2014•2015
FACULTEIT INDUSTRIËLE INGENIEURSWETENSCHAPPEN
*master in de industriële wetenschappen: biochemie*

# Masterproef

## Development of a high-throughput PCR & sequencing approach for SNP identification in *Leguminosae* species


Promotor :
ir. Myriam MEYERS

Promotor :
prof. dr. MARCOS EGEA-CORTINES
prof. dr. JULIA WEISS

Ine Maes


*Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: biochemie*

Gezamenlijke opleiding Universiteit Hasselt en KU Leuven

2014•2015

# Faculteit Industriële ingenieurswetenschappen

*master in de industriële wetenschappen: biochemie*

# Masterproef

## Development of a high-throughput PCR & sequencing approach for SNP identification in L*eguminosae* species

Promotor :
ir. Myriam MEYERS

Promotor :
prof. dr. MARCOS EGEA-CORTINES
prof. dr. JULIA WEISS

Ine Maes

*Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: biochemie*

universiteit hasselt | KU LEUVEN

## Preface

This thesis is made as a completion of my master's degree of industrial engineering in biochemistry at KULeuven/UHasselt. I had the opportunity to go on Erasmus and to make my master's thesis at the Universidad Politécnica de Cartagena in Spain.
Going on Erasmus was a real adventure for me and I did not know what to expect. Afterwards, I am really happy that I took this opportunity. Besides learning a lot in the lab, I also learned to be responsible and independent. My English and Spanish also improved a lot. And last but not least, I made a lot of new friends that made my four months abroad unforgettable. Looking back, I could not be happier that I took this chance to go abroad.
I would like to thank my internal supervisor ir. Myriam Meyers for all the help and support she has given me the past four months. She was always very pleased to answer my questions and she really helped me to create the best version of this thesis.
I really want to thank my external supervisors Prof. Dr. Julia Weiss and Prof. Dr. Marcos Egea-Cortines. They also were a very big support. They not only helped me with the experiments and the thesis, they also tried to made me feel at home. I could go to them with every question I had, even when it was not about the thesis. Thank you.
I also appreciate all the support I got from Marina, Victoria and Marta. You were always very pleased to help me in the lab if I had questions and you also assisted a lot in the beginning by giving a lot of information and tips about Cartagena. Thank you for that. I shared my Erasmus moments and lab experiences with Joke and I also want to thank her for being a support the past four months. And last but not least I want to thank my family and my friends at home. They have supported me and encouraged me to hold on during hard moments.

## Table of contents

## List of tables

## List of figures

## Abstract

Genotypes of *Pisum sativum* and *Vicia faba* with high protein content and resistance against diseases, heat and drought need to be identified. The genotypes may differ because of point mutations in the genome, single nucleotide polymorphisms (SNP's). 88 genotypes of *Pisum sativum* and 116 of *Vicia faba* should be examined to find the ones corresponding to the desired phenotype. This work can be minimalized by making a phylogenetic analysis based on SNP's in order to evaluate genetic variability.

The goal of this master's thesis is to optimize PCR and sequencing conditions to identify SNP's. Three thermocyclers are tested: TProfessional Thermocycler (Biometra), GeneAmp PCR System 9700 (Applied Biosystems) and Stratagene Robocycler (Agilent technologies). Three PCR kits are tested: MyTaq DNA Polymerase kit (Bioline), PrimeSTAR GXL DNA Polymerase kit (Clontech) and OneTaq DNA Polymerase kit (New England Biolabs). 60 SNP sequences for *Vicia faba* and 45 for *Pisum sativum* are selected and primers are designed. In a first PCR, these DNA fragments are amplified and the first adaptor is added. A second PCR is done to elongate the DNA fragment with a second adaptor sequence as a preparation for sequencing. SNP's are identified using Ion Torrent sequencing, which includes library preparation and emulsion PCR (emPCR). The ideal concentration of the amplicon library for emPCR is determined.

Optimal PCR conditions are obtained with the TProfessional Thermocycler and OneTaq DNA Polymerase kit. 35,5 % of the primers for *Pisum sativum* and 28,3 % for *Vicia faba* did not work and should be redesigned. The ideal library concentration for emPCR is 13 pM.

## Abstract in het Nederlands

Genotypen van *Pisum sativum* en *Vicia faba* met een hoog eiwitgehalte en resistentie tegen ziektes, warmte en droogte moeten geïdentificeerd worden. Deze genotypen kunnen verschillend zijn omwille van puntmutaties, single nucleotide polymorphisms (SNP's). 88 genotypen van *Pisum sativum* en 116 van *Vicia faba* moeten onderzocht worden om degene te vinden die overeenkomt met het gewenste fenotype. Het werk wordt versneld door een fylogenetische analyse te maken op basis van SNP's, zodat genetische variabiliteit geëvalueerd kan worden.

Het doel van deze masterproef is het optimaliseren van de PCR- en sequencingcondities om SNP's te identificeren. Drie thermocyclers worden getest: TProfessional Thermocycler (Biometra), GeneAmp® PCR System 9700 (Applied Biosystems), en Stratagene Robocycler (Agilent technologies). Drie PCR-kits worden getest: : MyTaq™ DNA Polymerase kit (Bioline), Primestar GXL kit (Clontech), en OneTaq® DNA Polymerase kit (New England Biolabs).
60 SNP-sequenties voor *Vicia faba* en 45 voor *Pisum sativum* zijn geselecteerd en primers zijn ontworpen. In een eerste PCR worden de DNA-fragmenten geamplificeerd en verlengd met een eerste adaptorsequentie. Een tweede PCR verlengt de fragmenten met een tweede adaptorsequentie, ter voorbereiding van het sequensen. De SNP's worden geïdentificeerd via Ion Torrent Sequencing, wat een bibliotheek voorbereiding en een emulsie PCR (emPCR) inhoudt. De optimale concentratie van de amplicon bibliotheek voor de emPCR wordt gezocht.

Optimale PCR-condities worden verkregen met de TProfessional Thermocycler en OneTaq DNA Polymerase kit. 35,5 % van de primers voor *Pisum sativum* en 28,3% voor *Vicia faba* werkten niet en moeten herontworpen worden. De ideale bibliotheek concentratie voor de emPCR is 13 pM.

# 1 Eurolegume

This research is part of the project 'Eurolegume'. The European Union founded this project to deliver an updated biochemical, nutritional and morphological description of valuable genotypes. Also new feed and food products and biological methods to enhance the nutritive value of the residual biomass need to be developed [1]. Figure 1 shows the official logo of the project 'Eurolegume'.


**Figure 1: Project 'Eurolegume' [1]**

Eurolegume involves eighteen partners coming from ten European countries. UPCT in Spain is also a participant in the project. UPCT contributes in different tasks of the project:

- Evaluation of genetic diversity by molecular methods;
- Biological nitrogen fixation focused investigations;
- Transcriptomic characterization in drought stress;
- Sourcing and characterization of genetic diversity of local resources of cowpea, faba bean and pea and phenotyping selected accessions for site specific biotic and abiotic stresses;
- Evaluation of the quality of new feed and food;
- The multiuse of all three species in development of innovative food/feed, with special emphasis on sensor quality;
- Development of new packaging and processing techniques;
- Soil characteristics to optimize fertilization of the tested crops [1].

This research is located in the first task. The overall goal of this task is to find genotypes of *Pisum sativum* and *Vicia faba* that are able to grow in the climate of Spain. Especially in the south of Spain, a very dry and hot climate limits the growth. *Pisum sativum* and *Vicia faba* are species of the *Leguminosae* family. The selected plants need to be rich in protein and resistant against diseases, heat and drought [1].
*Vicia faba* and *Pisum sativum* have a lot of genotypes. It is time consuming to investigate all these genotypes to find the ones corresponding to the desired phenotype. Some of the collected accessions may be identical. It would save a lot of time and money if analysis of identical genotypes could be avoided [1].

# 2 Leguminosae

## 2.1 Biological nitrogen fixation

*Leguminosae* is the third largest family of flowering plants, after *Orchidaceae* and *Asteraceae*. This family has an agricultural and economic importance. Many species of *Leguminosae* are harvested as crops for human and animal consumption, for example beans and peas [2].

A unique aspect of the *Leguminosae* family is the biological nitrogen fixation. Nitrogen fixation refers to the conversion of atmospheric nitrogen gas $N_2$ into organic fixed nitrogen. 80% of the atmosphere is nitrogen gas. Living organisms can not use this $N_2$ because of the triple bond between the two N-atoms that makes $N_2$ very inert. All organisms need a nitrogen source to make amino acids and nucleic acids. Biological nitrogen fixation is a bacterial process that reduces $N_2$ into $NH_3$ and this $NH_3$ is directly converted into amino acids or nucleic acids [3] [4]. This conversion only happens with the help from nitrogen fixing bacteria, diazotrophs. They contain the enzyme nitrogenase. Bacteria belonging to the genera *Rhizobium*, *Mesorhizobium*, *Sinorhizobium*, *Bradyrhizobium* and *Azorhizobium* are diazotrophs and contain the enzyme nitrogenase [4].

Figure 2 shows the process of biological nitrogen fixation. When *Leguminosae* plants are in low nitrogen conditions, they release flavonoids into the soil. The concentration of flavonoids is the highest at the root hair [5]. These flavonoids are recognized by the diazotrophs, which results in induction of *nod* genes in these bacteria. Expression of these *nod* genes leads to excretion of the Nod factor. This factor activates the infection process. First, the root hairs of the plants curl and the nitrogen fixing bacteria are enclosed in the curl. The plant cell wall degrades and the cell membrane invaginates. An intracellular tubular structure (infection thread) is formed. The nitrogen fixing bacteria can enter the root hair through this tubular structure. The infection thread grows and contains bacteria with a layer of wall material from the plant [6]. Once the bacteria reach the root, they stimulate cell divisions in the cortex (outermost layer of the root). This leads to formation of a nodule. The bacteria lose their cell wall and the cell morphology changes to form large, branching cells: bacteroids. They are completely dependent on the host plant for energy [7]. The reaction in figure 2 is the halfreaction that occurs.

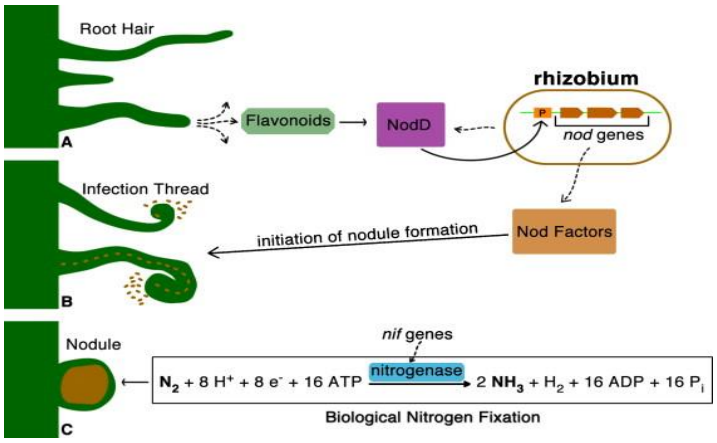



**Figure 2: Biological nitrogen fixation process [8]**

Figure 3 shows root nodules on root hairs, formed by nitrogen fixing bacteria.

**Figure 3: Root nodules formed by nitrogen fixing bacteria [4]**

The plant and the bacteria form a symbiotic relationship. The enzyme complex consists of two distinct proteins: dinitrogenase and dinitrogenase reductase. Dinitrogenase reductase contains iron. Dinitrogenase contains iron, molybdenum and an iron-molybdenum cofactor . The electron donor is ferredoxin or flavodoxin. These are low-potential iron-sulfur proteins [8]. Dinitrogenase reductase is reduced by receiving an electron from the electron donor, and then binds to 2 Mg•ATP. It forms a complex with dinitrogenase. It transfers the electron to dinitrogenase, with hydrolysis of 2 Mg•ATP to 2Mg•ADP + 2$P_i$. The complex dissociates and the process can be repeated. The reduction of $N_2$ to $NH_3$ demands six electrons. When dinitrogenase is completely reduced, dinitrogenase reduces $N_2$ to $NH_3$ [9].

Figure 4 shows the total reaction, catalyzed by nitrogenase.

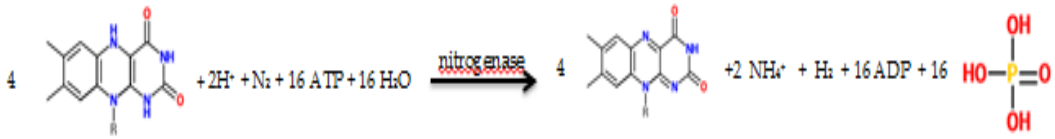



**Figure 4: Reaction catalyzed by nitrogenase [11] [12]**

Only six electrons are necessary to reduce $N_2$ to $NH_3$, but actually eight electrons are consumed. For each mole of $N_2$ that is reduced, two electrons are being lost as $H_2$. The reason for this reducing wastage is unknown [8].
The nitrogenase reaction requires a large amount of energy to convert $N_2$ into $NH_3$ (16 ATP). The plant supplies carbohydrates to the bacteria and in return, the plant receives some of the fixed nitrogen [5].

Some bacterium species are better in nitrogen fixation than others. Bacteria colonizing nodules of soybean, cowpea and faba bean are good nitrogen fixers. Therefore, these plants have high protein content [3].

## 2.2 *Glycine max* (soybean)

*Glycine max* or soybean is a species of the *Leguminosae* family. Soybean receives N from three sources: $N_2$ -fixation by *Bradyrhizobium*, $NO_3^-$ and $NH_4$ from the soil and N from fertilization [10]. It is the most important bean in the world, because of the high protein content. Dry soybean contains approximately 36% protein and the oil content is approximately 19%. It also contains 35% carbohydrates, 5% minerals and vitamins [11].

Soybean grows under warm conditions in tropical, subtropical and temperate climates. Growth rate decreases under 18°C and above 35°C. The minimum temperature for growth is 10°C. The crop can be grown on a wide range of soils, except very sandy soils [12]. Soybean is mostly produced in the United States. Other soybean producing countries are Brazil, Argentina, China and India [13].

Soybean meal is the by-product of the extraction of soybean oil. This soybean meal is the most important protein source used to feed farm animals. Animal feed plays an important role in the food chain. It can affect the quality of meat, eggs and milk. [14]. Using soybean for animal breeding leads to fast growth and profitable yields of these products [15]. The ingredients of animal feed can be divided in two groups: energy feed and protein feed. The most commonly used energy feed is corn [16].

Soybean also has some other applications. A first example is the use for human consumption. Soybean contains 18% oil. This oil can be used for cooking or frying. Another application is producing biodiesel out of soy oil through a transesterification

reaction. Soybean is also processed into soy milk, tofu, soy flour and other products. Because of the high protein content, soy products provide an alternative to meat [17].

The European Union imports up to 35 million tons of soybean per year. Most of it is imported from South America [18]. There is a huge demand of soybean in the EU, but soybean can not be cultivated under European climatic conditions [19].

## 2.3 *Pisum sativum* and *Vicia faba*

*Pisum sativum* is a species of the *Leguminosae* family. It is widely grown for its legume, peas. *Pisum sativum* is flowering from May to September, and the seeds ripe from July to October [20]. The flowers are hermaphrodite. Depending on the cultivar, flower color can be white, pink, lavender, blue or purple. The protein content of *Pisum sativum* is highly variable and is influenced by genetic and environmental factors. Protein contents can vary from 15,5 to 39,9 % [21]. The plant grows between 10°C and 20 °C. *Pisum sativum* is intolerant to drought, especially when it occurs in the flowering period [22].

*Vicia faba* is also a species of the *Leguminosae* family. The beans can be used as human food or animal feed. The dried seeds contain approximately 25% protein, 1.5% fat and 49% carbohydrate. The flowers are white or white with black/dark purple spots. The plant grows between 18 and 27°C [23].

Figures 5 & 6 show the plants of *Vicia faba L.* and *Pisum sativum L.*

**Figure 5:** ***Vicia faba*** **[50]**

**Figure 6:** ***Pisum sativum*** **[27]**

# 3 Quantitative genetics and molecular markers

## 3.1 Quantitative genetics

Phenotypic variation in a population is due to genetic and environmental differences. Quantitative genetics studies the nature of the genetic differences, the influences of the genetic and environmental factors and how the phenotypic variation translates into evolutionary change [24].
Those traits that are controlled by several genes in a quantitative way are said to be controlled by loci (QTL's). Examples of traits are: height, weight, number, disease presence or absence,...
In case of a quantitative trait, more than one locus is responsible for the variation between genotypes and these traits do not follow a Mendelian pattern of segregation [24].

It is important to find the QTL's to understand the variation between different genotypes. Molecular markers of Mendelian inheritance can help to localize a QTL on a genetic map.

## 3.2 Molecular markers

A molecular marker gives information about allelic variation at a given locus. It is a particular segment of DNA, representative for the differences at the genome level, found at specific spots in the genome. These markers may correlate with the phenotypic expression of a trait. Molecular markers can help to clear genetic variation. Detection and analysis of genetic variation can help to understand the molecular basis of various phenomena. It is expensive to sequence the genome of all genotypes, that is why molecular markers can help to reduce the work [25].

Since the markers and the genes they mark are close together, they tend to stay in linkage when a new generation of plants is produced. If the place of the markers and the distance to specific genes is identified, a genetic linkage map can be made. Such a linkage map can give a detailed analysis of associations between important traits and genes . Molecular markers are 'landmarks' in the genome. They are transmitted from one generation to the next one [26].

## 3.3 SNP's

DNA from 88 genotypes of *Pisum sativum* and 116 genotypes of *Vicia faba* collected in the south of Spain is available in the lab at the UPCT in Spain. The genotypes may differ because at certain places in the DNA sequence, a nucleotide is changed. These spots are called SNP's: single nucleotide polymorphisms. The polymorphism needs to have a frequency of at least 1% in a population before it can be considered as a SNP. In theory, SNP's can be bi-, tri-, or tetra-allelic polymorphisms but the chance that tri- and tetra-allelic polymorphisms appear is very low [27]. Figure 7 shows SNP's on

chromosomes of different genotypes (chromosome 1 is from genotype 1, chromosome 2 is from genotype 2,...). The genotypes are almost the same, except for certain spots where a nucleotide is changed: SNP's. The SNP's on figure 7 are all biallelic (only two possibilities for nucleotide). The purine nucleotide is changed for the other purine nucleotide (A for G or G for A), and the pyrimidine nucleotide is changed for the other pyrimidine nucleotide ( C for T, T for C) [28] .

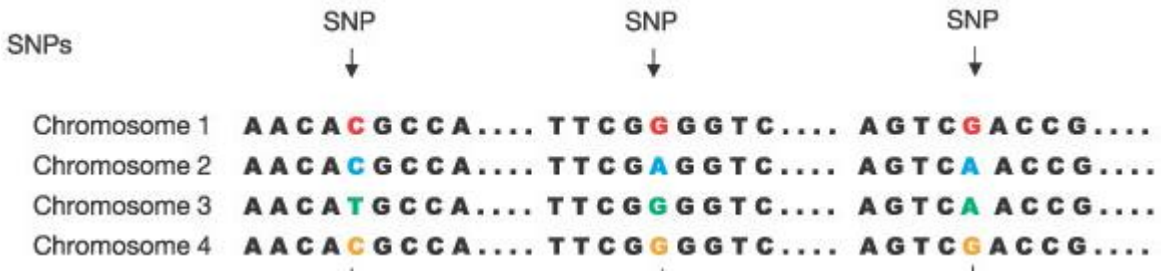



**Figure 7: SNP's on chromosomes from different genotypes [32]**

## 3.4 SNP's as molecular markers

SNP's play an important role in population genetics. The genotypes of *Vicia faba* and *Pisum sativum* with genes for high protein content and resistance against diseases, heat and drought need to be identified. It is very important to understand the relationship between these phenotypes (high protein content and resistance) and their individual genes. The position of these traits has to be known [29].

Modern linkage mapping with polymorphic DNA markers, such as SNP's, can help to find QTL's due to their Mendelian pattern of segregation. Polymorphic DNA markers are very useful for several reasons. They are variable, so a linkage map of DNA markers for high protein content and resistance can be constructed. This linkage map can then be used to determine the genetic distance between the markers. With DNA markers, the co-inheritance between a marker and a variable phenotypic trait can be traced to determine the genetic distance between marker and the gene for a trait. The polymorphic DNA markers can serve as landmarks in the search for a specific gene [30].

# 4 Goals

Europe does not want to depend on South America anymore for soybean. That is why genotypes of *Vicia faba* and *Pisum sativum* with a high protein content and resistance against diseases, heat and drought have to be identified. The plants need to be able to grow in the climate of Spain. Especially in the south of Spain, a very dry and hot climate limits the growth.

As explained above, it demands a lot of work to examine all genotypes to find the one corresponding to the desired phenotype. Some of the collected accessions may be identical. Phylogenetic analysis based on SNP's can be done to minimalize the work. In this analysis, evolutionary relationships are estimated. The result of this analysis is a treelike diagram that represents a pedigree of the relationships between the genotypes. The relationships between the genotypes can be based on the SNP's in the DNA sequence. If the DNA sequences from two genotypes have almost the same sequence, with the same SNP's, then they are a part of the same race. If two sequences have many different SNP's, they are not part of the same race [31].

The goal of this master's thesis is to optimize PCR and sequencing conditions to identify SNP's in the genome. If the SNP's are identified, a phylogenetic analysis can be made. This phylogenetic analysis can be coupled to phenotypic analysis and this should speed up the work.

SNP's are identified using a PCR based technique followed by Ion Torrent sequencing. In a genomic PCR, DNA fragments with the chosen SNP will be amplified and elongated with a common sequence. In a following barcoding PCR, the amplicons from the genomic PCR are elongated with an adaptor sequence as a preparation for DNA sequencing. DNA sequencing includes library preparation, emulsion PCR and sequencing. In the emulsion PCR, the amplicons from the barcoding PCR are amplified onto beads.

