Towards the development of new CAPS markers, enabling fine-mapping and size reduction of the IL5B introgression region

–

Research report

–

Clarinde van Hierden
920721338030

Supervisors: Dr. A.G. Bovy & Dr. Ir. A.W. van Heusden
PBR-80436 MSc Thesis at the department Plant Breeding - WUR
March 2014

Index

Foreword 4

Abbreviations 5

Abstract 6

Introduction - Social and scientific relevance 7
Towards an improved quality 7
Demand for improved nutritional value of tomatoes and the influence of semi-polar metabolites, particularly flavonoids 7
Better understanding of beneficial traits, derived from wild Solanum relatives, using introgression lines, focussed on the IL5B region 8
Research objective and experimental design 8

Material and methods 10
1.1 Materials 10
1.2 Developing new CAPS markers 11

Results 13
2.1 Bioinformatics – towards the design of new CAPS markers 13
2.2 First design of new CAPS markers 13
2.3 Selecting new developed CAPS markers 14
2.4 Determining the optimal test condition 15
2.4.1 Gradient PCR 15
2.4.2 HindIII restriction pattern of the selected primer pairs 16
2.5 Verification of new CAPS markers on BC5 progeny 18
2.6 Additional designed CAPS markers 20

Discussion 21
3.1 Deviating results 21
3.2 Relevance to continue with related research 22
3.2.1 Future research objectives 22
3.2.1.1 MYB and CHI regions 22
3.2.1.2 Up- and downstream areas 23
3.2.2 Metabolic analysis 23

Conclusion 24
General conclusion 24
Recommendations 24

Acknowledgements 26

References 27
Annex I - DNA isolation CTAB method 29
Annex II - PCR reaction (DT polymerase and buffer) 30
Annex III - HindIII restriction 30
Annex IV - Primer pairs [first batch] 31
Annex V - Additional designed CAPS markers 33

Foreword

Developing a well-functioning so called Cleaved Amplified Polymorphic Sequences (CAPS) based screening method, tested on several introgression lines obtained from the cross between commercial material *S. lycopersicum* cv. 'Moneymaker' (MM) and the wild accession *S. chmielewskii* LA1840 (CHM) , is the research objective of my master thesis. Commissioned and supported by the department Plant Breeding (PBR) from Wageningen University and Research Centre (WUR) this study has been designed and performed. The development of such a technique will allow fine-mapping and size reduction of the IL5B region. This fine-mapping is essential because the studied area contains genes (including CHI and MYB) coding for interesting metabolites (such as flavonoids) which are related to an improved taste and nutritional value of tomato fruits.

The involvement in scientific research during this master thesis will be combined with some additional learning objectives, mainly focussed on further orientation both within as well after completion of my study. Certain aspects, like defining quality traits, improvement of these factors, investigating metabolomics, regarding both their composition as well as their involvement (with)in pathways, but also marker development caught my attention. Therefore, more knowledge about these disciplines, but also the interaction between them, interested me. Hence, participation in current studies within these areas will provide better understanding of plant breeding research. Discovering the aspects and subjects having my interest, as well as studying interactions between certain disciplines, will challenge and motivate me. This kind of advanced orientation will support my decisions made within my further study as well as encourage future career orientation.

Abbreviations

°C	degrees Celsius
BC	Backcross
bp	base pairs
CAPS	Cleaved Amplified Polymorphic Sequences
CHI	Chalcone-flavanone Isomerase
CHM or CH	*Solanum chmielewskii*
CTAB	cetly trimethylammonium bromide
cv.	cultivar
DNA	deoxyribonucleic acid
EtBr	Ethidium Bromide
F	Forward
FAO	Food and Agricultural Organization of the United Nations
H	Heterozygous
HPLC	High-Performance Liquid Chromatography
ILs	Introgression Lines
kb	kilo bases
MAS	Marker Assisted Selection
MM	*Solanum lycopersicum* cv. 'Moneymaker'
mQTL	metabolite Quantitative Trait Loci
MS	Mass Spectrometer
MYB	Transcription factor
NarCh	Naringenin Chalcone
NILs	Near Isogenic Lines
PBR	Plant Breeding
PCR	Polymerase chain reaction
R	Reverse
S	Selfing
SGN	Solanaceae Genome Network
SNP	Single Nucleotide Polymorphisms
Tm	melting Temperature
VHL	van Haeringen lab
WHO	World Health Organisation
WUR	Wageningen University and Research centre

Abstract

During this research several Cleaved Amplified Polymorphic Sequences (CAPS) markers have been developed and tested on introgression lines, to determine the origin of certain introgression regions. Eventually discovery of candidate genes within these areas, coding for interesting metabolites (such as flavonoids) related to an improved taste and nutritional value was the ultimate goal.
Unfortunately the predicted fine-mapping and size reduction of certain segments, including the IL5B region, through CAPS markers has not been achieved during this master thesis. Nevertheless the recommendations obtained from this research might serve as a basis for the design of new studies focussed on the development of a well-functioning (CAPS) based technique.

Overall the most remarkable results obtained during this master thesis were the specific reactions of the tested primer pairs, as well as contradictory results between the findings obtained from this research and the available data. For several regions it appeared to be impossible to obtain proper amplification, whereas for other segments the genetic profile determined did not match the previously obtained results by van Haeringen lab (VHL) as well as the known pedigree information.

Finally, due to inconclusive results, several aspects in the design of the CAPS technique are recommended to modify and study (separately or in combination with each other) during future researches:

1) Use fresh (isolated) material to repeat the experiment
2) Perform gel electrophoresis with Ethidium Bromide and add larger amounts of HindIII treated polymerase chain reaction (PCR) samples to increase the intensity of the signal. The Ethidium Bromide (EtBr) will improve the ease of scoring the right fragment size.
3) Specificity and intensity might be correlated to melting temperatures (Tm). Therefore testing of primer pairs under different Tm conditions is advised.
4) If parent specific primers are necessary, due to small sequence differences, these parental specific primers should be tested in two separate mixtures instead of a combination of them.
5) To decrease the number of false-positive introgression data it can be tested if reduction of the PCR fragment size does increase the specificity and will indeed improve the method.
6) Alter the primer design by developing primers that will include the single nucleotide polymorphism (SNP) itself. This will result in two parental specific primers as well as size reduction of the PCR fragment. Detection of each parent will still be based on presence or absence of HindIII restriction, due to malfunctioning of the restriction site in one of both parents.
7) A final solution might be performance of Infinium Bead Array analysis, although this will increase the research costs.

Overall the main advice for future studies is to focus solely on the development of a well-functioning technique, instead of immediately incorporate screening of interesting plant populations. This latter should be regarded as a separate research, after which metabolic studies and linkage of the various data obtained can be performed.

Concluding; only when more research has been done and a well-functioning method has been developed, breeding regarding flavonoid composition thereby improving tomato quality can be accomplished. After all knowledge about the pathways involved as well as available markers will allow marker assisted selection (MAS), resulting in tomatoes with an improved taste and nutritional value, due to their metabolic composition. Incorporation of such quality aspects within vegetable breeding will be beneficial for human health, especially with regard to the emerging new disease patterns and consumer demands of upcoming years.

Introduction

Social and scientific relevance

Towards an improved quality

Cultivation and consumption of tomatoes (*Solanum lycopersicum*) takes place worldwide, making it the most important non cereal crop (Ballester et al. 2013). According to the Food and Agriculture Organization of the United Nations (FAO) the production in 2011, exceeded the 15 million tonnes (FAO, 2013 http://faostat.fao.org/). This diversity results in a crop having a high economic value and rich history in breeding activities. It has served as model organism for freshly fruit plants and the complete genome sequence of tomato is made available through the Solanaceae Genome Network (SGN; http://solgenomics.net)(Consortium 2012). This new development will stimulate breeding research and allows incorporation of new technologies, like improved marker assisted selection (MAS). Such new improvements illustrate that even today the process of (modern) breeding is still evolving, in which each period has its own characteristics, both in techniques as well as in goals and selection criteria.

In the past the major selection goal during the breeding of tomatoes was focused on yield, colour and shape. Due to extensive domestication a shift in the composition of the genetic material used for cultivation has occurred, from a mixture of landraces to pure homozygous lines. This reduced genetic variation, whereby selection has been focussed on a minority of characteristics, has a negative effect on several other traits.

These days the main breeding goal for tomatoes is improvement of their quality, especially in terms of taste and healthy components, resulting into an improved nutritional value. However quality is a broad concept and involves many different components of the production chain like; growers, manufactures, logistics and consumers, each of them have their own 'selection criteria' to judge the quality of a tomato.

Demand for improved nutritional value of tomatoes and the influence of semi-polar metabolites, particularly flavonoids

Nowadays consumers are more and more focussed on better tasting products, with a proper nutritional value and also the demographic composition is and will shift towards a more aged population. Research from the World Health Organisation (WHO) (Aging 2011) proved a correlation between the number of disabled individuals and an increasing number of older people. Such new disease patterns and consumer demands are emerging both in the developed as developing world, and results in the need for healthy, attractive and tasteful vegetables, including tomatoes.

Breeding for improved quality aspects requires incorporation of new knowledge and technologies during the breeding process. This is for example necessary in order to discover the genes underlying these, often complicated, traits but also to understand the working mechanisms involved in these processes or to select the right plant material.

Regularly improvement of the nutritional value is influenced by semi-polar metabolites such as flavonoids, phenolic acids and alkaloids. Loci responsible for these health-related compounds, present in tomato, are roughly identified by Ballester et al. (2013) through studying of the semi-polar metabolite composition in red-ripe tomato fruits. This study resulted into new metabolite quantitative trait loci (mQTL) influencing the metabolite content, which is the first step in the identification of genes underlying the synthesis of these semi-polar metabolites. For example, robust mQTLs responsible for changes in flavonol glycosides were identified on chromosome 5 (Ballester et al. 2013). In tomato, flavonoids include a wide variation of polyphenolic secondary metabolites, generated by specific mutations in flavonoid-coding genes, which results in quantitative and qualitative differences in metabolic profiles (Bovy et al. 2007; Micka 2013). Flavonoids can have different beneficial effects both in plant as in humans. In plants, flavonoids are involved in many

biological processes such as pigmentation, protection against ultraviolet light as well as pathogenic microorganisms, plant fertility and germination of pollen (Bovy et al. 2007). While for humans the mode of action of flavonoids is based on their anti-oxidant activity (Kong et al. 2003) which results in scavenging of the free radicals derived from fatty acids and oxygen (Greenberg et al. 1994). This anti-oxidant capacity results in decreased risk of coronary heart disease and inhibition of cholesterol accumulation in blood serum. Combined with their positive influence on vascular cell walls, flavonoids are capable to prevent several chronic diseases, like cardiovascular risk and atherosclerosis. The beneficial effect of flavonoids on these diseases, which are often reported amongst aging individuals, makes them interesting to improve the nutritional value of certain (vegetable) products.

