGS) for mutant detection in TILLING projects. By using NGS, TILLING becomes *in silico* procedure, which could be time saving. Also with NGS, the pools of plants could be as high as 40 - to 50 - fold on some NGS instruments with high throughput and at reasonable cost (Chen et al. 2014).

Figure 3. Overview of the steps taken in TILLING. *Source: (Till, Reynolds, et al. 2003). Parts of plants whose cells carry mutant alleles are depicted in red. Therefore, in M1 population plants are chimeric, and in the M2 population, plants consists of homogeneous either wild type or mutant plants.*

Further relation of mutation characteristics found by NGS and phenotypic changes occurring in mutant plants can be used in functional analysis of a mutated gene (Till, Colbert, et al. 2003).

1.1.6 Kompetitive Allele Specific PCR (KASP)

Currently the most used uniplex, one marker at a time, SNP genotyping method in our lab (Wageningen UR, department of Plant Breeding) is the TaqMan™ technique (Livak 1999). With the high evolvement in technical improvements of the last 20 years, it is argumentative that more recent techniques perform better. Based on literature it is a very cost-effective and fast technique compared to TaqMan™ (Kumpatla et al. 2012). Also, LGC Genomics states that KASP achieves higher assay design succes rates (98-100%) and conversion to a successful working assays (93-94%) compared to TaqMan™ (Respectively 72 and 61%).

KASP (LGC Genomics) is an easy designable, trustworthy and cost-effective SNP genotyping technique (Semagn et al. 2014). The technique is based on the competition between two allele specific forward primers that both harbor a common location related part and an unique tail sequence. The unique tail sequences correspond to a universal FRET (Fluorescence Resonant Energy Transfer) cassette. One cassette is labelled with a FAM™ dye and one cassette is labelled with a HEX™ dye. Based on the fluorescence, you know which allele specific primer bound to your template DNA.

The KASP technique is of special interest to us, because we decoded the FRET specific oligo's In this way, we are able to design the used primers on our own. This increases the cost-effectiveness even more compared to the scenario where LGC designs your primers.

1.1.7 Chimerism

The tissue of a plant shoot is derived from three germ layers of the shoot meristem(Laux and Mayer 1998) (Figure 4). Each layer presumably contains 1–3 stem cells (sc, darkly shaded) within the apical part of the central zone (Stewart and Dermen 1970). The stem cells are at the ultimate source of the whole shoot tissue. In total there are therefore about 3*3=9 stem cells.

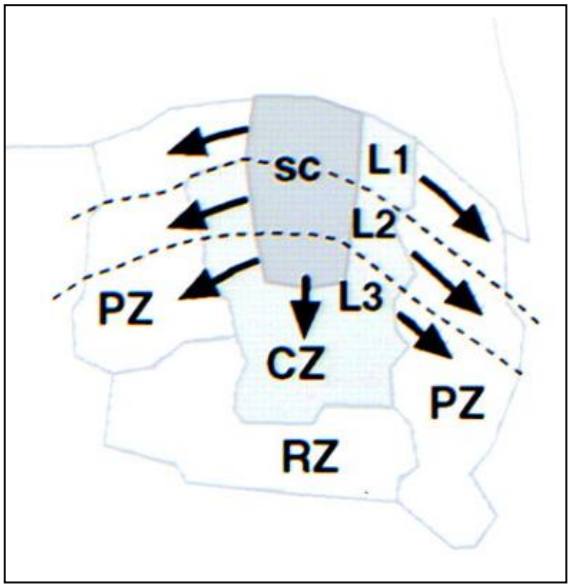


Figure 4. Organization of the plant shoot meristem. Abbreviations: SC=Stem Cell, CZ=Central Zone, PZ=Peripheral zone, RZ=Rib zone, L1/L2/L3=layer. *Source: (Mayer et al. 1998)*

A mutagen, e.g. EMS, mutates at random. Because of the random mutation, every cell within an organism is different. If you apply mutagenesis at the seed stage, a chimeric plant will arise. A chimera is a single organism composed of genetically distinct cells. Since only mutations in the stem cells will end up in the shoot, a lot of mutations will not be visible in the shoot. If a mutation arises in one of the stem cells, it will only end up in the part of the plant that derives from that one stem cell. So if a mutation arises in a stem cell, only about 10% (1/9=0.11) of the shoot cells contains your mutation. Since the mutant cells are heterozygous, our expectations are that only about 5% of the plant's genome will be of a mutant origin. The big problem is that you also only have 10% chance that your mutation ends up in the reproductive organs and in the next generation, the M2 mutant generation. Most tissue cells is not involved in the production of the next generation, therefore it could be the case that your mutation is not heritable. To overcome this problem, the mutant population is selfed, giving rise to the M2 population, before DNA is isolated (Figure 3). When your mutation is found in the M2 population, you know for sure that the mutation is in the heritable tissue.

1.2 Previous work

Tomato seeds (cv. MicroTom) have been treated with EMS giving rise to an M1 mutant population consisting of 2,432 plants (Tuinen et al. 2012). Leaf material was collected from all parts of each M1 plant. Then 38 pools each containing equal amounts of leaf material of 64 plants were prepared. Genomic DNA was extracted from each of the 38 pools.

25 candidate genes were selected based on research carried in several tomato projects dedicated to tomato fruit quality, such as EU-SOL and CBSG, and from related literature (Table 1).The list of selected genes includes genes that are involved in tomato fruit perception and quality, like glycosyltransferases, malate transporters, fruit ripening genes, and others. PCR primers for the genes of interest were designed (Chapter 7.1) and complete genes or, for some genes, fragments covering a whole genes were amplified from each of the 38 pools. The amplified gene DNA were subjected to Illumina Solexa high-throughput sequencing.

The sequencing reads were then mapped onto cDNA sequences of the candidate genes derived from tomato genome (www.solgenomics.net) and frequencies of the most prominent EMS nucleotide transitions (G->A, C->T) were calculated using CLC Genomic Workbench 7 software. Mutations which appear with a frequency of 0.4-1.0% and that satisfying for quality requirements were selected and subjected to a protein functional effect analysis using several on-line tools: SIFT, Polyphen-2, Panther and SNAP. Mutations which were predicted to affect protein's function or caused stop-codons in corresponding cDNAs were selected (Table 2).

1.3 Aims of this project

- Confirm the previously found mutations in the M1 generation by resequencing
- Develop a method to determine mutant plants in a large TILLING population.
- Find individual mutant plants within the mutant pools for a further functional analysis of mutated candidate genes.
- Screen for existing variation in Solanum lycopersicum accessions.

Table 1. Tomato fruit quality gene list

Solyc ID	Gene name	Enzyme family	trait
Solyc04g081830	GGT15	glycosyltransferase	VOC, secondary metabolite glycosides
Solyc01g067350	GGT19	glycosyltransferase	VOC, secondary metabolite glycosides
SL1.00sc07184_419.1.1	GGT11	glycosyltransferase	VOC, secondary metabolite glycosides
Solyc11g007460	GGT6	glycosyltransferase	VOC, secondary metabolite glycosides
Solyc03g078240	GGT13	glycosyltransferase	VOC, secondary metabolite glycosides
Solyc09g059170	GGT1	glycosyltransferase	VOC, secondary metabolite glycosides
Solyc10g085230	GAME5	glycosyltransferase	Alkaloids
Solyc06g072910	MalT1	Aluminum-activated malate transporter	Sour/Malate/Citrate
Solyc06g072920	MalT2	Aluminum-activated malate transporter	Sour/Malate/Citrate
Solyc06g074100	MalT3	malate transporter	Sour/Malate/Citrate
Solyc06g072580	PDH	Pyruvate dehydrogenase s.u.E1	Sour/Malate/Citrate
Solyc06g008740	Zinc finger	Zinc finger transcription factor	Phenyl Ethanol
Solyc07g006220	GAE1	UDP-glucoronate 4-epimerase 1	Candidate gene in a fruit firmness QTL
Solyc05g050990	GAE3	UDP-glucoronate 4-epimerase 3	Candidate gene in a fruit firmness QTL
Solyc05g051350	RHGL	Rhamnogalacturonate lyase	Candidate gene in a fruit firmness QTL
Solyc07g006140	CYP72A54	cytochrome P450	Alkaloids
Solyc07g006890	CYP94A6	cytochrome P450	Alkaloids
Solyc07g007460	CYP71D48	cytochrome P450	Alkaloids
Solyc05g012020	RIN	rin transcription factor	Fruit ripening
Solyc10g050160	MT Guaiacol	methyl transferase	QTL for guaiacol
Solyc01g107820	Twi-1	glycosyltransferase	VOC, secondary metabolite glycosides
Solyc11g030600	MET1	DNA (Cytosine-5)-methyltransferase	Epigenetic gene regulation
Solyc01g006540	LOX1.2	lipoxygenase	Lipid VOC
Solyc01g105890	TPS5	Terpene synthase	QTL for terpenes
Solyc08g005640	TPS20	Terpene synthase	QTL for terpenes

Table 2. Overview of the potentially interesting mutations that were found in the first sequencing experiment with Illumina Solexa high-throughput sequencing

Solyc ID	Gene name	Plant pool	Reference Position	Transition	Count	Coverage	Frequency	protein effect prediction score	Amino acid change
SL1.00sc04684_66	false GGT11	B5_R1	922	C>T	6	812	0.739	4	R308W
Solyc07g006220	GAE1	E2_R2	275	C>T	8	772	1.036	3	T92I
Solyc10g085230	GAME5	D3_R1	125	C>T	42	5374	0.782	4	T42I
SL1.00sc07184_419	GGT11	D6_R2	65	G>A	13	1506	0.863	2	S22N
Solyc03g078240	GGT13 (A)	B2_R1	865	C>T	46	9627	0.478	5	Q289 stop
Solyc03g078240	GGT13 (B)	D1_R1	782	C>T	27	5610	0.481	4	S261F
Solyc03g078240	GGT13	D1_R2	782	C>T	31	5311	0.584	4	S261F
Solyc01g067350	GGT19	A1_R1	572	C>T	23	5278	0.436	5	Q180 stop
Solyc06g074100	MalT3	E1_R1	646	C>T	10	1749	0.572	2	P216S
Solyc11g030600	MET1 (A)	D6_R1	1967	C>T	16	2557	0.626	1	S656F
Solyc11g030600	MET1 (B)	E2_R1	3871	G>A	3	690	0.435	4	G1291S
Solyc10g050160	GMT10 (A)	A6_R1	67	C>T	1	235	0.426	5	Q23 stop
Solyc10g050160	GMT10 (B)	D2_R1	136	G>A	59	6750	0.874	2	E46K
Solyc10g050160	GMT10	D2_R2	136	G>A	58	6512	0.891	2	E46K
Solyc05g051350	RHGL	E5_R2	1810	C>T	11	2688	0.409	5	Q604 stop
Solyc05g012020	RIN	E1_R1	445	C>T	3	334	0.898	2	S64F
Solyc05g012020	RIN	E1_R2	451	C>T	3	317	0.946	2	T66I
Solyc01g107820	Twi-1 (A)	A2_R1	804	C>T	36	8640	0.417	5	Q184 stop
Solyc01g107820	Twi-1 (B)	E1_R1	1281	C>T	3	524	0.573	5	Q343 stop
Solyc06g008740	Zinc finger	F2_R2	256	C>T	3	595	0.504	5	Q12 stop

2 Material and Methods

2.1 Plant material and previous work

2,432 tomato, *Solanum lycopersicum,* seeds (cv. MicroTom) were treated with EMS. EMS, a mutagenic compound, created at random mutations throughout the cells of the tomato seeds. Germination of the seeds gave rise to the first mutant population (M1). Several leafs were collected at as uniform as possible of every individual M1 plant and DNA was isolated. The seeds that the M1 generation produced were collected and stored. The isolated DNA was randomly divided over 38 pools of 64 samples. The 38 pools were sequenced using Illumina Solexa high-throughput sequencing. The potentially interesting mutations that were found in the 25 candidate gene list (Table 1) are in Table 2.

2.2 DNA amplification and preparation for high-throughput sequencing

To amplify candidate gene sequences a standard PCR was performed using AccuTaq™ DNA polymerase. DNA amplification was performed using the Biosystems® 2720 Thermal Cycler using a standard 35-cycle run with a denaturation temperature of 95 °C, annealing temperature of 57 °C and an elongation temperature of 68 °C. A single PCR reaction (15 µl) consisted of 0.5 µl DNA (20-50ng/µl concentration), 1.5 µl AccuTaq™ buffer (10x), 0.75 µl Forward Primer (10µM), 0.75 µl Reverse Primer (10µM), 0.3 µl AccuTaq™ polymerase and 11.2 µl Milli-Q. Water-controls were included by replacing DNA with Milli-Q. The primers were designed by others for the first sequencing round described in the chapter 'previous work'. Primer sequences can be found in the appendix (Chapter 7.1)

PCR products stained with loading buffer were separated by electrophoresis on a 1% agarose gel using 0.5x TAE buffer and 0.1% Ethidium Bromide for about 30 minutes at 80V with the 1kb GeneRuler™ ladder for reference. Gels were photographed and images were examined to identify a proper PCR product. To determine the concentration of amplified DNA, the Qubit™ Assay method was used. Amplification success and DNA concentrations can be found in the appendix (Chapter 7.2)

T4 ligase and kinase were added to form one long strands of DNA which were then sequenced using Illumina solexa Sequencing.

2.3 Resequencing and SNP confirmation

The resequence data were manually analyzed. We checked whether the earlier found mutations were still there and if they acquired reliable quality requirements. Reliable quality requirements are a frequency between 0.4 and 1.0%, a high coverage and count and a forward/reverse balance between 0.4 and 0.6. If the found SNPs are not in agreement with the quality requirements, it could be a good indication that something went wrong with sequencing and therefore the SNP is not trustworthy. Because we only sequenced one gene per pool, the reads coverage is much higher. Subsequently the reliability of the sequence data got higher compared to the first sequence experiment.

2.4 SNP genotyping using Kompetitive Allele Specific PCR (KASP)

2.4.1 The standard KASP procedure

To genotype all 64 samples, we used Kompetitive Allele Specific PCR (KASP) (LGC genomics). The primer sequences that were used are in the appendix (Chapter 7.3). A Bio-Rad CFX-96 rt-PCR machine was used to measure the expression of the different fluorescent dyes. A single KASP reaction (10 µl) consisted of 1.0 µl DNA (5-50 ng/ µl concentration), 5.0 µl 2x KASP MasterMix, 1.0 µl Forward WT primer (1.0 µM), 1.0 µl Forward Mutant primer (1.0 µM), 1.0 µl reverse primer (1.0 µM) and 1.0 µl Milli-Q. KASP try-outs have been performed with amplicon lengths between 73 and 112 bp and primer concentrations between 0.05 µM and 0.3 µM. Tomato (cv. Micro-Tom) DNA with a known genotype for the rhamnosyl transferase gene were used for the KASP try-outs

The thermocycling conditions were followed according to the protocol provided by LGC genomics (see table 3). An extensive home-made KASP protocol is attached in the appendix.

Table 3. Thermal cycling conditions used in the qPCR performed by Bio-Rad CFX qPCR machine based on the KASP genotyping kit user guide. Step 5 and 6 are optional steps for additional thermal cylcing.