Several tasks need to be fulfilled:

- Selecting SNP's that are distributed over the entire genome;
- Primer design for DNA fragments that contain a SNP;
- Find the right conditions to perform a genomic and barcoding PCR;
- Amplify all DNA fragments with the SNP;
- Find the right concentration of amplicon library to perform an emulsion PCR.

In this research, 6 genotypes of *Pisum sativum* and 6 genotypes of *Vicia faba* are used.

# 5 Materials and methods

Linkage maps, developed for *Pisum sativum* and *Vicia faba* using SNP markers, were used as a basis for the selection of polymorphic markers [32] [33]. In case of *Vicia faba*, the linkage map was related to QTL mapping for Ascochyta Blight (a common and destructive disease of bean caused by the fungus *Ascochyta fabae)*. In case of *Pisum sativum*, the linkage map was related to QTL mapping for salt resistance [32] [33].

A series of SNP sequences are selected to determine the phylogenetic relation among *Pisum sativum* and *Vicia faba* accessions. *Pisum sativum* has nine linkage groups. A linkage group is a set of genes on a chromosome that tend to be transmitted together. On chromosome five and six, there are two linkage groups for salt resistance [33]. *Vicia faba* has twelve linkage groups. On chromosome one, there are six linkage groups for Ascochyta Blight and on chromosome three, there are two linkage groups [32] [33].

Five SNP sequences per linkage group are chosen, and primers for these sequences are designed (45 primersets for *Pisum sativum* and 60 primersets for *Vicia faba*) [32] [33].

## 5.1 Design primers and optimization of PCR conditions

### 5.1.1 Primer design

All the primers ( to amplify the DNA fragment with the chosen SNP) have an additional common sequence 'C C T C T C T A T G G G C A G T C G G T G A T T' at the 5'end, followed by a specific primer sequence of fourteen till sixteen nucleotides and with a melting temperature of 49-51°C for only this specific sequence. The common sequence is the same for the forward and reverse primers and is necessary for the next, barcoding PCR.
The primersets are designed and ordered online, at Invitrogen [34]. For each chosen sequence, a forward and a reverse primer need to be designed.

Table 1 shows the parameters (for specific sequence) used online for to design the primers.

**Table 1: Parameters designed primers**

| | Min | Opt | Max |
|---|---|---|---|
| Primer size (bases) | 14 | 15 | 16 |
| Primer Tm (°C) | 49 | 50 | 51 |
| Primer % GC | 40 | / | 60 |
| Product size | 100 | / | 160 |

### 5.1.2 Optimization of PCR conditions

#### *5.1.2.1 Thermocyclers*

Three thermocyclers are tested: TProfessional Thermocycler from Biometra, GeneAmp® PCR System 9700 from Applied Biosystems (Life Technologies) and Stratagene Robocycler from Agilent technologies.

The used temperature program depends on the used PCR kit.

#### *5.1.2.2 PCR kits*

Three PCR kits are tested and compared: MyTaq™ DNA Polymerase kit from Bioline, Primestar GXL (TaKaRa) kit from Clontech and OneTaq® DNA Polymerase kit from New England Biolabs.

MyTaq™ DNA Polymerase kit

This kit contains buffer and MyTaq™ DNA Polymerase.
Table 1 shows the used volumes that are put together in a PCR tube.

**Table 2: Content for one PCR tube, using the MyTaq DNA polymerase kit**

| | V(µl) |
|---|---|
| MyTaq™ DNA Polymerase | 0,25 |
| Buffer (including dNTP's) | 5 |
| Primermix | 1 forward + 1 reverse |
| DNA | 2 |
| $H_2O$ | 15,75 |

Table 3 shows the temperature program that should be followed.

**Table 3: Program used on GeneAmp thermocycler for the MyTaq DNA polymerase kit**

| Step | T (°C) | time | |
|---|---|---|---|
| Initial denaturation | 95 °C | 1 min | 1 cycle |
| Denaturation | 95 °C | 15 sec | 10 cycles |
| Annealing | 50->45°C (-0.5 per cycle) | 15 sec | |
| Extension | 75°C | 10 sec | |
| Denaturation | 95 °C | 15 sec | 25 cycles |
| Annealing | 62°C | 15 sec | |
| Extension | 72°C | 10 sec | |

Primestar GXL (TaKaRa) kit

The kit includes buffer, dNTP's and Primestar GXL DNA polymerase.

Table 4 shows the used volumes that are put together in a PCR tube and table 5 shows the temperature program that is used on the GeneAmp thermocycler.

**Table 4: Content for one PCR tube, using the Primestar GXL DNA Polymerase kit**

| | V(µl) |
|---|---|
| TaKaRa Taq DNA Polymerase | 0,5 |
| Buffer | 5 |
| Primermix | 0,5 forward + 0,5 reverse |
| DNA | 2 |
| $H_2O$ | 14,5 |
| dNTP's | 2 |

**Table 5: Program used on GeneAmp thermocycler for the Primestar GXL DNA Polymerase kit**

| Step | T (°C) | time | |
|---|---|---|---|
| Denaturation | 98 °C | 10 sec | 10 cycles |
| Annealing | 55 °C | 15 sec | |
| Extension | 68°C | 15 sec | |
| Denaturation | 98 °C | 10 sec | 30 cycles |
| Annealing | 60°C | 15 sec | |
| Extension | 68°C | 15 sec | |

OneTaq® DNA Polymerase kit

The kit contains buffer, dNTP's and OneTaq DNA polymerase.

Table 6 shows the used volumes that are put together in a PCR tube.

**Table 6: Content one PCR tube, using the OneTaq® DNA Polymerase kit**

| | V(µl) |
|---|---|
| OneTaq® DNA Polymerase | 0,125 |
| Buffer | 5 |
| Primermix | 0,5 forward + 0,5 reverse |
| DNA | 2 |
| $H_2O$ | 16.375 |
| dNTP's (used from TaKaRa kit) | 0.5 |

Table 7 shows the used temperature program for the OneTaq® DNA Polymerase kit.

**Table 7: Program used on GeneAmp thermocycler for the OneTaq® DNA Polymerase kit**

| Step | T (°C) | time | |
|---|---|---|---|
| Denaturation | 98 °C | 10 sec | 10 cycles |
| Annealing | 55 °C | 15 sec | |
| Extension | 68°C | 15 sec | |
| Denaturation | 98 °C | 10 sec | 30 cycles |
| Annealing | 60°C | 15 sec | |
| Extension | 68°C | 15 sec | |

#### *5.1.2.3 Testing the quality of a PCR*

##### 5.1.2.3.1 Gel electrophoresis

The quality of the PCR is tested by gel electrophoresis on a 1.5-2.5% agarose gel (higher concentrations of agarose are used for separation of small DNA's, low concentrations are used for larger DNA's). 1.5-2.5 g agarose is mixed with 100 ml of a tris-acetate-EDTA buffer. This has to be heated in the microwave until the gel is formed and completely melted. Afterwards, 7 µl ethidium bromide solution (10 mg/mL in $H_2O$) is added. Ethidium bromide is a DNA interchelator. It inserts itself into the spaces between the base pairs of the double stranded DNA and then generates a fluorescent signal [35]. After cooling, the gel can be poured in a tray and has to solidify at room temperature.

After solidifying, the tray can be removed and the gel can be put in the electrophoresis chamber filled with tris-acetate-EDTA buffer( 0,04 M Tris acetate and 1 mM EDTA, pH 8.3). Samples of DNA mixed with loading buffer (0,25% bromophenol blue,15% Ficoll in $H_2O$ and 0.25% xylene cyanol) can then be applied to the gel. The lid is put on the apparatus and a current has to be applied. The DNA migrates to the positive electrode [36]. Figure 8 shows the ladder that is used to determine the size of the DNA fragments.

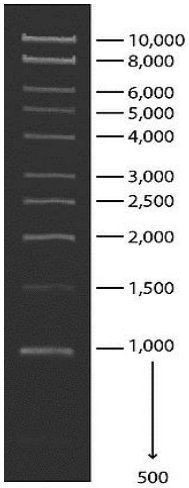



**Figure 8: Ladder used for gel electrophoresis [44]**

Figure 9 shows the different steps in gel electrophoresis.

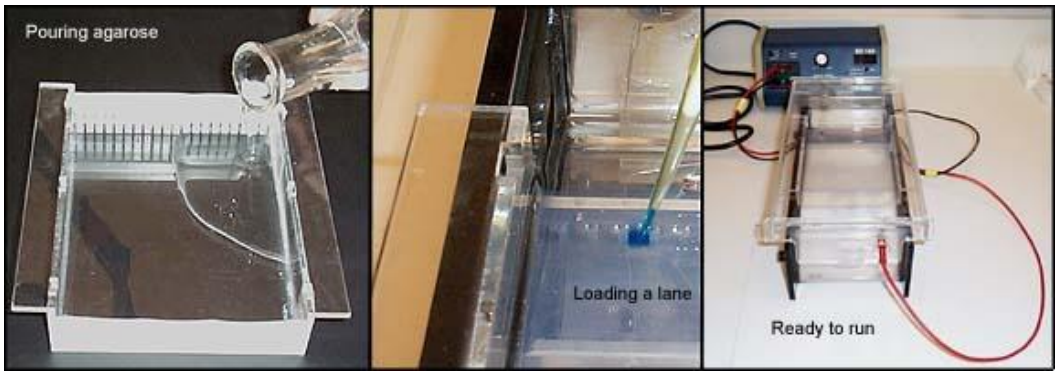
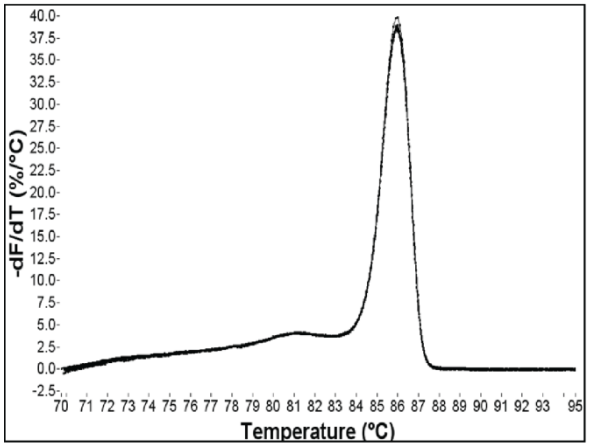



**Figure 9: Different steps in gel electrophoresis [43]**

After migration of DNA, the DNA is visualized in an ultraviolet lightbox by interchelating with EtBr. If the PCR was successful, the PCR product will be visible on the gel as a discrete band [36].

#### 5.1.2.3.2 Melting peak analysis

The quality of the PCR can also be analyzed by melting peak analysis instead of gel electrophoresis. This is done by observing the melting curve of the PCR product in a Stratagene Mx3000P thermocycler. At temperatures below the melting temperature, SYBR Green can bind to the double stranded DNA generating a fluorescent signal. If the melting temperature is reached, double stranded DNA becomes single stranded and SYBR Green is released. This causes a decrease in fluorescence intensity. For the melting or dissociation curve, the derivative of fluorescence (-dF/dT) vs temperature is plotted. The temperature at which the peak occurs corresponds with the melting temperature of the amplicon. An optimal PCR gives a discrete and sharp melting peak [37].
Figure 10 shows a melting curve of an optimal PCR.




A. CFTR Exon 17b

**Figure 10: Example of a melting curve [46]**

Table 8 shows the content of the well for the qPCR on the Mx3000P qPCR system.

**Table 8: Content well for qPCR on Mx3000P qPCR system**

| | V (µl) |
|---|---|
| 1:1000 Sybrgreen (Molecular probes, 10 000 concentrate in DMSO) | 1 |
| Buffer Gotaq 5x (Promega) | 1 |
| PCR-product | 5 |
| $H_2O$ | 3 |

Table 9 shows the temperature program that is used on the Mx3000P qPCR system

**Table 9: Used temperature program for qPCR on Mx3000P qPCR system**

| | | |
|---|---|---|
| Denaturation | 95 °C | 1 min |
| Measurement melting curve | 50 °C→95°C | 30 sec/°C |

## 5.2 PCR

### 5.2.1 Genomic PCR

A genomic PCR is performed to amplify the DNA fragment with the chosen SNP. Primers with the additional common sequence at the 5′ end followed by a 3′ specific sequence are added with the DNA. The common sequence is necessary for the barcoding PCR. Figure 11 shows primer annealing steps (a) in the genomic PCR. After elongation (b), a DNA fragment with the common sequence at 5′ end is obtained.

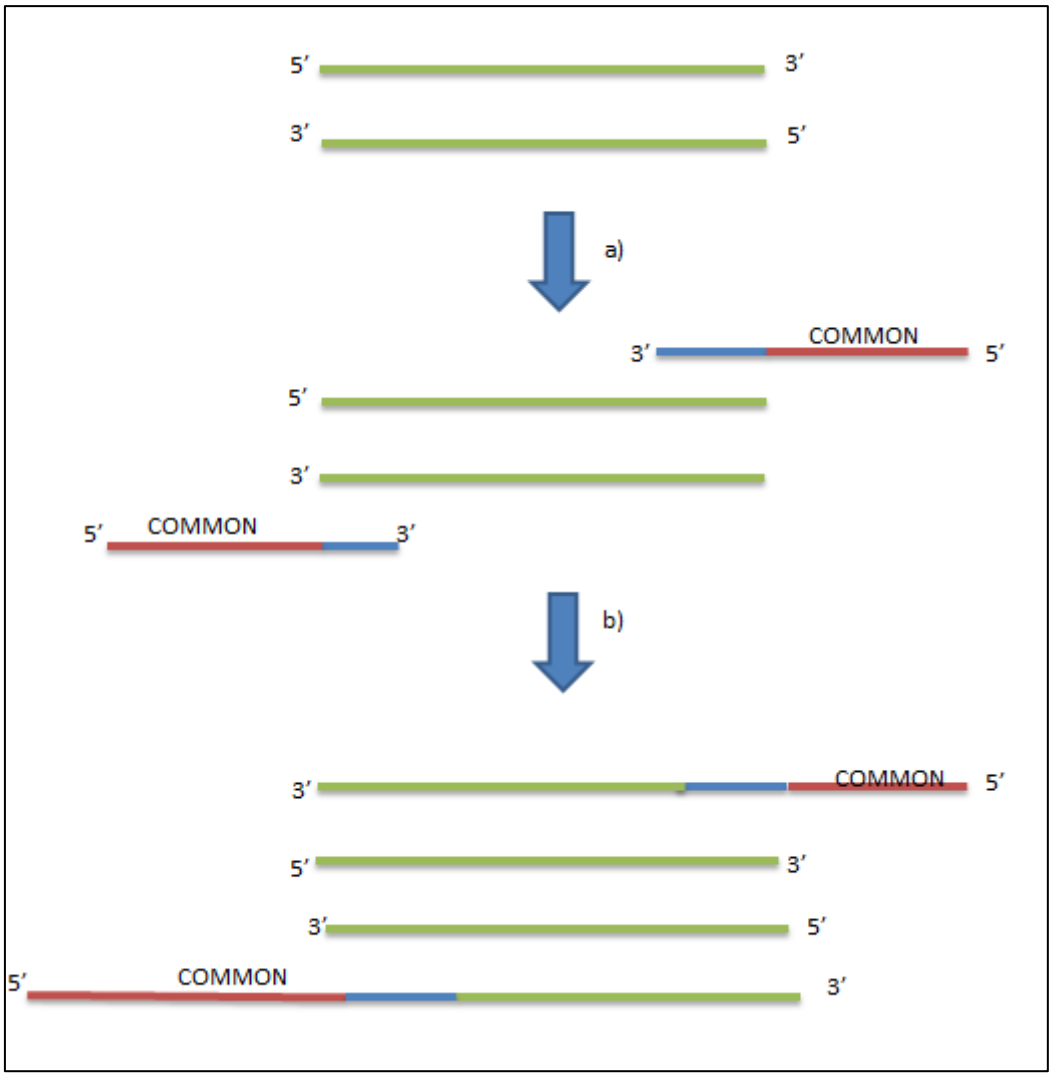



**Figure 11: First round genomic PCR**

Figure 12 shows the annealing and elongation step in the second PCR round (c). After the genomic PCR, double-stranded DNA fragments with the common sequence on 5′ end and the complementary common sequence on the 3′ end are obtained.

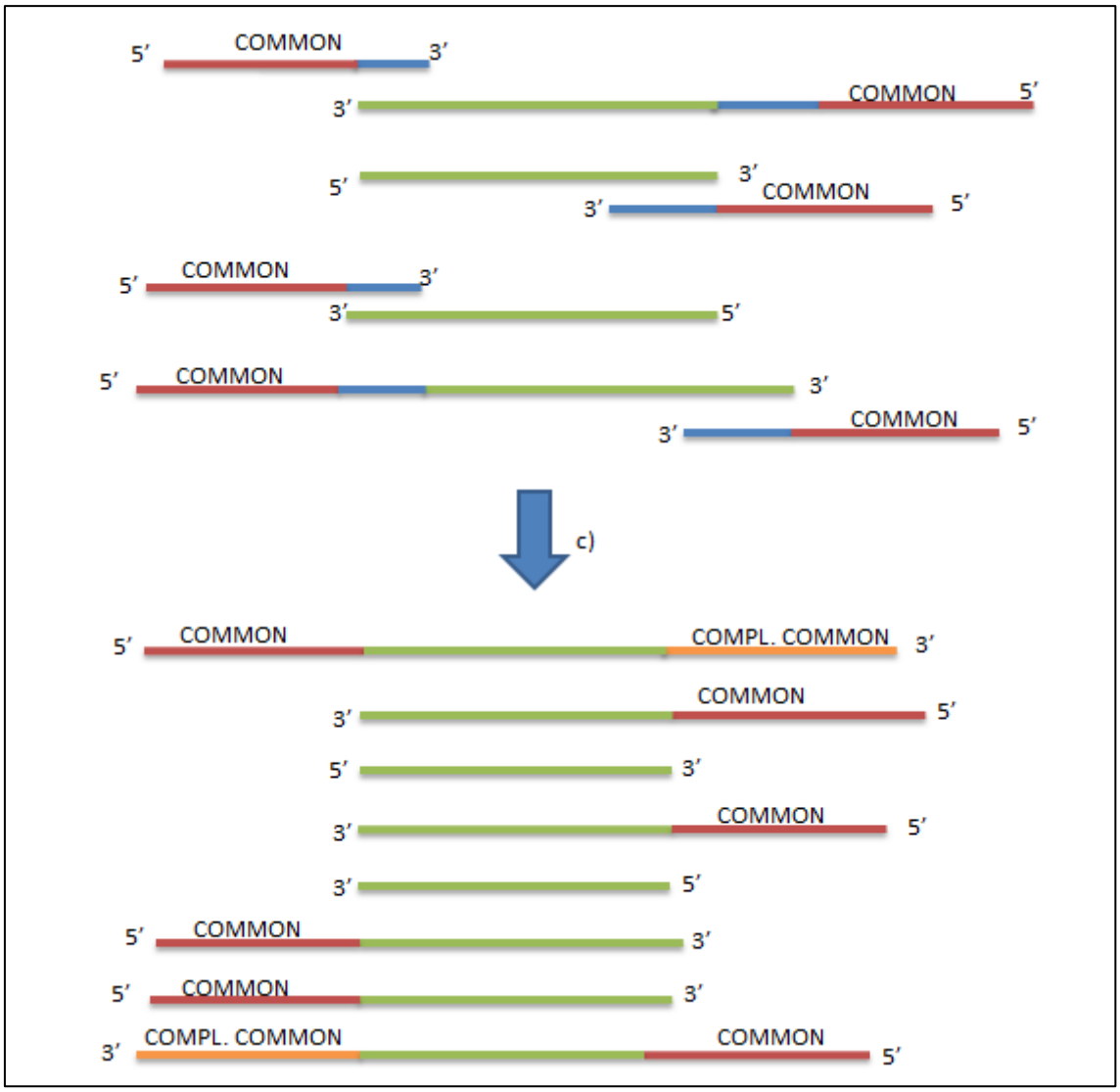



**Figure 12: Second round genomic PCR**

The PCR program for the genomic PCR has two phases: a first one with the annealing step on a lower melting temperature (55°C) at which the specific sequence of the primers binds to the corresponding sequence within the template. The second phase has an annealing step at a higher melting temperature(60°C) because then the whole primer sequence, including the common sequence, can bind.

The amount of volume of each reagent that has to be put together and the temperature program depends on the chosen PCR kit and thermocycler.

### 5.2.2 Barcoding PCR

The barcoding PCR is a preparation for the DNA sequencing. In the genomic PCR, a forward and reverse primer is added but in the barcoding PCR, only one primer is added. The primer for the barcoding PCR has an adaptor sequence at the 5′ end followed by a key sequence and the common sequence at the 3′ end. With the common sequence, the primer binds to the complementary common sequence at 3′ end. For the

sequencing, an Ion 314™ semi-conductor chip with one million wells is used. DNA from twelve genotypes ( six from *Pisum sativum*, six from *Vicia faba*) is put together on one chip. The adaptor sequences for all these twelve genotypes are the same, but the key sequences are different, to differentiate between the twelve genotypes. The common sequence of the primer is the specific sequence in this PCR for amplification of the DNA fragment (amplicon from the genomic PCR).

Figure 13 shows the first round of the barcoding PCR.