In tomato the major flavonoid present is Naringenin chalcone (NarCh), but also quercetin-rutinoside (rutin) and kaempferol-rutinoside are important (Verhoeyen et al. 2002). Formation of these compounds occurs during ripening of the tomato fruits and accumulation of it takes place in the peel tissue and only on small scale in the flesh of the fruit (Adato et al. 2009; Verhoeyen et al. 2002). In general the accumulation of flavonoids in cultivated tomato fruits is very low. Whereas in a few wild Lycopersion accessions, larger amounts of flavonoids are noticed, in which occasionally the genes involved in the flavonol biosynthetic pathway are even expressed in the fruit flesh (Willits et al. 2005).

Previous research (Adato et al. 2009; Ballester et al. 2010; Ballester et al. ; Bovy et al. 2007; Jez et al. 2002; Micka 2013; Muir et al. 2001; Verhoeyen et al. 2002) resulted in more detailed information about the biochemical composition of the flavonoid pathway. And through analysis on molecular and biochemical basis, some clues regarding the biosynthesis and rate-limiting steps of flavonoid accumulation in the tomato peel could be discovered.

Better understanding of beneficial traits, derived from wild *Solanum* relatives, using introgression lines, focussed on the IL5B region

Due to recurrent breeding for tomato cultivars sharing many appreciated phenotypic traits within a narrow gene pool, reduction of the available genetic variation has occurred. Therefore incorporation of new traits, such as increased flavonoid content (both in peel and flesh of the fruits) requires introduction of wild relatives into the breeding program in order to (re)use the original gene pool.

Incorporation of the beneficial characteristics of wild crossable *Solanum* material often occurs through the development of introgression lines (ILs). These ILs are derived by crossing a wild accession into a tomato background. Ideally each of the lines derived contains a single defined chromosome segment from the wild accession, which substitutes its homologous region in the tomato background (Ballester et al. 2013).

Such an IL population, derived from a cross between the commercial material *S. lycopersicum* cv. 'Moneymaker' and the wild accession *S. chmielewskii* LA1840, is available and resulted into the discovery of mQTLs correlated to the synthesis of semi-polar metabolites (Ballester et al. 2013). Not all 33 ILs analysed in the previous research consists out of a single *S. chmielewskii* introgression, for instance the genotype IL5B (located at the top of chromosome 5) carried four introgression segments (on chromosome 2, 5, 11 and 12). Metabolic research of the IL5B segment revealed presence of flavonoids. But in order to discover the candidate genes linked to the synthesis of these metabolites, it is necessary to fine-map the IL5B introgression region and to separate the four introgression segments. This is possible through additional crosses and development of new markers, which can be used to screen the progeny during future breeding programmes, focussed on introgression of genes (and traits) of interest into commercial tomato cultivars.

Research objective and experimental design

The ultimate goal of this research is fine-mapping and size reduction of the IL5B region derived from the cross between the commercial material *S. lycopersicum* cv. 'Moneymaker' (MM) and the wild accession *S. chmielewskii* LA1840 (CHM), through the development of new so called Cleaved

Amplified Polymorphic Sequences (CAPS) markers. Fine-mapping of the IL5B region is essential because it contains genes (including CHI and MYB) coding for interesting metabolites (such as flavonoids) which are related to an improved taste and nutritional value of tomato fruits. Eventually the new developed CAPS markers and selected recombinants will enable more specific mapping of the IL5B region. Also they will contribute to the discovery of more candidate genes as well as the selection of the progeny in future breeding programmes.
Accomplishment of the research objective requires the performance of several experiments. Roughly this experimental section can be subdivided into three parts:

I. Development of new well-functioning CAPS markers for fine-mapping of the IL5B region up- and downstream the CHI and MYB segment (table 1). This will result in determination of the introgression region present and enables recombinant screening among seedlings for metabolic analysis.

Table 1 Genomic composition of several plants (PV-numbers) for the IL5B region, including the MYB and CHI segments (Micka 2013). Boxes indicate introgression segments from CHM which will be fine-mapped using the new developed CAPS markers. Md7559 refers to the genomic position 6,461,353 at chromosome 5 and serves as control to interpret recombination events.

DNA \ Assay	**Chr. 5** *MYB*	**Chr. 5** *CHI*	**Chr. 5** *md7559*
Control CHM	T:T	G:G	T:T
Control MM	C:C	A:A	C:C
PV121453	C:C	G:G	T:T
PV121454	C:C	A:A	C:C
PV 121457-1	T:T	A:A	C:C
PV 121457-20	T:T	A:A	C:C
PV 121457-32	T:T	A:A	C:C

II. The genetic knowledge will be combined with the metabolic profile to determine the influence of the CHI and MYB region on the flavonoid pathway, separately. Screening the plants (table 2) with new developed CAPS markers will provide more detailed information about the introgression region present, which allows fine-mapping of the candidate gene location.

Table 2 Genomic composition of several plants (PV-numbers) for the IL5B region, including the MYB and CHI segment (Micka 2013). Boxes indicate introgression segments from CHM. These plants will recombine for the CHI region (becoming homozygous MM or CHM) combined with homozygous CHM for MYB. Md7559 refers to the genomic position 6,461,353 at chromosome 5 and serves as control to interpret recombination events.

DNA \ Assay	**Chr. 5** *MYB*	**Chr. 5** *CHI*	**Chr. 5** *md7559*
Control CHM	T:T	G:G	T:T
Control MM	C:C	A:A	C:C
PV121453	C:C	G:G	T:T
PV121454	C:C	A:A	C:C
PV 121457-15	T:T	G:A	T:T
PV 121457-28	T:T	G:A	T:T
PV 121457-70	T:T	G:A	T:T

III. Identification and separation of the four introgression regions (chromosome 2, 5, 11 and 12) present using available marker data (obtained from Infinium Bead Array analysis performed by van Haeringen lab (VHL) located in Wageningen) as well as new developed CAPS markers during this research.

Material and method

1.1 Materials

Plant material:

- Both parents, *S. lycopersicum* cv. 'Moneymaker' [MM] and the wild accession *S. chmielewskii* LA1840 [CHM], of the IL population (Ballester et al. 2010)
- A selected BC_4 line [PV101318_A1] which is tested homozygous MM for the MYB region and heterozygous for the CHI introgression segment. Additional heterozygous introgression regions on the chromosomes 2, 11 and 12 were also determined (table 3)
- Selected BC_4S_1 line [PV121454] and BC_4S_2 line [PV121453] from the cross between MM and CHM, carrying CHM introgression regions on chromosomes 2, 5, 11 and 12 (figure 1 and 3)
- Three selected BC_4S_3 lines [PV131510], [PV131513] and [PV131521] containing an introgression segment between the MYB and CHI region (figure 2). The parents from these plants were homozygous CHM for MYB and heterozygous CHM for CHI (table 2), enabling outcrossing and recombination for CHI in the progeny.
- Three selected BC_5 lines [PV131567A/B] and [PV131568A] and their seedlings (figure 1) containing a CHM introgression segment downstream the CHI region (table 1)
- Seedlings from six selected BC_5 lines [PV131504A/B], [PV131506A/B] and [PV131507A/B] (figure 2) containing a CHM introgression segment upstream the MYB region (table 1)

Restriction enzyme:

- Both parental genomes (MM and CHM) were screened for single nucleotide polymorphisms (SNPs) in the restriction site of the enzyme HindIII [5'…A↓AGCTT…3' ; 3'…TTCGA↑A…5'].

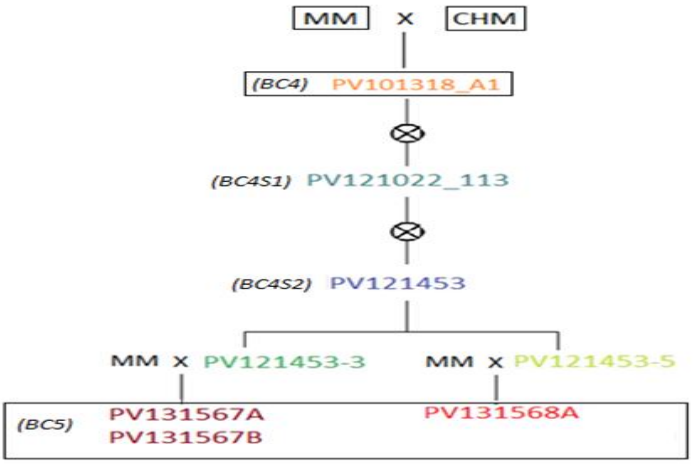


Figure 1 Pedigree of the plant number [PV121453]. Plant numbers that are indicated with boxes will be studied during this research to develop new CAPS markers

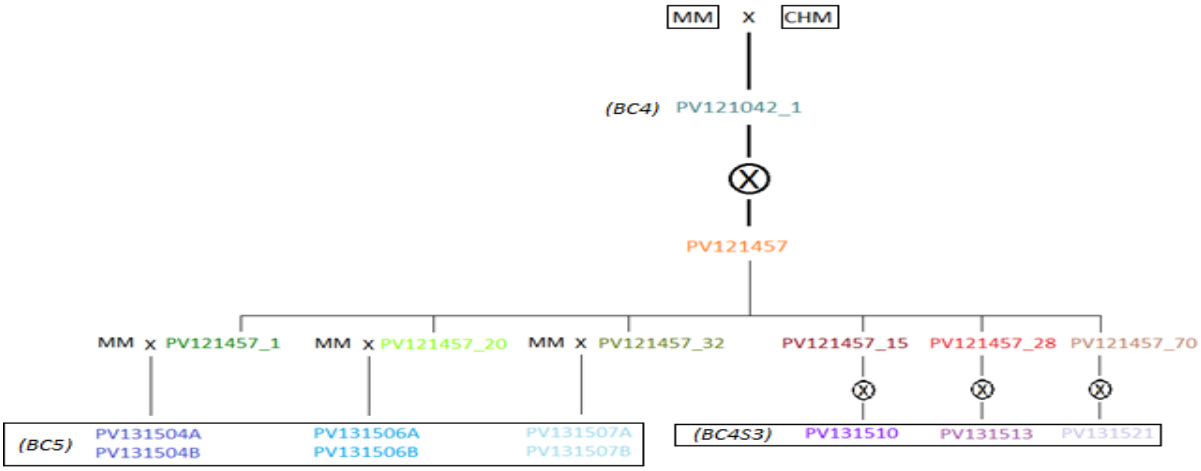


Figure 2 Pedigree of the plant number [PV121457]. Plant numbers that are indicated with boxes are useful for future research.

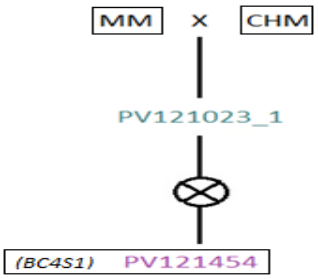


Figure 3 Pedigree of the plant number [PV121454]. Plant numbers that are indicated with boxes are useful for future research.