Step	Procedure	Temperature (°C)	Time	Number of cycles per step
1	Activation	94°C	15 min	1
2	Denaturation	94°C	20 sec	10
	Annealing/Elongation	61°C (ΔT: -0.6°C /cycle)	60 sec	
3	Denaturation	94 °C	20 sec	26
	Annealing/Elongation	55 °C	60 sec	
4	Reading	37 °C	60 sec	1
5	Denaturation	94 °C	20 sec	3
	Annealing/Elongation	55 °C	60 sec	
6	Re-Reading	37 °C	60 sec	1

2.4.2 Modification of KASP for detection of alleles with a low frequency

Next to the normal KASP approach, we also conducted a KASP assay without primer competition. We performed a KASP assay where only one type of forward primer was added to the sample mix. Or the forward WT primer or the forward mutant primer. The absence of one of the primers was compensated by the addition of extra Milli-Q. So in this case a single KASP reaction (10 µl) consisted of 1.0 µl DNA (5-50 ng/ µl concentration), 5.0 µl 2x KASP MasterMix, 1.0 or 0.0 µl Forward WT primer (1.0 µM), 1.0 or 0.0 µl Forward Mutant primer (1.0 µM), 1.0 µl reverse primer (1.0 µM) and 2.0 µl Milli-Q

2.5 Cloning of Twi-1 gene sequence fragments

2.5.1 General introduction

The cloning of Twi-1 DNA sequence was performed using the Golden braid 2.0 method (Sarrion-Perdigones et al. 2013). The forward primer to amplify the DNA fragment was: GCGCCGTCTCGCTCGGGTATGGCTCCCATGATTAAA. The reverse primer that was used to amplify the DNA fragment was: GCGCCGTCTCGCTCGTTAACGATATGAAGTTATGTCTTGTA . In red you see the BsmB1 recognition site, in yellow the actual cutting site. The used plasmid vector, pUPD, contains an ampicillin resistance and the LacZ operon.

2.5.1 Gene Transformation to E. coli

DNA fragments were amplified using earlier described methods with the above mentioned forward and reverse primers. Next the DNA was cleaned using the Illustra MicroSpin™ G-50 columns. Then the amplified DNA was cut and ligated into the pUPD vector using a 25-steps thermocycling program. 2 minutes at 37 °C followed by 5 minutes at 16 °C. The reaction mix contained 0.5 µl DNA (40 ng/µl), 1.0 µl pUPD (75 ng/µl), 0.6 µl T4 ligase (5 U/µl), 0.5 µl BsmB1 (10 U/µl), 1.0 µl T4 Ligase Buffer, 1.0 µl DTT (10 mM) and 5.4 µl Milli-Q.

We used the heat shock transformation method (http://www.embl.org). 5 µl of the ligation reaction mix was added to a Eppendorf tube that contained 30 µl of competent E. coli cells and stored on ice for 15 minutes. Apply the heat shock by placing the Eppendorf tube in a waterbath at 42 °C for 45 seconds. After the heat treatment, place the Eppendorf tubes as fast as possible on ice. Next we added SOC medium, about 465 µl so the total volume in the Eppendorf tube is 500 µl. Then the Eppendorf tube is placed in the stove (37 °C) for 1 hour, the tubes were mixed every 10 minutes. Because we did not know the success rate of the experiment, we created two different concentrations of competent cells. The first group contains 20% of the competent cells, by pouring 100 µl of the total 500 µl to a plate. The residual 400 µl were centrifuged at 6800 rpm for 1 minute. 300 µl supernatant was discarded and the residual 100 µl, which contains 80% of the competent cells, was poured to another plate. Every plate contained 12.5 ml nutrient rich LB medium with 12.5 µl ampicillin and 5 µl X-Gal. The plates were incubated overnight at room temperature.

2.5.3 Colony enrichment and SNP genotyping

The next day, we ticked off 90 individual colonies and placed them in a 15 ml tube containing 2.5 ml LB medium and 2.5 µl ampicillin. Since we expect about 5% of the DNA to be mutant DNA, we expect about 4 or 5 colonies out of 90 to contain the mutant SNP. The 90 15 ml tubes were placed in a incubator (37

°C) for overnight, in this way the individual colonies could increase in quantity. The last step before the colonies could be genotyped, is to transfer 100 µl of the LB medium containing the competent E.coli cells in a 96-wells plate. After spinning of the plate, the supernatant was discarded and 50 µl of Milli-Q was added to the residual. After killing the E. coli cells by placing the 96-wells plate for 10 minutes at 94 °C, the DNA was ready for genotyping.

KASP was used to genotype single colonies as described above. The KASP genotyping technique was performed twice for all 90 samples.

2.6 *In silico* association study

Genotypic data were collected from the 150 Tomato Genome ReSequence Project (http://www.tomatogenome.net/). The promoter region of the 8 glycosyl transferase genes (table 1) have been screened for genetic polymorphism in 54 *Solanum lycopersicum* accessions. Genetic polymorphisms are all deviations from the reference genomes, like SNPs and InDels. metabolomic data of all 54 *Solanum lycopersicum* accessions are collected with the use of GC-MS and LC-MS, this data is confidential.

The Kruskal-Wallis rank sum test with the use of the MapQTL 6 software (Van Ooijen 2009) was used to statistically test the association between a certain polymorphism and a certain compound.

3 Results

3.1 Resequence data

The 17 found mutations after sequencing were subject of a resequencing experiment. DNA amplification of the mutant pools was performed, only for the gene that contains a possible mutant. Table 4 shows the results of the mutations that were found back in the second sequencing round. From the initial 17 SNPs that were found, 8 SNPs were confirmed at the quality conditions chosen (see Materials & Methods). Because in the second sequence experiment only one gene pool was amplified instead of all genes, as we did with the first sequence experiment, we ended up with a higher amplification density per gene. Because the amplification density per gene was higher, the sequence coverage was higher in the second sequence experiment compared to the first sequence experiment. The higher the coverage of a sequence is, the more reliable the result is. The mutant frequency ($Frequency = \frac{Count}{Coverage} * 100\%$) should be 0.78% ($Frequency = \frac{1}{128} * 100\%$) because in a pool of 64 plants (128 chromosomes) only 1 plant is heterozygous (1 chromosome), so 1 chromosome out of 128 is expected to be mutated. However, all observed mutation frequencies were lower than the theoretical 0.78 (table 4).

As can be seen in Table 4, only C>T and G>A transitions were selected. We selected only these transitions because EMS only creates C>T or G>A transitions in >99% of the cases (Greene et al. 2003).

The genes that contain a mutation belong to different gene families. GAE1 is a gene involved in tomato fruit firmness. MET1 is a gene involved in transferring methyl groups. Methyl groups are important for the methylation of DNA and therefore it can affect expression of genes. GMT10 is also a methyl transferase, but is a candidate gene found in a QTL responsible for elevated guaiacol quantities. GGT13, GGT19 and Twi-1 are all different glycosyltransferase genes. As is described in the introduction, glycosyltransferase genes are important for the production of flavonoids. Flavonoids have a potential health-protecting effect. Also, the expression of Twi-1 responds rapidly to wound- and pathogen related signals(O'Donnell et al. 1998) indicating a function in plant's defense. As the protein effect prediction score (based on *in silico* prediction software) showed, some mutations could highly alter the protein functional properties. The mutation in the MET1 gene and in the GMT10 gene on the contrary do not score very high, relatively 1 and 2. It is therefore doubtful whether the mutation really affects the plant's phenotype. Four mutations change a Glutamine (Q) to a stop codon (*). A stop codon definably changes the translated protein. The mutations resulting in a stop codon are therefore of high interest, since a changed phenotype is expected.

Table 4. Resequence data confirms 8 out of 17 SNPs that were found after the first sequence round.

Solyc ID	Mutant name	Reference Position	transition	Count	Coverage	Frequency	Protein effect prediction score	code
Solyc07g006220	GAE1	275	C>T	1559	269489	0.58	3	T92I
Solyc03g078240	GGT13 (A)	865	C>T	1382	491722	0.28	5	Q289*
Solyc03g078240	GGT13 (B)	782	C>T	923	514502	0.18	4	S261F
Solyc01g067350	GGT19	572	C>T	1050	481164	0.22	5	Q180*
Solyc11g030600	MET1	1967	C>T	1885	329865	0.57	1	S656F
Solyc10g050160	GMT10 (B)	136	G>A	871	163362	0.53	2	E46K
Solyc01g107820	Twi-1 (A)	804	C>T	242	316207	0.08	5	Q184*
Solyc01g107820	Twi-1 (B)	1281	C>T	843	245949	0.34	5	Q343*

3.2 The effect of different conditions on the KASP performance

There are a lot of SNP genotyping techniques. Within our research group (Department of Plant Breeding at the Wageningen University) the TaqMan™ method (Livak 1999) or High Resolution Melting (HRM) (Liew et al. 2004) are the currently used SNP genotyping techniques. Since these techniques are relatively old, late '90s/begin '00s, it is argumentative that there are newer techniques that function better. A more recent SNP genotyping technique is the Kompetitive Allele Specific PCR (KASP) by LGC Genomics.. Chapter 3.3 show the results of experimenting with some variable features of the KASP technique.

3.2.1 Differences in amplicon length

A known rhamnosyltransferase (RT) (Solyc01g094980) mutant gene, which was studied in a different project, was used to test different KASP conditions. First the effect of different amplicon length was tested (figure 5). Primers for RT gene were designed to produce amplicons of three different lengths. Figure 5A shows the result of a KASP assay with an amplicon length of 73 bp, Figure 5B shows the result of an amplicon size of 101 bp and Figure 5C shows the result of a KASP assay that was performed with an amplicon length of 112 bp. The RT genotypes are in all three cases correctly assigned. The data show clear genotyping clusters. Although there are differences visible, especially figure 5C is different from the rest. Although clear genotyping clusters were found, the clusters are not as isolated compared to the results that are shown in Figure 5A and 5B. If you work with samples of an unknown genotype, you could more easily make mistakes with assigning a genotype to a sample. A slight deviation from the cluster where it actually belongs to, could lead to a false assigned genotype.

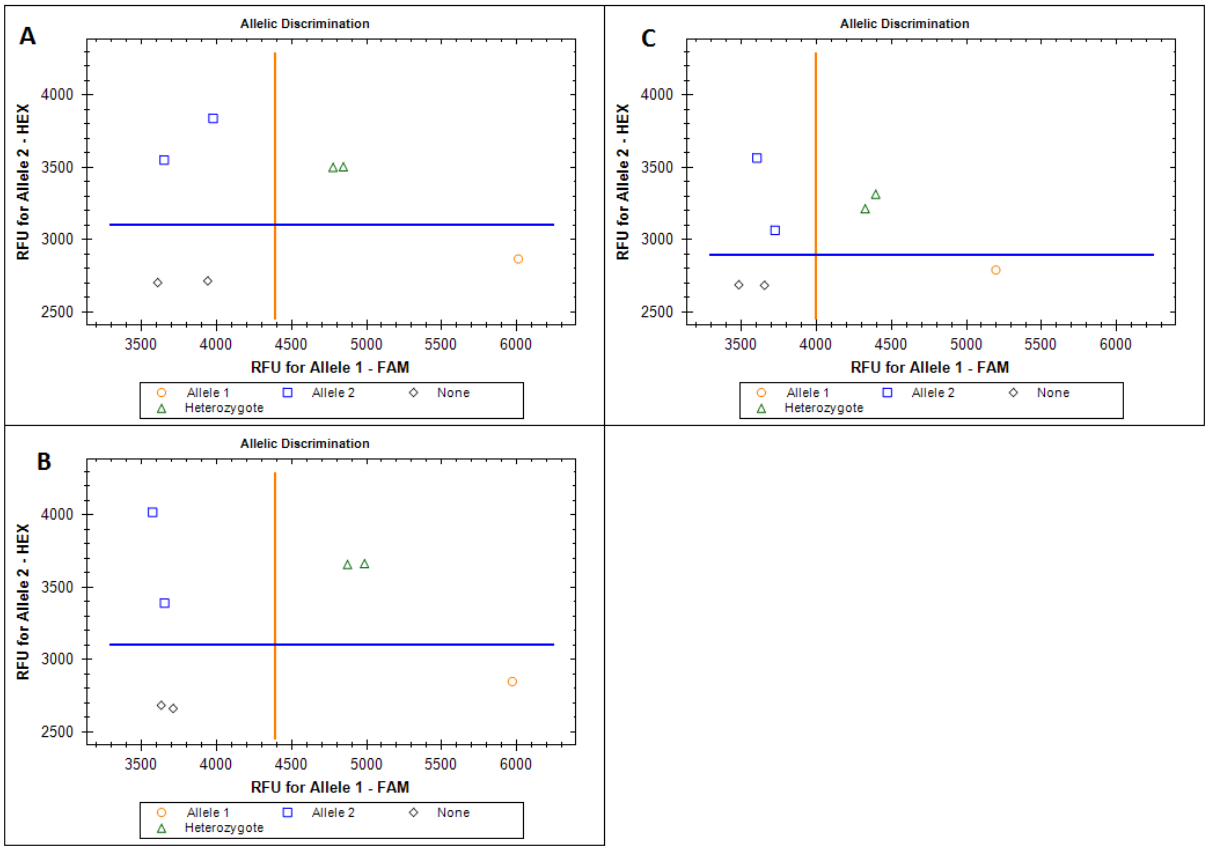


Figure 5. Three KASP results with positive control samples of the Rhamnosyl Transferase gene. A) amplicon length of 73 bp; B) amplicon length of 101 bp; C) Amplicon length of 112 bp (primer quantity 1.5 µM). The colors and shapes of the data points are assigned automatically by the KASP data analysis software. Black rhombus – non template controls, blue squares – homozygous mutant allele, green triangles – Heterozygotes, orange circles – homozygous Wild Type allele.

3.2.2 The optimum primer quantity

Figure 6 visualizes the results of the second KASP experiment, the search for the optimal primer concentration for a KASP assay. The figure 6A to 6F shows the addition of the increasing concentrations: 0.05 µM, 0.075 µM, 0.1 µM, 0.15 µM, 0.2 µM and 0.3 µM of all three primers of the rhamnosyl transferase gene. Amplicon length that was used was 73 bp long. Figure 6A makes clear that 0.05 µM is too low for a proper KASP assay. Figure 6B shows that 0.075 µM of the primers is also too low. The absolute amount of RFUs is low compared to the others, below 4500 RFU. The heterozygote sample and the mutant sample that show the same fluorescent expression (figure 6B) is probably the result of a technical error of the KASP technique, as the technique is not 100% solid. The Figures 6C-6E all show correct genotypic clusters, however the differences between the cluster of homozygous wild type and the heterozygotes is small. Still it can be stated that the optimal primer concentration is within the range of 0.05 and 0.2. . In figure 6F, primer concentration of 0.3, the assigned genotype by the software is not in agreement with the reality. Unfortunately, this cannot be made clear by the Figure itself, but a good indication that the KASP result is not really trustworthy is the fact that the No Template Controls (NTC) are not in the correct corner. Possible explanations of a dysfunctional KASP with high primer concentrations is further discussed in the discussion section.