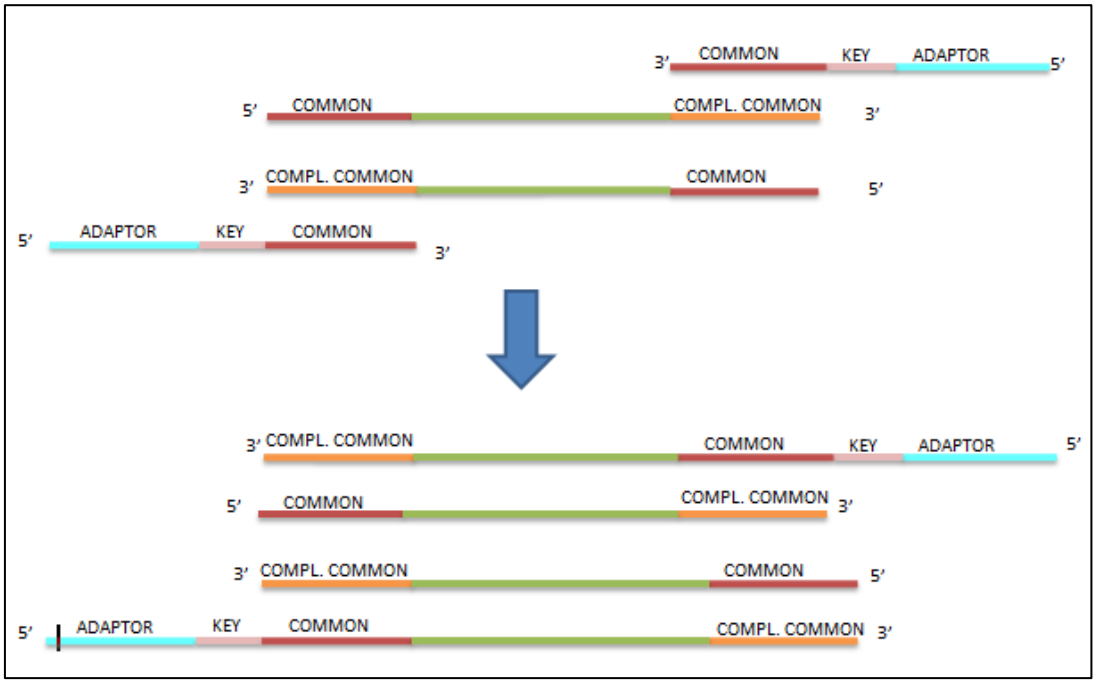



**Figure 14: First round barcoding PCR**

Figure 14 shows the second round of the barcoding PCR.

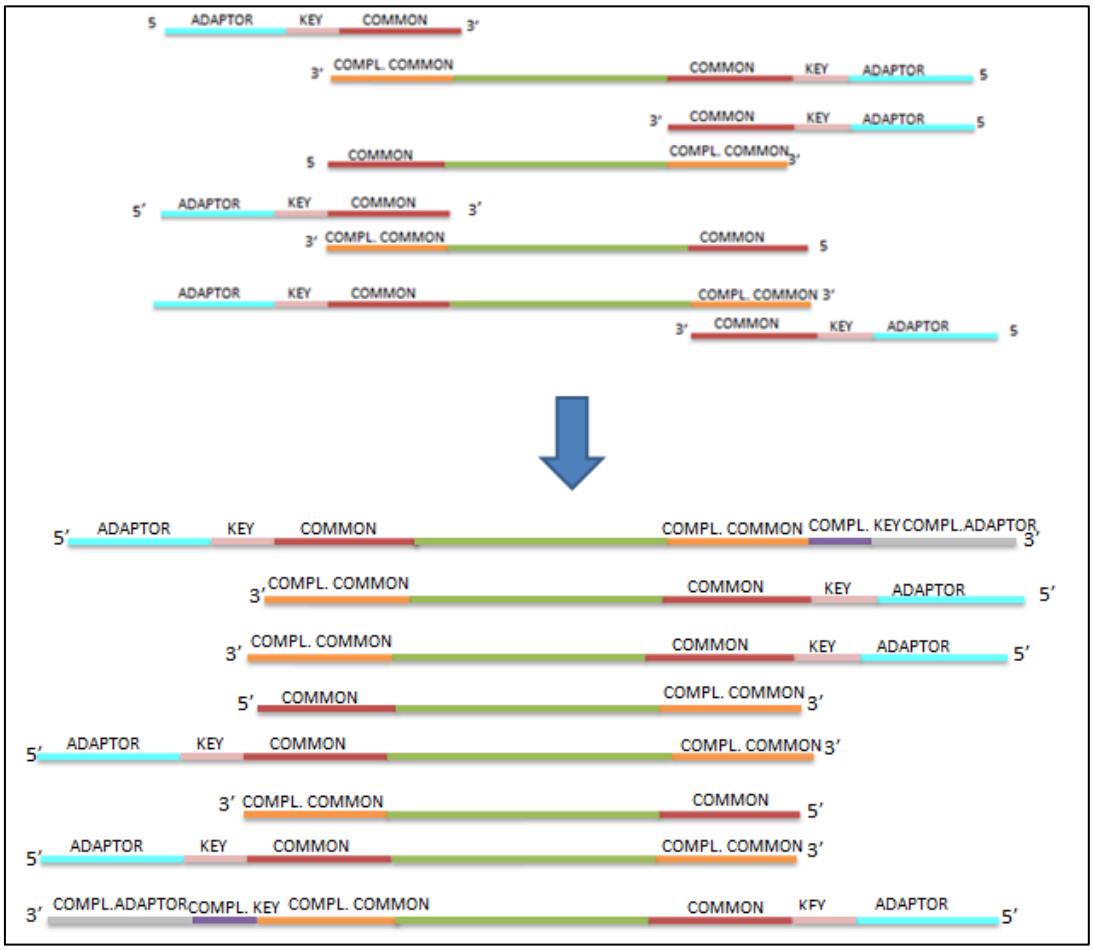



**Figure 15: Second round barcoding PCR**

Amplicons are obtained with at 3′ end the complementary common, key and adaptor sequence and at 5′ end the common, key and adaptor sequence.

Sense:

5′ CCATCTCATCCCTGCGTGTCTCCGAC-TCAG-CCTCTCTATGGGCAGTCGGTGATT-DNAsequence-AATCACCGACTGCCCATAGAGAGG-CTGA-CGGAGACACGCAGGGATGAGATGG 3′

Antisense:

3′GGTAGAGTAGGGACGCACAGAGGC-AGTC- GGAGAGATACCCGTCAGCCACTAA- complementary DNAsequence- TTAGAGGCTGACGGGTATCTCTCC-GACT-GCCTCTGAGCGTCCCTACTCTACC5′

The amount of volume of each reagent that has to be put together and the temperature program depends on the chosen PCR kit and thermocycler.

## 5.3 DNA Sequencing

The protocol for Ion torrent technology includes different steps:

- Library preparation;
- Emulsion PCR;
- Semiconductor sequencing.

### 5.3.1 Library preparation

During the genomic PCR and the barcoding PCR, the DNA fragment is prepared for sequencing. All amplicons from the genomic PCR (six genotypes from *Pisum sativum*, six genotypes from *Vicia faba*) need to be put together in one pool. First, the amplicons should be purified and adapted to the right concentration before it can be put together in one pool.

Purification

For the purification, an Agencourt AM PureXP Reagent is used. This reagent is added to the DNA in a proportion 1:1,8. So if 50 µl of the amplicon is used, 90 µl Agencourt AM PureXP Reagent is necessary. Agencourt AM PureXP Reagent contains an optimized buffer that selectively binds DNA fragments of 100 base pairs and larger to paramagnetic beads. Primers, nucleotides, salts, enzymes are removed from the sample. With a DynaMag™-2 from Life Technologies, the beads are separated from the DNA fragments. The beads and DNA fragments are washed twice with 70% ethanol [38]. Figure 15 shows the DynaMag™ that is used during the purification [39].

**Figure 16: DynaMag™-2 Magnet [50]**

Preparing samples of equimolar concentrations

All twelve samples that are going to be put together in one pool need to have the same concentration. The concentration of the twelve samples is measured using the Qubit 2.0 fluorometer from Invitrogen. This fluorometer uses fluorescent dyes that emit signals only when bound to the specific target molecules [40].

When the concentration is known, the samples are diluted to the lowest concentration. When each sample has an equimolar concentration, the twelve samples can be put together.

Converting concentration pool to 26 pM

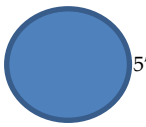
According to the Ion Amplicon Library Preparation (Fusion Method) user guide, the concentration of the pool has to be diluted to ±26 pM [41] . Two other concentrations, 13 pM and 6,5 pM, are also tested and compared in the emulsion PCR. The diluted amplicon library is now ready for template preparation.

### 5.3.2 Emulsion PCR

An emulsion PCR is a simple PCR reaction in an oil phase. Oil is mixed with a PCR mix, beads and single stranded DNA. Emulsion droplets are formed. These droplets act like microreactors and are very stable at high temperatures. One droplet contains one bead, one single stranded DNA-strand, primers and a PCR mix [42]. The amplicon library is highly diluted because there should be many beads for one single stranded DNA strand. This is done to make sure that only one single stranded DNA strand is captured in the emulsion droplet together with one bead. The beads contain the common sequence so the DNA strands bind with the complementary common sequence onto the beads.

Single stranded DNA strands with the adaptor sequence at 5' end and complementary adaptor sequence at 3' end were obtained in the barcoding PCR. The emulsion droplets are formed and contain one strand DNA (sense or antisense), primers and a PCR mix. The next step is the annealing step. The bead contains the common sequence. Depending on what strand has been captured in the emulsion droplet (sense or antisense), the strand anneals to the probe on the bead (figure 16). In the elongation step, the DNA polymerase elongates the common sequence on the bead. In the sequencing, two sequences (complementary to each other) are generated for one SNP sequence of one genotype.

Bead:




Denatured DNA:

SENSE

5'-ADAPTOR-KEY-COMMON-DNA SEQUENCE WITH SNP - COMPL.COMMON- COMPL.KEY-COMPL.ADAPTOR-3'

or::

5' CCATCTCATCCCTGCGTGTCTCCGAC-TCAG-CCTCTCTATGGGCAGTCGGTGATT-DNAsequence with SNP-AATCACCGACTGCCCATAGAGAGG-CTGA-CGGAGACACGCAGGGATGAGATGG 3'

ANTISENSE

3′- COMPL.ADAPTOR- COMPL.KEY- COMPL.COMMON- COMPL. DNA SEQUENCE- COMMON-KEY- ADAPTOR-5′

or:

3′GGTAGAGTAGGGACGCACAGAGGC-AGTC- GGAGAGATACCCGTCAGCCACTAA- compl. DNAsequence-TTAGAGGCTGACGGGTATCTCTCC-GACT-CAGCCTCTGAGCGTCCCTACTCTACC5′

Figure 16 shows the annealing step in the emulsion PCR where the complementary common sequence of the amplicon (sense or antisense strand) anneals to the probe on the bead.

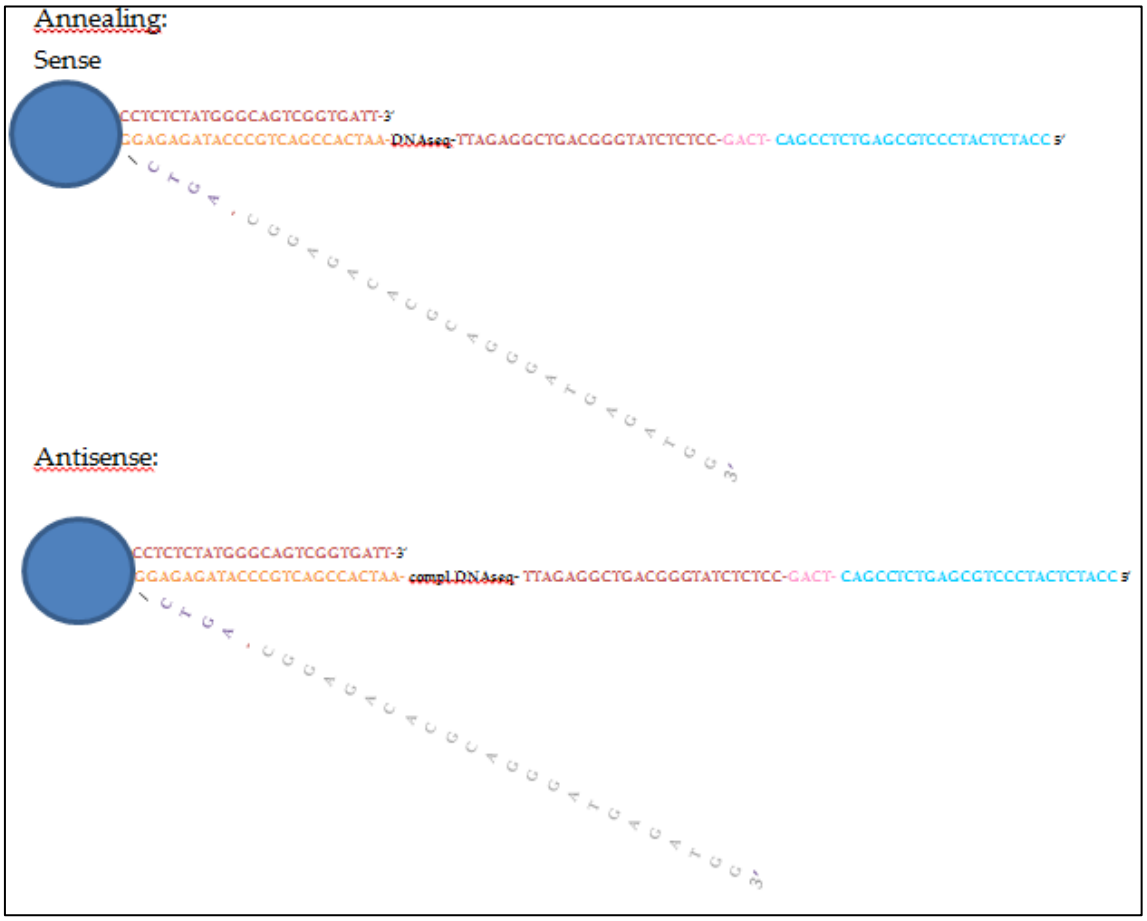



**Figure 17: Annealing step in emulsion PCR**

Figure 17  shows the next step, elongation of the probe by the DNA polymerase.

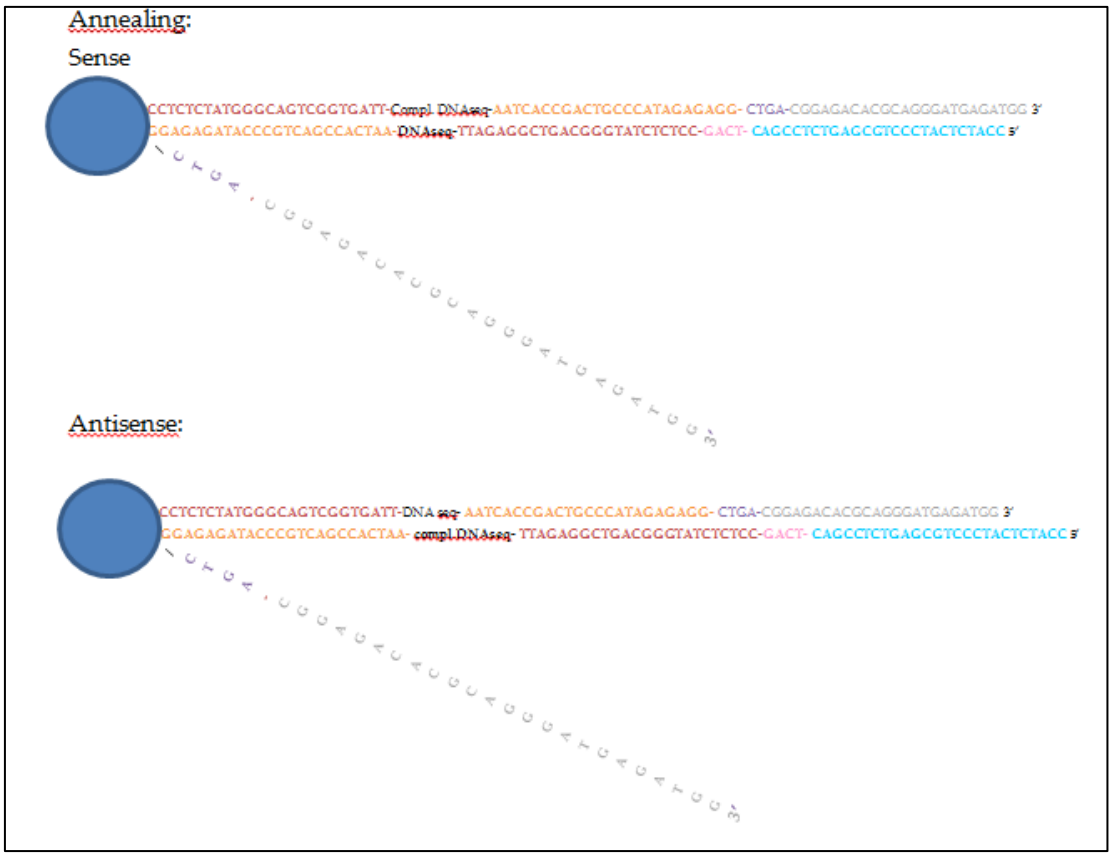



**Figure 18: Elongation step in emulsion PCR**

After the elongation step, double stranded DNA is formed on the beads. The next step is a denaturation step. The original (antisense or sense) strand is denatured from the bead.  The denatured strand anneals again to another probe on the bead for further amplification.  The primer (with biotine) anneals to the complementary of the adaptor sequence of the elongated probe on the bead. In the next amplification step, the polymerase elongates both strands: another probe on the bead and the biotine labeled strand.

The primers in the emulsion PCR have the adaptor sequence and bind to the complementary adaptor sequence. Figure  18 shows annealing of the primers to the complementary adaptor sequence.

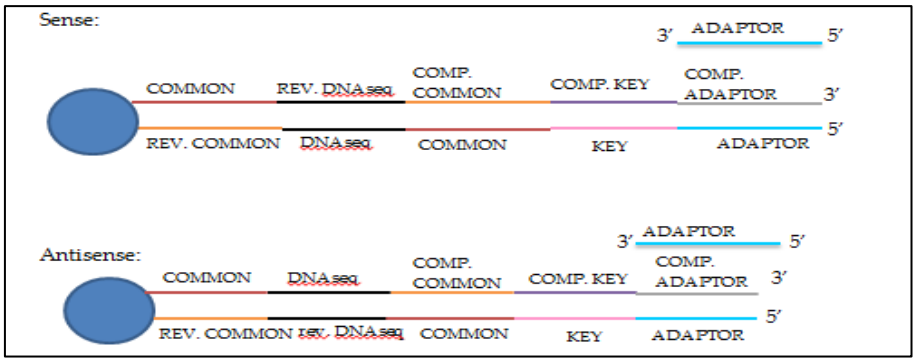



**Figure 19: Annealing primer to sense strand in emulsion PCR**

The last three steps are repeated for 30-60 cycles. The temperatures are 95°C for denaturation and 64°C for annealing and extension.

The emulsion PCR is done using the protocol from the Ion OneTouch 200 Template Kit v2 [43].

After the emulsion PCR, a bead selection is done to maximize the sequencing yield. This is done by adding streptavidin coated magnetic beads. The primer (adaptor sequence) contains a biotin label. Magnetic Streptavidin coated beads are added to the beads from the emulsion PCR and they bind to the biotin. Immobilization is obtained with a magnet. Denaturation with NaOH makes the DNA on the bead single stranded, loosens the DNA from the Streptavidin beads and the probe beads are collected for sequencing. This enrichment is done using the protocol from the Ion OneTouch 200 Template Kit v2 [43].

In this research, three amplicon library concentrations (26 pM, 13 pM and 6,5 pM) are tested and compared. The quality of the enrichment is tested by the Ion Sphere Quality Control assay [43]. The Ion Sphere Particles (or ISP's, beads with double stranded DNA obtained in emulsion PCR) are labeled with two different fluorophores: Alexa Fluor 488 and Alexa Fluor 647. Alexa Fluor 488 binds to the probe site on the bead and measures all the ISP's. Alexa Fluor 647 binds to the adaptor site and measures only the ISP's with extended templates.
The ratio of all ISP's to the template ISP's gives the % templated ISP's[43].

### 5.3.3 DNA sequencing

The DNA sequencing uses an Ion Torrent technology. This technology uses an Ion 314™, 316™ or 318™ semi-conductor chip with one million, six million or eleven million wells. Each well of the chip is filled with about a million copies of single stranded DNA (antisense or sense) attached to beads and a DNA polymerase. The chip is flooded by one specific deoxyribonucleotide triphosphate after another. If the flooded deoxyribonucleotide triphosphate is complementary with the nucleotide of the strand of the DNA next to the annealed and formerly elongated probe, the monophosphate nucleotide is incorporated by the DNA polymerase. Pyrophosphate is

released of the deoxyribonucleotide triphosphate and a hydrogen ion comes free. This changes the pH of the solution. This pH change is detected by an ion sensor, which translates the chemical information into digital information. It can be that two nucleotides are incorporated at the same time (shortly after each other during flooding one specific nucleotide), than the signal is double [44] [45].

Figure 19 shows the reaction that occurs when a nucleotide is incorporated.