1.2 Developing new CAPS markers

Three chromosomes (2, 5 and 11) were screened with new CAPS markers, allowing identification and separation of the existing introgression segments. The chromosome positions studied to design the new CAPS markers, were based on the previous results (table 3), derived from Infinium Bead Array analysis performed by VHL located in Wageningen, and were distributed over different genomic positions.

Table 3 Available marker data, analysis performed by VHL on the plants [PV101318_A1] (BC_4) and [PV121022_113] (BC_4S_1). For each seq-md number the genomic position, the MM sequence and the final result (*pink*/*light green*/*dark green*) is displayed. *Pink* indicates a region homozygous for MM, *light green* homozygous CHM whereas *dark green* corresponds with a heterozygous introgression region. An *empty* square referrers to a genomic position that has not been tested. While the *orange* box displays doubtful results obtained from VHL.

Chromosome	Position	Seq-md#	MM	PV101318_A1	PV121022_113
2	28.520.428	5719	G:G		
	42.549.116	6683	C:C		
	48.094.332	8825	C:C		
	48.838.336	8159	A:A		
5	1.860.303	5370	T:T		
	6.461.255	7559	C:C		
	20.527.748	1377	C:C		
	43.744.992	1880	T:T		
11	5.279.506	8439	C:C		
	6.497.192	1693	G:G		
	46.408.368	2734	A:A		
	51.096.556	7969	C:C		
12	497.852	3161	A:A		
	2.087.551	6489	C:C		
	16.701.745	4416	G:G		
	32.162.204	851	T:T		
	52.930.616	449	A:A		
	62.021.900	3504	G:G		

Primer pairs were developed using the Primer3Plus software (Untergasser et al. 2007). When necessary (due to SNPs) the designed primer pair was altered, resulting in two separate primers each of them specific to one of the parents.
Eventually 23 specific primer pairs (chromosome **2**: [*B1-MM*, *B1-CHM*, *B2-MM*, *B2-CHM*, *B3-MM*, *B3-CHM*, *B4-MM* and *B4-CHM*] chromosome **5**: [*A1-M/C*, *A2-MM*, *A2-CHM*, *A3-MM*, *A3-CHM*, *A4-MM* and *A4-CHM*] chromosome **11**: [*A1-MM*, *A1-CH*, *A2-MM*, *A2-CHM*, *A3-MM*, *A3-CHM*, *A4-MM* and *A4-CHM*]) were designed (overview is given in table 7).
Also mixtures of specific-primer pairs, corresponding to a similar genomic position tested, were developed. In total 11 of these primer mixtures (chromosome **2**: [*B1-M/C*, *B2-M/C*, *B3-M/C* and *B4-M/C*] chromosome **5**: [*A2-M/C*, *A3-M/C* and *A4-M/C*] chromosome **11**: [*A1-M/C*, *A2-M/C*, *A3-M/C* and *A4-M/C*]) were made and tested within this research.

The distinctiveness character for the parental alleles of each designed primer pair (specific or mixture) were tested firstly, using isolated deoxyribonucleic acid (DNA) from both parents, MM and CHM. DNA isolation has been performed according to the cetly trimethylammonium bromide (CTAB) method described in annex-I. After isolation of the DNA, each primer pair was added and a polymerase chain reaction (PCR), using Dreamtaq (DT) polymerase and buffer, has been performed (procedure described in annex-II). The obtained PCR products were digested with HindIII (protocol added as annex-III) for 1h at 37 degrees Celsius (°C), after which the results were made visible through gel-electrophoresis using an Ethidium bromide gel (EtBr) and a 1 kilo base (kb) Plus DNA ladder from GeneRuler™. The obtained restriction pattern of each primer pair, for MM as well as for CHM, was checked with the expected pattern. When a proper correlation was observed the selected primer pair was used for testing of the BC_5 lines ([PV131567A], [PV131567B] and [PV131568A]) to determine the origin and location of the introgression segment present.

Results

2.1 Bioinformatics – towards the design of new CAPS markers

Bioinformatics (Finkers 2013) provided an overview of the genomic positions at which an altered restriction site of the HindIII enzyme between both parents (MM and CHM) was present. The positions listed indicated a SNP only present in CHM, which limited restriction by HindIII at that specific position. For each region of interest the nearest altered restriction site was chosen from the list and via Tomato variant explorer (WUR and BGI 2013) powered by JBrowse 1.10.12 the genomic sequence adjacent to this position was obtained. This genomic sequence allowed primer design using the Primer3Plus software (Untergasser et al. 2007).

Important criteria correlated to the design of these primers were; an optimal product size of 1 kb as amplified region in which the altered restriction site was not centrally located, exclusion of multiple HindIII restriction sites within the amplicon and generation of digested segments with a proper detectable length of 800-200 base pairs (bp). For each primer pair designed the expected restriction pattern (length of fragments in bp) for both parental alleles was predicted in advance (listed in tables 4, 5 and 6 annex IV).

2.2 First design of new CAPS markers

The position and (homozygous or heterozygous) character of CHM introgression segments on the chromosomes 2, 5, 11 and 12 previously determined, were verified through use of new developed CAPS markers. The designed CAPS markers corresponding to the positions already investigated as well as markers covering regions not studied before (grey cells) are listed in table 7. Only the new developed CAPS markers on chromosome 12 were not tested during this research due to time constrictions.

Table 7 First batch of new developed CAPS markers, allowing verification of the available introgression data displayed in the column 'Seq-md number [position]' as well as in table 3. The HindIII sites correlated to these and new (grey coloured cells) CAPS marker positions are listed as well. For each position the used forward (F) and reverse (R) primer are displayed and if necessary divided into a MM or CHM specific primer. A " means specification for each parent was not needed according to the genomic sequences displayed in JBrowse.

Chromosome	Position (HindIII site)	Seq-md# [position]	Name primer pair	F primer	R primer
2	4,949,427	None	B1- MM	TCAACAAGTCAAGGCAACCA	ACCAGTCCGTCAATTTCCAG
			CHM	"	ACAAGTCCGTCAACTCCCAG
	20,002,287	None	B2- MM	GAGGATGGTGCCAATTTTCT	TATCTCGGCCCTCGTATTTG
			CHM	"	TAGCTCGGCCCTCGTATTTG
	28,464,819	5719 [28,520,428]	B3- MM	CGGAGGATTATGTTCGGTTG	CATGTTTCGATAGATTGTTCCGTA
			CHM	AGAAGGATTATGTTCGGTTG	CATGTTTCAATAGATTGTTCCTA
	42,589,256	6683 [42,549,116]	B4- MM	GGAGGATAATTTAGGTGAACCA	TGTTTGTTGTCATGAACCCATT
			CHM	GAAGGATAATTTAGGTGAACCA	"
5	2,995,300	None	A1- MM	GCAAATCAGCATCAACCAGA	TGCTAAGTTTGCTGGTCCAA
			CHM	"	"
	6,467,256	7559 [6,461,255]	A2- MM	TGTGCACGCTTAGAAGCATAA	TGCTACCAAACGGGTCCTTA
			CHM	"	TCCTTCCAAACGAGTCCTTA
	20,548,353	1377 [20,527,748]	A3- MM	TGAAGCAAGGCTTGGAATCT	TGTCCCTTTTCGCATGTTTT
			CHM	TGAATCAAGGCTTGAAATCT	CTTCCCTTATAACATGATTT
	39,994,430	1880 [43,744,992]	A4- MM	GCGAGAATTCTTTTGGTGGA	CAAACTTCGAACCGGACAAT
			CHM	GCGAGAAGTCTTTTGGTGGA	CAAACTTCAAACCAGACAAT
11	2,004,495	None	A1- MM	ACGCATTCACACCGAAAAAG	TACTGATCCGACGTGCTCAA
			CHM	ACGCATTCACACTGGAAAAG	"
	5,276,912	8439 [5,279,605]	A2- MM	AGCGAGAGTGAGGGTTTTGA	CCAAATATCGAGCGGGATAAA
			CHM	AACGAGAGTGAGGGTTTTGA	TCAAATATCGAGCGGGAAAAA
	6,522,663	1693 [6,497,192]	A3- MM	TGCGGATGAGAAGGTAACAA	CTGCGCTTGCAGAGTTTCTA
			CHM	TGCGGATGAGAAGATAACAA	CTACGCTTGTAGAGTTTGTA
	51,096,119	7969 [51,097,323]	A4- MM	GTCTCTGGCGCTTCGTTTAG	CGAGATAAATGGAGCTATGAGG
			CHM	"	TGAGATAAATGGAGCTATGAGG
12	191,352	None	A1- MM	TACATTCCTTTGCCCGTCTC	TGCAACATCCCTAACCACCT
			CHM	TACATTCCTTTGCCCATCTC	TGCAACATCCATAACCACCT
	504,270	3161 [497,852]	A2- MM	GGAGCATATGCCAGGGATTA	AAGAGAGCCCGGTAAGGAAG
			CHM	TGAGCATATGCCAGGGATTA	"
	16,718,829	4416 [16,701,745]	A3- MM	TCATGTTCATCGCTTTTTGC	AAGGTGCACTCCCATTGAAC
			CHM	TCATGTTCGTCGCTTTTTTG	AAGGTGCACTTCCATTGAAC
	32,187,026	851 [32,162,204]	A4- MM	TGCTTCCGTCCTTTGAAGAT	TAGGGTTTCCCACTCCTCCT
			CHM	TGCTTCTGTCCTTTGAAGAT	"
	52,928,548	449 [52,930,616]	A5- MM	TGCAGGAACCCATCAATAGAG	CCTCTTGTTGCATTTTGAGC
			CHM	"	CCTCTTGTTGCGTTTTAAGA
	62,026,195	3504 [62,021,900]	A6- MM	TTCCCATGTAGGCTTTCGTC	CCGAACACACCATTGATGAC
			CHM	"	"

2.3 Selecting new developed CAPS markers

According to tables 4, 5 and 6 annex IV, in which the expected and observed fragment pattern for each desgined primerpair (without HindIII restriction) has been compared, it appeared that not each primerpair functioned well enough. Multiple primer pairs did not produce the fragments according to the expected size (approximately 900-1000 bp) when they were tested on isolated parental DNA. Only 14 primer pairs functioned as expected (summarized in table 8) and were selected.

Table 8 The abbreviation of the 14 selected primer pairs of which the restriction pattern was studied and information about introgression regions was obtained during this study.