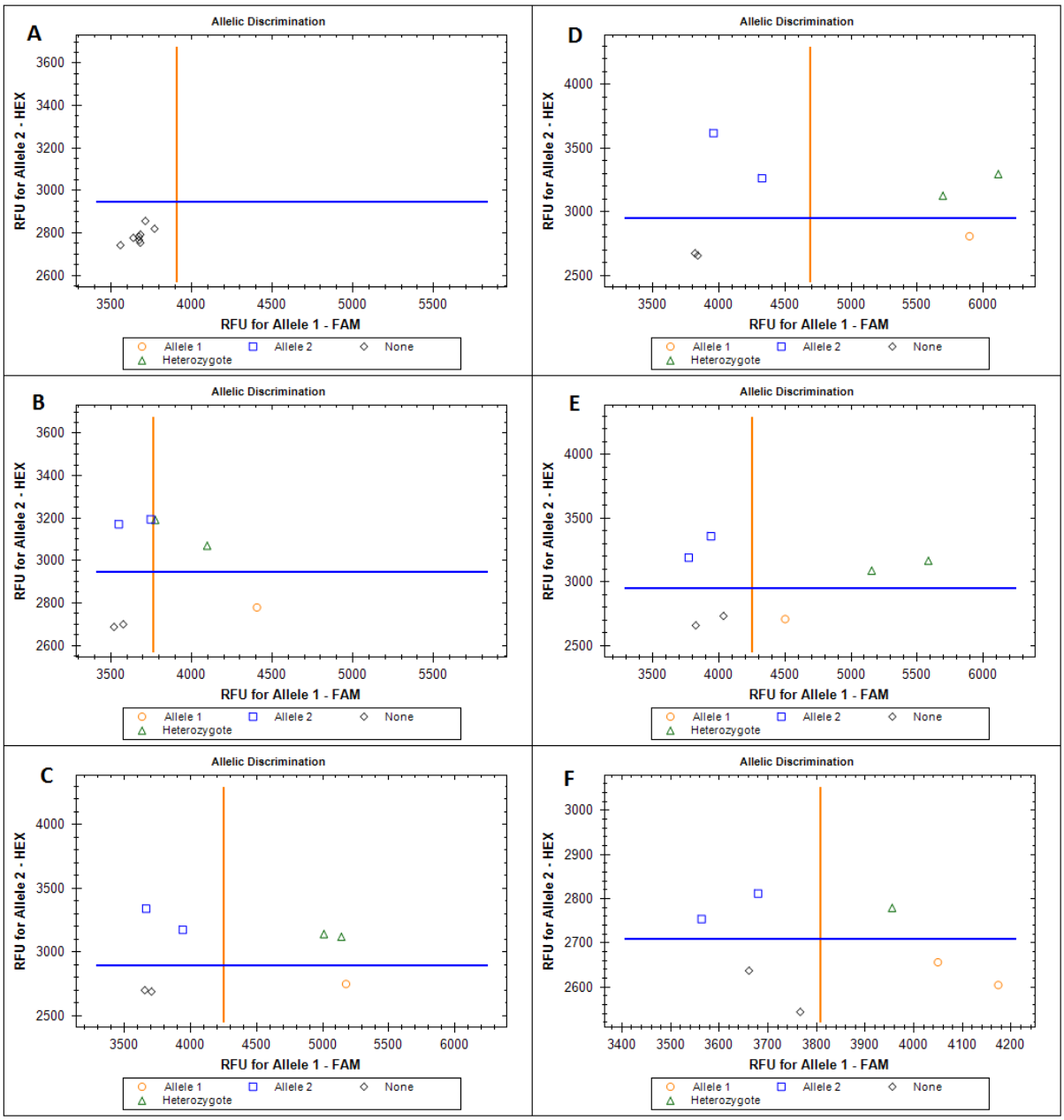


Figure 6. KASP assay of samples with a known genotype (positive controls) of the rhamnosyl transferase gene with six different primer concentrations. A) 0.05 µM; B) 0.075 µM; C) 0.1 µM; D) 0.15 µM; E) 0.2 µM and F) 0.3 µM. The colors and shapes of the data points are assigned automatically by the KASP data analysis software. Black rhombus – non template controls, blue squares – homozygous mutant allele, green triangles – Heterozygotes, orange circles – homozygous Wild Type allele. All data points were classified as the right genotype except for the data points in A) and F) with respectively 0.05 µM and 0.3 µM of the primer concentration.

3.2.3 The effect of extra cycling

We also tested the effect of performing additional thermocycles (Figure 7). Figure 7A shows the result of a normal KASP assay (amplicon length 101, primer concentration 0.15 µM). Figure 7B shows the same KASP assay with 3 additional Polymerase Chain Reactions cycles. Figure 7C shows the same KASP assay with again 3 additional cycles, so in total 6 extra cycles. Already after a normal KASP protocol the samples could easily be genotyped (Figure 7A). Extra cycles increased the absolute RFUs. For example the homozygote allele 1 sample moves from 6000 RFU for FAM™ expression, via 7000 RFU after 3 additional cycles (Figure 7B) to 7600 RFU after 6 additional cycles (Figure 7C). Because all samples increase in RFU, even the NTCs, extra cycles do not make the total image much clearer. Extra cycles do not improve the ability to assign a genotype to a sample.

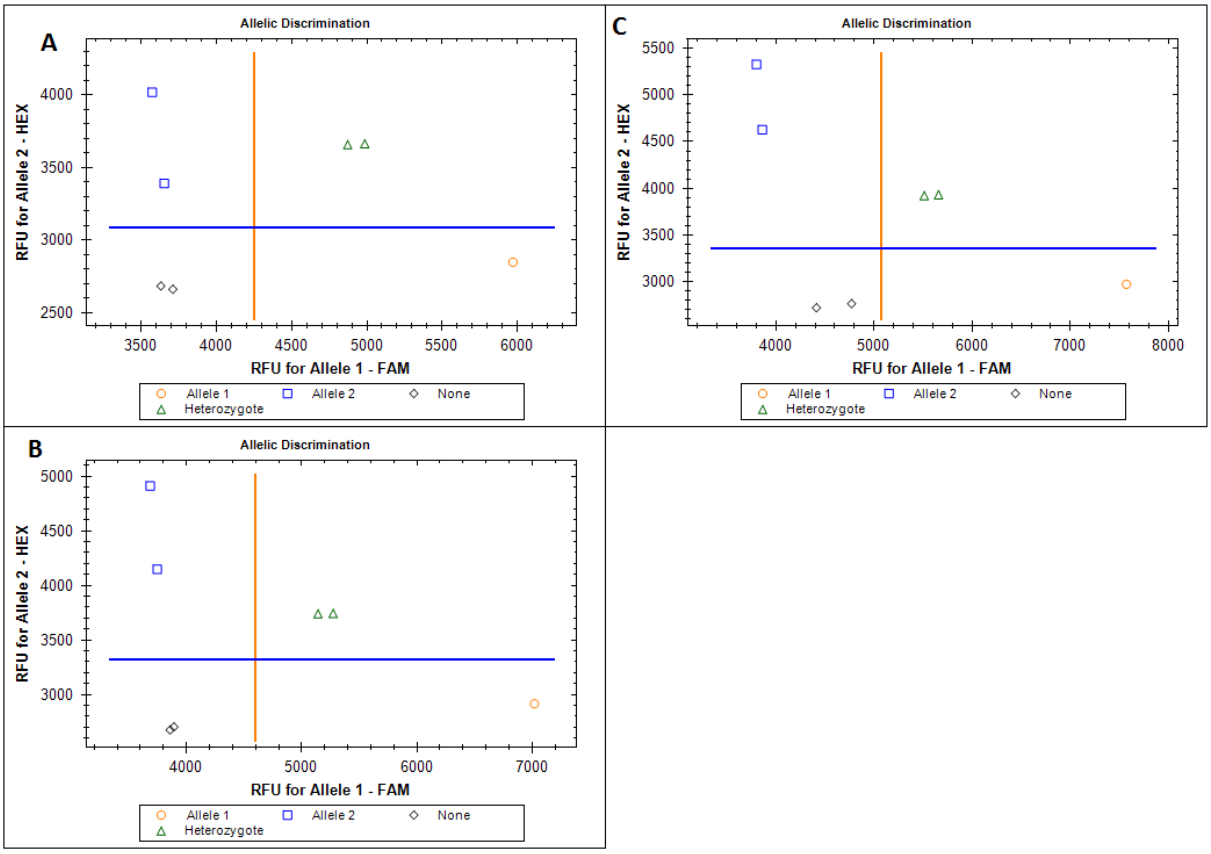


Figure 7. KASP assay on Rhamnosyl Transferase gene. Primer concentration 1.5 µM, amplicon length 101 bp. A) no additional cycles; B) 3 additional cycles; C) 6 additional cycles. The colors and shapes of the data points are assigned automatically by the KASP data analysis software. Black rhombus – non template controls, blue squares – homozygous mutant allele, green triangles – Heterozygotes, orange circles – homozygous Wild Type allele

3.3 SNP genotyping of the mutant plant pools with the use of KASP

The sequence results described in section 3.1 are based on plant material of a pool consisting of 64 different plants. However, only 1 plant contains the mutation. To identify this one plant out of 64 plants, a SNP detection method called KASP is performed. Of the one mutant plant, the M2 seeds will be grown giving rise to the M2 mutant population. All those plants are the result of selfing the mutant M1 plant. In the M2 generation, the plants will segregate according to Mendelian segregation. The homozygous mutant plants can be selected and phenotyped. By genotyping all 64 plants in the mutant pool, you only need to grow the M2 generation of the true mutant instead of the M2 generation of all 64 plants.

Four different mutations: Twi-1(B), GMT10(B), GGT13(A) and GGT13(B) were subjected to KASP assay in order to find the individual mutant M1 plants in the corresponding plant pools. Due to lack of time, we were only able to subject these four mutations to KASP. Figure 8 shows the result of 4 mutant pools that have been genotyped to identify the one mutant plant out of the 64 plants within the pool. Figure 8A shows the KASP assay results of the pool which contains the Twi-1 (B) mutation. The Figure clearly shows one plant that exceeds in HEX values, the mutant alele signal, compared to the other plants. The single plant with the high HEX signal was suggested to be a mutant plant. A mutant plant was also found for GMT10(B) mutation (Figure 8B). no mutant plants were detected by KASP in the GGT13(A) and GGT13(B) pools (Figure 8C and 8D).

For further experiments on mutant plant detection, Twi-1 (B) was selected. This choice was due to following reasons. First, Twi-1 is a very interesting gene that can be involved in the flavonoid biosynthesis. Second, the SNP changes a Glutamine to a stop codon. The change to a stop codon makes it highly probable that eventual phenotype will be altered. And finally the mutations found in Twi-1 in the first sequencing experiment were confirmed by the re-sequencing, which increased the probability that the mutations are indeed present in one of the M1 plants of the corresponding plant pool.

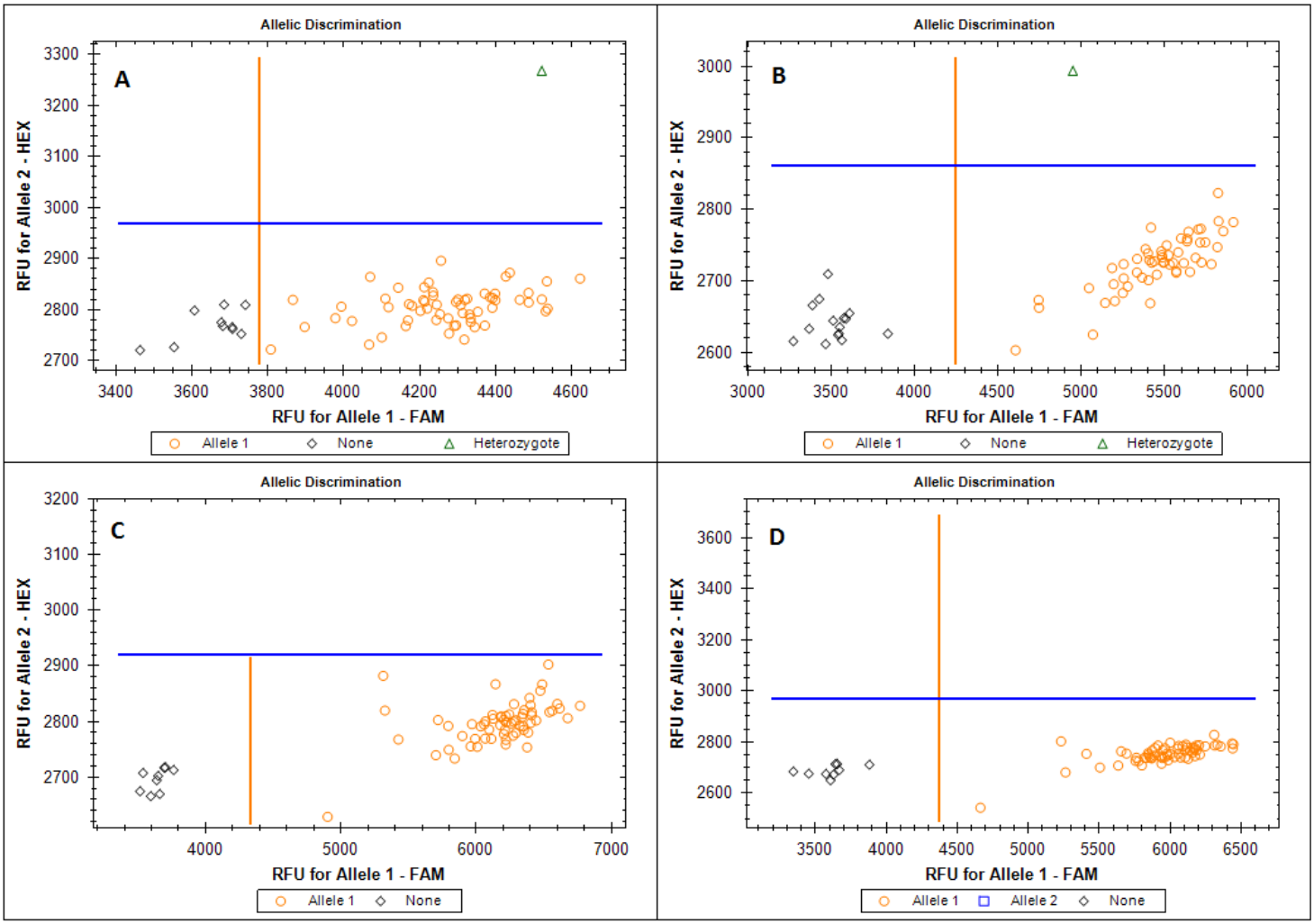


Figure 8. KASP result of the mutant pools for A: Twi-1 (B), B: GMT10 (B), C: GGT13 (A) and C: GGT13 (B). The colors and shapes of the data points are assigned automatically by the KASP data analysis software. Black rhombus – non template controls, green triangle – Heterozygote (mutant) allele, orange circle – homozygous Wild Type allele

3.4 KASP assay with low mutant DNA quantities

Despite the reproducible results of sequencing of M1 plant pools: a half of the mutations found in the first sequencing experiment were found in the second with much higher read coverage, the KASP assay was not able to consistently confirm these results. This could be due to chimerism of M1 plant material we used in this study, which leads to a low concentration of a mutant allele. To investigate the performance of KASP assay on detection of low frequency alleles we performed an allele dilution experiment, in which a mutant allele of the rhamnosyl transferase gene was diluted by different amounts of its wild type version. Figure 9 shows the KASP result of an experiment where we diluted the amount of mutant rhamnosyl transferase DNA. What this experiment has made clear, is the fact that KASP is probably not suitable when the mutant allele is present in really low quantities. When there is 5% or 1% of the mutant allele and 95% respectively 99% of the WT allele, you are not able to discriminate between the samples containing mutant DNA or WT DNA.

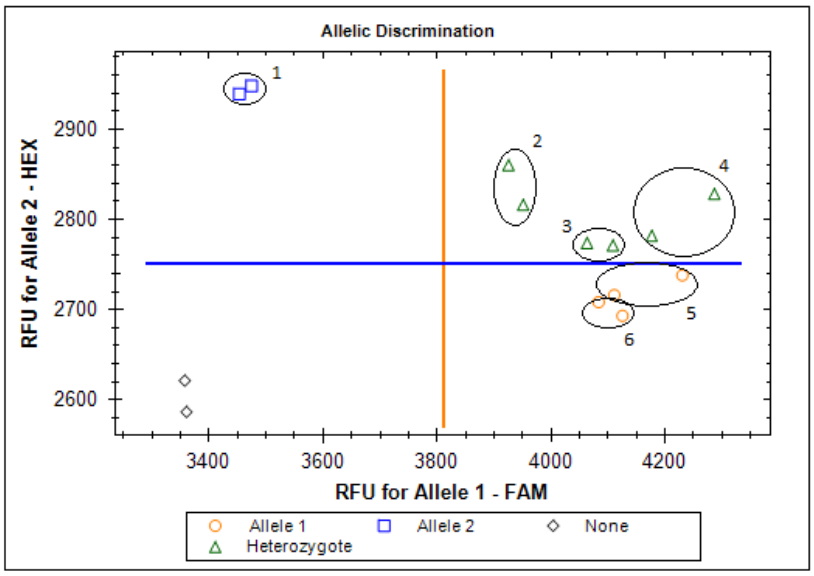


Figure 9. Mutant DNA dilution analysis. 1) 100%; 2) 50%; 3) 25%; 4) 12.5%; 5) 5% and 6) 1% mutant DNA. DNA is diluted with WT DNA. The colors and shapes of the data points are assigned automatically by the KASP data analysis software. Black rhombus – non template controls, blue squares – mutant allele, orange circle –Wild Type allele.