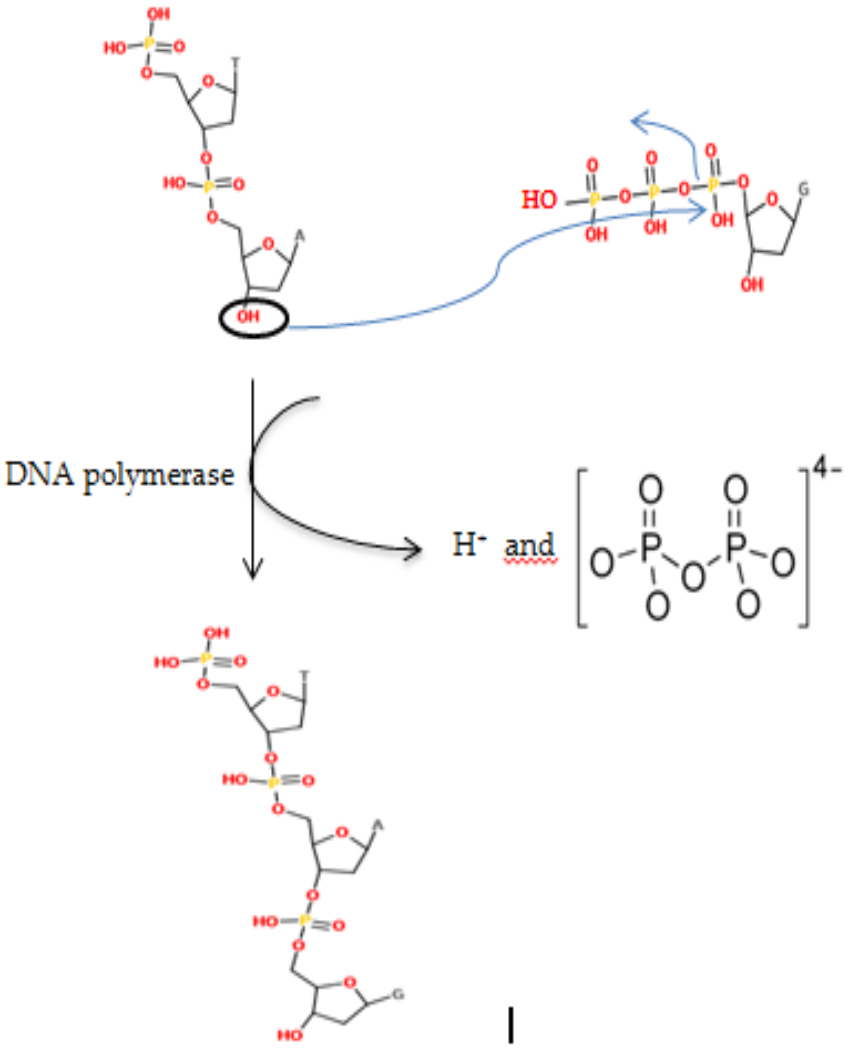



**Figure 19: Reaction of elongation [46]**

The DNA sequencing needs a PGM (personal genome machine). Figure 20 shows a PGM machine and figure 21 shows an ionogram. The first four responses are from the key sequence. The key sequence is used to difference between the different plants, but also to normalize the signal. It is programmed that only one nucleotide is incorporated for the key sequence, so the light intensity signal corresponds with incorporation of one nucleotide. If afterwards the signal is higher, this means that more nucleotides are incorporated at the same time [42].

Figure 20: PGM machine [42]

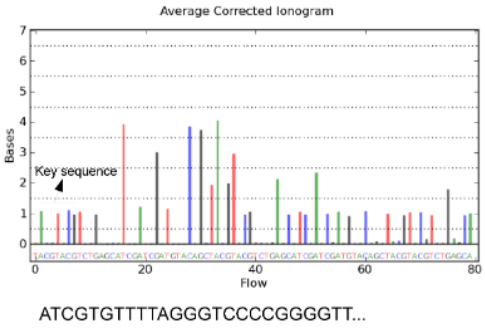



Figure 21: Example of an ionogram [42]

The system needs to be prepared for measuring pH changes during sequencing. Four tubes in the front are filled with the dNTP's. On the side, a bottle W1 is filled with 350µl 100mM NaOH , a bottle W2 is filled with two liters freshly prepared milliQ water and a bottle W3 is filled with 50 ml buffer.
Different steps are followed:

1. The system measures the pH of buffer W3. This is the reference pH.
2. The system measures the pH of solution W2. When this pH is not correct, the system pumps an amount of W2 in bottle W1, and this is mixed by bubbling argon through it.
3. A volume of W1 solution is pumped into W2 and mixed with argon. The pH is again measured. This step is repeated until the pH is correct.
4. The four tubes with dNTP's are filled with W2 solution and mixed with argon gas. The pH of these four tubes should be the same as the W2 solution.
5. Now the system is ready for sequencing and the chips can be loaded with the beads. Maximum one bead per well is allowed [42].

The sequencing is done using the protocol from the Ion PGM Sequencing kit [47].

# 6 Results and discussion

## 6.1 Experiment design

### 6.1.1 Primer design

45 SNP's for *Pisum sativum* (nine linkage groups) and 60 SNP's for *Vicia faba* (twelve linkage groups) were chosen and primersets were designed. They were designed online at Invitrogen [34]. All primers have the common sequence at 5' end and then the specific sequence. Appendix 1 contains the designed primers.

The primers of *Vicia faba* were numbered from F01 to F120, and the ones of *Pisum sativum* from P01 to P90. F01 is the forward primer for the first SNP sequence, F02 is the reverse primer for the first SNP sequence, F03 is the forward primer for the second SNP sequence, and so on.

### 6.1.2 PCR kit

Three PCR kits were tested and compared: MyTaq™ DNA Polymerase kit from Bioline, Primestar GXL DNA Polymerase kit from Clontech and OneTaq® DNA Polymerase kit from New England Biolabs. The testing and comparison was done by doing the genomic PCR on selected primersets.
MyTaq™ DNA Polymerase kit and Primestar GXL DNA Polymerase kit were tested with primers F01/02 and P01/02. OneTaq® DNA Polymerase kit was tested with primers F07/08 and P09/10.
The PCR's were done with a GeneAmp® PCR System 9700.
In the negative control, DNA was replaced by an equivalent volume of water. An amplification of this negative control indicates that there was contamination in the samples, and the samples with DNA were most likely contaminated as well.

#### *6.1.2.1 Results & discussion*

Figure 22 shows the result of the gel electrophoresis of the OneTaq® DNA Polymerase kit.

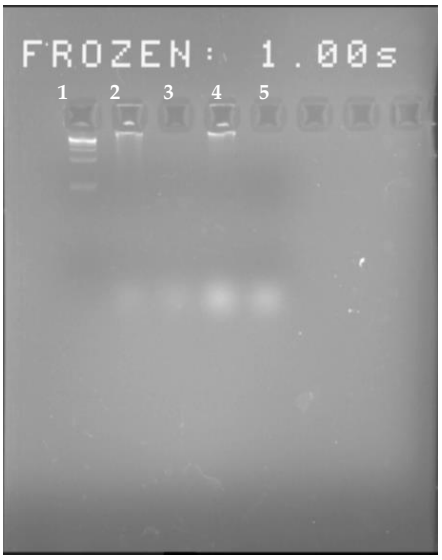



**Figure 22: Gel electrophoresis MyTaq™ DNA Polymerase kit**

1) Ladder
2) F01/F02 : band slightly visible at 100-200 bp
3) Negative control: indication of contamination
4) P01/P02: band slightly visible at 100-200 bp
5) Negative control: contamination

Number five gives a clear band and number three also gives a slightly visible band. These are the negative controls and this indicates contamination. Nothing can be concluded about the DNA samples, because they were contaminated as well. At the top of number two and four, a band is visible. This is a band from the genomic DNA.

Figure 23 shows the result of the gel electrophoresis of the Primestar GXL DNA Polymerase kit.


**Figure 23: Gel electrophoresis for Primestar GXL DNA Polymerase kit**

1) Ladder
2) F01/F02 : nothing visible
3) Negative control
4) P01/P02: band visible at 100-200 bp
5) Negative control

Number three and five show that there was no contamination. There was no amplification for primers F01/02 (2), but there was amplification for P01/02 (4). In both number two and four, a band from the genomic DNA is visible. It could be that the primers for *Vicia faba* are not good designed. If they are not good designed, they are not selective. It could also be that the installed melting temperature was too low. A band is visible at 100-200 bp for *Pisum sativum.* It is desired to obtain DNA fragments from 100 till 160 basepairs long, therefore this is a good result.

For the PCR with OneTaq® DNA Polymerase, two different DNA samples for *Vicia faba* and *Pisum sativum* were used. This PCR was done three weeks after the first two PCR's. To be sure negative results would not be dedicated to degradation of stored DNA, two new DNA samples were added. On the control gel (figure 24), other samples were also tested (only numbers 4,5,12 and 13 are relevant).

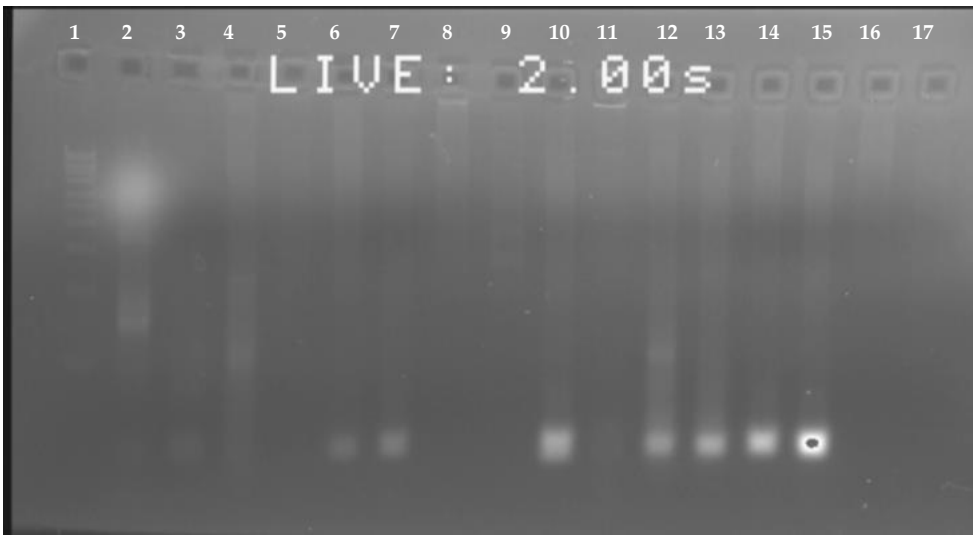



**Figure 24: Gel electrophoresis for OneTaq® DNA Polymerase kit**

4) new DNA F07/08 : nothing visible
5) old DNA F07/08: nothing visible
12) new DNA P09/10: distinct band at 100-200 bp
13) old DNA P09/10: distinct band at 100-200 bp

Primers P09/10 gives clear distinct bands of 100-200 bp. For *Vicia faba,* there was no amplification.

#### *6.1.2.2 Conclusion*

When the three PCR kits are compared, there can be concluded that both the OneTaq® DNA Polymerase kit and Primestar GXL DNA Polymerase kit gave amplification for *Pisum sativum.* The bands from the OneTaq® DNA Polymerase kit (figure 24) are much more clear than the bands from the Primestar GXL DNA Polymerase kit. Therefore, the OneTaq® DNA Polymerase kit gives a better amplification of the DNA fragment. That is why this kit is chosen to do the PCR's.

### 6.1.3 Thermocycler

Three thermocyclers were tested: TProfessional Thermocycler from Biometra, GeneAmp® PCR System 9700 from Life Technologies and Stratagene Robocycler from Applied Biosystems.
The Stratagene Robocycler was tested with the TaKaRa Taq DNA Polymerase.
The TProfessional Thermocycler and GeneAmp® PCR System 9700 were tested with the TaKaRa Taq DNA Polymerase kit and with OneTaq® DNA Polymerase kit.

#### *6.1.3.1 Results & discussion*

Figure 25 shows the result of the gel electrophoresis of the amplicons obtained with the Stratagene Robocycler, using the TaKaRa Taq DNA Polymerase.


**Figure 25: Gel electrophoresis Stratagene Robocycler**

2) F01/02 – nothing visible
3) F03/04 – nothing visible
4) F05/06 – nothing visible
5) F07/08 – nothing visible
6) F09/10– nothing visible
8) P01/02- band slightly visible at 100-200 bp
9) P03/04- band slightly visible at 100-200 bp
10) P05/06- band slightly visible at 100-200 bp
11) P07/08- band slightly visible at 100-200 bp
12) P09/10- band slightly visible at 100-200 bp

For *Vicia faba*, there was almost no amplification because there is nothing visible on the gel (number 2-5). F09/10 gives a band around 100-200 bp, but the band is slightly visible. For *Pisum sativum*, there was a really small amplification because the bands are slightly visible (number 8-12).

Figure 26 shows the result of the gel electrophoresis of the amplicons obtained with the GeneAmp® PCR System 9700 and TProfessional Thermocycler , using the TaKaRa Taq DNA Polymerase and OneTaq® DNA Polymerase. A qPCR Stratagene Mx3000P thermocycler is also used to test if the PCR worked if the melting temperature was 50°C instead of 55°C. Primers F07/08 and P09/10 were used.

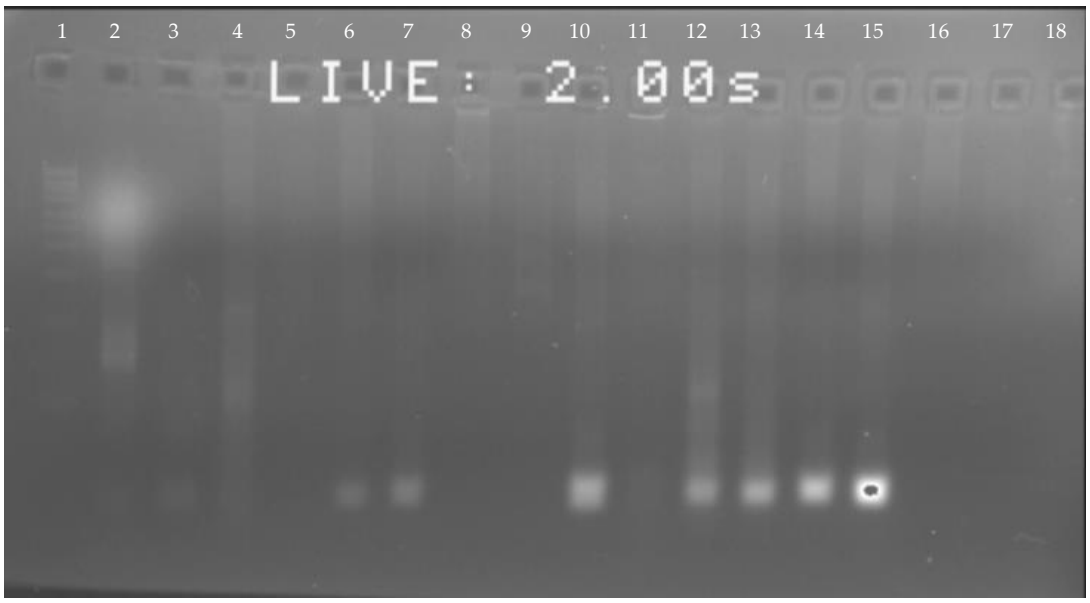



**Figure 26: Gel electrophoresis GeneAmp® PCR System 9700 and TProfessional Thermocycler**

1) Ladder - slightly visible (bad quality picture)
2) *Vicia faba* new DNA on Stratagene Mx3000P – smear visible
3) *Vicia faba* old DNA on Stratagene Mx3000P – band of 100-200 bp, almost not visible
4) *Vicia faba* new DNA on GeneAmp® PCR System 9700, using OneTaq® DNA Polymerase – smear visible
5) *Vicia faba* old DNA on GeneAmp® PCR System 9700, using OneTaq® DNA Polymerase - nothing visible
6) *Vicia faba* new DNA on TProfessional Thermocycler, using OneTaq® DNA Polymerase – distinct band (at 100-200 bp)
7) *Vicia faba* old DNA on TProfessional Thermocycler, using OneTaq® DNA Polymerase – distinct band (at 100-200 bp)
8) *Vicia faba* new DNA on TProfessional Thermocycler, using TaKaRa Taq DNA Polymerase - nothing visible
9) *Vicia faba* new DNA on TProfessional Thermocycler, using TaKaRa Taq DNA Polymerase - nothing visible
10) *Pisum sativum* new DNA on Stratagene Mx3000P – distinct band (at 100-200 bp)
11) *Pisum sativum* old DNA on Stratagene Mx3000P - nothing visible
12) *Pisum sativum* new DNA on GeneAmp® PCR System 9700, using OneTaq® DNA Polymerase - distinct band (at 100-200 bp)
13) *Pisum sativum* old DNA on GeneAmp® PCR System 9700, using OneTaq® DNA

Polymerase - distinct band (at 100-200 bp)
14) *Pisum sativum* new DNA on TProfessional Thermocycler, using OneTaq® DNA Polymerase - distinct band (at 100-200 bp)
15) *Pisum sativum* old DNA on TProfessional Thermocycler, using OneTaq® DNA Polymerase - distinct band (at 100-200 bp)
16) *Pisum sativum* new DNA on TProfessional Thermocycler, using TaKaRa Taq DNA Polymerase - nothing visible
17) *Pisum sativum* old DNA on TProfessional Thermocycler, using TaKaRa Taq DNA Polymerase - nothing visible

The PCR on Stratagene Mx3000P at 50°C instead of 55°C (numbers 2,3,10,11 ) worked for some samples. Number two gave smear on the gel, DNA fragments with different sizes were obtained. Therefore, The specificity of the primers was not good. Number three gives a slightly visible band at 100-200 bp. The concentration of the amplicon is very little so there was only minor amplification. Number ten gives a distinct band at 100-200 bp. Number eleven gives no distinct band. This is due to evaporation of the product during PCR because the plate was not well covered with the sticker.

#### *6.1.3.2 Conclusion*

In general, *Pisum sativum* always gives more distinct bands than *Vicia faba.* Using the OneTaq® DNA Polymerase kit on a GeneAmp® PCR System 9700 or TProfessional Thermocycler gives clear distinct bands for *Pisum sativum* . Also for *Vicia faba,* two bands are visible if OneTaq® DNA Polymerase is used on a TProfessional Thermocycler.

Using TaKaRa Taq DNA Polymerase on a TProfessional Thermocycler gives no amplification for both *Pisum sativum* and *Vicia faba.*

Because using a TProfessional Thermocycler results in distinct bands for both *Pisum sativum* and *Vicia faba,* this thermocycler is chosen to perform the PCR's.

Optimizing PCR conditions was done by doing genomic PCR's. The TProfessional Thermocycler and OneTaq® DNA Polymerase kit were supposed to be used in the barcoding PCR also. That is why a test was done to test if amplicons are obtained if the barcoding PCR is done.
Figure 27 shows the result of the gel electrophoresis of the barcoding PCR with the TProfessional Thermocycler and OneTaq® DNA Polymerase kit.

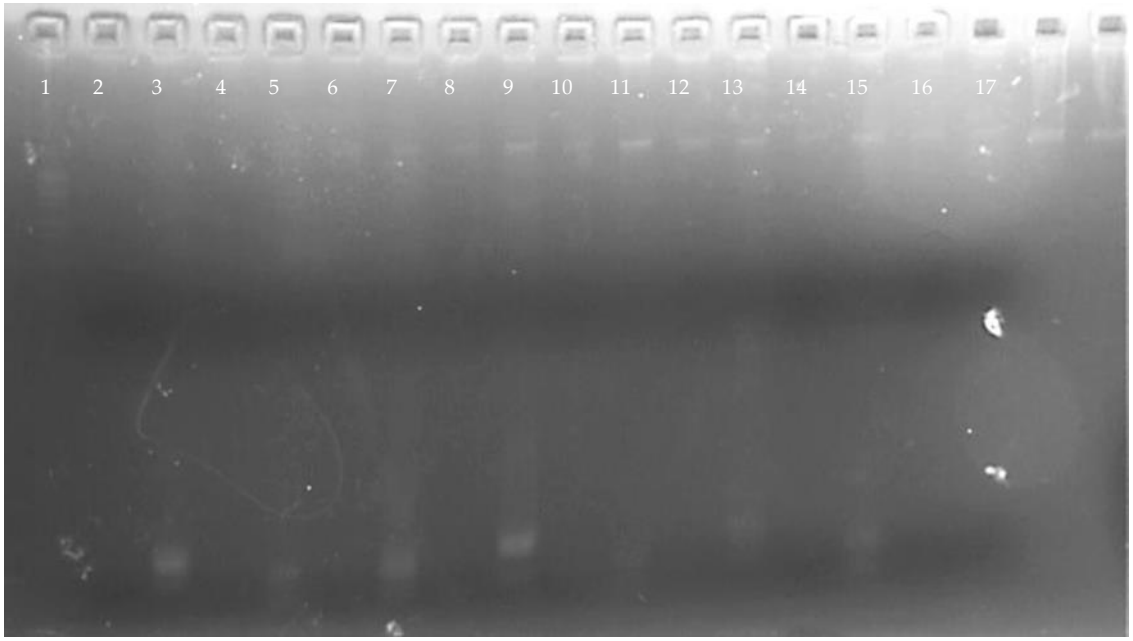



**Figure 27:gel electrophoresis barcoding & genomic PCR with TProfessional Thermocycler & OneTaq® DNA Polymerase kit.**

1) Ladder
2) Negative control
3) F01/02, barcoding PCR
4) F01/02, genomic PCR
5) F03/04, barcoding PCR
6) F03/04, genomic PCR
7) F07/08, barcoding PCR
8) F07/08, genomic PCR
9) F09/10, barcoding PCR
10) F09/10, genomic PCR
11) P01/02, barcoding PCR
12) P01/02, genomic PCR
13) P07/08, barcoding PCR
14) P07/08, genomic PCR
15) P09/10, barcoding PCR
16) P09/10, genomic PCR
17) Negative control

The negative controls don't show a band, so there was no contamination. The products of the genomic PCR are also put on the gel. If bands at the same height for both genomic and barcoding PCR are obtained, this means that there was no barcoding PCR because the amplicons from the barcoding PCR should have a bigger length than the amplicons from the genomic PCR. Number 3,5,7 and 9 give a distinct band around 200 bp, and there is no band visible at that height for number 4,6,8 and 10. This means that there really was an amplification in the barcoding PCR. Number 11,13 and 15 give slightly visible bands and for the genomic PCR, there are no bands visible.