Chromosome 2	Chromosome 5	Chromosome 11
B1 – MM	A1 – MM	A1 – CHM
B1 – CHM	A3 – MM	A2 – MM
B2 – MM	A4 – CHM	A2 – CHM
B2 – CHM		A3 – MM
		A3 – CHM
		A4 – MM
		A4 – CHM

Besides verification of the parent specific primers, also the primer mixtures were tested. A notable finding was the absence of amplification for each multiple-primer mix (results not shown in order to promote the clarity of the report). Even if both MM or CHM specific primers worked, a combination of these specified primers did not result into PCR-products of the expected length.

2.4 Determining the optimal test condition

Testing the 14 selected primers on both parents, resulted in 28 different combinations. From these combinations only five couples (*05A1MM* on CHM, *05A3MM* on MM, *11A2MM* on MM, *11A3MM* on MM and *11A4MM* on MM) performed well enough, at a melting temperature (Tm) of 57.0°C. Wherease the other 23 often demonstrated aspecific amplification, resulting in more and deviating fragments, and/or a reduced intensity of the desired amplified product (approximately 900-1000 bp), thereby hampering the scoring of the results (tables 4, 5 and 6 annex IV).
These 23 combinations were divided into three groups, listed below, based on the characteristic(s) that needed improvement. Therefore out of the 23 left, ten combinations were selected to increase the intensity as well as specificity of the desired fragments. For nine samples only the specificity of the amplified PCR product should be enhanced, allowing reduction of the number of aspecific amplified fragments. Whereas the last four combinations only required an increased intensity of the preferred PCR fragment.

Both specificity and intensity:

- *02B1MM* on CHM
- *02B2MM* on CHM
- *02B2CH* on CHM
- *05A3MM* op CHM
- *11A2MM* on CHM
- *11A2CH* on MM and CHM
- *11A3MM* on CHM
- *11A3CH* on MM and CHM

Specificity:

- *02B1MM* on MM
- *02B1CH* on MM and CHM
- *02B2MM* on MM
- *02B2CH* on MM
- *05A1MM* on MM
- *05A4CH* on MM and CHM
- *11A1CH* on MM

Intensity:

- *11A1CH* on CHM
- *11A4MM* on CHM
- *11A4CH* on MM and CHM

In order to reduce the aspecific amplification but also to increase the intensity of the gel electrophoresis results, gradient PCRs for the selected primer pairs have been performed, allowing determination of the optimum Tm.

2.4.1 Gradient PCR

In general the gradient PCR (Tm 50.0 up to 60.0°C) of the three primers on chromosome 5 (figure 4) indicates better amplification occurs at higher gradient temperatures, which resulted in fragments of the expected 1 kb length and less aspecific binding, causing a reduction in the number of additional fragments. Because small differences in intensity as well as specificity between 57.4°C [8] and 60.0°C

[12] are observed (figure 4), it was decided to perform the HindIII restriction of the selected 14 primer pairs, on parental and progeny DNA, at a higher Tm of 60.0°C instead of 57.0°C.

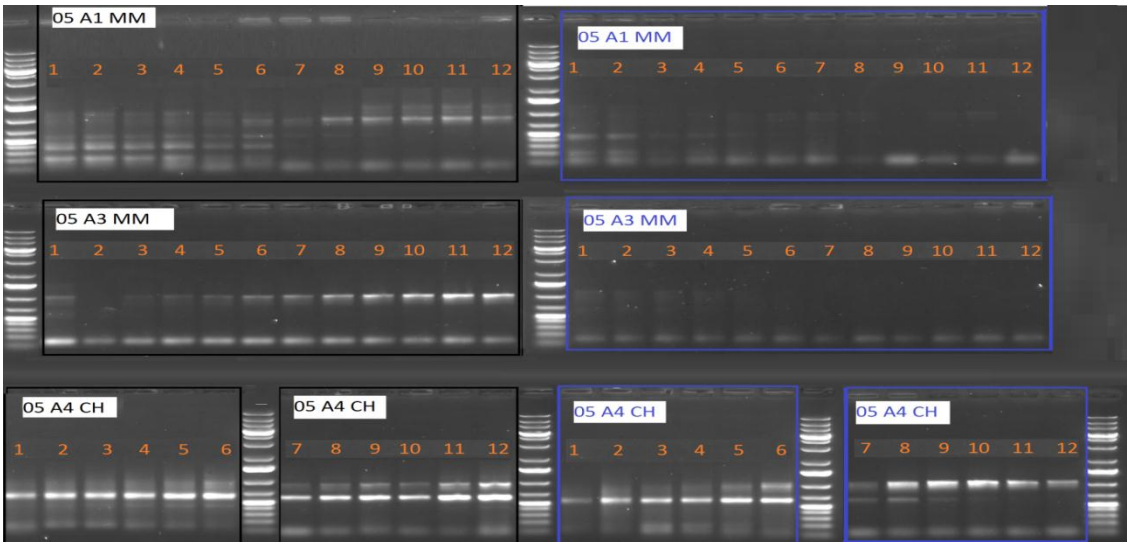
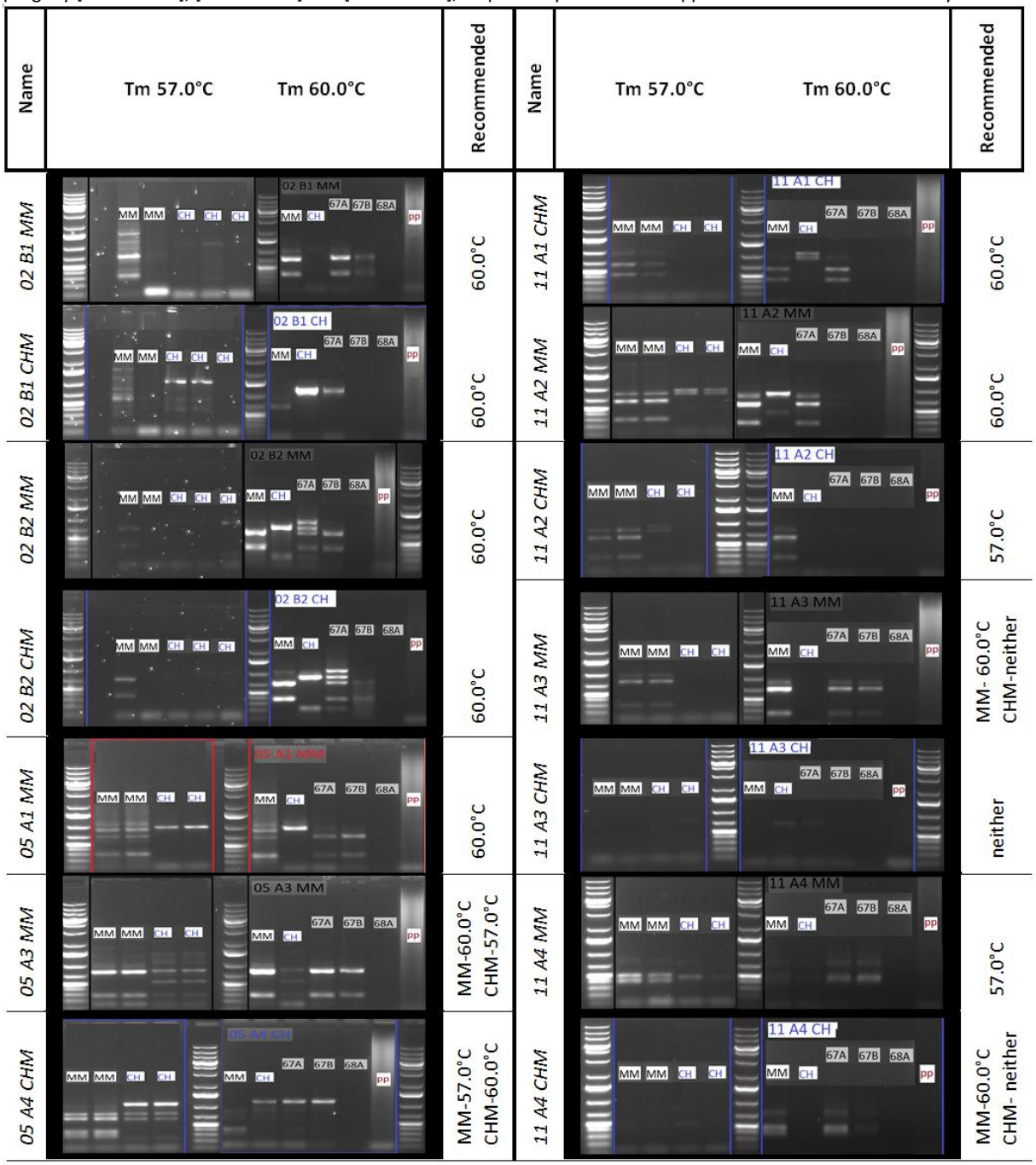


Figure 4 Gradient PCR of the three selected primer pairs of chromosome 5 (*A1–MM* ; *A3–MM* ; *A4–CH*). The black boxes refers to MM DNA while the blue ones are tested on CHM. Numbers 1 to 12 corresponds with the gradient, Tm from 50.0°C till 60.0 °C; [1] 50.0°C, [2] 50.3°C, [3] 50.9°C, [4] 51.7°C, [5] 52.8°C, [6] 54.3°C, [7] 56.0°C, [8] 57.4°C, [9] 58.5°C, [10] 59.3°C, [11] 59.8°C and [12] 60.0°C.

2.4.2 HindIII restriction pattern of the selected primer pairs

The restriction pattern for all 14 primer pairs at Tm 57.0°C versus 60.0°C are compared in table 9. For each combination the most optimal temperature, based on the gel results, is listed as 'recommended'.

Table 9 Gel electrophoresis results after HindIII restriction, for the 14 selected primer pairs during two different conditions. Left picture corresponds to a melting temperature of 57.0°C while at the right a Tm of 60.0°C has been applied. For each lane the sample present is listed. Whereby *MM* and *CH* refer to the parents, while *67A*, *67B* and *68A* corresponds to the BC_5 progeny [PV131567A], [PV131567B] and [PV131568A], respectively. The last lane *pp* serves as blanc and lacks only DNA.



Remarkable are the results for the three primer pairs of chormosome 5, espacially when they are compared to the gradient PCR data, a schematic overview of this comparison is listed below.

Primer	Tested on	Results Gradient PCR	Results Tm 57.0°C and 60.0°C PCR
05 A1 MM	MM	Tm 60.0°C resulted in less by products and a higher intensity	Similar results at Tm 57.0°C and 60.0°C, only the presence of additional fragments was slightly reduced at Tm 60.0°C
	CHM	The Tm 60.0°C condition provides a single band but overall the amplification is very low.	Both conditions provided specifc results, but the fragment was more intense at Tm 60.0°C
05 A3 MM	MM	Both Tm 57.4°C as well as Tm 60.0°C performed well, although Tm 60.0°C is preferred due to a more clear amplification product	The expected fragment was observed at both Tms tested, altough a higher intensity at Tm 60.0°C was observed
	CHM	Hardly any bands were visible, only the Tm of approximately 50.0°C resulted in very light fragments	The Tm 57.0°C condition resulted in a more intense hit, although multiple additional fragments were present compared to Tm 60.0°C
05 A4 CH	MM	A rise in Tm corresponded with an increased intensity of the amplified fragment, but also of the additional fragments	A Tm of 57.0°C increased the intensity of the fragments compared to a Tm of 60.0°C, where hardly any bands were visible
	CHM	A Tm of 57.4°C provided a more intense fragment compared to 60.0°C. Although Tm 60.0°C displayed a single product whereas 57.4°C multiple fragments.	Tm 60.0°C did result in slightly less intense fragments compared to Tm 57.0°C. But at 60.0°C no additional fragments are present.