3.5 KASP genotyping of the potential Twi-1 (B) mutant plant, another approach

As shown in figure 9, a normal KASP assay is not very suitable in cases where the quantity of mutant DNA is low. The expression of the HEX (mutant) dye does not differ enough from normal HEX background noise. It is therefore not possible to discriminate between the mutant plant and a WT plant. To overcome the problem, we changed the KASP assay slightly. We took out the 'competition' part of the KASP assay by splitting the original three primer reaction into two separate reactions: reaction 1 with the common reverse primer and the WT-specific (FAM) primer and reaction 2 with the common primer and the Twi-1(B) specific (HEX) primer. As a template we used the one plant that appeared to be the mutant after the first KASP assay and two at random chosen plants that were in the 'cloud' of WT plants (Figure 8A,B). Figure 10A shows the KASP assay where only the forward WT primer is added. It is clearly visible that the potential mutant shows less FAM expression, indicating the presence of lower quantities of WT DNA in the sample. Figure 10B shows the KASP assay where only the forward mutant primer is added. The possible WT samples show a comparable expression pattern as the NTC (black diamonds), indicating no amplification at all. The possible mutant sample on the contrary shows some elevated HEX fluorescent values, indicating binding of the forward mutant primer which contains a HEX specific oligo.

The results shown in Figure 10 are another clue that the one sample that popped up as 'mutant' in the normal KASP approach (Figure 8A), is really the mutant. It is however still unclear why the results shown in Figure 10A were not reproducible. Therefore the result shown in Figure 10 is another clue, but still no true confirmation that this one plant really contains the mutant Twi-1 (B) mutation.

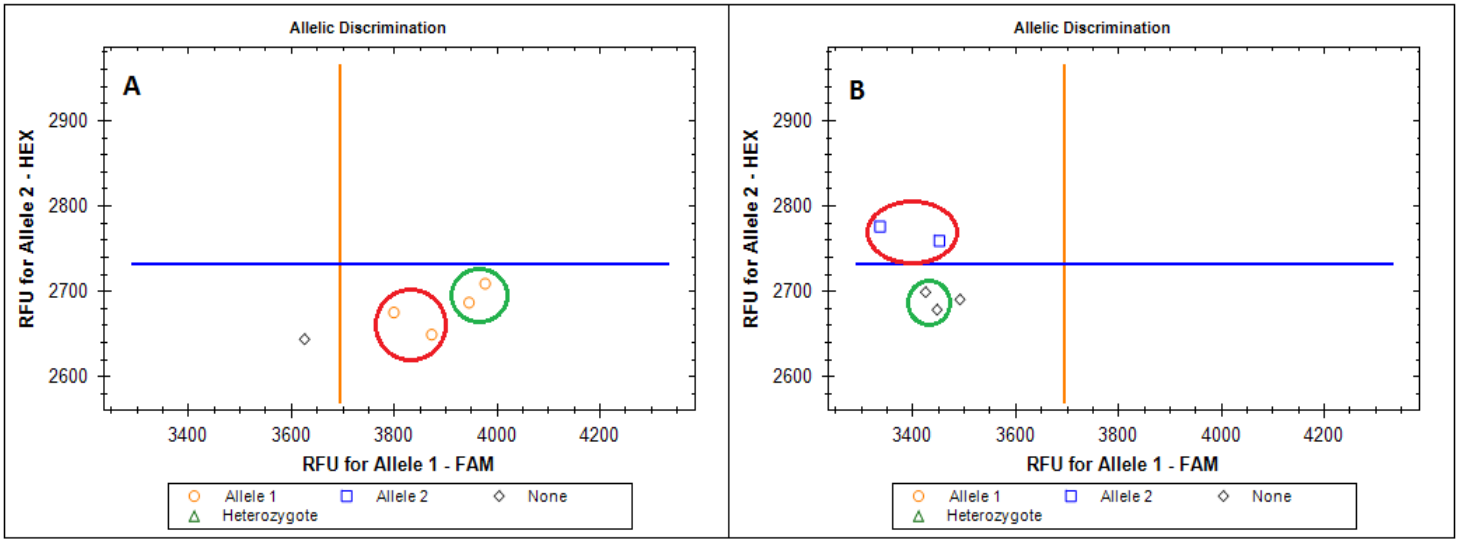


Figure 10. KASP result of Twi-1(B) potential mutant (red encircled) and potential WT (green encircled) A: KASP assay with only WT forward primer; B: KASP assay with only mutant forward primer. Allele 1, indicated by FAM expression is the WT allele and allele 2 indicated by HEX expression is the mutant allele. The colors and shapes of the data points are assigned automatically by the KASP data analysis software. Black rhombus – non template controls, blue squares – mutant allele, orange circle –Wild Type allele

3.6 KASP of the potential Twi-1 (B) mutant clones

The ultimate proof of mutations to be present are sequence data. But since the estimation of present mutant DNA is only 5%, the mutation will not be visible with normal sequence techniques. To overcome this problem, the DNA of the potential mutant plant can be cloned to *E. coli*. An *E. coli* colony only contains pure DNA sequences. In a pure DNA sequence a single colony, only one particular allele can be found. So DNA isolated from a colony is 100% identical. There are no heterozygous or chimeric in E. coli colonies. For this reason KASP could be a proper SNP genotyping technique. This has been performed for the potential mutant Twi-1(B) plant.

Figure 11 shows the KASP results of the clone population. The DNA from the potential mutant sample from within the mutant pool is transformed to E. coli, resulting in this clone population. The population consists of 90 colonies. Every colony contains one DNA template and not a mixture of DNA collected from different cells from different leafs from a plant. In that way, you could say that the DNA in every colony is pure. Cloning the Twi-1 locus to E. coli was a way to overcome the 'dilution' of mutant DNA due to chimerism. Concluding, the population contains 90 E. coli colonies in which the Twi-1 locus of the potential mutant is transformed.

The KASP experiment with the clone populations has been performed two times, and in both cases there was a comparable outcome (Figure 11). Three groups were identified: 1) non template control group, 2) a large group consisted of most of the samples and 3) a group of two samples located between the non-template control group and group 2. The fluorescence of those samples was in both KASP experiments different compared to the rest of the samples. The DNA of the two colonies in group 3 show less FAM expression compared to the colonies of group 2. Two colonies encircled in subgroup 2.1 showed the highest HEX (mutant allele) signal in both replicate KASP assays.

Although the result of this KASP assay did not find colonies which could be unambiguously classified as carrying the mutant alele, we selected colonies of group 3 and 2.1 for further sequencing to find out what alleles of Twi-1 they carry.

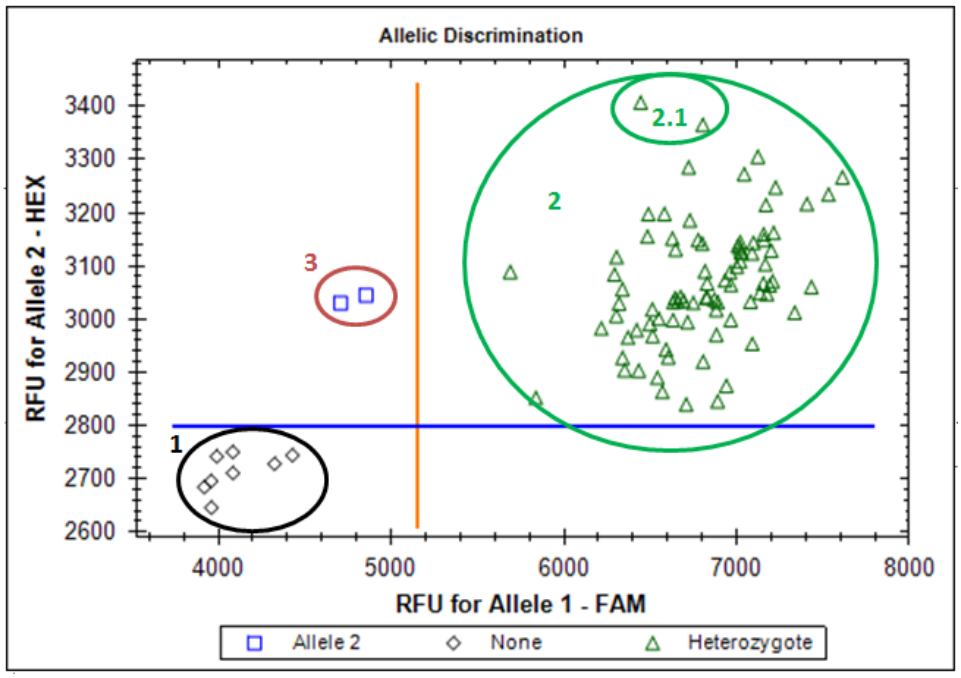


Figure 11. KASP result of clones of the potential Twi-1 (B) mutant plant. The plant expression patterns can be divided into 3 groups. 1) No Template Controls (NTC); 2) the large cloud; 3) Samples in between the large cloud and the NTCs. The two samples within circle 2.1 are samples from the cloud with a higher than average HEX (mutant allele) expression. The colors and shapes of the data points are assigned automatically by the KASP data analysis software. Black rhombus – non template controls, blue squares – homozygous mutant allele, green triangles – Heterozygotes, orange circles – homozygous Wild Type allele.

3.7 Assessing the existing variation *in S. lycopersicum* cultivars

The goal of this TILLING project is to create *de novo* mutations that could possibly improve the tomato fruit quality. However, there is also variation in tomato fruit quality in the existing tomato cultivars (Tikunov et al. 2005; Ortiz-Serrano and Gil 2007; Birtic et al. 2009). To get a better insight in the relationship between this existing phenotypic variation, we tried to link it to genotypic data. The source of the genotypic data is the 150 Tomato Genome ReSequencing Project (http://www.tomatogenome.net/). We first focused on the promoter region of eight important glycosyltransferase genes, involved in the flavonoid synthase pathway, since we got clues that regulative differences are involved rather than protein structure differences (Muir et al. 2001; Bovy et al. 2002; Adato et al. 2009). We considered the promoter region the region between the start of a gene of interest up to 1000 bp upstream.

168 polymorphism compared to the reference genome Heinz in the promoter region of the eight glycosyltransferase genes were found in 54 S. lycopersicum tomato accessions. The phenotypic database contains the quantitative metabolomic data of all 54 tomato accessions for 129 volatiles, 43 primary metabolites and 124 secondary metabolites. The rank sum test of Kruskal-Wallis with the use of MapQTL 6 software resulted in a lot of positive associations between one of the 168 polymorphism and one of the 296 compounds. Therefore we needed to take some measures.

The biggest problem was polymorphism that occurred in only one accession. All positive and negative deviations from the mean were significantly linked to that one particular polymorphism. To reduce this lack of replicates, we only took polymorphisms that occurred in three or more accessions. This led to the reduction of 168 genetic polymorphism to 14 genetic polymorphism. We also cut in the amount of compounds to test. Most compounds are hardly effecting the tomato fruit quality. Therefore we made a selection out of the 296 compounds. We took only 12 volatiles and 48 non-volatile secondary metabolites. No primary metabolites were selected since the candidate genes were not likely to affect their accumulation. The last measure we took to reduce the amount of false positive is to set the threshold Kruskall-Wallis K statistic value for df=1 to 6, where normally a K value of 3.84 is the threshold for P=0.05 significant result according to the Chi-square distribution.

After all restrictions to decrease the chance of a false positive result, we found 51 significant links between a polymorphism and a compound (Chapter 7.4). The 51 found links are distributed over 18 compounds, which indicates that there are several compounds that have multiple significant links. Sometimes the different links refer to polymorphism in the promoter region of the same gene, but in most cases a compound was significantly linked to multiple glycosyl transferase genes.

Taking a closer look at the 18 compounds that were found to be significantly associated to a certain polymorphism in the promoter region in one of the 5 genes, a few interesting associations were found. Associations were found between Quercetin and Kampferol aglycones to promotor polymorphisms in the Twi-1 gene (Solyc01g107820), with different glycosyl-donor-groups. We also found associations to different forms of eugenol and its precursor guaiacol to promotor polymorphism in the GGT13 gene (Solyc03g078240) and the GGT15 gene (Solyc04g081830). Significant links between a esculeosides and tomatines were also found. Most compounds shoed links to mutations in the promoter region of the Twi-1, GGT13 and GG15 genes.

The putative associations found should be studied further with gene expression and reverse genetic analyses, as described in chapter 5 Further Research.

4 Discussion

4.1 Reverse Genetics vs. Forward Genetics

TILLING is used as a tool for studying gene function, but also as a tool for crop improvement (Chen et al. 2014). For crop improvement, a TILLING population is used as a source of new genetic variation. The newly obtained genetic variation can be assessed in two different ways. One could make use of a so called reverse genetic approach or a forward genetic approach. In short, a reverse genetic approach first analyses the genetic variation and then checks whether or not this new genetic variation results in an interesting new phenotype. A forward genetic approach on the contrary starts with analyzing the phenotypes of the mutant population followed by screening the genome for the responsible new derived SNPs.

We performed the reverse genetics approach. We first mapped the potentially interesting mutations obtained by TILLING of candidate genes that could be involved in affecting tomato fruit quality traits. Next step is to test the newly derived phenotypes in the offspring generation. An advantage of such approach is the reduced amount of plants to be grown, since M2 seed of only M1 plants that carry a mutation would be grown and genotyped for this mutation. A big disadvantage of TILLING is the small amount of mutations that have a positive effect on tomato fruit quality traits. With protein prediction software you can already predict if a SNP alters a protein and therefore a phenotype. But how the new phenotype will look like is really uncertain. Potentially no effect will be observed because most mutations are neutral. If an effect will be observed, the chance of altering the phenotype in a positive way is really low as well. In bacteria, it was found that about 1% of the mutations resulted in a selective advantage (Perfeito et al. 2007). Of course, it is not proven that bacteria and plants are comparable in this way. Also the described mutants involve fitness, which is a total different trait than fruit quality. But still, it can be stated that the chance of ending up with a positive effect on tomato fruit quality is low.

A forward genetic approach takes much more space and time and therefore money. One needs to grow the offspring generation of all plants. In this experiment, the M1 generation consists already of 2,432 individual plants. The M2 generation will consists of at least a 20 fold of plants, since we need to compensate for losses due to chimerism and Mendelian segregation. To collect phenotypic data is an extreme amount of work. Also, growing all those plants takes a lot of space. So out of practical reasons, the forward genetic approach is not very promising.

Another problem to which you should deal with when working in a TILLING population, is the high amount of mutations that are present in every plant. Treatment with 1% EMS creates about 1 mutation per 147-1237 kb in the tomato genome (Gady et al. 2009; Minoia et al. 2010). If you find an interesting phenotype in the M2 population, all mutation throughout the genome of that plant are potential causes. But the question: 'which SNP is the real cause to this altered phenotype?' is yet unanswered. The other way around is also the case, when you find an interesting SNP and you want to analyse the phenotype of the plant containing that SNP. Because of all the mutations in that plant, it's phenotype will definitely be altered compared to a reference plant. But it is uncertain which altered trait in the phenotype is really caused by your found SNP.

4.2 The 25 selected genes for Tomato fruit quality

Mutagenesis creates mutations throughout the whole genome at random at high rates. In a population that consists 2,432 plants you end up with a big amount of mutations. To reduce the work, we tried to focus on mutations in only a small part of the genome. 25 genes that are involved in tomato fruit quality are selected and screened. These 25 genes are selected based on research carried in several tomato projects dedicated to tomato fruit quality, such as EU-SOL and CBSG, and from related literature (Table 1). The big advantage of focussing on only a small amount of the genome, is the reduction of work. It is practically impossible to screen the whole genome for mutations. On the other hand, these 25 genes are selected based on experience, but they are just a small amount of genes that are involved in tomato fruit quality. One should keep in mind that the genome of the tomato plants of the M1 generation is full of mutations and by focussing on just 25 genes, a lot of mutations are not taken into account. A lot of mutations in other genes are not used in the experiment, and are therefore 'lost'.