From figure 27, there can be concluded that the combination of TProfessional Thermocycler and OneTaq® DNA Polymerase kit also works for the barcoding PCR.

## 6.2 Genomic PCR

The first six plants of *Pisum sativum* (genotypes 259,305,304,258,283,281 and 102) and the first six plants of *Vicia faba (*genotypes 102,140,104,112,100 and 184) were used. The genomic PCR was done on plates with 96 wells. Figure 28 shows the plate of the PCR for the first set of primers.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 259 F01/ 02 | 305 F01/ 02 | 304 F01/ 02 | 258 F01/ 02 | 283 F01/ 02 | 281 F01/ 02 | 102 P01/ 02 | 140 P01/ 02 | 104 P01/ 02 | 112 P01/ 02 | 100 P01/ 02 | 184 P01/ 02 |
| 259 F03/ 04 | 305 F03/ 04 | 304 F03/ 04 | 258 F03/ 04 | 283 F03/ 04 | 281 F03/ 04 | 102 P03/ 04 | 140 P03/ 04 | 104 P03/ 04 | 112 P03/ 04 | 100 P03/ 04 | 184 P03/ 04 |
| 259 F05/ 06 | 305 F05/ 06 | 304 F05/ 06 | 258 F05/ 06 | 283 F05/ 06 | 281 F05/ 06 | 102 P05/ 06 | 140 P05/ 06 | 104 P05/ 06 | 112 P05/ 06 | 100 P05/ 06 | 184 P05/ 06 |
| 259 F07/ 08 | 305 F07/ 08 | 304 F07/ 08 | 258 F07/ 08 | 283 F07/ 08 | 281 F07/ 08 | 102 P07/ 08 | 140 P07/ 08 | 104 P07/ 08 | 112 P07/ 08 | 100 P07/ 08 | 184 P07/ 08 |
| 259 F09/ 10 | 305 F09/ 10 | 304 F09/ 10 | 258 F09/ 10 | 283 F09/ 10 | 281 F09/ 10 | 102 P09/ 10 | 140 P09/ 10 | 104 P09/ 10 | 112 P09/ 10 | 100 P09/ 10 | 184 P09/ 10 |
| 259 F11/ 12 | 305 F11/ 12 | 304 F11/ 12 | 258 F11/ 12 | 283 F11/ 12 | 281 F11/ 12 | 102 P11/ 12 | 140 P11/ 12 | 104 P11/ 12 | 112 P11/ 12 | 100 P11/ 12 | 184 P11/ 12 |
| 259 F13/ 14 | 305 F13/ 14 | 304 F13/ 14 | 258 F13/ 14 | 283 F13/ 14 | 281 F13/ 14 | 102 P13/ 14 | 140 P13/ 14 | 104 P13/ 14 | 112 P13/ 14 | 100 P13/ 14 | 184 P13/ 14 |
| 259 F15/ 16 | 305 F15/ 16 | 304 F15/ 16 | 258 F15/ 16 | 283 F15/ 16 | 281 F15/ 16 | 102 P15/ 16 | 140 P15/ 16 | 104 P15/ 16 | 112 P15/ 16 | 100 P15/ 16 | 184 P15/ 16 |

**Figure 28: First plate from genomic PCR**

The quality of the PCR was tested with a qPCR on a Mx3000P qPCR System from Agilent Technologies. Not all amplicons were tested with melting curve analysis, just some to check if the thermocycler worked properly and if no mistakes were made.

Also a gel electrophoresis was done to check the quality of the PCR.

Table 10 shows the amount of volume of each reagent of the OneTaq DNA Polymerase kit that was put together for the genomic PCR. Table 11 shows the used temperature program on the TProfessional Thermocycler.

**Table 10: Content tube genomic PCR**

| | V(µl) |
|---|---|
| 5X One*Taq* Standard Reaction Buffer | 5 |
| dNTP's (2,5 mM) | 0,5 |
| Primer forward | 0,5 |
| Primer reverse | 0,5 |
| OneTaq DNA polymerase | 0,125 |
| $H_2O$ | 16,375 |
| DNA | 2 |

**Table 11: Temperature program genomic PCR on TProfessional Thermocycler**

| Step | T (°C) | time | |
|---|---|---|---|
| Denaturation | 98 °C | 10 sec | 10 cycles |
| Annealing | 55 °C | 15 sec | |
| Extension | 68°C | 15 sec | |
| Denaturation | 98 °C | 10 sec | 30 cycles |
| Annealing | 60°C | 15 sec | |
| Extension | 68°C | 15 sec | |

### 6.2.1 Results & discussion

Figures 29 and 30 show the melting curves of the amplicons obtained with the first set of primers F01-F16 and P01-P16. The melting curves were always made with the amplicons of genotypes 304 (*Vicia faba)* and 104 *(Pisum sativum)*

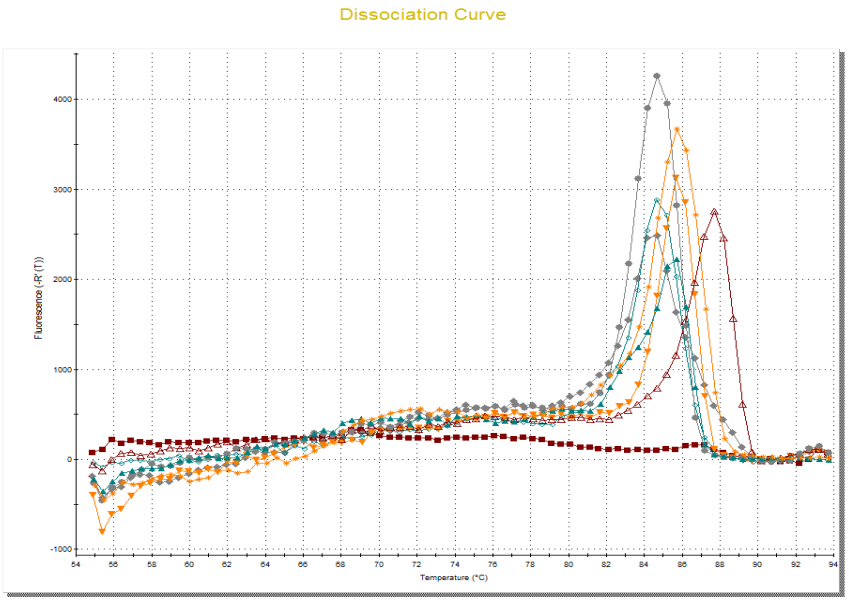



**Figure 29: Melting curve *Pisum sativum* P01-P16**

■ : P1-2 : did not work

♦ : P3-4

° : P5-6

▼ : P7-8

Δ : P9-10

• : P11-12

▲ : P13-14

✱ : P15-16

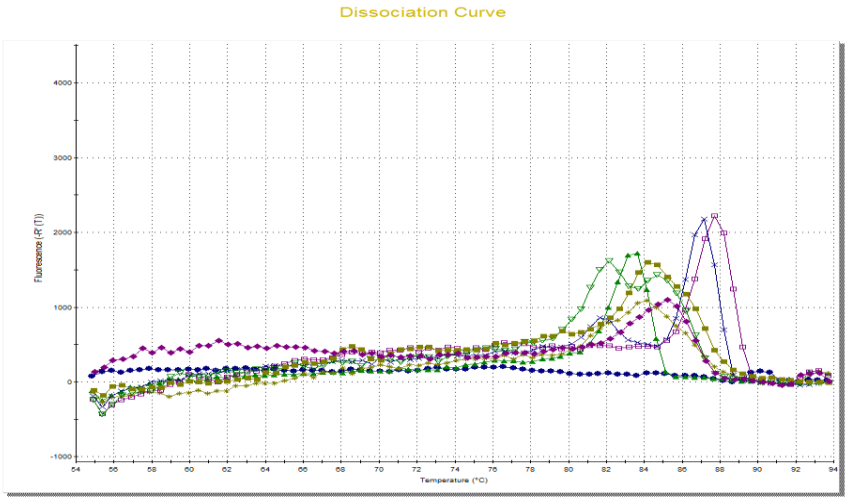



**Figure 30: Melting curve *Vicia faba* F01-F16**

• : F1-2 : did not work

▲ : F3-4

✱ : F5-6

❑: F7-8

✕: F9-10

≤ (reversed empty triangle): F11-12

■ : F13-14

♦ : F15-16

Primers F01/02 did not work for *Vicia faba*. Primer P01/02 did not work for *Pisum sativum*. They also did not work during the optimization of the PCR conditions. It is possible that the primers are not properly designed, therefore the selectivity of the primers is very low and they do not anneal to the single stranded DNA. Another possibility is that the used melting temperature of the temperature program is not the melting temperature of the used primer, therefore the primer will also not anneal to the single stranded DNA. The melting temperature of P1 is 50,33 °C and P2 is 49,08 °C. The melting temperature of F1 is 49,07 °C and F2 is 50,12 °C. The used melting temperature is 55°C in phase one, and 60 °C in phase two. The melting temperature was set on 55°C to increase selectivity of the PCR. It is possible that the melting temperature was too high and the primer could not anneal to its target DNA.

Primer 09/10 ans 11/12 both show two peaks. Two different amplicon products were obtained. It could be that the primer bound on two different places on the DNA fragments, therefore the specificity of the primers is not good.

Appendix B contains melting curves of some other sets of primers for the genomic PCR.

Figure 31 shows the gel electrophoresis done after the PCR with primers F1-F16 and P1-P16.

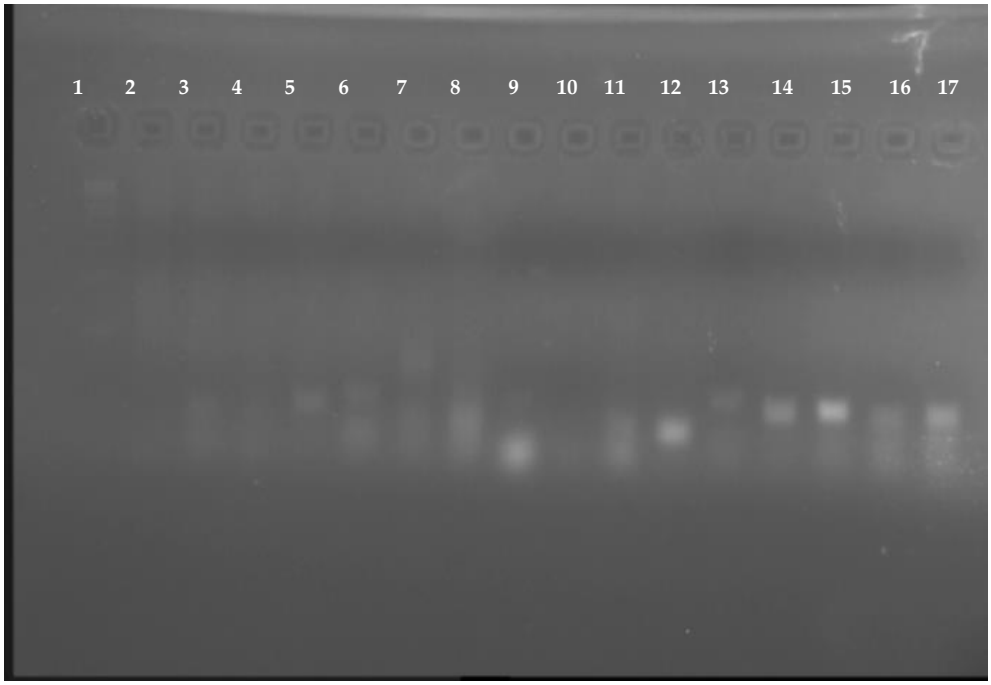



**Figure 29: Gel electrophoresis *Pisum sativum* and *Vicia faba* F01-F16 and P01-P16**

1) Ladder
2) F01/02 : nothing visible
3) F03/04 : band slightly visible at 100-200 bp
4) F05/06 : band slightly visible at 100-200 bp
5) F07/08 : band slightly visible at 100-200 bp
6) F09/10 : band slightly visible at 100-200 bp
7) F11/12 : band slightly visible at 100-200 bp
8) F13/14 : band at 100-200 bp
9) F15/16: band at 100-200 bp
10) P01/02: nothing visible
11) P03/04: band at 100-200 bp
12) P05/06: band at 100-200 bp
13) P07/08: band at 100-200 bp
14) P09/10: band at 100-200 bp
15) P11/12: band at 100-200 bp
16) P13/14: band at 100-200 bp
17) P15/16: band at 100-200 bp

For the gel electrophoresis, the results for *Pisum sativum* are again better than the results for *Vicia faba*. There is no distinct band visible for primers F01/02 and P01/02 (number 2 and 10). Also F05/06 does not give a clear band, but it gives a little peak in the melting curve, so there was amplification.

Table 12 shows an overview of the primers that did not work during the genomic PCR (melting curves can be found in appendix B).

**Table 12: Primers that did not work in the genomic PCR**

| *Vicia Faba* | *Pisum sativum* |
|---|---|
| F01-02 | P01-02 |
| F33-34 | P27-28 |
| F39-40 | P37-38 |
| F35-36 | P53-54 |
| F49-50 (really small peak) | P55-56 |
| F61-62 (double peak) | P59-60 |
| F65-66 | P61-62 (really small peak) |
| F57-68 | P63-64 (really small peak) |
| F73-74 | P65-66 |
| F77-78 | P67-68 |
| F83-84 (really small peak) | P75-76 |
| F89-90 (really small peak) | P77-78 |
| F81-82 (really small peak) | P81-82 |
| F91-92 (really small peak) | |
| F113-114 | |
| F111-112 (really small peak) | |
| F117-118 (really small peak) | |
| F119-120 (really small peak) | |

## 6.3 Barcoding PCR

In this PCR, the adaptor and key sequences were added to the DNA fragments. The barcoding PCR was done using two different methods and both methods were tested. Each PCR can be done separately, this was a first method. 60 PCR's for *Vicia faba* and 45 PCR's for *Pisum sativum* were done. The shown melting curves below and in appendix C are the melting curves from the amplicons obtained with this separate PCR's.

A second method to do this barcoding PCR is to first put all amplicons from the genomic PCR together. For *Pisum sativum,* there were 45 genomic PCR's. 1 µl of each amplicon of the genomic PCR (with common sequence on 5' end and the complementary common sequence on the 3' end) was added together. From this 45 µl, 2 µl was taken to do the PCR. For *Vicia faba* , 1 µl of each of the 60 amplicons from the genomic PCR was put together and from this 60 µl, 2 µl was also taken to do the PCR. Twelve different PCR's were done ( adaptor 7-12 for *Pisum sativum* and adaptor 7-12 for *Vicia faba).*

Table 13 shows the amount of volume of each reagent of the OneTaq DNA Polymerase kit that was put together for the separate barcoding PCR. Table 14 shows the amount

of each reagent that was put together for the collective barcoding PCR. Table 15 shows the used temperature program on the TProfessional Thermocycler.

**Table 13: Content tube barcoding PCR**

| | V(µl) |
|---|---|
| 5X One*Taq* Standard Reaction Buffer | 5 |
| dNTP's (2,5 mM) | 0.5 |
| Adaptor | 1 |
| PCR | 1 |
| $H_2O$ | 17,375 |
| OneTaq DNA polymerase | 0,125 |

**Table 14: Content PCR tube when PCR products are put together**

| **Reagent** | **Volume (µl)** |
|---|---|
| 5X One*Taq* Standard Reaction Buffer | 5 |
| dNTP's (2,5 mM) | 0.5 |
| Adaptor | 1 |
| PCR | 2 |
| $H_2O$ | 16,375 |
| OneTaq DNA polymerase | 0,125 |

**Table 15: Temperature program barcoding PCR on TProfessional Thermocycler**

| Step | T (°C) | time | |
|---|---|---|---|
| Initial denaturation | 94 °C | 1 min | 1 cycle |
| Denaturation | 94 °C | 30 sec | 10 cycles |
| Annealing | 60 °C | 30 sec | |
| Extension | 68°C | 30 sec | |
| Denaturation | 94 °C | 30 sec | 30 cycles |
| Annealing | 68°C | 15 sec | |
| Extension | 68°C | 5 min | |

The quality of the PCR was tested with a qPCR on a Mx3000P qPCR System from Agilent Technologies and with gel electrophoresis.

### 6.3.1 Results & discussion

Figures 32 and 33 show the melting curves of the separate amplicons of the first set of primers F01-F16 and P01-P16. Appendix C shows the melting curves of other primers.

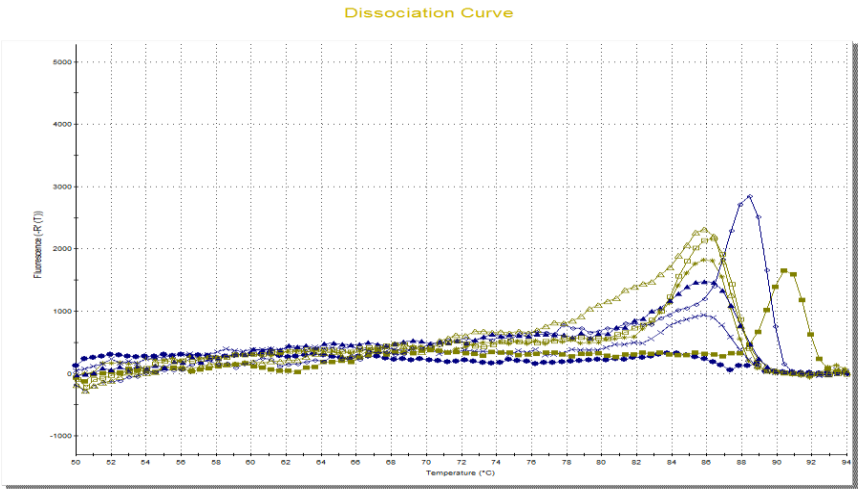



**Figure 30: Melting curve *Vicia faba* F1-F16**

• : F1-2
✱ : F3-4
✕: F5-6
■ : F7-8
°: F9-10
Δ: F11-12
▲ : F13-14
❑: F15-16

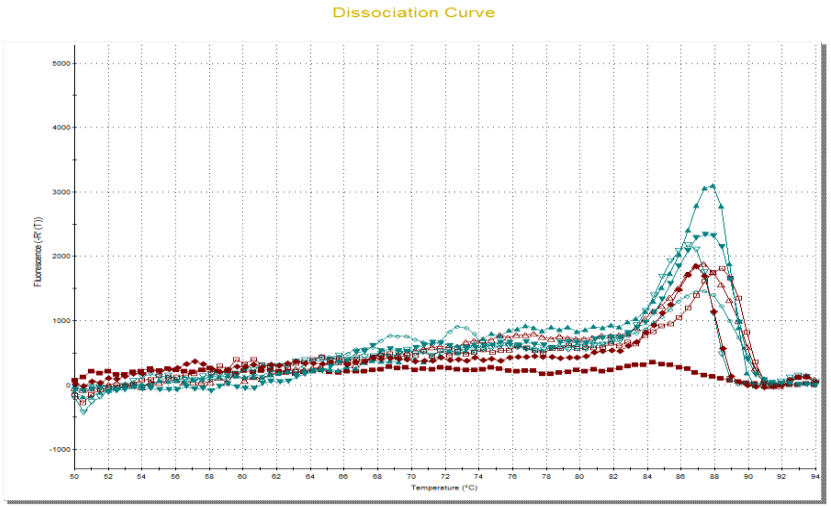



**Figure 31: Melting curve *Pisum sativum* P01-P16**

■ : P1-2
°: P3-4
Δ: P5-6
▲ : P7-8
❑: P9-10
≤ (reversed empty triangle): P11-12
♦ : P13-14
▼ : P15-16

Primers F01-02 did not work for *Vicia faba.* This is a confirmation of the results in the genomic PCR because the specific primers also did not work in the genomic PCR, so there was no amplicon for the barcoding PCR and the adaptor sequence could not be added to the amplicon.

Primers P01-02 did not work for *Pisum sativum.* This is also a confirmation of the results in the genomic PCR.

Figure 34 shows the gel electrophoresis done on the amplicon obtained with the genomic PCR if all products were put together first.

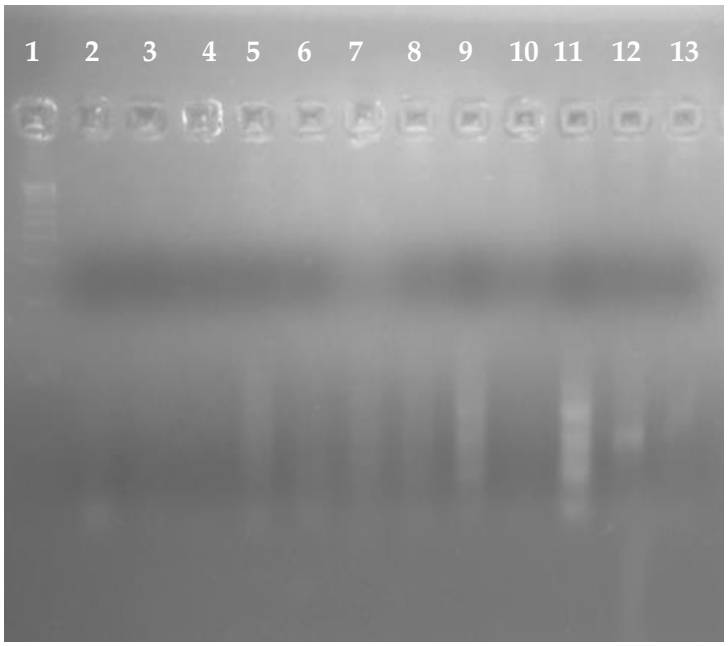



**Figure 32: gel electrophoresis barcoding PCR, all products from the genomic PCR put together**

1) Ladder
2) *Vicia faba*-Adaptor 1
3) *Vicia faba* -Adaptor 2
4) *Vicia faba*- Adaptor 3
5) *Vicia faba*- Adaptor 4
6) *Vicia faba*- Adaptor 5
7) *Vicia faba*- Adaptor 6
8) *Pisum sativum*- Adaptor 1
9) *Pisum sativum*- Adaptor 2
10) *Pisum sativum*- Adaptor 3
11) *Pisum sativum*- Adaptor 4
12) *Pisum sativum*- Adaptor 5
13) *Pisum sativum*- Adaptor 6

The gel electrophoresis gives a good result. Each adaptor gives amplification. Smear is visible, but this is expected, because it was a PCR of 60/45 SNP amplicons.