This overview indicate that for some situations (*05A1MM* on CHM, *05A3MM* on CHM and *05A4CH* on MM) the results of the gradient PCR and of the Tm 57.0/60.0°C PCRs do not match. In case of *05A1MM* on CHM, a difference in intensity at Tm 60.0°C between the gradient PCR and the single PCR was detected. For primer pair *05A3MM* tested on CHM, there was no amplification observed during the gradient PCR (figure 4), while both single PCRs did result in several amplified fragments (table 9). Also contradictory results between the gradient PCR and the Tm 57.0/60.0°C PCR for *05A4CH* on MM, which clearly indicates a higher intensity at 57.0°C instead of 60.0°C, were obtained. Despite some contradictions in the results between the gradient and single PCRs performed, the pictures in table 9 indicates an improved amplification when a Tm of 60.0°C is used instead of 57.0°C.

2.5 Verification of new CAPS markers on BC_5 progeny

Testing the 14 selected markers on isolated DNA from both parents as well as on the BC_5 progeny plants [PV131567A], [PV131567B] and [PV131568A] at a Tm of 60.0°C, resulted in a higher specificity and intensity compared to the 57.0°C situation (table 9). Also increasing the amount of PCR-product used during gel electrophoresis improved the intensity of the results. Although, for the sake of clarity, not all gel results from these tests are shown, the restriction pattern observed for the parental genotypes corresponded in most cases to the expected product sizes with (MM) and without (CHM) HindIII restriction, which are listed for each primer pair in table 4, 5 and 6 annex IV. Only for the primers *05A1MM*, *11A2MM*, *11A2CH*, *11A3MM*, *11A4MM* and *11A4CH* in the MM samples, sometimes also in the progeny, an additional larger fragment than expected was noticed. Also there was no amplified PCR product observed for the CHM samples of the primers *11A3MM*, *11A3CH*, *11A4MM* and *11A4CH*. Another remarkable finding was a slightly larger fragment than predicted for the MM pattern of primer *05A3MM*.
Despite the three BC_5 plantnumbers, also DNA isolated from [PV101318_A1] (BC_4) was incorporated in the test allowing verification, through this CAPS marker based methode, of the data obtained from analysis by VHL (dark square tabel 10).

Overall the testsresults provided the following information (table 10) about the location of heterozygous and homozygous introgression regions within the studied plantnumbers. Completion of

this table has been acchieved from information of several separate tests, in which not all samples performed equally well, therefore the data dispalyed in this table is a summary of multiple tests.

Table 10 Comparison of the available marker data of plants [PV101318_A1] (BC_4) and [PV121022_113] (BC_4S_1) [dark square], with new CAPS marker data tested on the progeny [PV131567A], [PV131567B] and [PV131568A] (BC_5) and as control on [PV101318_A1] (BC_4) material. For each seq-md number the genomic position and the final result (*pink*/*light green*/*dark green*) is displayed. *Pink* indicates a region homozygous for MM, *light green* homozygous CHM whereas *dark green* corresponds with a heterozygous introgression region. When a certain combination has not been tested a '-' sign is visible in the table. In case the combination has been tested but no amplified product was obtained a '*' is placed.

Chromosome	Position (HindIII site)	Seq-md# [position]	PV101318_A1	PV121022_113	Parental specific primer (name)	PV101318_A1		PV131567A		PV131567B		PV131568A	
2	4,949,427	None	-	-	B1-MM CHM	M H	**H**	M C	**H**	M H	**H**	* *	*
	20,002,287	None	-	-	B2-MM CHM	* H	**H**	H H	**H**	M M	**MM**	M *	**M***
	28,464,819	5719 [28,520,428]			B3-MM CHM	- -	-	- -	-	- -	-	- -	-
	42,589,256	6683 [42,549,116]			B4-MM CHM	- -	-	- -	-	- -	-	- -	-
5	2,995,300	None	-	-	A1-MM CHM	M M	**MM**	M M	**MM**	M M	**MM**	* *	*
	6,467,256	7559 [6,461,255]			A2-MM CHM	- -	-	- -	-	- -	-	- -	-
	20,548,353	1377 [20,527,748]			A3-MM CHM	M M	**MM**	M M	**MM**	M M	**MM**	* *	*
	39,994,430	1880 [43,744,992]		-	A4-MM CHM	* *	*	H C	**H**	H C	**H**	* *	*
11	2,004,495	None	-	-	A1-MM CHM	* *	*	M M	**MM**	M M	**MM**	* *	*
	5,276,912	8439 [5,279,605]			A2-MM CHM	* *	*	M C	**H**	M *	**M***	* *	*
	6,522,663	1693 [6,497,192]			A3-MM CHM	* *	*	M *	**M***	M *	**M***	* *	*
	51,096,119	7969 [51,097,323]			A4-MM CHM	* *	*	M M	**MM**	M M	**MM**	* *	*

Analysing the results from table 10 reveals some extraordinary findings:

- Firstly, regarding the comparison with the VHL data:
  - Only one position, chromosome 5 - 20,527,748 bp, out of the eight tested by VHL could be verified through this new CAPS based method
  - Unfortunately this only hit 'MM' did not match the 'H' data for [PV101318_A1] obtained by VHL at this position
- In addition some of the introgression origin for the progeny [PV131567A] and [PV131567B] did not correspond with the expected pattern, predicted through the VHL data and the known pedigree of this population (figure 1). For instance primer pair:
  - ***05A3***: 'H' is expected for [PV131567A] and [PV131567B] instead of 'MM'
  - ***05A4***: [PV131567A] and [PV131567B] should be 'MM' instead of 'H'
  - ***11A2***: the VHL data itself was already contradictory (orange box table 3). In case the [PV101318_A1] data 'MM' is indeed correct we expect a 'MM' pattern for [PV131567A] and [PV131567B] instead of 'H'. But if the [PV121022_113] 'CC' was

true, than ‘H’ is expected for [PV131567A] and [PV131567B], which is still possible for ‘M*’
    - ***11A3***: prediction is ‘H’ for [PV131567A] and [PV131567B], which is still plausible for ‘M*’ especially when amplification of CHM did not perform well
    - ***11A4***: expect ‘H’ for [PV131567A] and [PV131567B] instead of ‘MM’
- For [PV131568A] only one hit was obtained ‘M*’ while for all other primer combinations no amplification has been observed

2.6 Additional designed CAPS markers

To obtain more detailed information about introgression segments in multiple regions of the chromosomes previously studied, new CAPS markers were developed according to the same method used before. Due to difficulties in testing of the first collection of CAPS markers, it was not possible to study this second group of CAPS markers during this master thesis. Nevertheless their position and corresponding primer pairs are listed in table 11 annex V.

Discussion

Fine-mapping and size reduction of the IL5B region derived from the cross between the commercial material *S. lycopersicum* cv. 'Moneymaker' and the wild accession *S. chmielewskii* LA1840, through the development of CAPS markers was the main goal of this research. Such new CAPS markers would allow selection of seedlings based on the presence or absence of certain introgression segments, making determination and separation of those regions possible.
Initially the research focus was mainly laid on determination of the target introgression regions, containing candidate genes coding for interesting metabolites, making screening of them among the progeny possible. Also certain phenotypic characteristics (like metabolic composition) of the selected plants would be tested in this study, which would allow a rough pre-screening of the recombinants available and provided information about logical up-following research goals.
Due to these main interests the development of the new CAPS method seemed to be of less importance and was expected to be quite simple. Unfortunately during this study it became clear that developing of such a CAPS based method was not as easy as thought before. Several drawbacks have been faced, alternatives are investigated but, no correct working CAPS method could be designed. Nevertheless performance of this research did result in several recommendations, based on the characteristics tested and the results obtained. Some of these findings, like for instance; reducing the size of the PCR fragments, start over a new trial by using fresh materials or alteration of the primer design method, will be described in more detail together with corresponding examples as well as substantiations during the next paragraphs. Also the research possibilities with the plant populations still available will be described more specific in the final parts of this report.
This master thesis will be completed through summarization of the main recommendations and description of the overall conclusion. Thereby the findings from this study might serve as a fundament in order to develop new and more extensive future researches within this area.

3.1 Deviating results

In general the CAPS based method did not provided as clear results as expected and several drawbacks occurred. These findings raised the question if large scale application of a new CAPS based screening method will be indeed functional. Not only specification of the test conditions, Tm and parental specific sequences, for each investigated primer pair seemed to be a limiting factor (figure 4 and table 9) in order to obtain detectable PCR fragments. Also the comparison of the introgression outcomes from this research, with the data of Infinium Bead Array analysis by VHL, did not match. Especially when the known pedigree of the plants [PV131567A] and [PV131567B] (figure 1) was taken into account, it was obvious that certain hits (primers: *05A3*, *05A4*, *11A2*, *11A3* and *11A4*) were not logical. Within the result section of this report, the expected and observed introgression pattern for these deviating primer pairs has been described.
Another remarkable finding was the lack of amplification for the third BC_5 plant [PV131568A]. A possible explanation for this phenomenon could be a low quality of the DNA isolated from [PV131568A]. Therefore the advice is to repeat this test with fresh isolated material and to use a known tomato household gene as positive control. Explanation of the single hit [PV131568A] gave for primer pair *02B2MM*, is thought to be due to overflow of material from the adjacent MM sample.

The main question based on these extraordinary results is which detection method can be regarded as the most reliable one, the new CAPS based method or the Infinium Bead Array analysis performed by VHL? In addition it is uncertain if the differences in introgression results can indeed be prescribed towards the use of two detection methods? If it does, might alteration of the CAPS based method lead to better performance of the technique, making analysis of introgression segments possible? After all a fundamental difference between the Infinium Bead Array analysis, with regard to the method used during this study, is the length of the amplified fragments. Whereas the goal in this

research has been amplification of 900-1000 bp, VHL set a target range of 41-95 bp product size for the Infinium Bead Array analysis. This reduction in PCR fragment size will induce a relative higher specificity to detect sequence alterations due to SNPs. Therefore scaling towards 1 kb products might be in favour of detecting homozygous segments instead of heterozygous introgressions. Whereas VHL could be capable to distinguish better between the amplification of MM and CHM material, resulting in an improved detection of both true homozygotes and heterozygotes, instead of multiple false positives. Therefore alteration of the CAPS based method, towards amplification of smaller PCR fragments might be worthwhile to investigate in more detail. Perhaps this hypothesis of increased specificity due to size reduction, could be verified and will result in an improved detection method.