The 25 genes that were selected is just a small selection of important genes for tomato fruit quality. It is argumentative that the tomato fruit quality is not dependent of only 25 genes, but on much more. An

example of a gene that is not part of our experiment is the chalcone flavone isomerase gene (Solyc05g010320)(Hanson et al. 2014). This locus is a QTL for high rutin content in the ripe full tomato fruit. Increasing rutin content in the full tomato fruit would increase the antioxidative effect of a tomato fruit. For this reason the gene could be of importance for tomato fruit quality. This is just an example, but indicates that the screened part of the genome is far from complete and that a lot of potential interesting new derived variation has not been subject of the experiment (yet).

4.3 TILLING for crop improvement assessed, pros and cons of the technique

TILLING has already been used for crop improvement in a lot of different species (see 'Introduction'). It is a very useful way of increasing genetic variation in your germplasm. This variation is inherited stable. It is also easy to crossbreed with plants from your mutant population compared to interspecific crosses with some wild relatives. In a lot of cases, the modern cultivars are so different from a wild relative that breeding is hardly possible. All kinds of high-tech techniques have been developed to widen the possibilities with crossbreeding, like embryo rescue. But making crosses between your modern cultivar and a plant from your mutant population which contains an interesting SNP is much easier. Another advantage of using TILLING for crop improvement is the fact that it is stated to be non-transgenic (Slade et al. 2005). This are two very important advantages that make it easy for a crop breeder to use a mutant population for crop improvement.

Recent improvements in sequencing (NGS) make TILLING again more popular. Mutation detection used to be extremely time consuming. High-throughput NGS make mutation detection extremely efficient and fast. This made the TILLING technique again more promising for crop improvement, since it increased the ease of use. Since mutation detection is not the limiting step in the process anymore, the reverse genetic TILLING approach has become more interesting than the forward genetic approach. It has become easier to genotype a huge amount of samples than to grow a huge amount of plants and phenotype all of them.

On the other hand, the technique also got downsides. Probably the most important downside is the fact that the technique it is really a long shot. The chance of finding a new SNP that improves your phenotype is really low. Due to this small chance, you need to mutagenize a substantial amount of seeds. In our case we mutagenized 2,432 seeds, resulting in an equally large M1 population. Growing all those plants takes a lot of space and time. Isolating DNA from all those plants takes a lot of time. Phenotyping all those plants takes a lot of time. All this effort and money for an uncertain goal. So, you could say that TILLING is a risk and it is uncertain if your effort, time and money, is paid back.

In a crop species where the amount of genetic variation is very low due to intensive breeding programs like in tomatoes, TILLING could be an important way of increasing the genetic variation. It gets harder and harder to push the boundaries of the genetic potential of your modern cultivar. In spite of the effort it takes, TILLING could be a last option to increase plants potential without a transgenic approach

4.4 TILLING in the M1 generation and chimerism

When you use the often used TILLING approach (Figure 3), you see that in most cases the M2 generation is screened for mutations. The M1 generation, the generation of which the seeds are treated with a mutagen, is selfed giving rise to the M2 generation. Via this approach you are for sure that the found mutation is heritable and the mutation is dispersed to all cells in the plant. The big disadvantage is the huge amount of work. Every M1 plant, in our case 2,432 plants, will produce many seeds. So your M2 population is a 10-fold or even a 20-fold bigger than your M1 population.

In our experiment we screened for SNPs in the M1 population. This approach saved a lot of time, space and money. On the other hand, the plants are chimera. The amount of mutated cells in your mutant plant is low, which made it difficult to detect with the used SNP detection methods. Also, if you confirm that a M1 plant contains a specific mutation in which you are interested, it is uncertain that the mutation is transferred to the next generation. Probably not all produced seeds will contain the mutation of interest. So, if you find yourself a mutation, you still need to be lucky that the mutation will show up in the next generation. Due to normal Mendelian segregation you lose some M2 plants, since the mutation in the M1 generation is always heterozygous. But the biggest loss is probably due to chimerism, that your seeds do not contain your mutation of interest, but just contain the WT, unmutated allele.

To get an idea of the amount of cells that will contain the mutant allele compared to cells that contain WT allele, estimations were made. Based on research carried out by Stewart and Dermen (1970), we could expect about 10 % of the cells to be mutated. Since the mutated cells are heterozygous for the mutation, about 5% of the total amount of DNA is supposed to be mutated. If you look at the sequence data for the Twi-1 (B) mutant, you come to other conclusions. In a pool of 64 plants, 1 plant is mutated, which is heterozygous. So you expect 1 out of 64*2=128 chromosomes to contain the mutant allele. This is a frequency of 0.78%. However, observed sequence data were lower. In the first sequence round we found a frequency of 0.57% and in the second sequence round we observed a frequency of 0.34% to be mutated. Based on these sequence data, you expect that about half of the cells is heterozygous mutated and half of the cells is full mutant. But, estimations about chimerism based on sequence data is unreliable. In a pool of 64 plants, it is unlikely that all DNA is covered in equal amounts. So to our opinion, the estimation based on Stewart and Dermen (1970) that about 5% of the DNA in a plant is mutated is the best estimation.

4.5 Suitability of KASP for low frequency SNP detection

We had problems with finding the mutant plant in the pool of 64 using KASP (Figure 8). For Twi-1 (B) and for GMT10 (B) we could find a sample that showed a heterozygous expression. But, when we tried to confirm this result, we could not find the same result for Twi-1 (B). In a second and a third try out no heterozygous showed up. For GGT13 (A) and GGT13 (B), we could not find a heterozygous sample at all. Does this indicate that the mutation really is not there? Two rounds of sequencing showed the same result, these mutations exist. Of course it is possible that the sequence is not correct, but the chance that this same mutation was found two times by accident is really low. This has led to the question whether or not our SNP genotyping technique is suitable for our experiment. Figure 9 shows the KASP result of an experiment where we diluted the amount of mutant DNA. What this experiment has made clear, is the fact that KASP is probably not suitable when the mutant allele is present in really low quantities. When there is 5% or 1% of the mutant allele and 95% respectively 99% of the WT allele, you are not able to discriminate between the samples containing mutant DNA or WT DNA. So it is not possible to distinguish between a WT and a sample that contains chimeristic mutant DNA in low quantities. As discussed before, we expect there to be about 5% of mutant DNA in our mutant plant. Combining these facts make you wonder if the KASP technique was a suitable technique for this project.

There are alternative SNP genotyping techniques available that deal with the low quantity of mutant DNA. One technique that deals with low mutation abundancy is called castPCR™ (http://www.lifetechnologies.com). CastPCR™ blocks the annealing and elongation steps of the allele that you are not interested in (Roma et al. 2013). A second technique that suits the detection of low abundant mutations is ddPCR™ (http://www.bio-rad.com). ddPCR™ enables the absolute quantitation of nucleic acids in a sample. It proved to detect a rare mutant within a 100,000 fold excess of WT background (Hindson et al. 2011). These techniques seem to suit an experiment with low quantities of the mutant allele better.

4.6 *In silico* assessment of existing genetic polymorphism and metabolomic data

This assay is based on an *in silico* statistics experiment. We tried to link genotypic data to phenotypic data. The genotypic data that we used for this experiment were gathered via the 150 Tomato Genome Resequencing Project (www.tomatogenome.net). We focused on the promoter region of some important glycosyltransferase genes, involved in the flavonoid synthase pathway, since we got clues that regulative differences are involved rather than protein structure differences (Muir et al. 2001; Bovy et al. 2002; Adato et al. 2009). Data were collected of 54 *Solanum lycopersicum* accessions. Since *S. lycopersicum* is easily self-pollinated, genetic homogeneity within the modern cultivars is high to increase phenotypic uniformity. As a result, the genome is highly conserved and does not contain a lot of polymorphisms. A consequence of the low amount of polymorphisms, is the low amount of replicates per polymorphism. The low amount of replicates is a downside spoken in statistical terms. In this way it decreases the statistical power. We observed multiple times a SNP that only occurred in one accession. A consequence of a SNP that only occurs in one accession, is that it quickly links to a certain compound. If a SNP occurs in one accession, and that accession accidently got high values for a set of compounds, a statistical analysis will link all those compounds to your SNP. Of course, this linkage could be a correct linkage, however one can argue whether those significant links are real links or just false positives. So the consequence of one SNP in only one accession, is that you end up with a lot of compounds that are linked to that SNP.

The other way around is of course also possible. If a compound is high in quantity in just 1 accession, it is clear that that accession got some interesting genotypic variation that causes this. However when you perform a test statistic, all SNPs that this accession contains, are linked to this compound. In this way you end up with one compound that is linked to a lot of SNPs. It could be the case that one or a set of those SNPs is really causing the high quantity of that compound, but it is also argumentative that a lot of links between a SNP and a compound are false positives.

To overcome above mentioned problems, we needed to have a critical look at all found significant links. What we need is multiple accessions that contain the same SNP and high values. To draw a line, we only used polymorphism that occurred in at least 3 accessions. We are aware of the fact of the chance of deleting true links. Nevertheless we believe that this approach is the best approach, because the initial amount of significant links that we found was unrealistic high. We reduced from 168 polymorphism to 14 polymorphism, an enormous reduction. There were 18 compounds linked to these 14 polymorphism dispersed over 5 genes, which is a more realistic amount.

The most interesting associations that were found were with Quercetin, Kampferol and guaiacol. Both, Quercetin and Kampferol, are flavonoids which are thought to be health-promoting (Duthie and Crozier 2000). Eugenol and guaiacol tend to be health-promoting since they have anti-bacterial activity and slow down tumor cell growth (Aggarwal and Shishodia 2006; Burt 2004; Sakihama et al. 2002). Next, we also found strong associations with several esculeosides. A study has shown that esculeosides got beneficial activities in the human body including anti-osteoporosis, anti-menopausal disorder and anti-tumor activities (Manabe et al. 2011). The last group of important tomato secondary metabolites to which we found an association with are the tomatines. Tomatines also tend to have beneficial health promoting effects (Friedman 2013). Most associations were found with the *GGT15* and *Twi1* genes.

This association study was very quick and dirty. None of the found associations between a certain polymorphism and a compound are on forehand true associations. But these *in silico* found results could be a start for further research. We know in which accessions we can find which genetic polymorphism. If you cross the accessions which contains the genetic polymorphism of interest with an accession that does not contain the polymorphism of interest, your polymorphism of interest will segregate in the offspring population. SNP genotyping the offspring population, e.g. with the use of KASP, and collecting phenotypic data could give confirmation of the influence of that one particular polymorphism on the plant's phenotype.

5 Further Research

This TILLING project is an enormous investment and the project is still in the beginning stages. Therefore, there is still a lot to do.

1. Detect the one Twi-1(B) mutant plant within the pool of 64 plants that contains the mutant sequence.
    a. Sequence the most interesting colonies to get full confirmation
    b. If step a did not work: Dilute the colony DNA, since there is probably too much DNA present. Because there is too much DNA present, the primers get saturated in the KASP assay
    c. Perform a new KASP assay with diluted amounts of colony DNA
    d. Sequence the most interesting colonies to get full confirmation

2. Detect the one mutant plant within the pool of 64 plants that contains the mutant sequence for the other 7 mutations found after two sequencing experiments.
    a. Because KASP showed to be not very efficient in detecting low quantities of mutant DNA in a solution, you could make use of the castPCR™ or the ddPCR™ technique. These two mutation detection techniques are especially designed for the detection of low mutation quantities
    b. Sequence the mutated genes of the most interesting plants to get full confirmation

3. Once you have determined which M1 plants contain the mutation in the gene of interest, you grow the M2 generation of that plant to get homozygous mutant plants.
    a. Grow a large amount (at least 50 if possible) of M2 seeds, seeds obtained by selfing the M1 plant. Due to chimerism, about 95% of the M2 plants will not contain the mutation. The residual 5% of the seeds will segregate following Mendelian laws. This means that only 25% of the 5% is homozygous mutated.
    b. Genotype all M2 plants for the mutation of interest. A normal KASP assay can be used, because in the M2 population there are no problems regarding chimerism.
    c. Select the plants that are homozygous mutants.
    d. If you did not receive homozygous mutants, self the heterozygote M2 plants and grow the M3 generation. In this case it is not necessary to grow a large amount of seeds, since 25% of the M3 generation derived from a heterozygous M2 plant is homozygous mutant. Then repeat steps b and c.

4. Phenotype the homozygous mutant plants and link the *de novo* mutation to a phenotypic trait.
    a. Phenotype the homozygous mutant plant extensively. Because you do not know exactly what the result of the mutation will be, you need to phenotype all possible traits of interest.
    b. Make sure that several homozygous mutant plants are phenotyped to increase your statistical power.
    c. Simple test statistics (student's T-test) could be sufficient to determine whether the mutation of interest has a significant effect on the plant's phenotype.

5. Test mutations other than in the 25 selected genes for tomato fruit quality.
    a. The mutation population contains mutations all over the genome. In this project we only focused on 25 tomato fruit quality related genes, but of course there are many more mutations in other loci that can be of interest to a breeder. Sequencing would provide you the genotypic data. The particular SNP can function as a marker. The next step is to check whether or not your phenotypic trait co-segregates with your marker/SNP

The ultimate goal of the TILLING project is to increase the variation available in the existing germplasm. If you are able to find a mutation and link it to a specific phenotypic trait, a breeder could use it to improve its own modern cultivar. So, the TILLING population could be a source of new breeding material to improve modern cultivars.

6 References

Adato, A., T. Mandel, S. Mintz-Oron, I. Venger, D. Levy, M. Yativ, E. Dominguez, et al. 2009. Fruit-Surface Flavonoid Accumulation in Tomato Is Controlled by a SIMYB12-Regulated Transcriptional Network. *Plos Genetics* 5 (12): 23. http://dx.doi.org/10.1371/journal.pgen.1000777.

Aggarwal, B. B., and S. Shishodia. 2006. Molecular targets of dietary agents for prevention and therapy of cancer. *Biochemical Pharmacology* 71 (10): 1397-1421. http://dx.doi.org/10.1016/j.bcp.2006.02.009.

Ahloowalia, B. S., M. Maluszynski, and K. Nichterlein. 2004. Global impact of mutation-derived varieties. *Euphytica* 135 (2): 187-204. http://dx.doi.org/10.1023/B:EUPH.0000014914.85465.4f.

Amsterdam, A., and N. Hopkins. 2006. Mutagenesis strategies in zebrafish for identifying genes involved in development and disease. *Trends in Genetics* 22 (9): 473-478. http://dx.doi.org/10.1016/j.tig.2006.06.011.

Baldwin, E. A., J. W. Scott, C. K. Shewmaker, and W. Schuch. 2000. Flavor trivia and tomato aroma: Biochemistry and possible mechanisms for control of important aroma components. *Hortscience* 35 (6): 1013-1022.

Birtic, S., C. Ginies, M. Causse, C. Renard, and D. Page. 2009. Changes in Volatiles and Glycosides during Fruit Maturation of Two Contrasted Tomato (Solanum lycopersicum) Lines. *Journal of Agricultural and Food Chemistry* 57 (2): 591-598. http://dx.doi.org/10.1021/jf8023062.

Bovy, A., R. de Vos, M. Kemper, E. Schijlen, M. A. Pertejo, S. Muir, G. Collins, et al. 2002. High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. *Plant Cell* 14 (10): 2509-2526. http://dx.doi.org/10.1105/tpc.004218.

Burt, S. 2004. Essential oils: their antibacterial properties and potential applications in foods - a review. *International Journal of Food Microbiology* 94 (3): 223-253. http://dx.doi.org/10.1016/j.ijfoodmicro.2004.03.022.