Table 16 gives an overview of the primers that did not work during the barcoding PCR. Appendix C contains the melting curves of the other primers.

**Table 16: Products of genomic PCR that did not give amplification in barcoding PCR**

| *Vicia faba* | *Pisum sativum* |
|---|---|
| F01-02 | P01-02 |
| F21-22 | P27-28 |
| F29-30 | P37-38 |
| F33-34 | P47-48 |
| F35-36 | P49-50 |
| F39-40 | P51-52 |
| F49-50 | P53-54 |
| F65-66 | P59-60 |
| F67-68 | P65-66 |
| F73-74 | P67-68 |
| F77-78 | P75-76 |
| F81-82 | P77-78 |
| F83-84 | P81-82 |
| | P83-84 |
| | P87-88 |
| | P89-90 |

Most of the primers that did not work during the barcoding PCR also did not work in the genomic PCR. P49-50, P51-52, F81-82,P83-84, P87-88 and P89-90 worked in the genomic PCR but not in the barcoding PCR. They gave a really small peak in the genomic PCR so the reason that they did not work in the barcoding PCR is probably due to low amplification in the genomic PCR.

35,5 % of the primers for *Pisum sativum* and 28,3 % for *Vicia faba* did not work and should be redesigned. All amplicons, also those where the primer did not work, were used for the DNA sequencing.

Appendix D shows the primers that worked in both genomic PCR and barcoding PCR.

## 6.4 DNA-sequencing

### 6.4.1 Library preparation

After the barcoding PCR, four different pools for each DNA sample were obtained ( see 5.3.2.1):

- Pool 1: Amplicons *Vicia faba* from separate barcoding PCR (six samples, adaptors 1-6 were used);
- Pool 2: Amplicons *Vicia faba* from collective barcoding PCR (six samples, adaptors 7-12 were used);
- Pool 3: Amplicons *Pisum sativum* from separate barcoding PCR (six samples, adaptors 1-6 were used);
- Pool 4: Amplicons *Pisum Sativum* from collective barcoding PCR (six samples, adaptors 7-12 were used);

These pools were purified, using the protocol from LifeTechnologies. Afterwards, the concentration of the tubes was measured (see 5.3.2.1).
Table 17 shows the concentration of the samples of pool one and three.

**Table 17: Concentration samples pool 1&3**

| DNA sample | Concentration (ng/ml) |
|---|---|
| 259 | 1860 |
| 305 | 700 |
| 304 | 668 |
| 258 | 796 |
| 281 | 806 |
| 283 | 632 |
| 102 | 754 |
| 140 | 488 |
| 104 | 502 |
| 112 | 392 |
| 100 | 1060 |
| 184 | 732 |

Table 18 shows the concentration of the samples of pool 2 and 4.

**Table 18: Concentration samples pool 2 & 4**

| DNA sample | Concentration (ng/ml) |
|---|---|
| 259 | 932 |
| 305 | 807 |
| 304 | 1250 |
| 258 | 578 |
| 281 | 1000 |
| 283 | 1380 |
| 102 | 818 |
| 140 | 1400 |
| 104 | 352 |
| 112 | 1160 |
| 100 | 2260 |
| 184 | 980 |

Afterwards, the concentration was converted to nM, using an online converter (size DNA: 130 bp). Tables 19 and 20 show the concentration of the pools in nM.

**Table 19: Concentration samples pool 1&3**

| DNA sample | Concentration (nM) |
|---|---|
| 259 | 22,05 |
| 305 | 8,3 |
| 304 | 7,92 |
| 258 | 9,43 |
| 281 | 9,55 |
| 283 | 7,49 |
| 102 | 8,94 |
| 140 | 5,78 |
| 104 | 5,95 |
| 112 | 4,65 |
| 100 | 12,56 |
| 184 | 8,68 |

**Table 20: Concentration samples pool 2&4**

| DNA sample | Concentration (nM) |
|---|---|
| 259 | 11,05 |
| 305 | 9,57 |
| 304 | 14,82 |
| 258 | 6,85 |
| 281 | 11,85 |
| 283 | 16,36 |
| 102 | 9,7 |
| 140 | 16,59 |
| 104 | 4,17 |
| 112 | 13,75 |
| 100 | 26,79 |
| 184 | 11,62 |

Before putting all samples together to obtain one pool with all samples, the samples first had to be diluted to the lowest concentration. The lowest concentration was the one from DNA sample 104, with a concentration of 4,17 nM.
Tables 21 and 22 show the amount of volume used of each sample to obtain a pool where each sample has an equimolar concentration.

**Table 21: Volume of samples of pool 1&3**

| DNA sample | Volume used in pool (µl) |
|---|---|
| 259 | 3,40 |
| 305 | 9,04 |
| 304 | 9,48 |
| 258 | 7,96 |
| 281 | 7,86 |
| 283 | 10,02 |
| 102 | 8,40 |
| 140 | 12,99 |
| 104 | 12,62 |
| 112 | 16,14 |
| 100 | 5,86 |
| 184 | 8,65 |

**Table 22: Volume of samples of pool 2&4**

| DNA sample | Volume used in pool (µl) |
|---|---|
| 259 | 6,79 |
| 305 | 7,84 |
| 304 | 5,06 |
| 258 | 10,96 |
| 281 | 6,33 |
| 283 | 4,59 |
| 102 | 7,74 |
| 140 | 4,52 |
| 104 | 18 |
| 112 | 5,46 |
| 100 | 2,80 |
| 184 | 6,46 |

One pool was obtained with all samples together. The concentration of this pool was measured again, and the concentration was 350 ng/ml. The concentration of this pool was converted to 26 pM ( 1 µl pool with 160,5 µl of TE)**.** Dilutions of 13 and 6,5 pM were also made.

### 6.4.2 Emulsion PCR

Table 23 gives the results of the quality control after enrichment of the samples obtained with the emulsion PCR. The quality was tested by the Ion Sphere Quality Control assay (see 5.3.2.2 ).

**Table 23: Results quality control after emulsion PCR**

| Concentration | AF 488 | AF647 | NC AF 488 | NC AF 647 | % template ISP's |
|---|---|---|---|---|---|
| 26 pM, unenriched | 341,9 | 219,6 | 22,6 | 3,6 | 82,03 % |
| 26 pM enriched | 220,5 | 187,5 | 22,6 | 3,6 | 112,68 % |
| 13 pM unenriched | 221,2 | 90,9 | 16,8 | 3,4 | 54,58 % |
| 13 pM enriched | 69,7 | 38,9 | 16,8 | 3,4 | 81,38 % |
| 6,5 pM unenriched | 55,1 | 28,3 | 18,6 | 3,8 | 81,39% |
| 6,5 pM enriched | 53,6 | 29,3 | 18,6 | 3,8 | 88,35% |

With a concentration of 26 and 13 pM, the highest fluorescence is obtained. The protocol says that the fluorescence of the unenriched beads (AF 488) has to be higher than 100 [43]. It also says that the ideal % unenriched template ISP's is between 10 and 30% [43]. The % template ISP's with 26 pM is 82,03% and that is too high. A concentration of 13 pM also gives a too high percentage, but it is closer to 30%. The fluorescence of 6,5 pM is very low therefore there are very few beads. It is possible that the beads got lost during the steps of the enrichment or that the library was too diluted,

and the emulsion PCR did not work. The concentration of 6,5 pM should be tested again to be sure what concentration is the most ideal. Relying on these results, 13 pM is the most ideal concentration to use for Ion Torrent Sequencing.

# 7 Conclusion

A combination of the OneTaq® DNA Polymerase kit on a TProfessional Thermocycler works the best for the genomic and barcoding PCR.

64,5 % of the primers for *Pisum sativum* and 71,7% of the primers for *Vicia faba* worked in both the genomic and barcoding PCR. The ones that did not work should be redesigned.

For the emulsion PCR, it is ideal to use an amplicon library with a concentration of 13 pM. The obtained percentage template ISP's is 54,58 %. A concentration of 26 pM gave a higher fluorescence but the % template ISP's is closer to the optimal range of 10-30% if 13 pM is used. Therefore, A concentration of 13 pM is more ideal to use for Ion Torrent Sequencing. The sequencing will give two responses for one SNP sequence, complementary to each other.

For the future, all primers that did not work should be redesigned and tested again. Also, the amplicon library concentration of 6,5 pM should be tested again. If a low fluorescence is obtained again, it is sure that 13 pM is the most ideal concentration to use for the Ion Torrent Sequencing.

## Bibliography


[1] European Union Seventh Framework, "Eurolegume," [Online]. Available: http://www.eurolegume.eu/. [Accessed 06 03 2015].

[2] M. Wojciechowski, J. Mahn and B. Jones, "Fabaceae legumes," 14 06 2006. [Online]. Available: http://tolweb.org/Fabaceae/21093. [Accessed 16 04 2015].

[3] W. Lindemann and C. R. Glover, "Nitrogen Fixation by Legumes," College of Agriculture and Home Economics, 05 1990. [Online]. Available: http://www.csun.edu/~hcbio027/biotechnology/lec10/lindemann.html. [Accessed 15 04 2015].

[4] B. Meentemeye, "Legume-Rhizobium," Kenyon College , 25 04 2012. [Online]. Available: https://microbewiki.kenyon.edu/index.php/Legume-Rhizobium. [Accessed 15 04 2015].

[5] Boundless, "The Legume-Root Nodule Symbiosis," Boundless, 14 11 2014. [Online]. Available: https://www.boundless.com/microbiology/textbooks/boundless-microbiology-textbook/microbial-ecology-16/microbial-symbioses-196/the-legume-root-nodule-symbiosis-993-7114/. [Accessed 14 04 2015].

[6] B. Biswas and H. Das, Plant-Microbe Interactions, New York: Springer Science + Bussiness Media, 1998.

[7] S. C. Wagner, "Biological Nitrogen Fixation," The Nature Education, 2011. [Online]. Available: http://www.nature.com/scitable/knowledge/library/biological-nitrogen-fixation-23570419. [Accessed 01 05 2015].

[8] M. Madigan, D. Clark, D. Stahl and J. Martinko, Brock Biology of Microorganisms 13th Edition, Benjamin Cummings, 2010, pp. 365-366.

[9] G. Stacy, R. Burris and H. Evans, Biological Nitrogen Fixation, United States of America : Chapman and Hall, 1992.

[10] Plant Nutrition Today, "DO HIGH YIELDING SOYBEANS NEED TO BE FERTILIZED WITH NITROGEN?," 05 2012. [Online]. Available: http://www.ipni.net/ipniweb/pnt.nsf/5a4b8be72a35cd46852568d9001a18da/f2bcb909db1a0267062579ab0066e840/$FILE/%234%20IPNI%20PNT%20Spring2012.pdf. [Accessed 15 04 2015].

[11] A. Dixit, X. Antony, N. Sharma and R. Tiwari, "Soybean constituents and their functional benefits," *Opportunity, Challenge and Scope of Natural,* pp. 367-383, 2011.

[12] FAO Water Development and Management Unit, "Crop Water Information: Soybean," FAO, [Online]. Available: http://www.fao.org/nr/water/cropinfo_soybean.html. [Accessed 12 04 2014].

[13] SoyaTech, "Soy facts," SoyaTech, [Online]. Available: http://www.soyatech.com/soy_facts.htm. [Accessed 10 04 2015].

[14] Food Standards Agency, "Animal Feed," [Online]. Available: https://www.food.gov.uk/business-industry/farmingfood/animalfeed. [Accessed 08 04 2015].

[15] H. Burley, "What's feeding our food?," 12 2008. [Online]. Available: http://www.foe.co.uk/sites/default/files/downloads/livestock_impacts_summary.pdf. [Accessed 20 03 2015].

[16] L. Bragg, M. Miller, J. Hillman, S. Koplan, T. Askey and T. J. Okun, Industry & Trade Summary- Animal feeds, Washington: United Stated International Trade Commission.

[17] The North Carolina Soybean Producers Associatio, "Uses of soybeans," The North Carolina Soybean Producers Associatio, [Online]. Available: http://ncsoy.org/media-resources/uses-of-soybeans/. [Accessed 20 03 2015].

[18] F. Kallenberg, "Natuurramp soja," MILIEUDEFENSIE MAGAZINE , Nederland, 2009.

[19] É. Giraud-Héraud and A. Fuentes Espinoza, "International market for soybeans coexistence of gm and non-gm products," The École polytechnique and Institut National de la Recherche Agronomique (I, France, 2012.

[20] Plant for a future, "Pisum sativum - L.," [Online]. Available: http://www.pfaf.org/user/plant.aspx?LatinName=Pisum+sativum. [Accessed 04 04 2015].

[21] R. Reichert and S. L. MacKenzie, "Composition of peas (Pisum sativum) varying widely in protein content," *J. Agric. Food Chem,* vol. 30, no. 2, pp. 312-317, 1982.

[22] D. J. Frame, "Pisum sativum L.," FAO, [Online]. Available: http://www.fao.org/ag/agp/AGPC/doc/GBASE/Data/pf000493.HTM. [Accessed 05

04 2014].

[23] V. Heuzé, G. Tran, R. Delagarde, F. Lebas and M. Lessire, "Faba bean (Vicia faba)," Feedipedia, 01 12 2014. [Online]. Available: http://www.feedipedia.org/node/4926. [Accessed 05 04 2015].

[24] B. Walsh, "Quantitative Genetics," *Encyclopedia of life sciences,* pp. 1-7, 2001.

[25] M. Agarwal, N. Shrivastava and H. Padh, "Advances in molecular marker techniques and their applications," *Plant Cell Rep,* pp. 617-631, 2007.

[26] K. Semagn, Å. Bjørnstad and M. N. Ndjiondjop, "An overview of molecular marker methods for plants," *African Journal of Biotechnology,* vol. 25, no. 5, pp. 2540-2568, 2006.

[27] A. J. Brookes, "The essence of SNPs," *Gene,* vol. 234, no. 2, pp. 177-186, 1999.

[28] "International HapMap Project," [Online]. Available: http://hapmap.ncbi.nlm.nih.gov/whatishapmap.html. [Accessed 19 03 2015].

[29] S. Flint-Garcia, J. Thornsberry and E. Buckler, "Structure of linkage disequilibrium in plants," *Annu. Rev. Plant Biol.,* vol. 2003, no. 54, pp. 357-374, 2003.

[30] "Modern Linkage Mapping with polymorphic DNA," *Los Alamos Science,* no. 20, pp. 94-99, 1992.

[31] A. Baxevanis and F. Ouellette, "PHYLOGENETIC ANALYSIS," in *Bioinformatics: a practical guide to the analysis of genes and proteins, second edition*, John Wiley & Sons, Inc., 2001.

[32] S. Kaura, R. Kimbere, N. Cogana, M. Materne, J. Forster and J. Paull, "SNP discovery and high-density genetic mapping in faba bean (Viciafaba L.) permits identification of QTLs for ascochyta blight resistance," *Elsevier.*

[33] A. Leonforte, S. Sudheesh, N. Cogan, P. Salisbury and N. Marc, "SNP marker discovery, linkage map construction".

[34] "Custom Primers - OligoPerfect™ Designer," Invitrogen, [Online]. Available: https://tools.lifetechnologies.com/content.cfm?pageid=9716. [Accessed 05 03 2015].

[35] "Ethidium Bromide Staining," National diagnostics, [Online]. Available: https://www.nationaldiagnostics.com/electrophoresis/article/ethidium-bromide-staining. [Accessed 02 05 2015].

[36] "Agarose Gel Electrophoresis of DNA," Biotechnology and genetic engineering, 15 01 2000. [Online]. Available: http://arbl.cvmbs.colostate.edu/hbooks/genetics/biotech/gels/agardna.html. [Accessed 11 04 2014].

[37] J. Liu, M. Yan, Z. Wang, L. Wang, Y. Zhou and B. Xiao, "Molecular diagnosis of $\alpha$-thalassemia by combining real-time PCR with SYBR Green1 and dissociation curve analysis," *Translational Research,* vol. 148, no. 1, pp. 6-12, 2006.

[38] "Instructions for use Agencourt AMPure XP," Beckman Coulter, 08 2013. [Online]. Available: https://www.beckmancoulter.com/wsrportal/techdocs?docname=B37419AA. [Accessed 01 05 2015].

[39] "DynaMag™-2 Magnet," LifeTechnologies, [Online]. Available: https://www.lifetechnologies.com/order/catalog/product/12321D. [Accessed 03 05 2015].

[40] "Qubit 2.0 Fluorometer," Georgia Genomics Facility, [Online]. Available: http://dna.uga.edu/services/qubit-2-0-fluorometer/. [Accessed 28 05 2015].

[41] Van Allen Way, "Ion Amplicon Library Preparation (Fusion method)," 4 11 2011. [Online]. Available: http://www.edgebio.com/sites/default/files/Amplicon%20Fusion%20Primer%20Design.pdf. [Accessed 30 04 2015].

[42] A. Vierstraete, *Next Generation Sequencing,* Ghent: Ghent University, 2012.

[43] L. Technologies, "Ion OneTouch™ 200 Template Kit v2 DL," Life Technologies Corporation, 2013.

[44] Life Technologies, "Ion Torrent-Amplicon sequencing," Life Technologies, 2011. [Online]. Available: https://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_094273.pdf. [Accessed 12 04 2015].

[45] Life Technologies, "Ion Torrent™ next-gen sequencing technology," Youtube, 19 05 2014. [Online]. Available: https://www.youtube.com/watch?v=WYBzbxIfuKs. [Accessed 22 03 2015].

[46] "eMolecules," [Online]. Available: https://www.emolecules.com/. [Accessed 09 04 2015].

[47] Life Technologies, "Ion PGM Sequencing 200 kit," Life Technologies Corporation, 2014.

[48] The Beehive, "Soil and Climatic Conditions for Soybeans," The Beehive, [Online]. Available: http://india.thebeehive.org/content/480/1191. [Accessed 15 04 2015].

[49] M. Laranjo, A. Alexandre and S. Oliveira, "Legume growth-promoting rhizobia: an overview on the Mesorhizobium genus," *Microbiological research,* vol. 17, no. 160, p. 02, 2014.

[50] O. W. Thomé, Flora von Deutschland Österreich und der Schweiz, Germany, 1885.

[51] "Robocycler Gradient 96 Stratagene 000595," TRIAD, [Online]. Available: http://www.triadsci.com/?site=preowned&item=3818&menu=7407. [Accessed 10 04 2014].

[52] "Major Laboratory Equipment," Laboratory of Biotransformation, [Online]. Available: http://www.biotrans.cas.cz/equipment/. [Accessed 10 04 2014].

[53] Scientific equipment repair, [Online]. Available: http://www.scientific-equipment.com/equipment-for-sale/thermalcycling-pcr/geneamp-pcr-system-9700-with-96-well-sample-block-module?ln=2679. [Accessed 10 04 2015].

[54] N. Downey, "Interpreting Melt Curves: An Indicator, Not a Diagnosis," Integrated DNA technologies, [Online]. Available: https://eu.idtdna.com/pages/decoded/decoded-articles/core-concepts/decoded/2014/01/20/interpreting-melt-curves-an-indicator-not-a-diagnosis. [Accessed 03 04 2014].

[55] "Information on EC 1.19.6.1 - nitrogenase (flavodoxin)," BRENDA, The comprhenseive Enzyme Information System, 01 2015. [Online]. Available: http://www.brenda-enzymes.org/enzyme.php?ecno=1.19.6.1. [Accessed 22 05 2015].

[56] "Markers Selection Guide," Sigma Aldrich, [Online]. Available: http://www.sigmaaldrich.com/technical-documents/articles/biology/marker-selection-guide.html. [Accessed 02 06 2015].