3.2 Relevance to continue with related research

As has been described in the material and methods, several other plant populations are available for investigation of the introgression segments present. Although a suitable technique for this has not been developed during this study, the potential research goals of the populations left for future investigations will be discussed as well. These possibilities will show the relevance to develop a well-functioning (CAPS based) screening technique through which such plant populations can be studied in more detail. The resulting information, about the origin of introgression segments present, as well as the additional findings about the metabolic pathways involved, will allow development of markers correlated to certain metabolic characteristics within this process. This will enable MAS, resulting in breeding of tomatoes with for instance an improved flavonoid content.

3.2.1 Future research objectives

Roughly two different goals to design new markers, concerning the available plant populations, can be identified:

1) To improve the determination and eventually separation of introgression segments distributed over several chromosomes. Therefore future studies should be focussed on chromosome 2, 5 and 11 which requires further in-depth research. Whereas the introgression on chromosome 12 also need to be investigated.
2) Secondly, the introgression present on chromosome 5, particularly surrounding the IL5B region, should be further investigated through new developed CAPS markers. This IL5B area can be divided into investigation of the CHI and MYB regions, although the segments up- and downstream these areas should be studied as well.

3.2.1.1 MYB and CHI regions

Recombinants of the three selected BC_4S_3 lines [PV131510], [PV131513] and [PV131521] (figure 2), containing an introgression segment between the MYB and CHI region, can be used in future research to study the influence of each region on the flavonoid pathway. Especially since all selected recombinants are CHM-MYB but varies in CHI (MM or CHM) (table 12 [1] and [2]).
During such investigations [PV121454] can serve as MM control (for MYB and CHI) while [PV121453] will fulfil a role as MM-MYB and CHM-CHI control (table 12).

Table 12 Overview of the different selected BC lines available to fine-map the MYB and CHI segments of the IL5B region. Positive sign (*+*) indicates homozygosis for the listed introgression segment. The research aim of each genotype is shortly described in the last column.

Plant number	**CHM**		**MM**		**Study metabolic effect of**
	MYB	*CHI*	*MYB*	*CHI*	
PV121453		+	+		CHM-CHI
PV121454			+	+	MM control
PV131510, PV131513, PV131521 [1]	+	+			CHM-MYB and CHM-CHI
PV131510, PV131513, PV131521 [2]	+			+	CHM-MYB

3.2.1.2 Up- and downstream areas

Recombinant screening of the CHM introgression segment downstream the CHI region can be done by screening of the seedlings from [PV131567A], [PV131567B] and [PV131568A] (figure 1) with new CAPS markers. While studying the seedlings from [PV131504A/B], [PV131506A/B] and [PV131507A/B] (figure 2) might reveal information about recombination of the CHM introgression segment upstream the MYB region. Combined CAPS research of these two plant populations will result in fine-mapping of the IL5B region and selection of interesting recombinants. Selfing of such recombinants will provide proper sets of Near Isogenic lines (NILs) for these interesting introgression areas.

3.2.2 Metabolic analysis

When well-functioning markers have been developed, the following research subject will be determination of the metabolic profile of the studied plants. Obtaining such metabolic data for each line containing interesting introgression regions and combining it with the genomic data from CAPS researches, will provide information about genotype-phenotype linkage. Analysis techniques like the High-Performance Liquid Chromatography Mass Spectrometer (HPLC-MS) enables studying of (differences in) the metabolic profile between tomato fruits from several selected lines. Output from such analysis can been compared with known mass-values from a computer database to characterise the available metabolites. Determination of significant differences in relative metabolic content between two samples is possible through comparison of the area under the peak (using SPSS, two sample t-test $df=n_1-n_2-2$ when $n_1 = n_2$).

These kind of studies will encourage the discovery of candidate genes influencing certain metabolic pathways and compositions. This knowledge as well as available CAPS markers, will stimulate selection for these improved or altered metabolic compositions during future breeding programmes.

Conclusion

General conclusion

The initial focus of this master thesis intended to be determination of the origin and location of multiple introgression regions containing candidate genes on three chromosomes, as well as the metabolic influence these regions have. Although afterwards, it became clear that the research goal has shifted during this investigation, indicating that development of a well-functioning CAPS based method should be the first priority in order to allow determination of the regions searched for. Designing such a technique was thought to be quite simple, but obviously this research component has been underestimated at the beginning of this study. Therefore it can be stated that development of an indeed well-functioning method, based on CAPS, requires further investigation.

In order to do this it is recommended to base the design of future studies solely on the development of such a technique. In which the primary research goal have to be well-functioning of the method, eventually allowing screening of the available plant populations.
Multiple aspects of these populations can be studied through such a technique. Not only additional introgression segments on several chromosomes can be determined, also fine-mapping of certain interesting regions such as the IL5B will be possible through proper functioning CAPS markers.
When these markers indeed function the way they should, it is possible to obtain a wide variation of genetic information for several interesting positions. This available genetic knowledge will allow linkage of it with, through metabolic analysis obtained, phenotypic information. Such more advanced studies will be the basis of unravelling the metabolic pathways involved and provide clues for the genes interacting in these processes.

Knowledge about this process and the availability of markers correlated to the genes responsible for these metabolic characteristics, allows performance of MAS. Which will stimulate breeding of tomatoes, focussed on their metabolic composition, beneficial for human health.

Recommendations

Although development of a well-functioning CAPS based technique did not succeeded during this research, several findings throughout this study might function as recommendations which can be altered during future investigations;

1) Fresh materials
   Starting a new experiment using fresh material might improve the obtained results. Perhaps new isolation of DNA from [PV131568A] does lead to amplification of the tested primers.
2) Intensity of gel electrophoresis results
   In order to improve the intensity of the gel results it is advised to increase the amount of PCR product placed on gel. This will provide more clear results and is beneficial for scoring of the obtained fragments. Also Ethidium Bromide gels instead of Gel Red loading buffer improved the ease of scoring the right fragment size of the PCR products obtained.
3) Separately test parental specific primers, instead of using primer pair mixtures
   As has been showed during this study, mixing of parental specific primers into one mixture did not result into amplification. Whereas separate testing of both parental primers did provided PCR fragments. Therefore it is suggested to continue with separate parental specific primer mixtures, although this will increase the amount of work.
4) Reduce the size of the PCR fragment
   As with the analysis performed by VHL, leading towards proper introgression results. It is suggested to study in more detail the effect of a reduced PCR fragment size. Testing products of only a few hundreds of base pairs might indicate if size reduction does indeed improve the specificity to detect the origin of introgression segments. Hopefully this phenomenon will

reduce the number of false positive hits and will provide a proper estimation of the homo- as well as heterozygous regions present

5) Alter the primer design

   As addition to the size reduction of the PCR fragments, it is also possible to incorporate aspects of the primer design used by VHL. Although each SNP specific primer is fluorescently labelled during the Infinium Bead Array analysis, the detection of homo- and heterozygous samples for the new CAPS technique should be based on the restriction pattern. Therefore a consistent primer design with a SNP, between the parents MM and CHM, located in the restriction site of an enzyme is required. Development of two primers, one for each SNP-variant will allow amplification for the MM as well as the CHM type. For one of these two different amplification products restriction is expected to occur, while the other one will remain undigested. Based on the pattern observed it is still possible to discriminate each of the three possible genotype outcomes, without the need of a specific fluorescent detector.

   The main difference between this altered technique and the method used within this research, will be the location of the primer. Whereas now the primer was located several hundreds of base pairs in front of the SNP, the new variant will use a primer which includes the SNP itself, thereby causing two separate primers specific to either one of the parents, as well as size reduction of the PCR fragment.

6) Change of technique or even outsourcing of the research

   In case alteration of the CAPS based method will not improve the results, making detection of the present introgression segments impossible, another technique should be applied to obtain the necessary information. Infinium Bead Array analysis as has been performed previously by VHL might be an alternative, although such techniques will increase the research costs. In addition, using this method for detection does no longer require SNPs within a restriction site of an enzyme. The only requirement left will be occurrence of the SNP in only one of both parents.

When more research will be performed and a well-functioning CAPS based method will be designed, it is possible to determine candidate genes and their influences on the metabolic pathway, concerning interesting flavonoids. These studies will allow discovery of markers which will enable large scale performance of such a screening technique through MAS. This development might be the first step towards breeding for increased flavonoid content and will encourage the development of new tomato cultivars with an improved taste and nutritional value.

Acknowledgements

First of all I would like to thank Arnaud Bovy and Sjaak van Heusden for the fast arrangement of a possibility to perform my master thesis research at this group within the department Plant Breeding. Within a few days I already received an enthusiastic reaction of them, in which they described the research subject they came up with. Although multiple new aspects came along during this research, which gave me the opportunity to learn a lot, Sjaak and Arnaud were always available for questions, discussions or recommendations. Due to this attitude, their comments, time and support, it was very pleasant to cooperate with them during this research.

A special acknowledgement goes to Richard Finkers, who developed the file containing the HindIII sites differing between the two parents. Due to this bioinformatics data it was possible to develop CAPS markers easier than through manual screening of certain regions.

Thanks to Marcela Viquez-Zamora, Fien Meijer-Dekens and Jos Molthoff I became familiar with performance of new techniques in the laboratory. Although several times the lab results were not as expected, they always assisted in finding a proper solution and they came up with new ideas to test certain separate elements in more detail. Unfortunately not all 'mysteries' could be solved, nevertheless they have taught me a lot and I am very thankful for their supervision and participation during this research.

Also involvement in the weekly research meetings was very interesting. The explanations and advises provided during these meetings did stimulate a more critical stance. And in addition I obtained more knowledge about the factors and processes involved in quality improving breeding and research. These meetings as well as several other aspects of my master thesis, certainly did contribute to an improved orientation within my further study but also with regards to a future career.