Chen, L., L. G. Hao, M. A. J. Parry, A. L. Phillips, and Y. G. Hu. 2014. Progress in TILLING as a tool for functional genomics and improvement of crops. *Journal of Integrative Plant Biology* 56 (5): 425-443. http://dx.doi.org/10.1111/jipb.12192.

Choi, S. U., S. Y. Ryu, S. K. Yoon, N. P. Jung, S. H. Park, K. H. Kim, E. J. Choi, and C. O. Lee. 1999. Effects of flavonoids on the growth and cell cycle of cancer cells. *Anticancer Research* 19 (6B): 5229-5233.

Dalmais, M., J. Schmidt, C. Le Signor, F. Moussy, J. Burstin, V. Savois, G. Aubert, et al. 2008. UTILLdb, a Pisum sativum in silico forward and reverse genetics tool. *Genome Biology* 9 (2): 12. http://dx.doi.org/10.1186/gb-2008-9-2-r43.

Dong, C. M., J. Dalton-Morgan, K. Vincent, and P. Sharp. 2009. A Modified TILLING Method for Wheat Breeding. *Plant Genome* 2 (1): 39-47. http://dx.doi.org/10.3835/plantgenome2008.10.0012.

Duthie, G., and A. Crozier. 2000. Plant-derived phenolic antioxidants. *Current Opinion in Lipidology* 11 (1): 43-47. http://dx.doi.org/10.1097/00041433-200002000-00007.

FAOSTAT. 2012. Food and Agriculture Organisation of the United Nations, Statistical Database. In http://www.faostat3.fao.org.

Friedman, M. 2013. Anticarcinogenic, Cardioprotective, and Other Health Benefits of Tomato Compounds Lycopene, alpha-Tomatine, and Tomatidine in Pure Form and in Fresh and Processed Tomatoes. *Journal of Agricultural and Food Chemistry* 61 (40): 9534-9550. http://dx.doi.org/10.1021/jf402654e.

Gachon, C. M. M., M. Langlois-Meurinne, and P. Saindrenan. 2005. Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends in Plant Science* 10 (11): 542-549. http://dx.doi.org/10.1016/j.tplants.2005.09.007.

Gady, A. L. F., F. W. K. Hermans, M. Van de Wal, E. N. van Loo, R. G. F. Visser, and C. W. B. Bachem. 2009. Implementation of two high through-put techniques in a novel application: detecting point mutations in large EMS mutated plant populations. *Plant Methods* 5: 14. http://dx.doi.org/10.1186/1746-4811-5-13.

Gilchrist, E. J., N. J. O'Neil, A. M. Rose, M. C. Zetka, and G. W. Haughn. 2006. TILLING is an effective reverse genetics technique for Caenorhabditis elegans. *Bmc Genomics* 7: 16. http://dx.doi.org/10.1186/1471-2164-7-262.

Greene, E. A., C. A. Codomo, N. E. Taylor, J. G. Henikoff, B. J. Till, S. H. Reynolds, L. C. Enns, et al. 2003. Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in arabidopsis. *Genetics* 164 (2): 731-740.

Hanson, P., R. Schafleitner, S. M. Huang, C. W. Tan, D. Ledesma, and R. Y. Yang. 2014. Characterization and mapping of a QTL derived from Solanum habrochaites associated with elevated rutin content (quercetin-3-rutinoside) in tomato. *Euphytica* 200 (3): 441-454. http://dx.doi.org/10.1007/s10681-014-1180-7.

Hertog, M. G. L., and P. C. H. Hollman. 1996. Potential health effects of the dietary flavonol quercetin. *European Journal of Clinical Nutrition* 50 (2): 63-71.

Hindson, B. J., K. D. Ness, D. A. Masquelier, P. Belgrader, N. J. Heredia, A. J. Makarewicz, I. J. Bright, et al. 2011. High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry* 83 (22): 8604-8610. http://dx.doi.org/10.1021/ac202028g.

Hollman, P. C. H., and M. B. Katan. 1999. Dietary flavonoids: Intake, health effects and bioavailability. *Food and Chemical Toxicology* 37 (9-10): 937-942. http://dx.doi.org/10.1016/s0278-6915(99)00079-4.

Hunt, G. M., and E. A. Baker. 1980. PHENOLIC CONSTITUENTS OF TOMATO FRUIT CUTICLES. *Phytochemistry* 19 (7): 1415-1419. http://dx.doi.org/10.1016/0031-9422(80)80185-3.

Jander, G., S. R. Baerson, J. A. Hudak, K. A. Gonzalez, K. J. Gruys, and R. L. Last. 2003. Ethylmethanesulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. *Plant Physiology* 131 (1): 139-146. http://dx.doi.org/10.1104/pp.102.010397.

Janssen, K., R. P. Mensink, F. J. Cox, J. L. Harryvan, R. Hovenier, P. C. Hollman, and M. B. Katan. 1998. Effects of the flavonoids quercetin and apigenin on hemostasis in healthy volunteers: results from an in vitro and a dietary supplement study. *The American journal of clinical nutrition* 67 (2): 255-62.
Kumpatla, S. P., R. Buyyarapu, I. Y. Abdurakhmonov, and J. A. Mammadov. 2012. "Genomics-Assisted Plant Breeding in the 21st Century: Technological Advances and Progress." In *Plant Breeding*.
Lamour, K. H., L. Finley, O. Hurtado-Gonzales, D. Gobena, M. Tierney, and H. J. G. Meijer. 2006. Targeted gene mutation in Phytophthora spp. *Molecular Plant-Microbe Interactions* 19 (12): 1359-1367. http://dx.doi.org/10.1094/mpmi-19-1359.
Laux, T., and K. F. X. Mayer. 1998. Cell fate regulation in the shoot meristem. *Seminars in Cell & Developmental Biology* 9 (2): 195-200. http://dx.doi.org/10.1006/scdb.1997.0208.
Le Gall, G., M. S. DuPont, F. A. Mellon, A. L. Davis, G. J. Collins, M. E. Verhoeyen, and I. J. Colquhoun. 2003. Characterization and content of flavonoid glycosides in genetically modified tomato (Lycopersicon esculentum) fruits. *Journal of Agricultural and Food Chemistry* 51 (9): 2438-2446. http://dx.doi.org/10.1021/jf025995e.
Liew, M., R. Pryor, R. Palais, C. Meadows, M. Erali, E. Lyon, and C. Wittwer. 2004. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical Chemistry* 50 (7): 1156-1164. http://dx.doi.org/10.1373/clinchem.2004.032136.
Livak, K. J. 1999. Allelic discrimination using fluorogenic probes and the 5 ' nuclease assay. *Genetic Analysis-Biomolecular Engineering* 14 (5-6): 143-149. http://dx.doi.org/10.1016/s1050-3862(98)00019-9.
Manabe, H., Y. Murakami, M. El-Aasr, T. Ikeda, Y. Fujiwara, M. Ono, and T. Nohara. 2011. Content variations of the tomato saponin esculeoside A in various processed tomatoes. *Journal of Natural Medicines* 65 (1): 176-179. http://dx.doi.org/10.1007/s11418-010-0443-4.
Marlatt, C., C. T. Ho, and M. J. Chien. 1992. STUDIES OF AROMA CONSTITUENTS BOUND AS GLYCOSIDES IN TOMATO. *Journal of Agricultural and Food Chemistry* 40 (2): 249-252. http://dx.doi.org/10.1021/jf00014a016.
Mayer, K. F. X., H. Schoof, A. Haecker, M. Lenhard, G. Jurgens, and T. Laux. 1998. Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem. *Cell* 95 (6): 805-815. http://dx.doi.org/10.1016/s0092-8674(00)81703-1.
Mba, C. 2013. Induced Mutations Unleash the Potentials of Plant Genetic Resources for Food and Agriculture. *Agronomy* 3 (1): 200-231.
Menda, N., Y. Semel, D. Peled, Y. Eshed, and D. Zamir. 2004. In silico screening of a saturated mutation library of tomato. *Plant Journal* 38 (5): 861-872. http://dx.doi.org/10.1111/j.1365-313X.2004.02088.x.
Minoia, S., A. Petrozza, O. D'Onofrio, F. Piron, G. Mosca, G. Sozio, F. Cellini, A. Bendahmane, and F. Carriero. 2010. A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC research notes* 3: 69. http://dx.doi.org/10.1186/1756-0500-3-69.
Muir, S. R., G. J. Collins, S. Robinson, S. Hughes, A. Bovy, C. H. R. De Vos, A. J. van Tunen, and M. E. Verhoeyen. 2001. Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nature Biotechnology* 19 (5): 470-474. http://dx.doi.org/10.1038/88150.
Ness, A. R., and J. W. Powles. 1997. Fruit and vegetables, and cardiovascular disease: A review. *International Journal of Epidemiology* 26 (1): 1-13. http://dx.doi.org/10.1093/ije/26.1.1.
Noble, A. C. 1996. Taste-aroma interactions. *Trends in Food Science & Technology* 7 (12): 439-444. http://dx.doi.org/10.1016/s0924-2244(96)10044-3.
O'Donnell, P. J., M. R. Truesdale, C. M. Calvert, A. Dorans, M. R. Roberts, and D. J. Bowles. 1998. A novel tomato gene that rapidly responds to wound- and pathogen-related signals. *Plant Journal* 14 (1): 137-142. http://dx.doi.org/10.1046/j.1365-313X.1998.00110.x.
Oleykowski, C. A., C. R. B. Mullins, A. K. Godwin, and A. T. Yeung. 1998. Mutation detection using a novel plant endonuclease. *Nucleic Acids Research* 26 (20): 4597-4602. http://dx.doi.org/10.1093/nar/26.20.4597.
Ortiz-Serrano, P., and J. V. Gil. 2007. Quantitation of free and glycosidically bound volatiles in and effect of glycosidase addition on three tomato varieties (Solanum lycopersicum L.). *Journal of Agricultural and Food Chemistry* 55 (22): 9170-9176. http://dx.doi.org/10.1021/jf0715673.
Parry, M. A. J., P. J. Madgwick, C. Bayon, K. Tearall, A. Hernandez-Lopez, M. Baudo, M. Rakszegi, et al. 2009. Mutation discovery for crop improvement. *Journal of Experimental Botany* 60 (10): 2817-2825. http://dx.doi.org/10.1093/jxb/erp189.
Perfeito, L., L. Fernandes, C. Mota, and I. Gordo. 2007. Adaptive mutations in bacteria: High rate and small effects. *Science* 317 (5839): 813-815. http://dx.doi.org/10.1126/science.1142284.
Piron, F., M. Nicolai, S. Minoia, E. Piednoir, A. Moretti, A. Salgues, D. Zamir, C. Caranta, and A. Bendahmane. 2010. An Induced Mutation in Tomato eIF4E Leads to Immunity to Two Potyviruses. *Plos One* 5 (6): 10. http://dx.doi.org/10.1371/journal.pone.0011313.
Roma, C., C. Esposito, A. M. Rachiglio, R. Pasquale, A. Iannaccone, N. Chicchinelli, R. Franco, et al. 2013. Detection of EGFR Mutations by TaqMan Mutation Detection Assays Powered by Competitive Allele-Specific TaqMan PCR Technology. *Biomed Research International*: 9. http://dx.doi.org/10.1155/2013/385087.
Rozanski, A., J. A. Blumenthal, and J. Kaplan. 1999. Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Circulation* 99 (16): 2192-2217.
Saito, T., T. Ariizumi, Y. Okabe, E. Asamizu, K. Hiwasa-Tanase, N. Fukuda, T. Mizoguchi, et al. 2011. TOMATOMA: A Novel Tomato Mutant Database Distributing Micro-Tom Mutant Collections. *Plant and Cell Physiology* 52 (2): 283-296. http://dx.doi.org/10.1093/pcp/pcr004.
Sakihama, Y., M. F. Cohen, S. C. Grace, and H. Yamasaki. 2002. Plant phenolic antioxidant and prooxidant activities: phenolics-induced oxidative damage mediated by metals in plants. *Toxicology* 177 (1): 67-80. http://dx.doi.org/10.1016/s0300-483x(02)00196-8.

Sarrion-Perdigones, A., M. Vazquez-Vilar, J. Palaci, B. Castelijns, J. Forment, P. Ziarsolo, J. Blanca, A. Granell, and D. Orzaez. 2013. GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *Plant Physiology* 162 (3): 1618-1631. http://dx.doi.org/10.1104/pp.113.217661.
Schijlen, E., C. H. R. de Vos, S. Martens, H. H. Jonker, F. M. Rosin, J. W. Molthoff, Y. M. Tikunov, et al. 2007. RNA interference silencing of Chalcone synthase, the first step in the flavonoid biosynthesis pathway, leads to parthenocarpic tomato fruits. *Plant Physiology* 144 (3): 1520-1530. http://dx.doi.org/10.1104/pp.107.100305.
Semagn, K., R. Babu, S. Hearne, and M. Olsen. 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding* 33 (1): 1-14. http://dx.doi.org/10.1007/s11032-013-9917-x.
Slade, A. J., S. I. Fuerstenberg, D. Loeffler, M. N. Steine, and D. Facciotti. 2005. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nature Biotechnology* 23 (1): 75-81. http://dx.doi.org/10.1038/nbt1043.
Slade, A. J., C. McGuire, D. Loeffler, J. Mullenberg, W. Skinner, G. Fazio, A. Holm, et al. 2012. Development of high amylose wheat through TILLING. *Bmc Plant Biology* 12. http://dx.doi.org/10.1186/1471-2229-12-69.
Steinmetz, K. A., and J. D. Potter. 1996. Vegetables, fruit, and cancer prevention: A review. *Journal of the American Dietetic Association* 96 (10): 1027-1039. http://dx.doi.org/10.1016/s0002-8223(96)00273-8.
Stewart, A. J., S. Bozonnet, W. Mullen, G. I. Jenkins, M. E. J. Lean, and A. Crozier. 2000. Occurrence of flavonols in tomatoes and tomato-based products. *Journal of Agricultural and Food Chemistry* 48 (7): 2663-2669. http://dx.doi.org/10.1021/jf000070p.
Stewart, R. N., and H. Dermen. 1970. DETERMINATION OF NUMBER AND MITOTIC ACTIVITY OF SHOOT APICAL INITIAL CELLS BY ANALYSIS OF MERICLINAL CHIMERAS. *American Journal of Botany* 57 (7): 816-&. http://dx.doi.org/10.2307/2441339.
Tieman, D., P. Bliss, L. M. McIntyre, A. Blandon-Ubeda, D. Bies, A. Z. Odabasi, G. R. Rodriguez, et al. 2012. The Chemical Interactions Underlying Tomato Flavor Preferences. *Current Biology* 22 (11): 1035-1039. http://dx.doi.org/10.1016/j.cub.2012.04.016.
Tikunov, Y., A. Lommen, C. H. R. de Vos, H. A. Verhoeven, R. J. Bino, R. D. Hall, and A. G. Bovy. 2005. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiology* 139 (3): 1125-1137. http://dx.doi.org/10.1104/pp.105.068130.
Tikunov, Y. M., J. Molthoff, R. C. H. de Vos, J. Beekwilder, A. van Houwelingen, J. J. J. van der Hooft, M. Nijenhuis-de Vries, et al. 2013. NON-SMOKY GLYCOSYLTRANSFERASE1 Prevents the Release of Smoky Aroma from Tomato Fruit. *Plant Cell* 25 (8): 3067-3078. http://dx.doi.org/10.1105/tpc.113.114231.
Till, B. J., T. Colbert, R. Tompa, L. C. Enns, C. A. Codomo, J. E. Johnson, S. H. Reynolds, et al. 2003. High-throughput TILLING for functional genomics. *Plant Functional Genomics* 236: 205-220.
Till, B. J., J. Cooper, T. H. Tai, P. Colowit, E. A. Greene, S. Henikoff, and L. Comai. 2007. Discovery of chemically induced mutations in rice by TILLING. *Bmc Plant Biology* 7: 12. http://dx.doi.org/10.1186/1471-2229-7-19.
Till, B. J., S. H. Reynolds, E. A. Greene, C. A. Codomo, L. C. Enns, J. E. Johnson, C. Burtner, et al. 2003. Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Research* 13 (3): 524-530. http://dx.doi.org/10.1101/gr.977903.
Till, B. J., S. H. Reynolds, C. Weil, N. Springer, C. Burtner, K. Young, E. Bowers, et al. 2004. Discovery of induced point mutations in maize genes by TILLING. *BMC plant biology* 4: 12. http://dx.doi.org/10.1186/1471-2229-4-12.
Tuinen, A. v., E. H. Schipper, U. Kaur, A. Rochette, R. G. Meijer-Dekens, and Y. Bay. 2012. An EMS population of tomato cultivar Micro-Tom. In search of durable disease resistance. Wageningen University and Research Centre, laboratory of Plant Breeding.
Uauy, C., F. Paraiso, P. Colasuonno, R. K. Tran, H. Tsai, S. Berardi, L. Comai, and J. Dubcovsky. 2009. A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *Bmc Plant Biology* 9: 14. http://dx.doi.org/10.1186/1471-2229-9-115.
Van Ooijen, J. W. 2009. MapQTL 6, Software for the mapping of quantitative trait loci in experimetnal populations of diploid species. Wageningen, Netherlands: Kyazma B.V.
Verhoeyen, M. E., A. Bovy, G. Collins, S. Muir, S. Robinson, C. H. R. de Vos, and S. Colliver. 2002. Increasing antioxidant levels in tomatoes through modification of the flavonoid biosynthetic pathway. *Journal of Experimental Botany* 53 (377): 2099-2106. http://dx.doi.org/10.1093/jxb/erf026.
Wang, T. L., C. Uauy, F. Robson, and B. Till. 2012. TILLING in extremis. *Plant Biotechnology Journal* 10 (7): 761-772. http://dx.doi.org/10.1111/j.1467-7652.2012.00708.x.
Willits, M. G., C. M. Kramer, R. T. N. Prata, V. De Luca, B. G. Potter, J. C. Steffens, and G. Graser. 2005. Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. *Journal of Agricultural and Food Chemistry* 53 (4): 1231-1236. http://dx.doi.org/10.1021/jf049355i.
Winkler, S., A. Schwabedissen, D. Backasch, C. Bokel, C. Seidel, S. Bonisch, M. Furthauer, et al. 2005. Target-selected mutant screen by TILLING in Drosophila. *Genome Research* 15 (5): 718-723. http://dx.doi.org/10.1101/gr.3721805.
Wu, J. L., C. J. Wu, C. L. Lei, M. Baraoidan, A. Bordeos, M. R. S. Madamba, M. Ramos-Pamplona, et al. 2005. Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Molecular Biology* 59 (1): 85-97. http://dx.doi.org/10.1007/s11103-004-5112-0.
Xin, Z. G., M. L. Wang, N. A. Barkley, G. Burow, C. Franks, G. Pederson, and J. Burke. 2008. Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *Bmc Plant Biology* 8: 14. http://dx.doi.org/10.1186/1471-2229-8-103.