## Appendix A: Designed primers

### *Vicia faba Forward Primers*

| Locus_Name | Commun sequence +SNP1for |
|---|---|
| SNP_50000404 | CCTCTCTATGGGCAGTCGGTGATTCAAAACTTGGAAAGATGA |
| SNP_50001182 | CCTCTCTATGGGCAGTCGGTGATTCCCTTTTCTCTCTCAAAA |
| SNP_50002158 | CCTCTCTATGGGCAGTCGGTGATTTCATTTGGATAATCCTTTC |
| SNP_50002306 | CCTCTCTATGGGCAGTCGGTGATTGTGGTGGCCAAATCT |
| SNP_50002318 | CCTCTCTATGGGCAGTCGGTGATTGCAGCGTTGATTATGTT |
| | |
| SNP_50000197 | CCTCTCTATGGGCAGTCGGTGATTTGCTTGCTCATACGC |
| SNP_50000487 | CCTCTCTATGGGCAGTCGGTGATTCAGAACCAGTGGCAGT |
| SNP_50000557 | CCTCTCTATGGGCAGTCGGTGATTCTCCATTGAGCAGCA |
| SNP_50001252 | CCTCTCTATGGGCAGTCGGTGATTCCGTGGATTCCTCAC |
| SNP_50000760 | CCTCTCTATGGGCAGTCGGTGATTTTACCATGGGCCTCT |
| | |
| SNP_50000125 | CCTCTCTATGGGCAGTCGGTGATTGAAAACCCCTGCAAA |
| SNP_50000432 | CCTCTCTATGGGCAGTCGGTGATTGCTGGGACAGACTCC |
| SNP_50001365 | CCTCTCTATGGGCAGTCGGTGATTTTGGCCTTGATAGCC |
| SNP_50001987 | CCTCTCTATGGGCAGTCGGTGATTGGTTTTGTGGGCATT |
| SNP_50002207 | CCTCTCTATGGGCAGTCGGTGATTCCCCACTGTCGTTTT |
| | |
| SNP_50000307 | CCTCTCTATGGGCAGTCGGTGATTGAAGTTGCGGAAAGC |
| SNP_50001793 | CCTCTCTATGGGCAGTCGGTGATTCAACCCATGCTCCTT |
| SNP_50001828 | CCTCTCTATGGGCAGTCGGTGATTTTCGGTTATGGAAAGC |
| SNP_50001916 | CCTCTCTATGGGCAGTCGGTGATTCAGTAGCTTCCATACCG |
| SNP_50000886 | CCTCTCTATGGGCAGTCGGTGATTTGCATGCAGCTTTAGA |
| | |
| SNP_50000057 | CCTCTCTATGGGCAGTCGGTGATTGGGTCTTCTCGACCTT |
| SNP_50000069 | CCTCTCTATGGGCAGTCGGTGATTAAAATGGCTGAAACAAA |
| SNP_50000225 | CCTCTCTATGGGCAGTCGGTGATTGCAGATGAGAGGTGGA |
| SNP_50000911 | CCTCTCTATGGGCAGTCGGTGATTGGCAGCAAACACCTT |
| SNP_50001150 | CCTCTCTATGGGCAGTCGGTGATTCAATGGCCTTTGGAG |
| | |
| SNP_50000347 | CCTCTCTATGGGCAGTCGGTGATTGCCGAGTTCATAGCC |
| SNP_50000965 | CCTCTCTATGGGCAGTCGGTGATTCAGTTGTGTGATGAGCA |
| SNP_50001040 | CCTCTCTATGGGCAGTCGGTGATTCCGGAAGAACAGTGG |
| SNP_50001146 | CCTCTCTATGGGCAGTCGGTGATTGAGAAGTTGCAGCAAGA |
| SNP_50000089 | CCTCTCTATGGGCAGTCGGTGATTTGCATTCAGCAAAGC |
| | |
| SNP_50000217 | CCTCTCTATGGGCAGTCGGTGATTGCATTGTGGGAGAAGA |
| SNP_50000310 | CCTCTCTATGGGCAGTCGGTGATTTTGTAGCTGAGCCTGTT |
| SNP_50000505 | CCTCTCTATGGGCAGTCGGTGATTTTTCCAGCAAATCCAT |
| SNP_50000739 | CCTCTCTATGGGCAGTCGGTGATTCCAGGGAGTTGGACA |
| SNP_50000787 | CCTCTCTATGGGCAGTCGGTGATTACCCTGTCCACAAGC |

| | |
|---|---|
| | |
| SNP_50000084 | CCTCTCTATGGGCAGTCGGTGATTCAAAAACAAATGGTAGGA |
| SNP_50000285 | CCTCTCTATGGGCAGTCGGTGATTGGTTTTGTCCCAATGA |
| SNP_50000440 | CCTCTCTATGGGCAGTCGGTGATTAACTGAGCTTTCGAACAT |
| SNP_50001281 | CCTCTCTATGGGCAGTCGGTGATTTCCCAAACAAAGCAA |
| SNP_50002450 | CCTCTCTATGGGCAGTCGGTGATTGGCCGATACTTCTGC |
| | |
| SNP_50000763 | CCTCTCTATGGGCAGTCGGTGATTACTCTCAATCCTGCACTT |
| SNP_50001516 | CCTCTCTATGGGCAGTCGGTGATTCTGCAACTTCCATGAAT |
| SNP_50001531 | CCTCTCTATGGGCAGTCGGTGATTCTGGCAACATTCATCA |
| SNP_50001647 | CCTCTCTATGGGCAGTCGGTGATTTGGAGTTGTTCAGGGTA |
| SNP_50001880 | CCTCTCTATGGGCAGTCGGTGATTTGGAGAACCATTACCAA |
| | |
| SNP_50000181 | CCTCTCTATGGGCAGTCGGTGATTAACTCCATCGGCAAC |
| SNP_50000203 | CCTCTCTATGGGCAGTCGGTGATTCAGAAAACTGCATAGCC |
| SNP_50000402 | CCTCTCTATGGGCAGTCGGTGATTGGGTTTCCAAATCTCC |
| SNP_50000824 | CCTCTCTATGGGCAGTCGGTGATTCAAAAAGACTCCAGTGC |
| SNP_50001586 | CCTCTCTATGGGCAGTCGGTGATTTCCCAGGTGATTTCAT |
| | |
| SNP_50000208 | CCTCTCTATGGGCAGTCGGTGATTGAGGAATGAAAGGCTGT |
| SNP_50001889 | CCTCTCTATGGGCAGTCGGTGATTCCCTGGCTTCTCAAG |
| SNP_50002378 | CCTCTCTATGGGCAGTCGGTGATTTCCAGCTCTGTTACCC |
| SNP_50001498 | CCTCTCTATGGGCAGTCGGTGATTTTAACCAACGCCAAA |
| SNP_50002155 | CCTCTCTATGGGCAGTCGGTGATTCCCTCAGTCGGAGAA |
| | |
| SNP_50000022 | CCTCTCTATGGGCAGTCGGTGATTTCAAAGCCTCCAGGT |
| SNP_50000127 | CCTCTCTATGGGCAGTCGGTGATTTTGTTCAATTTTTCAATCA |
| SNP_50000436 | CCTCTCTATGGGCAGTCGGTGATTTTTCACAATGGGAACAT |
| SNP_50001725 | CCTCTCTATGGGCAGTCGGTGATTTGGGATCTGACAAGGA |
| SNP_50002190 | CCTCTCTATGGGCAGTCGGTGATTAGCATTTGCGATTACC |

***Vicia faba* reverse primers**

| Locus_Name | Communsequence+SNP1rev |
|---|---|
| SNP_50000404 | CCTCTCTATGGGCAGTCGGTGATTCGGTGACAAACCAACT |
| SNP_50001182 | CCTCTCTATGGGCAGTCGGTGATTACAATAGCCACAACCAA |
| SNP_50002158 | CCTCTCTATGGGCAGTCGGTGATTATGGGAACGTTGTGAG |
| SNP_50002306 | CCTCTCTATGGGCAGTCGGTGATTATCAAATGAGCCTTGC |
| SNP_50002318 | CCTCTCTATGGGCAGTCGGTGATTCCAATGCCTCAGTCC |
| | |
| SNP_50000197 | CCTCTCTATGGGCAGTCGGTGATTTGGGCCAATATGATTC |
| SNP_50000487 | CCTCTCTATGGGCAGTCGGTGATTAAGGTGGAGGAAAAATG |
| SNP_50000557 | CCTCTCTATGGGCAGTCGGTGATTTGGTCAAAAGCACCA |
| SNP_50001252 | CCTCTCTATGGGCAGTCGGTGATTTTGGCAACAATGTCAG |
| SNP_50000760 | CCTCTCTATGGGCAGTCGGTGATTTGGTATCAATTCCCTTG |
| | |
| SNP_50000125 | CCTCTCTATGGGCAGTCGGTGATTGTCGCTGAGCAGGTT |
| SNP_50000432 | CCTCTCTATGGGCAGTCGGTGATTCCTGAACATCTGACTCG |
| SNP_50001365 | CCTCTCTATGGGCAGTCGGTGATTGGGTTTGCATTTCGT |
| SNP_50001987 | CCTCTCTATGGGCAGTCGGTGATTACCAGCATTAGCATGTC |
| SNP_50002207 | CCTCTCTATGGGCAGTCGGTGATTGCATGTTGTGGCAAA |
| | |
| SNP_50000307 | CCTCTCTATGGGCAGTCGGTGATTTGGAGCAGCAAAGTG |
| SNP_50001793 | CCTCTCTATGGGCAGTCGGTGATTGCATGCAAATGCAAA |
| SNP_50001828 | CCTCTCTATGGGCAGTCGGTGATTTTCGGTTATGGAAAGC |
| SNP_50001916 | CCTCTCTATGGGCAGTCGGTGATTAGCTCTGCTGTGGAGA |
| SNP_50000886 | CCTCTCTATGGGCAGTCGGTGATTACCAACATGACGAATCA |
| | |
| SNP_50000057 | CCTCTCTATGGGCAGTCGGTGATTACCGACAATTTTACCG |
| SNP_50000069 | CCTCTCTATGGGCAGTCGGTGATTCCAGCTCTTCCACAAT |
| SNP_50000225 | CCTCTCTATGGGCAGTCGGTGATTTGGCTATCCAAGAAGC |
| SNP_50000911 | CCTCTCTATGGGCAGTCGGTGATTTGCGAAGAGGTTGATT |
| SNP_50001150 | CCTCTCTATGGGCAGTCGGTGATTTTCTTCCCCATGTCC |
| | |
| SNP_50000347 | CCTCTCTATGGGCAGTCGGTGATTCAGCGGTGCAGTAAA |
| SNP_50000965 | CCTCTCTATGGGCAGTCGGTGATTCGAAGCAACGAAAAA |
| SNP_50001040 | CCTCTCTATGGGCAGTCGGTGATTTGGGAATCCTTGGAA |
| SNP_50001146 | CCTCTCTATGGGCAGTCGGTGATTATGGAGCCAGTGGAA |
| SNP_50000089 | CCTCTCTATGGGCAGTCGGTGATTTGACAGTGGCAGTGAA |
| | |
| SNP_50000217 | CCTCTCTATGGGCAGTCGGTGATTGTGATCACCCTTCACAA |
| SNP_50000310 | CCTCTCTATGGGCAGTCGGTGATTCTCCATATGCCAGTGAC |
| SNP_50000505 | CCTCTCTATGGGCAGTCGGTGATTAGGCTTACCCGTTCA |
| SNP_50000739 | CCTCTCTATGGGCAGTCGGTGATTGATGCAGGAGCAGGT |
| SNP_50000787 | CCTCTCTATGGGCAGTCGGTGATTGAGGGACGGTTCTTG |
| | |

| | |
|---|---|
| SNP_50000084 | CCTCTCTATGGGCAGTCGGTGATTCGACAATGTCCGTCA |
| SNP_50000285 | CCTCTCTATGGGCAGTCGGTGATTGGCAGATCAACTCACAG |
| SNP_50000440 | CCTCTCTATGGGCAGTCGGTGATTCTTAAATTCCGCAAGG |
| SNP_50001281 | CCTCTCTATGGGCAGTCGGTGATTTCCCACCAACCAGAT |
| SNP_50002450 | CCTCTCTATGGGCAGTCGGTGATTAGGAGTCCCCAAACC |
| | |
| SNP_50000763 | CCTCTCTATGGGCAGTCGGTGATTCCACAAGCACCAAAA |
| SNP_50001516 | CCTCTCTATGGGCAGTCGGTGATTCCGAATCTTAGCCTCA |
| SNP_50001531 | CCTCTCTATGGGCAGTCGGTGATTTGGAGACCTTTTCTGC |
| SNP_50001647 | CCTCTCTATGGGCAGTCGGTGATTCAAACCACACGAGCA |
| SNP_50001880 | CCTCTCTATGGGCAGTCGGTGATTATAGGCCAACCTAATGG |
| | |
| SNP_50000181 | CCTCTCTATGGGCAGTCGGTGATTCCAGGAAGCTGTGTGT |
| SNP_50000203 | CCTCTCTATGGGCAGTCGGTGATTGCAGCAGTCCTTGGT |
| SNP_50000402 | CCTCTCTATGGGCAGTCGGTGATTAACTCCATCGGCAAC |
| SNP_50000824 | CCTCTCTATGGGCAGTCGGTGATTGGATCAGCAGGAAGC |
| SNP_50001586 | CCTCTCTATGGGCAGTCGGTGATTTCTAGGATAAGGGACCAA |
| | |
| SNP_50000208 | CCTCTCTATGGGCAGTCGGTGATTCTGAGGACGCTTGAAT |
| SNP_50001889 | CCTCTCTATGGGCAGTCGGTGATTAGCTTCAGCGATGACT |
| SNP_50002378 | CCTCTCTATGGGCAGTCGGTGATTCATGGGCAAATGAAGT |
| SNP_50001498 | CCTCTCTATGGGCAGTCGGTGATTCGTCCAGACGTGGAT |
| SNP_50002155 | CCTCTCTATGGGCAGTCGGTGATTATACCCGGATGTTGG |
| | |
| SNP_50000022 | CCTCTCTATGGGCAGTCGGTGATTTTCCATTTGGATCAGG |
| SNP_50000127 | CCTCTCTATGGGCAGTCGGTGATTTCCCTTCCGAGTTCA |
| SNP_50000436 | CCTCTCTATGGGCAGTCGGTGATTGCCATTCCAGAAACC |
| SNP_50001725 | CCTCTCTATGGGCAGTCGGTGATTCCCACCTGGGTTGTA |
| SNP_50002190 | CCTCTCTATGGGCAGTCGGTGATTTTGCCAGATCGTTTG |

### *Pisum sativum* forward primers

| Locus_Name | Commun sequence +SNP1for |
|---|---|
| SNP_100000272 | CCTCTCTATGGGCAGTCGGTGATTCCAGACCACCTCAGC |
| SNP_100000810 | CCTCTCTATGGGCAGTCGGTGATTAAGAGAGCGGCTTTG |
| SNP_100000629 | CCTCTCTATGGGCAGTCGGTGATTAGCTGCAAAGGGAGA |
| SNP_100000443 | CCTCTCTATGGGCAGTCGGTGATTTCCAGGTGCTCAACA |
| SNP_100000341 | CCTCTCTATGGGCAGTCGGTGATTCCCGAACCATCTGTT |
| | |
| SNP_100000460 | CCTCTCTATGGGCAGTCGGTGATTCCTGAAGCCAATGAAG |
| SNP_100000648 | CCTCTCTATGGGCAGTCGGTGATTCAGTCCCCAAACCTG |
| SNP_100000616 | CCTCTCTATGGGCAGTCGGTGATTTGCAATACACTTGCATAAT |
| SNP_100000543 | CCTCTCTATGGGCAGTCGGTGATTTAACGAGCGCAAAAG |
| SNP_100000664 | CCTCTCTATGGGCAGTCGGTGATTAGGCGAAATTGAAGC |
| | |
| SNP_100000360 | CCTCTCTATGGGCAGTCGGTGATTAAGGCTAGGAGCTAATCA |
| SNP_100000659 | CCTCTCTATGGGCAGTCGGTGATTAGGGAGAGCCAGAAAG |
| SNP_100000091 | CCTCTCTATGGGCAGTCGGTGATTCCTCACTCAATGACCAG |
| SNP_100000063 | CCTCTCTATGGGCAGTCGGTGATTTGGAATCCAGCACCT |
| SNP_100000288 | CCTCTCTATGGGCAGTCGGTGATTCATTGATGGTCTTTTTCA |
| | |
| SNP_100000395 | CCTCTCTATGGGCAGTCGGTGATTATTCACCCCACCTGA |
| SNP_100000035 | CCTCTCTATGGGCAGTCGGTGATTGTGCGCTCGCTTTA |
| SNP_100000786 | CCTCTCTATGGGCAGTCGGTGATTGACCACACGGAAACC |
| SNP_100000252 | CCTCTCTATGGGCAGTCGGTGATTCCTGACCACCGGTAT |
| SNP_100000071 | CCTCTCTATGGGCAGTCGGTGATTGGGACGTACTGGATCA |
| | |
| SNP_100000171 | CCTCTCTATGGGCAGTCGGTGATTCCAGAAGCTCTGATGC |
| SNP_100000408 | CCTCTCTATGGGCAGTCGGTGATTCACGGACCTCAAACC |
| SNP_100000335 | CCTCTCTATGGGCAGTCGGTGATTCCGGATGTGTGTGC |
| SNP_100000337 | CCTCTCTATGGGCAGTCGGTGATTACCCAGGTGACTGATG |
| SNP_100000653 | CCTCTCTATGGGCAGTCGGTGATTCCTGAGCGCCTCTAT |
| | |
| SNP_100000485 | CCTCTCTATGGGCAGTCGGTGATTGGAATTCCTGCCAAA |
| SNP_100000064 | CCTCTCTATGGGCAGTCGGTGATTGTACGGCAAGGTTCC |
| SNP_100000584 | CCTCTCTATGGGCAGTCGGTGATTCCTTGATTGGCTGGT |
| SNP_100000736 | CCTCTCTATGGGCAGTCGGTGATTGGGTGCGTAGTCGAT |
| SNP_100000301 | CCTCTCTATGGGCAGTCGGTGATTCCTCAAGCACGGTTT |
| | |
| SNP_100000639 | CCTCTCTATGGGCAGTCGGTGATTGTGTCCGAGGAAACAA |
| SNP_100000569 | CCTCTCTATGGGCAGTCGGTGATTGCTGCTTCCTTCTCAA |
| SNP_100000270 | CCTCTCTATGGGCAGTCGGTGATTTGCCACGAGTCAACA |
| SNP_100000594 | CCTCTCTATGGGCAGTCGGTGATTGTTGCGCTAGCAAGTT |
| SNP_100000702 | CCTCTCTATGGGCAGTCGGTGATTTCTCCCCAGGTACAAA |
| | |

| | |
|---|---|
| SNP_100000847 | CCTCTCTATGGGCAGTCGGTGATTTGGCGTAACGAACAC |
| SNP_100000655 | CCTCTCTATGGGCAGTCGGTGATTAGCGTGATCCCTGAG |
| SNP_100000654 | CCTCTCTATGGGCAGTCGGTGATTAGCACGGAGATGCTT |
| SNP_100000056 | CCTCTCTATGGGCAGTCGGTGATTGGGGATTCTGTTACCC |
| SNP_100000058 | CCTCTCTATGGGCAGTCGGTGATTGAAGCTTTGCCACCT |
| | |
| SNP_100000908 | CCTCTCTATGGGCAGTCGGTGATTCAGCTAGCCACCCTAA |
| SNP_100000703 | CCTCTCTATGGGCAGTCGGTGATTTTCGGCTCTAGACTGG |
| SNP_100000255 | CCTCTCTATGGGCAGTCGGTGATTTGATCCAACGATCCA |
| SNP_100000290 | CCTCTCTATGGGCAGTCGGTGATTAGCCAACTTGCCATC |
| SNP_100000332 | CCTCTCTATGGGCAGTCGGTGATTTGATGCTGCCACTGA |