References

Adato A, Mandel T, Mintz-Oron S, Venger I, Levy D, Yativ M, Domínguez E, Wang Z, Vos RCH de, Jetter R, Schreiber L, Heredia A, Rogachev I, Aharoni A (2009) Fruit-Surface Flavonoid Accumulation in Tomato Is Controlled by a SlMYB12-Regulated Transcriptional Network. PLoS Genet 5:e1000777

Aging WUnIo (2011) Global health and aging:32

Ballester A-R, Molthoff J, Vos R de, Hekkert BtL, Orzaez D, Fernández-Moreno J-P, Tripodi P, Grandillo S, Martin C, Heldens J, Ykema M, Granell A, Bovy A (2010) Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the Gene Encoding Transcription Factor SlMYB12 Leads to Pink Tomato Fruit Color. Plant Physiology 152:71-84

Ballester A-R, Tikunov Y, Molthoff J, Grandillo S, Vos R de, Bovy AG (2013) Identification of loci and putative genes affecting semi-polar metabolites in tomato fruit

Bovy A, Schijlen E, Hall RD (2007) Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics. Metabolomics 3:399-412

Consortium T (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635-641

Finkers R (2013)

Greenberg ER, Baron JA, Tosteson TD, Freeman Jr DH, Beck GJ, Bond JH, Colacchio TA, Coller JA, Frankl HD, Haile RW (1994) A clinical trial of antioxidant vitamins to prevent colorectal adenoma. New England Journal of Medicine 331:141-147

Jez JM, Bowman ME, Noel JP (2002) Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. Biochemistry 41:5168-5176

Kong J-M, Chia L-S, Goh N-K, Chia T-F, Brouillard R (2003) Analysis and biological activities of anthocyanins. Phytochemistry 64:923-933

Micka T (2013) The role of MYB transcription factor (Solyc05g009720.1.1) and Chalcone-flavonone isomerase 1 (Solyc05g010320.1.1) involved in phenylpropanoid pathway: Their effect on increase of polyphenolic compounds in tomato fruits.

Muir SR, Collins GJ, Robinson S, Hughes S, Bovy A, Vos CR de, Tunen AJ van, Verhoeyen ME (2001) Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. Nature biotechnology 19:470-474

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3 Nucleic Acids Research.

Verhoeyen M, Bovy A, Collins G, Muir S, Robinson S, Vos CR de, Colliver S (2002) Increasing antioxidant levels in tomatoes through modification of the flavonoid biosynthetic pathway. Journal of Experimental Botany 53:2099-2106

Willits MG, Kramer CM, Prata RT, Luca V de, Potter BG, Steffens JC, Graser G (2005) Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. Journal of agricultural and food chemistry 53:1231-1236

WUR, BGI (2013) Tomato variant explorer.

Annex-I DNA isolation CTAB method

1) Switch water bath on at 65°C.
2) Prepare buffer immediately before use [table below]. Put buffer at 65°C to solve ingredients.
3) Take the 96 wells plate from the -80°C put it on N2 and put the material in the plate grinder. Minimal 45 sec on at least 25 rpm. Put the plate on ice (not to long) and pre-warm the tube lids. Remove lids. **Be sure the material remains frozen**.
4) Add 450 µl buffer to the leave powder and vortex (or shake) very well, put it back into the plate grinder for 45 sec.
5) Put the plates in the 4-points clamps and incubate in a water bath at 65°C for 30 till 60 minutes. In between mix it a few times. Afterwards cool in ice water in the clamps!
6) In the fume hood, extract with 450 µl chloroform:iso-amylalcohol (24:1) (SEVAG) put it back in the clamps to keep the lids close when you shake them.
7) Mix well for 5 minutes.
8) Spin for at least 30 minutes at 2500 rpm. Longer seems better, between 30 min and 1 hour.
9) Pipet off 400-450 µl aqueous phase.
10) Add in the fume hood an equal volume cold isopropanl and invert tubes repeatedly until DNA precipitates. **DO NOT VORTEX!**
11) Spin for 30 minutes at 2500 rpm. Between 30 min and 1 hour.
12) Poor off supernatant into an isopropanol waste container.
13) Add 500 µl ice cold 70% ethanol and mix well by inverting the tubes.
14) Spin 30 minutes at 2500 rpm
15) Discard the ethanol.
16) Spin 5 minutes at 2500 rpm.
17) Careful remove the remaining ethanol.
18) Dry pellet at 37°C for approx. 15-30 minutes. Pellet should be dry so no ethanol is left.
19) Re-suspend DNA in 150 µl (pre-warmed) 10 mM Tris (pH 7,5-8,0) and RNAse (10µg/ml)
20) Put the solution in the 37°C for 30 minutes to dissolve the pellet.

Buffer stock solutions:

Extraction buffer stock: 0.35M Sorbitol 0.1M Tris-HCl pH 8.0 5mM EDTA pH 8.0	**Lysis buffer stock:** 0.2 M Tris-HCl pH 8.0 50 mM EDTA pH 8.0 2 M NaCl 2% CTAB	**Sarcosyl stock** 5% (W/V)
For 500 ml: 31.9 g Sorbitol 50 ml 1M Tris-HCl pH 8.0 5 ml 0.5M EDTA pH8.0 Fill up to 500 ml with mQ	**For 500 ml:** 100 ml 1M Tris-HCl pH 8.0 50 ml 0.5M EDTA pH8.0 200 ml 5 M NaCl 10 gr CTAB Fill up to 500 ml with mQ	**SEVAG**: For 500 ml: 240 ml Chloroform 10ml Iso-propanol. **TE buffer**: 10 mM Tris-HCl pH8.0 1mM EDTA pH8.0

Pipet solution in order as written below. After the extraction buffer, shake the tube well before adding the other two buffers.

		Use this one for 1 plate.
	40 ml - *80 samples*	**60 ml** - *120 samples*
Sodiumdisulfite *(=sodium metabisulfite)*	0.2 g	0.3 g
PVP-40 *(K29-32) Sigma*	0.8 g	1.2 g
Extraction buffer	16,7 ml	25 ml
Lysis buffer	16,7 ml	25 ml
Sarcosyl	6,7 ml	10 ml

Annex-II PCR reaction (DT polymerase and buffer)

1) Prepare primer pair mixtures according to the table below

	1x F + 1x R	1x F + 2x R	2x F + 1x R	2x F + 2xR
Stock F	10 ul	10 ul	10 ul (F-MM) 10 ul (F-CH)	10 ul (F-MM) 10 ul (F-CH)
Stock R	10 ul	10 ul (R-MM) 10 ul (R-CH)	10 ul	10 ul (R-MM) 10 ul (R-CH)
MQ	80 ul	70 ul	70 ul	60 ul
Total	100 ul	100 ul	100 ul	100 ul

2) Prepare a master mix for the number of samples necessary, according to the table below (the amount is given for one sample only, per primer mixture combination). The total volume of PCR product will be 20 ul eventually.

	1xF + 1xR	2xF + 1xR 1xF + 2xR	2x F + 2xR
DT buffer	2 ul	2 ul	2 ul
DT polymerase	0,5 ul	0,5 ul	0,5 ul
dNTPs	0,2 ul	0,2 ul	0,2 ul
MQ	15,3 ul	14,8 ul	13,3 ul
total	***18 ul***	***17,5 ul***	***17 ul***

3) Add to each master mixture the right amount of each primer pair mixture. For (1xF + 1xR) this means 1 ul per sample, for (2xF + 1xR) or (1xF + 2xR) 1,5 ul per sample and for (2xF + 2XR) 2ul per sample should be added.
4) After mixing, 19 ul of each mixture should be divided into each well of the PCR plate
5) Add to each well 1ul of genomic DNA (with a concentration of approximately 200 ng/ul)
6) The plate will be sealed, mixed and spin down. The PCR programme used is:
   94°C for 1:00 min
   [94°C for 0:30 min, 60°C for 0:30 min, 72°C for 0:30 min] **(30 cycles)**
   72°C for 2:00 min
   10°C for ∞
7) To load the PCR product on gel it is necessary to prepare a master mix of MQ and loading buffer (LB) appropriate for EtBr gel-electrophoresis. For a total volume of 10 ul the following proportions will be used for one sample: 2 ul LB + 5 ul of PCR product + 5 ul MQ. Mix well and place 10 ul from this mixture on gel (1% TAE EtBr gel)
8) 10 ul of genomic DNA can be used as reference in the proportion of: 1 ul gDNA + 2 ul LB + 9 ul LB and placed on gel as well. Run for approximately 1h at 95 V. Use a 1 kb Plus DNA ladder from GeneRuler™ as reference.

Annex-III HindIII restriction

1) Prepare a mastermix **(one ice)** of: 1 ul Buffer enzyme (Red) [10x] + 1 ul HindIII + 5 ul MQ for one sample. Put 7 ul in each well of a new PCR plate.
2) Add 3 ul PCR product to each well and mix
3) Spin down and place the new PCR plate for 1h at 37°C
4) Add 2 ul LB to each well, mix and put 10 ul from this mixture on a (1% TAE EtBr) gel. Run for approximately 1h at 95 V. Use a 1 kb Plus DNA ladder from GeneRuler™ as reference.

Annex IV Primer pairs [first batch]

Table 4 Chromosome 2, for each lane the origin of the DNA tested is indicated, either MM or CH. The expected fragment lengths are given in number of base pairs.

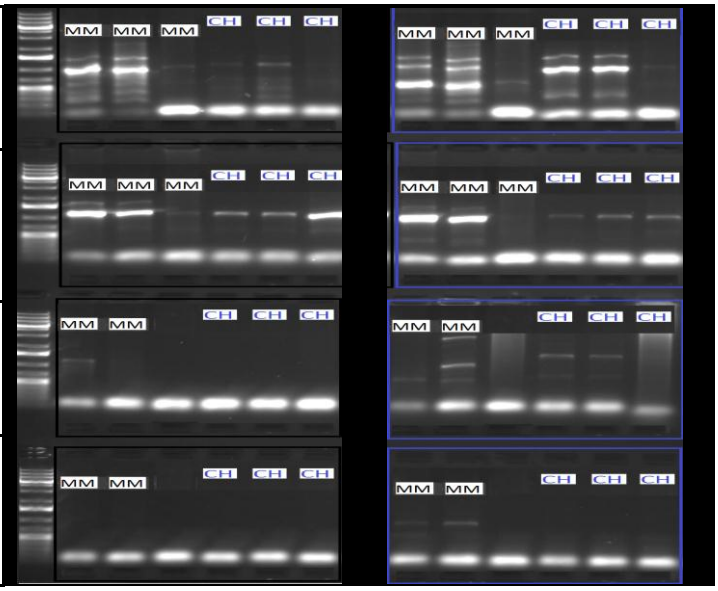
Name	Restriction pattern expected		Observed pattern of PCR product without HindIII restriction	
	MM	*CHM*	*MM specific primers*	*CHM specific primers*
B1	745 and 329	1074		
B2	721 and 358	1079		
B3	847 and 238	1085		
B4	713 and 203	916		

Table 5 Chromosome 5, for each lane the origin of the DNA tested is indicated, either MM or CH. The last lane of each square, referred as pp, only lacks DNA. The expected fragment lengths are given in number of base pairs.