7 Appendix

7.1 Primers for Resequencing

#	Gene	Primer combination	name	position	sequence	Primer length	GC%	Tm	amplicon size (bp)
1	FGGT11	8	For1	26	GTATACTTATGTTACCATGGCTAGC	25	40	60	1292
			Rev1	1318	GAGCCATCAACTCTTCCA	18	50	58	
2	GGT13 (A)	9	For1	31	GCAATGTTACCATGGCTAGC	20	50	61	1505
			Rev1	1536	GCATTTCAACAGAATCCCTCC	21	47	61	
3	GGT13 (B)	9	For1	31	GCAATGTTACCATGGCTAGC	20	50	61	1505
			Rev1	1536	GCATTTCAACAGAATCCCTCC	21	47	61	
4	GGT19	11	For1	20	GGGAGTTTTTCCATCATGGAG	21	47	60	1364
			Rev1	1384	GGTAGAAAGGTCCAACAACT	20	45	59	
5	GMT10 (A)	12	For1	3	GGCTGAGAAGAATGGTGAAG	20	50	60	1022
			Rev1	1025	GGATAGCAAGGGCTGTAGCA	20	55	65	
6	GMT10 (B)	12	For1	3	GGCTGAGAAGAATGGTGAAG	20	50	60	1022
			Rev1	1025	GGATAGCAAGGGCTGTAGCA	20	55	65	
7	MalT3	17	For1	21	ATCAAGCTGAAGTTGCTGAG	20	45	60	1122
			Rev1	1143	CATACCTTCAAGAAAAGAGCC	21	42	58	
8	Twi-1 (A)	18	For1	21	CTTCTTTCCCATGATGGCTC	21	50	60	1366
			Rev1	1387	TCAAAGTAGCCCATCCATTG	20	45	59	
9	Twi-1 (B)	18	For1	21	CTTCTTTCCCATGATGGCTC	21	50	60	1366
			Rev1	1387	TCAAAGTAGCCCATCCATTG	20	45	59	
10	GAE1	19	For1	97	AGTTATGCCTTCGTTAGAGGA	21	42	60	1278
			Rev1	1375	CAACTTTCCTTGATCATAGCC	21	42	58	
11	GAME5	21	For1	118	ACAGAGTGTAGTACTGGTGC	20	50	60	1540
			Rev1	1658	GATACACTTGCTTGCTATGTGG	22	45	61	
13	MET1 (B)	30	For1	13826	CTGATTGCTGCTGCTCTTC	19	52	60	1258
			Rev1	15084	AGAATGTGTGTGTGTACCTG	20	45	59	
14	RHGL	38	For1	6923	CATGGATTTGATTGATTGTGGCG	23	43	63	1069
			Rev1	7992	TTTGTTGTGGAGGTCCTTCC	20	50	61	
15	RIN	39	FOR1	44	AATGGGCCCTCCACGACACT	20	60	68	1099
			REV1	1143	ACATATGCGATCACAGATACC	21	42	59	
16	Zinc Finger	47	FOR1	324	TGGAGTTATACGGTGGGTCACA	22	50	64	1839
			REV1	2163	CATTACCGAGTAAGTGCTCC	20	50	59	
17	MET1 (A) alt.	29A	For1	12152	ATTCAGGTATAGGCTTGGGA	20	45	59	1524
			Rev1	13676	GTATTAGGTAGCACATACCTCC	22	45	59	

7.2 quantity of the amplified DNA, used for resequencing. Total volume of the sample to sequence is indicated at the right bottom of the table

#	Gene	Plant pool	Primer combination	Amount of DNA dilution(µl)	concentration DNA (ng/µl)	Total amount of DNA (ng)	DNA dilution to be added (%)	DNA dilution to be added (µl)
1	FGGT11	B5	8	45.6	9.670	440.85	0.10	4.6
2	GGT13 (A)	B2	9	50.0	25.170	1258.5	0.04	1.8
3	GGT13 (B)	D1	9	36.9	3.250	119.85	0.37	13.6
4	GGT19	A1	11	32.2	2.390	76.85	0.58	18.6
5	GMT10 (A)	A6	12	47.6	17.600	837.35	0.05	2.5
6	GMT10 (B)	D2	12	50.0	26.200	1310	0.03	1.7
7	MalT3	E1	17	44.3	7.470	330.85	0.13	5.9
8	Twi-1 (A)	A2	18	44.8	8.130	363.85	0.12	5.5
9	Twi-1 (B)	E1	18	42.5	5.650	239.85	0.18	7.8
10	GAE1	E2	19	50.0	28.840	1442	0.03	1.5
11	GAME5	D3	21	26.0	1.780	46.35	0.96	24.9
13	MET1 (B)	E2	30	38.5	3.710	142.85	0.31	12.0
14	RHGL	E5	38	25.5	1.740	44.35	1.00	25.5
15	RIN	E1	39	41.8	5.210	217.85	0.20	8.5
16	Zinc Finger	F2	47	33.8	2.630	88.85	0.50	16.9
17	MET1 (A) alt.	D6	29A	39.8	4.170	165.85	0.27	10.6
							Total:	161.9

7.3 Primers for KASP assay

#	Gene	Name	Startlocation	Sequence	Allele specific primer sequence (black)	GC%	Tm	Amplicon size (bp)
1	GAE1	For1	852	GAAGGTGACCAAGTTCATGCTCGGCGGAATGTCGGTGTTAGTTAC	24	54	59	85
		For2	852	GAAGGTCGGAGTCAACGGATTCGGCGGAATGTCGGTGTTAGTTAT	24	50	57	
		Rev1	937	CGATGTAAGTAAAATCCCGAGC	22	45	60	
2	GGT13 (A)	For1	840	GAAGGTGACCAAGTTCATGCTGTGGGTAATTAGATTCCCAGTTGGAC	26	42	58	86
		For2	840	GAAGGTCGGAGTCAACGGATTGTGGGTAATTAGATTCCCAGTTGGAT	26	38	56	
		Rev1	926	CCTCTGTTTCCAACTCTTTCTAGAAAACCC	30	43	60	
3	GGT13 (B)	For1	763	GAAGGTGACCAAGTTCATGCTGGGACTGAATATTTCCTATC	20	40	48	81
		For2	763	GAAGGTCGGAGTCAACGGATTGGGACTGAATATTTCCTATT	20	35	46	
		Rev1	844	CCCACAAGAAATTCACCTTGC	21	48	52	
4	GGT19	For1	545	GAAGGTGACCAAGTTCATGCTCCGCGGAAATGAACATTTCCATAGGACC	28	50	61	81
		For2	545	GAAGGTCGGAGTCAACGGATTCCGCGGAAATGAACATTTCCATAGGACT	28	46	60	
		Rev1	626	GTGTTGATCAAAATGATGTCATGGG	25	40	54	
5	MET1 (A)	For1	1945	GAAGGTGACCAAGTTCATGCTCCTAGAGCATCTATGGCGCCATC	23	57	59	81
		For2	1945	GAAGGTCGGAGTCAACGGATTCCTAGAGCATCTATGGCGCCATT	23	52	57	
		Rev1	2026	CCCCCCAGATTCTGTTGATTAGCC	24	54	59	
6	GMT10 (B)	For1	112	GAAGGTGACCAAGTTCATGCTCCTAGAGAGCCAGAATCCATGAAAG	25	48	58	89
		For2	112	GAAGGTCGGAGTCAACGGATTCCTAGAGAGCCAGAATCCATGAAAA	25	44	56	
		Rev1	201	CTGACCTTCATCTGCTGATGTTGTC	25	48	58	
7	Twi-1 (A)	For1	776	GAAGGTGACCAAGTTCATGCTGGAAATCATCTATCGATGAACACGCGTGC	29	48	62	92
		For2	776	GAAGGTCGGAGTCAACGGATTGGAAATCATCTATCGATGAACACGCGTGT	29	45	60	
		Rev1	868	CTGCTGTACTTCCAAAACAAAC	22	41	51	
8	Twi-1 (B)	For1	1254	GAAGGTGACCAAGTTCATGCTGGTTTAATCATAAGAGGATGGGCACCCC	28	50	61	83
		For2	1254	GAAGGTCGGAGTCAACGGATTGGTTTAATCATAAGAGGATGGGCACCCT	28	46	60	
		Rev1	1337	CCATCCACAATGAGTAACAAAAGCTCC	27	44	58	

7.4 All found significant associations between reliable polymorphisms and compounds of interest to the quality of the tomato fruit

#	#	Compound name	SolyID	Gene Name	position (kb)	SNP #	Reference	Alternative	K*	Df	Signif.	Nr inf.	Meanrank-1	Mean-a	Meanrank-h	Mean-h	Meanrank-b	Mean-b	Nr-h+Nr-b
1	1	Benzyl alcohol-hexose-pentose	Solyc04g081830	GGT15	63.314	71	G	A	8.39	1	****	50	23.7	631.834			45.8	1655.31	4
2		Benzyl alcohol-hexose-pentose	Solyc01g107820	Twi-1	86.996	81	A	G	7.33	2	**	50	23.7	632.249	36	826.535	43.5	1601.97	5
3		Benzyl alcohol-hexose-pentose	Solyc01g107820	Twi-1	86.996	83	G	T	7.33	2	**	50	23.7	632.249	36	826.535	43.5	1601.97	5
4		Benzyl alcohol-hexose-pentose	Solyc11g007460	GGT6	1.754	17	A	G	8.94	2	**	50	22.8	626.911	43	1581.62	38.3	1032.02	8
5	2	beta-tomatine II FA	Solyc03g078240	GGT13	43.776	34	A	C	10.4	1	****	50	23	142.069			43.5	698.219	6
6	3	Dehydroesculeoside A (FA)	Solyc03g078240	GGT13	43.776	34	A	C	8.91	1	****	50	23.2	3710.33			42.2	16006	6
7		Dehydroesculeoside A (FA)	Solyc01g107820	Twi-1	86.996	81	A	G	7.59	2	**	50	23.8	4235.44	25	2515.5	44.8	16545.1	5
8		Dehydroesculeoside A (FA)	Solyc01g107820	Twi-1	86.996	83	G	T	7.59	2	**	50	23.8	4235.44	25	2515.5	44.8	16545.1	5
9	4	Dehydrotomatine (S) I FA	Solyc09g059170	GGT1	49.174	158	C	T,G	7.51	2	**	50	27.4	39.6468	3	19	15.5	21.5	6
10	5	Esculeoside B (b)	Solyc03g078240	GGT13	43.776	34	A	C	9.86	1	****	50	23.1	61.6486			43	183.44	6
11		Esculeoside B (b)	Solyc04g081830	GGT15	63.314	71	G	A	8.01	1	****	50	23.8	53.3044			45.2	340.295	4
12	6	Eugenol malonyl dihexose-pentose	Solyc03g078240	GGT13	43.776	34	A	C	9.77	1	****	50	23.1	51.4527			42.8	719.36	6
13		Eugenol malonyl dihexose-pentose	Solyc04g081830	GGT15	63.313	58	G	A	9.6	2	***	50	24.8	76.5629	8.8	21	47.3	1030.92	5
14	7	Eugenol xylosyl diglucopyranoside	Solyc04g081830	GGT15	63.313	58	G	A	9.11	2	**	50	24.9	36.2546	7.5	20.5	46	623.596	5
15		Eugenol xylosyl diglucopyranoside	Solyc04g081830	GGT15	63.313	60	A	T	7.01	2	**	50	24.4	35.8795	12.5	22	46	623.596	4
16	8	Guaiacol malonyl xylosyl diglucoside	Solyc01g107820	Twi-1	86.996	81	A	G	7.17	2	**	50	23.7	206.733	42	654.473	42	1662	5
17		Guaiacol malonyl xylosyl diglucoside	Solyc01g107820	Twi-1	86.996	83	G	T	7.17	2	**	50	23.7	206.733	42	654.473	42	1662	5
18		Guaiacol malonyl xylosyl diglucoside	Solyc03g078240	GGT13	43.776	34	A	C	6.19	1	**	50	23.6	238.113			39.3	1021.41	6
19	9	Kaempferol 3-O-rutinoside	Solyc04g081830	GGT15	63.313	58	G	A	9.42	2	***	50	23.4	1266.14	49.5	16336.7	40.3	2812.54	5
20		Kaempferol 3-O-rutinoside	Solyc01g107820	Twi-1	86.995	78	C	T	6.64	2	**	50	23.9	1324.4	45	3720.99	43	11147.9	4
21		Kaempferol 3-O-rutinoside	Solyc01g107820	Twi-1	86.996	81	A	G	8.95	2	**	50	23.4	1266.14	45	3720.99	43.8	9347.49	5
22		Kaempferol 3-O-rutinoside	Solyc01g107820	Twi-1	86.996	83	G	T	8.95	2	**	50	23.4	1266.14	45	3720.99	43.8	9347.49	5
23	10	Kaempferol 3-O-rutinoside-7-O-glucoside	Solyc01g107820	Twi-1	86.996	81	A	G	10.8	2	****	50	23.2	300.317	48	3363.14	45.2	3027.39	5
24		Kaempferol 3-O-rutinoside-7-O-glucoside	Solyc01g107820	Twi-1	86.996	83	G	T	10.8	2	****	50	23.2	300.317	48	3363.14	45.2	3027.39	5
25		Kaempferol 3-O-rutinoside-7-O-glucoside	Solyc04g081830	GGT15	63.313	58	G	A	10.9	2	****	50	23.2	300.317	48.5	5520.03	44	1477.56	5
26		Kaempferol 3-O-rutinoside-7-O-glucoside	Solyc04g081830	GGT15	63.314	71	G	A	8.19	1	****	50	23.8	361.715			45.5	3087.02	4
27		Kaempferol 3-O-rutinoside-7-O-glucoside	Solyc01g107820	Twi-1	86.995	78	C	T	8.83	2	**	50	23.7	308.407	48	3363.14	45.7	3812.37	4
28	11	Kaempferol-hexose-deoxyhexose, -hexose, -C10H8O3 (176)	Solyc01g107820	Twi-1	86.995	78	C	T	6.96	2	**	50	23.9	53.5286	48	173.119	42.3	274.729	4
29		Kaempferol-hexose-deoxyhexose, -hexose, -C10H8O3 (176)	Solyc01g107820	Twi-1	86.996	81	A	G	7.36	2	**	50	23.7	53.4098	48	173.119	40.5	220.765	5
30		Kaempferol-hexose-deoxyhexose, -hexose, -C10H8O3 (176)	Solyc01g107820	Twi-1	86.996	83	G	T	7.36	2	**	50	23.7	53.4098	48	173.119	40.5	220.765	5
31	12	Kaempferol-hexose-deoxyhexose, -pentose	Solyc03g078240	GGT13	43.776	34	A	C	9.46	1	****	50	23.2	1373.79			42.7	3476.59	6
32		Kaempferol-hexose-deoxyhexose, -pentose	Solyc01g107820	Twi-1	86.996	81	A	G	6.52	2	**	50	23.8	1413.47	49	5743.52	38.8	2989.22	5
33		Kaempferol-hexose-deoxyhexose, -pentose	Solyc01g107820	Twi-1	86.996	83	G	T	6.52	2	**	50	23.8	1413.47	49	5743.52	38.8	2989.22	5
34	13	Lycoperoside H or Hydroxytomatine II FA (a)	Solyc01g067350	GGT19	68.204	77	A	G	7.26	2	**	50	24.1	42.2918	49	320.903	46.5	75.789	3
35	14	Methyl salicylate malonyl dihexose-pentose	Solyc01g107820	Twi-1	86.996	81	A	G	6.24	2	**	50	23.8	141.44	41	432.796	40.9	848.88	5
36		Methyl salicylate malonyl dihexose-pentose	Solyc01g107820	Twi-1	86.996	83	G	T	6.24	2	**	50	23.8	141.44	41	432.796	40.9	848.88	5
37	15	Quercetin 3-O-rutinoside-7-O-glucoside	Solyc01g107820	Twi-1	86.996	88	G	A	6.94	1	***	50	24.1	713.199			47	6548.98	3
38		Quercetin 3-O-rutinoside-7-O-glucoside	Solyc01g107820	Twi-1	86.997	122	A	T	6.94	1	***	50	24.1	713.199			47	6548.98	3
39		Quercetin 3-O-rutinoside-7-O-glucoside	Solyc01g107820	Twi-1	86.995	78	C	T	8.48	2	**	50	23.7	698.618	42	1383.89	47	6548.98	4
40		Quercetin 3-O-rutinoside-7-O-glucoside	Solyc01g107820	Twi-1	86.996	81	A	G	8.77	2	**	50	23.5	695.348	42	1383.89	44.2	5123.18	5
41		Quercetin 3-O-rutinoside-7-O-glucoside	Solyc01g107820	Twi-1	86.996	83	G	T	8.77	2	**	50	23.5	695.348	42	1383.89	44.2	5123.18	5
42	16	Quercetin 3-O-sophoroside	Solyc03g078240	GGT13	43.776	34	A	C	9.46	1	****	50	23.2	374.976			42.7	1714.86	6
43		Quercetin 3-O-sophoroside	Solyc04g081830	GGT15	63.314	71	G	A	7.98	1	****	50	23.8	390.518			45.2	2206.07	4
44		Quercetin 3-O-sophoroside	Solyc01g107820	Twi-1	86.996	81	A	G	6.35	2	**	50	23.8	395.079	50	4754.43	38	1063.78	5
45		Quercetin 3-O-sophoroside	Solyc01g107820	Twi-1	86.996	83	G	T	6.35	2	**	50	23.8	395.079	50	4754.43	38	1063.78	5
46	17	Quercetin-hexose-deoxyhexose, -C12H12O5 (236)	Solyc01g107820	Twi-1	86.996	88	G	A	7.46	1	***	50	24.1	31.7532			47.7	161.992	3
47		Quercetin-hexose-deoxyhexose, -C12H12O5 (236)	Solyc01g107820	Twi-1	86.997	122	A	T	7.46	1	***	50	24.1	31.7532			47.7	161.992	3
48		Quercetin-hexose-deoxyhexose, -C12H12O5 (236)	Solyc01g107820	Twi-1	86.995	78	C	T	7.72	2	**	50	23.9	31.8783	31.5	26	47.7	161.992	4
49		Quercetin-hexose-deoxyhexose, -C12H12O5 (236)	Solyc01g107820	Twi-1	86.996	81	A	G	7.95	2	**	50	23.7	31.9867	31.5	26	44.8	128.244	5
50		Quercetin-hexose-deoxyhexose, -C12H12O5 (236)	Solyc01g107820	Twi-1	86.996	83	G	T	7.95	2	**	50	23.7	31.9867	31.5	26	44.8	128.244	5
51	18	Quercetin-hexose-deoxyhexose, -pentose	Solyc03g078240	GGT13	43.776	34	A	C	7.38	1	***	50	23.4	3983.95			40.7	13261	6

8 KASP protocol for Bio-Rad CFX-96

By Frans Ketelaars (920320430120)

8.1 Short Introduction

KASP (Kompetitive Allele Specific PCR) is a PCR based genotyping technique, patented by LGC group. The technique enables you to discriminate between known bi-allelic SNPs or InDels. The technique is based on the affinity of a forward primer to bind to a position of interest on the genomic DNA. One forward primer ends complementary to one allele, the other forward primer ends complementary to the other allele. The reverse primer is a common one. Both forward primers also contain an oligo that is specific to a quenched fluorescent dye, HEX or FAM (Figure 1). Within our research group, these dye specific oligo's have been unravelled, therefore we were able to design our own primers (see 'primer design'). Due to the discovery of the dye specific oligo's it is possible to perform the KASP technique in your own lab, which reduces the cost enormously. This protocol will lead to a successful KASP assay with the Bio-Rad CFX-96 qPCR machine.

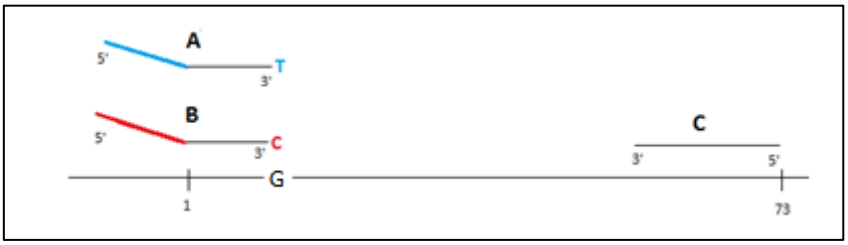


Figure 1. Overview of the primer orientation in a KASP assay. A) forward primer for allele 1, B) forward primer for allele 2 and C) common reverse primer. The numbers beneath the primers indicate the primer orientation. The numbers beneath the 'DNA' strand indicate the start and the end of the amplicon in base pairs (example)

8.2 Materials

- Bio-Rad CFX-96 with required software, CFX manager (However a wide range of qPCR instruments have been validated for KASP, check the LGC group website if your qPCR instrument is validated as well)
- Universal KASP master mix (2x concentrated) with a low ROX™ level
- Genomic DNA with known sequence
- The two forward primers and the common reverse primer (see 'primer design')

Primer design

As described above, there are three different primers involved in the KASP genotyping reaction: forward allele 1, forward allele 2 and a common reverse primer(see table 1 for an example). The forward alleles are build up from three different parts. A dye specific oligo (FAM™ or HEX™), a loci specific part that will bind to the DNA and at the end an allele specific part. The unravelling of the dye specific oligo is the part that made this 'home-made KASP assay' possible. They are universal and should function as the dye specific oligo's in every KASP assay.

The common reverse primer is a normal reverse primer as is used in a normal PCR procedure. However, the location of the common reverse primer is of high interest. Tests have shown that amplicon lengths above 100 base pairs (bp) are less reliable, so the position is of the common reverse primer is of great importance since the location of the forward primers is fixed .

It is also of importance to keep the normal primer design parameters in mind. Total DNA binding length between 20 and 30 bp; As equal melting temperature as possible for the gene specific part of the forward primers and the common reverse primer. Although these primer design standards are important, due to the fixed position of the forward primer, it is not always possible to fulfil them. Try to reach the primer design parameters within one his own discretions as good as possible.

Table 5. Example of KASP primers. The dye specific oligo's are highlighted in blue(FAM™) and red(HEX™). The DNA binding site of the primer is highlighted in black. In green you see the SNP specific binding place.

Forward primer allele 1	5'- GAAGGTCGGAGTCAACGGATT-GGATCAGTTATTTACTCTTCTTTCGGAA-A -3'
forward primer allele 2	5'- GAAGGTGACCAAGTTCATGCT-GGATCAGTTATTTACTCTTCTTTCGGAA-G -3'
Common reverse primer	5'-CCTAAAGCTAGTTCTTTTATCTGATCATC-3'

8.3 Methods

1. Start with preparing the samples in a qPCR plate. Every KASP genotyping reaction contains the components as is shown in table 2. It is the easiest to make a stock that contains all substances except for the DNA to test. Do not forget No Template Controls (NTCs), samples with H_2O. If possible, it would also be helpful to include positive controls, samples of known genotype, in your assay. Positive controls are especially helpful when you are working with low SNP frequencies. Seal the plate with an optically clear seal.

Table 6. Components of a KASP genotyping reaction and its quantities

Reagents	**1 KASP genotyping mix assembly (96-well plate)**
DNA (5-50 ng/ µl)	1 µl
2x KASP Master Mix	5 µl
Forward primer allele 1 (1 µM)	1 µl
Forward primer allele 2 (1 µM)	1 µl
Common reverse primer (1 µM)	1 µl
H_2O	1 µl
Total	**10 µl**

2. Thermal cycling conditions of a KASP assay for the qPCR instrument are detailed in table 3. The conditions are based on the KASP kit protocol provided by LGC group. Make sure you have a proper plate lay-out, define the location of your unknown samples and your negative controls. Use the 'select Fluorophores' button to select the FAM™ and HEX™ fluorophores. Highlight all wells of the plate that contain sample / negative controls and tick off each of the fluorophores. The fluorophore names will then be visible in the plate layout.

Table 7. Thermal cycling conditions for the KASP assay

Step	Description	Temperature	Time	Number of cycles per step
1	Activation	94 °C	15 min	1
2	Denaturation	94 °C	20 sec	10 cycles
	Annealing / Elongation	61-55 °C (drop 0.6 °C per cycle)	60 sec	
3	Denaturation	94 °C	20 sec	26 cycles
	Annealing / Elongation	55 °C	60 sec	
4	Reading	37 °C	60 sec	1

3. OPTIONAL. If you have not obtained clear genotyping clusters, the plate could be thermally cycled for an additional 3 cycles and read again. Table 4 shows the further thermal cycling conditions. This step, of further cycling and reading, can be performed multiple times until tight genotyping clusters have been obtained.

Table 8. Further thermal cycling conditions for the KASP assay

Step	Description	Temperature	Time	Number of cycles per step
1	Denaturation	94 °C	20 sec	3 cycles
	Annealing / Elongation	55 °C	60 sec	
2	Reading	37 °C	60 sec	1

4. When the qPCR machine has finished the thermal cycling program, you can view your results at the Bio-Rad CFX manager data analysis software. The data analysis software has several tabs, change to the 'Allelic Discrimination' tab since this is the relevant tab for KASP genotyping analysis. Next, you need to change the display mode from 'Cq' to 'RFU', clouds of data points will appear. Last, you can readjust the X (Blue) and Y (orange) axis to obtain the best scoring position for your assay.

8.4 Interpretation of the results

There are three genotypes, a sample could be homozygous for allele 1 or allele 2 or it could be heterozygous. Based on its genotype a forward primer does or does not bind to your genomic DNA in your sample. Because both primers got a different dye specific tail (HEX™ or FAM™), binding of a primer determines which dye will bind to your amplicon. Fluorescence occurs as the quenched HEX™ or FAM™ oligo is no longer quenched. Therefore the binding of the allele specific primer is the key step to discriminate between the different genotypes. When your sample is a homozygote for your allele of interest, it will only show the fluorescence of one of the two dyes and the other dye will stay in its quenched form.

If your sample is a heterozygote, it will contain both alleles in the same quantity. Both primers will bind in equal proportions to your genomic DNA and therefore the two dyes will be expressed in equal proportions. In this way, KASP is a very functional technique to discriminate between all three genotypes.

If everything went well, you will end up with three clearly different genotyping clusters of data points (see figure 2 for an example). Homozygous allele 1 with FAM™ expression, homozygous allele 2 with HEX™ expression and a mixture of expression of both dyes in case of a heterozygote. If distinction between the clouds is not very clear, you could decide to run some additional cycles. If extra cycling of your samples improves the results is to each his own opinion.

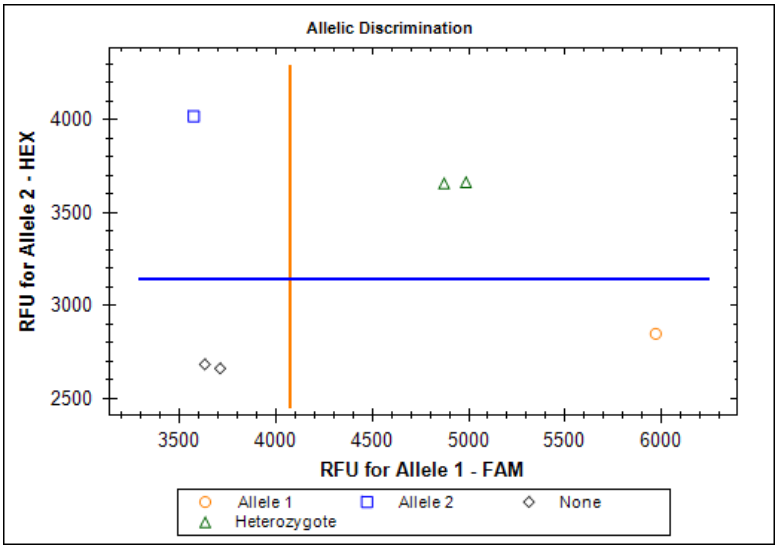


Figure 2. Example of the result of a successful KASP assay by Bio-Rad CFX-96

8.5 Final remarks

KASP genotyping is a technique that is very vulnerable for small deviations. Since we are not fully acquainted with the KASP technique yet, it is recommended to perform the KASP assay twice or use other techniques to confirm your results.
Be aware that this protocol does not suit the Bio-Rad CFX-384 qPCR machine. Tests have shown that a KASP assay with the use of the Bio-Rad CFX-384 qPCR machine did not work. Reasons for this failure are yet unclear. A KASP protocol which makes use of the Bio-Rad CFX-384 qPCR machine still needs to be developed.