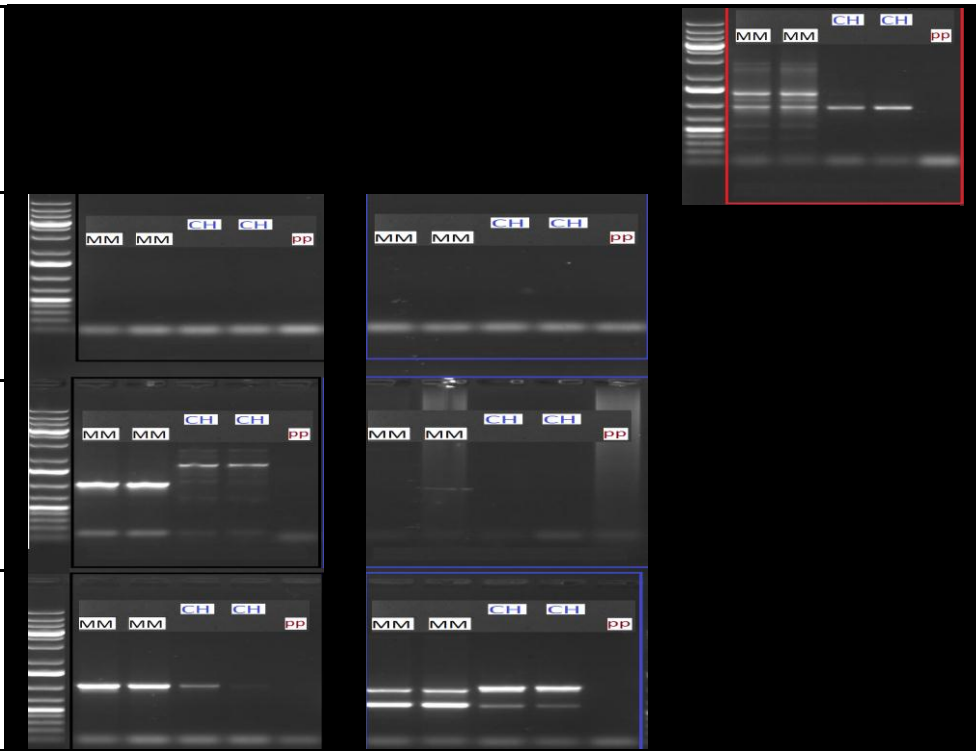
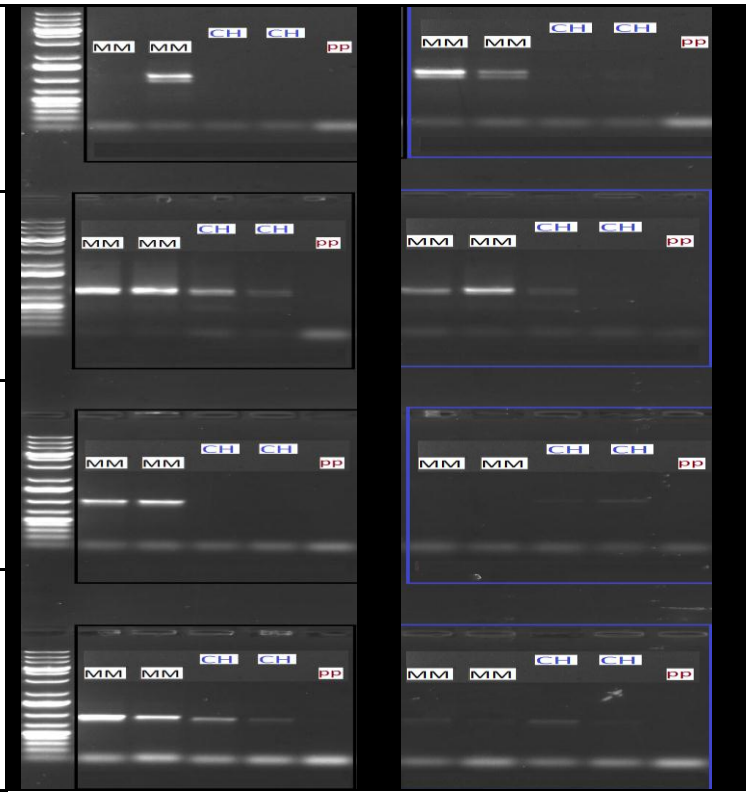
Name	Restriction pattern expected		Observed pattern of PCR product without HindIII restriction		
	MM	*CHM*	*MM specific primers*	*CHM specific primers*	*Non specific*
A1	687 and 274	961			
A2	599 and 325	924			
A3	783 and 272	1055			
A4	791 and 282	1073			

Table 6 Chromosome 11, for each lane the origin of the DNA tested is indicated, either MM or CH. The last lane of each square, referred as pp, only lacks DNA. The expected fragment lengths are given in number of base pairs.

Name	**Restriction pattern expected**		**Observed pattern of PCR product without HindIII restriction**	
	MM	*CHM*	*MM specific primers*	*CHM specific primers*
A1	657 and 433	1090		
A2	628 and 279	907		
A3	735 and 277	1012		
A4	539 and 392	931		

Annex V Additional designed CAPS markers

Table 11 New developed CAPS markers distributed among the chromosomes 2, 5, 11 and 12, allowing more detailed studying of certain genomic positions for presence of introgression segments. For each position the designed forward (F) and reverse (R) primer are displayed and if necessary divided into a MM or CHM specific primer. A " means specification for each parent was not needed according to the genomic sequences displayed in JBrowse. None of these new developed CAPS primers has been tested during this research due to time constrictions.

Chromosome	Position (HindIII site)	Name	F primer	R primer
2	30,000,898	C1- MM	TATGGTGAGCCTCCTTGCTT	TGACCGGACCTCATCATTTT
		CHM	"	TGACCGGATCTCATAATTTT
	32,085,310	C2- MM	ACCAGTCAACGGACCAGGTA	CCCTCGTGTGTTTGGTATCC
		CHM	ACCAGTCAAAGGACCAGGTA	"
	34,006,610	C3- MM	ACTTTGGAGTCGCTGGAAAA	CCAATTCGACAGGACAACCT
		CHM	"	"
	35,996,347	C4- MM	TTTTGGGGCATTAACAAAGG	CCATGATTTGTGTTGGATGG
		CHM	"	"
	38,007,491	C5- MM	CCTACACTCCTGGGATGGAA	CCTTTTTCATGCTGGCTCTC
		CHM	"	"
	40,055,397	C6- MM	TCCGACATCACTTCCCTTTC	CCGCCCATGTTCAGTTTATT
		CHM	"	"
5	3,494,843	B1- MM	CCACATTTTCGGCTATGCTT	CGCACGGATACTTGGAGAGT
		CHM	"	"
	4,027,675	B2- MM	TTCCCATTGTTCACATGGTC	GCAAAGTTGGAAAATGACGA
		CHM	"	GCAAAGTTGAAAAATGACTA
	4,506,316	B3- MM	GGGATAATGCCCAAGTACCTC	GGCTCCGCCACTATTTACAA
		CHM	GGGATAATGCCCAAGTACCCC	"
	5,022,687	B4- MM	AGCCAATGGCCTATCAAAAA	AACTGACCCCCGTTCTTTCT
		CHM	"	"
	5,503,493	B5- MM	CTTGAGCATGTTCAGCAACC	TAAGATACCGGGCTTGCCTA
		CHM	"	"
	6,000,083	B6- MM	TGTTGATTTTTCCCCTCAAG	GGTGGCTGCTCCATACAAAT
		CHM	TGTTGGTTTCTCCTCTTAAG	GGTGGCTGCTCCATATAAAT
	24,020,850	B7- MM	GCCACATAACCCCACTCAAT	GATGGCTTGTACTTGGTGATGA
		CHM	GCCCCATAACCCCACTTTAT	GATGGCTTGTACTTAGTTATGA
	28,001,603	B8- MM	GACCGCCAAAAGTTGAGAAG	GTGGTGTTTGCACCCAAGAT
		CHM	GACCATCAAAAGTCGAGAAG	"
	32,020,682	B9- MM	ATAACCATGCCTCCACGAAG	AGCGCATACAATAGCCCAAC
		CHM	"	"
	36,010,220	B10- MM	GCTTTCCCCCAGATCACATA	AGTGCCCAAGGCTCTAATCA
		CHM	"	"

11	5,487,953	B1- MM CHM	AAGGTGCTGGGAATTGTGTC "	TCCCCTCGGTTTTCTTCTTT "
	5,700,465	B2- MM CHM	CCGATTAATTTGTGGCGATT CCTATTAATTTGTGCTGATT	CGAACACATTAGGCATCCAG CGAACATATTAGGCATCCAG
	5,882,431	B3- MM CHM	GAACAACAGCGAATAAGTGAGC GAACAACAGCAAATAAGTGACT	CGTCGTCTTGCTTTTCTGG CGTCGTCTTGCTTTTCTAG
	6,125,181	B4- MM CHM	CGCCCATTTCATAAAGGGTA CACCCATTTCATAAAGAATA	CCCACTCCTAACCCATTTCA "
	6,350,889	B5- MM CHM	AAGGGACACTGGACTTGTGG AAGGGACACTAGACTTGTGG	CCAAACGACCCCTTATTTCA CCAAACGACCCCCTATTTCA
12	17,950,170	B1- MM CHM	GAGGGAGTTGAGGTTGATCG "	GGAAACCACCCAACTTTTGA "
	20,011,506	B2- MM CHM	TTTGACTTGGCGATCAAAGA "	GGTGTCCGAATTGGCTAAAA GGTGTCTGAATTGGCAAAAA
	22,011,644	B3- MM CHM	ATTAATGCCGACGAGGACAC TGTAATGCCGACAAGGCCAT	CCCGAAATCTTCCATAACCA "
	24,001,496	B4- MM CHM	TGAAGTGGGTGTTGTTTGGA TCAAGTGGGTGTTGTTTGGA	CCTATGATGCGCTGAGTCAA "
	25,981,069	B5- MM CHM	TGTTGTCTTAATGGGTGTGTCA TGTTGTCCCTATAGGTGTGTCA	ACTAGCACTCGTGGCTGGAT "
	27,997,659	B6- MM CHM	AGGCTTGACATGCTTTGGAG "	TGGGCGATCTCTCTTTGTCT TGGGCGGTCTCTATTTGTCT
	30,013,065	B7- MM CHM	TCATGCTCACAGGTCGTAGG "	CGATTTCATTGGCAGTAGCA "
	54,030,656	B8- MM CHM	GAATTTCCGGAGGTCTTTCC GAATTTCTGGAGGTCTTTCC	CTCTCACATGTCTCCGACCA "
	56,007,755	B9- MM CHM	TTGGATCTTGGATTCCCTTG "	ACATCAATCCAGGGCAGTGT ATATCATTCTAGGTCAGTGT
	57,972,553	B10- MM CHM	AAGATGGGCATAGTGGAACG "	CATCAGATGTCTCCGCTTCA "
	59,987,706	B11- MM CHM	AAAGAAGGGGAGGCTATGGA AAAGAAGGGGAGGCTGTGGA	CCCCGTGTTGATGATTTGAT "