***Pisum sativum* reverse primers**

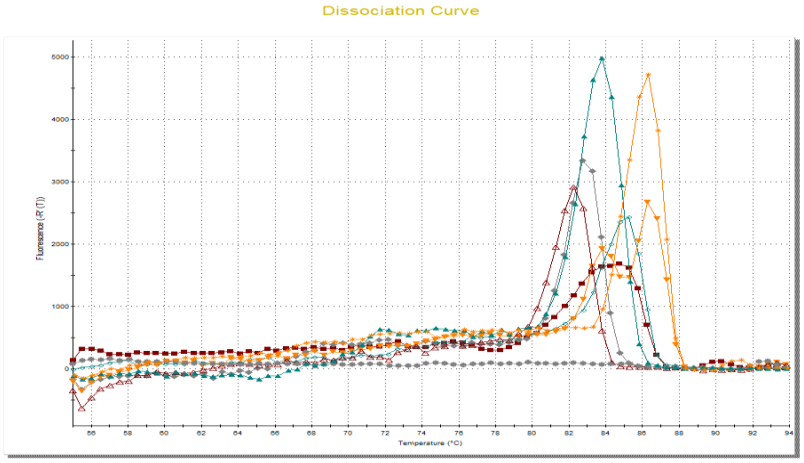
| Locus_Name | Commun sequence +SNP1rev |
|---|---|
| | |
| SNP_100000272 | CCTCTCTATGGGCAGTCGGTGATTGCCCATTTGAAGGTT |
| SNP_100000810 | CCTCTCTATGGGCAGTCGGTGATTTTCCTGCAGCTTTGA |
| SNP_100000629 | CCTCTCTATGGGCAGTCGGTGATTGGGCCTGGAAATGT |
| SNP_100000443 | CCTCTCTATGGGCAGTCGGTGATTAGGACCACCAGACCA |
| SNP_100000341 | CCTCTCTATGGGCAGTCGGTGATTGATGCCGGAGCTTAC |
| | |
| SNP_100000460 | CCTCTCTATGGGCAGTCGGTGATTATGCTGACCCAAGAAA |
| SNP_100000648 | CCTCTCTATGGGCAGTCGGTGATTTGTGATTCCAGGAGGA |
| SNP_100000616 | CCTCTCTATGGGCAGTCGGTGATTTCACGGGTTTTAGCC |
| SNP_100000543 | CCTCTCTATGGGCAGTCGGTGATTGTGGGTTCTTGTCAGC |
| SNP_100000664 | CCTCTCTATGGGCAGTCGGTGATTACTCTAACACCCCTTTCA |
| | |
| SNP_100000360 | CCTCTCTATGGGCAGTCGGTGATTCCCGGAACTTGGAG |
| SNP_100000659 | CCTCTCTATGGGCAGTCGGTGATTCGTTATCCCAAAATGC |
| SNP_100000091 | CCTCTCTATGGGCAGTCGGTGATTTGCTCATGGCAAGAC |
| SNP_100000063 | CCTCTCTATGGGCAGTCGGTGATTCAGTGGCACCAGACA |
| SNP_100000288 | CCTCTCTATGGGCAGTCGGTGATTGGCCATAACTTTGGAG |
| | |
| SNP_100000395 | CCTCTCTATGGGCAGTCGGTGATTAGCTGCACCACCAGT |
| SNP_100000035 | CCTCTCTATGGGCAGTCGGTGATTACCACTTTCGGTGTTG |
| SNP_100000786 | CCTCTCTATGGGCAGTCGGTGATTTCGTGCTCTTCGTGA |
| SNP_100000252 | CCTCTCTATGGGCAGTCGGTGATTCCTGATCCTTGCACTC |
| SNP_100000071 | CCTCTCTATGGGCAGTCGGTGATTGATTCAAGCAGGGTGA |
| | |
| SNP_100000171 | CCTCTCTATGGGCAGTCGGTGATTTCCATCCTCCAAAGC |
| SNP_100000408 | CCTCTCTATGGGCAGTCGGTGATTCATTCAGCCCATCCT |
| SNP_100000335 | CCTCTCTATGGGCAGTCGGTGATTACGCACTTCCTCTCAA |
| SNP_100000337 | CCTCTCTATGGGCAGTCGGTGATTATCCTCCGGCTCTTT |
| SNP_100000653 | CCTCTCTATGGGCAGTCGGTGATTGCAAAGCCCTTCATC |
| | |
| SNP_100000485 | CCTCTCTATGGGCAGTCGGTGATTCCAGGCTGTGGTGA |
| SNP_100000064 | CCTCTCTATGGGCAGTCGGTGATTTCCCAATTCAGCACA |
| SNP_100000584 | CCTCTCTATGGGCAGTCGGTGATTGTGCTCATGCTATTGCT |
| SNP_100000736 | CCTCTCTATGGGCAGTCGGTGATTACAAGGTGCCACAAAC |
| SNP_100000301 | CCTCTCTATGGGCAGTCGGTGATTATCCATAGCCGCAAC |
| | |
| SNP_100000639 | CCTCTCTATGGGCAGTCGGTGATTTGTCATGCCAGTTGC |
| SNP_100000569 | CCTCTCTATGGGCAGTCGGTGATTATGGGAAAGGAAAAGG |
| SNP_100000270 | CCTCTCTATGGGCAGTCGGTGATTAGAGCATGCGTGACA |
| SNP_100000594 | CCTCTCTATGGGCAGTCGGTGATTCGAATGATTGCATGG |
| SNP_100000702 | CCTCTCTATGGGCAGTCGGTGATTCCTGATGGTCCTCTCA |

| | |
|---|---|
| SNP_100000847 | CCTCTCTATGGGCAGTCGGTGATTTGACCTAAACGACGC |
| SNP_100000655 | CCTCTCTATGGGCAGTCGGTGATTCGTTACAGGGCGATT |
| SNP_100000654 | CCTCTCTATGGGCAGTCGGTGATTCGGAGCTCTTCTCGT |
| SNP_100000056 | CCTCTCTATGGGCAGTCGGTGATTTTGGCTGATTGACCA |
| SNP_100000058 | CCTCTCTATGGGCAGTCGGTGATTGGTGCCTCCTCTTCA |
| | |
| SNP_100000908 | CCTCTCTATGGGCAGTCGGTGATTACCGTTTCTGCGTCT |
| SNP_100000703 | CCTCTCTATGGGCAGTCGGTGATTACAGCAGCCCAAACT |
| SNP_100000255 | CCTCTCTATGGGCAGTCGGTGATTTGTGGACGCTACGAA |
| SNP_100000290 | CCTCTCTATGGGCAGTCGGTGATTGCCTCGCTTGAAAAT |
| SNP_100000332 | CCTCTCTATGGGCAGTCGGTGATTATCCCCGGAGGTATT |

## Appendix B : Melting curves genomic PCR

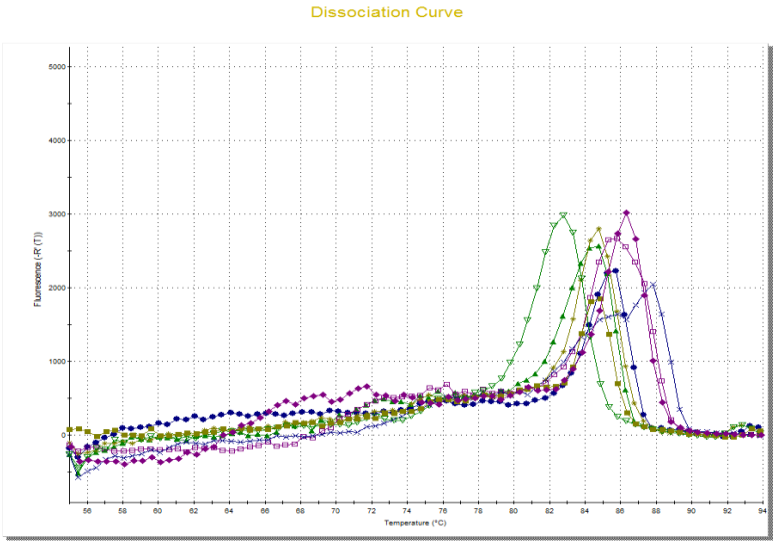
### Primers F17-F32 and P17-P32

***Pisum sativum***




■ : P17-18
♦ : P19-20
°: P21-22
▼ : P23-24
Δ : P25-26
• : P27-28
▲ : P29-30
✱ : P31-32

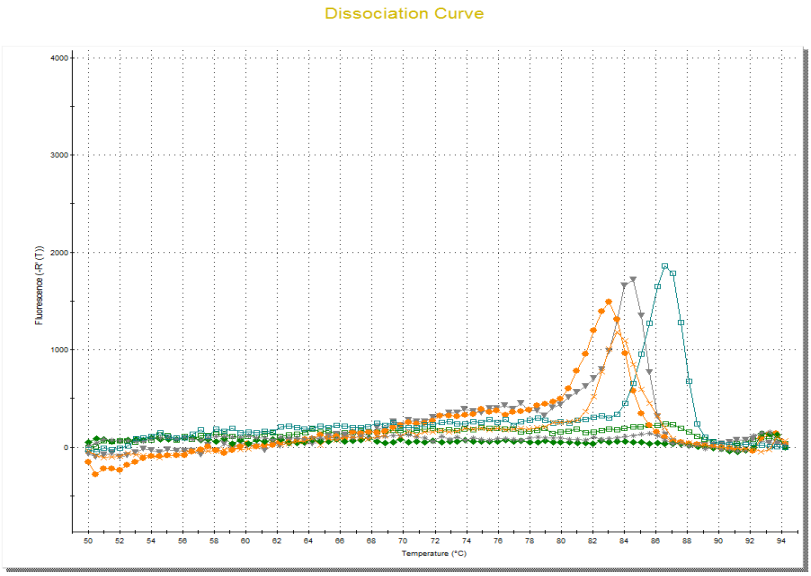
***Vicia faba***




• : F17-18
▲ : F19-20
✱ : F21-22
❑: F23-24
✕: F25-26
≤ (reversed empty triangle): F27-28
■ : F29-30
♦ : F31-32

**Primers F33-46 and P33-P42**

***Vicia faba***




❑: F33-34

♦ : F35-36

≤ (reversed empty triangle): F37-38

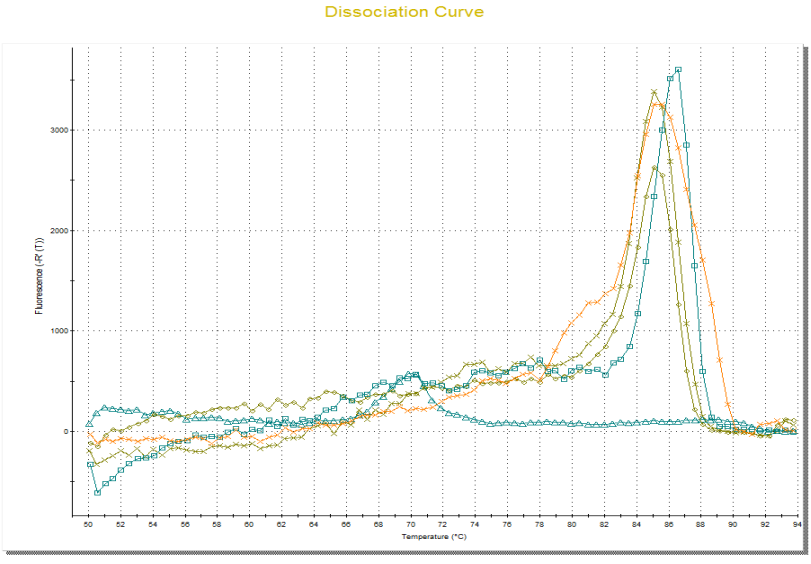
✱ : F39-40

❑: F41-42

• : F43-44

✕: F45-46

***Pisum sativum***




✕: P33-34

°: P35-36

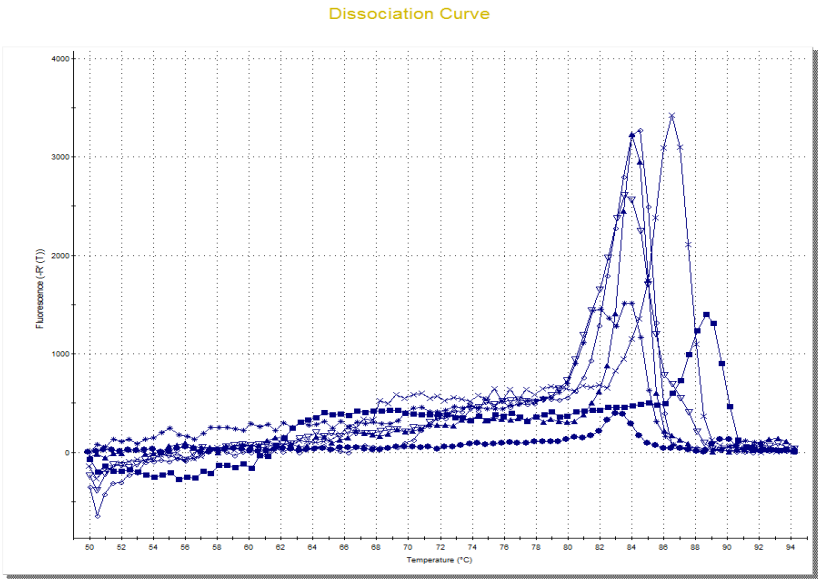
Δ : P37-38

❑: P39-40

✕: P41-42

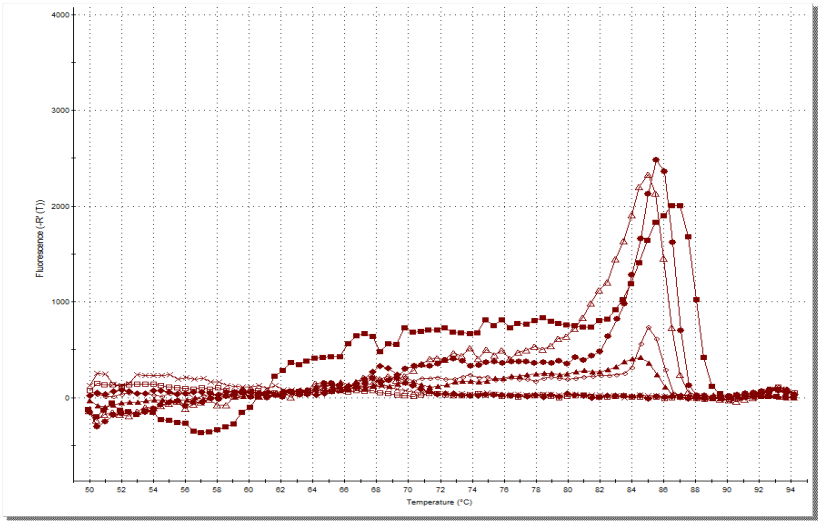
**Primers F49-64 and P49-64**

***Vicia faba***




• : F49-50
✕: F51-52
°: F53-54
▲ : F55-56
≤ (reversed empty triangle): F57-58
▼ : F59-60
✱ : F61-62
■ : F63-64

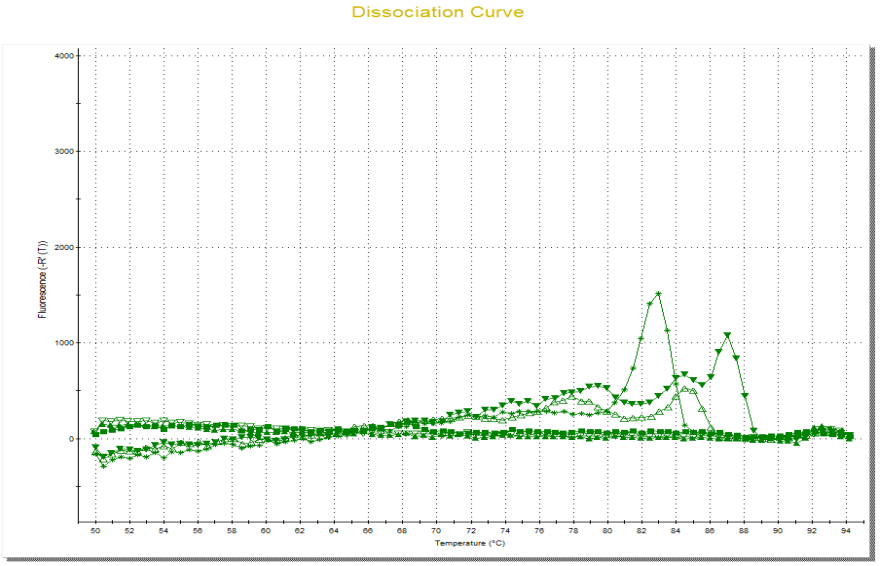
***Pisum sativum***




■ : P49-50
Δ : P51-52
❑: P53-54
♦ : P55-56
• : P57-58
✕: P59-60
°: P61-62
▲ : P63-64

**Primers F65-76 and P65-76**

***Vicia faba***




▲ : F65-67

≤ (reversed empty triangle): F67-68

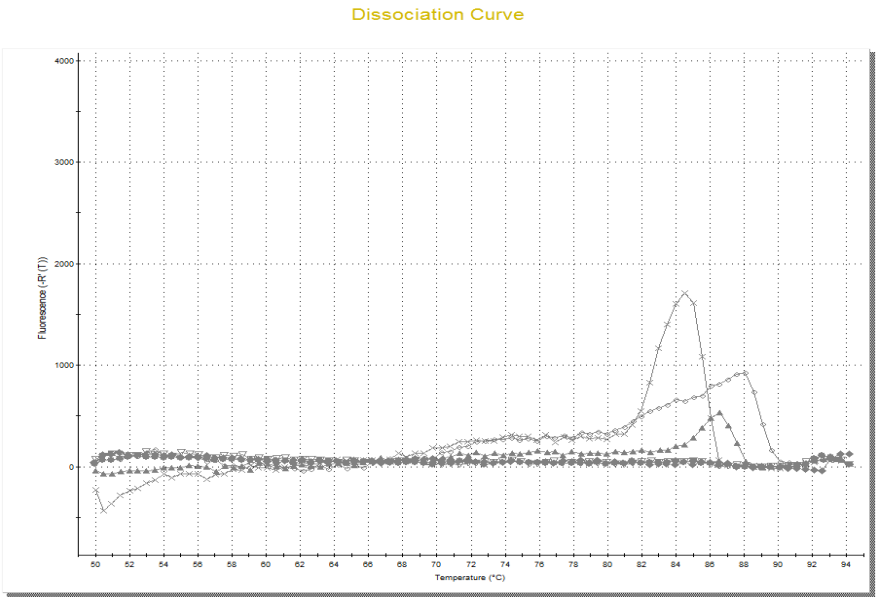
▼ : F69-70

✱ : F71-72

■ : F73-74

Δ : F75-76

***Pisum sativum***




♦ : P65-66

• : P67-68

✕: 69-70

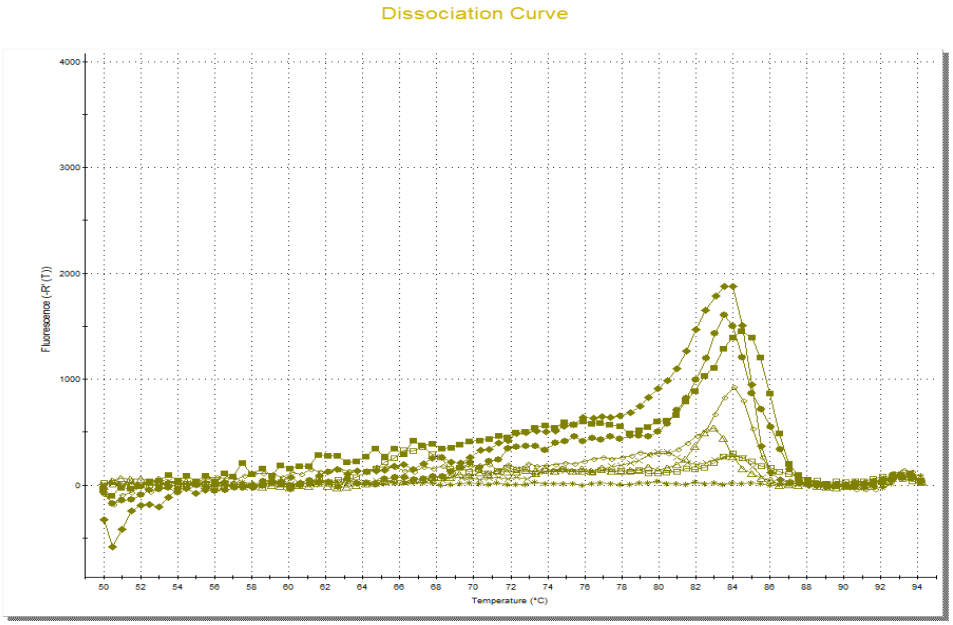
°: P71-72

▲ : P73-74

♦ : P75-76

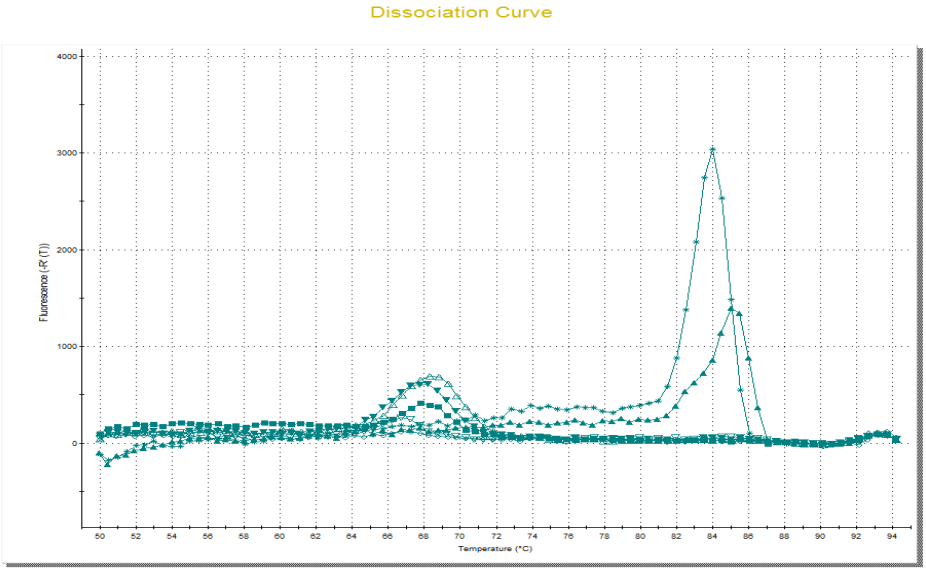
**Primers F77-92 and P77-90**

***Vicia faba***




✱ : F77-78
■ : F79-80
Δ : F81-82
❑: F83-84
♦ : F85-86
● : F87-88
✕: F89-90
°: F91-92

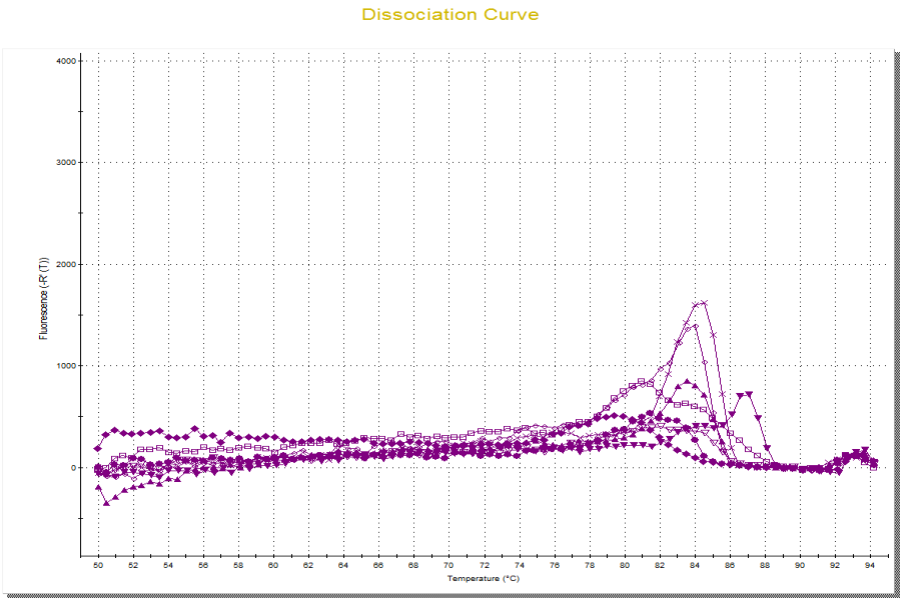
***Pisum sativum***




°: P77-78
▲ : P79-80
≤ (reversed empty triangle): P81-82
▼ : P83-84
✱ : P85-86
■ : P87-88
Δ : P89-90

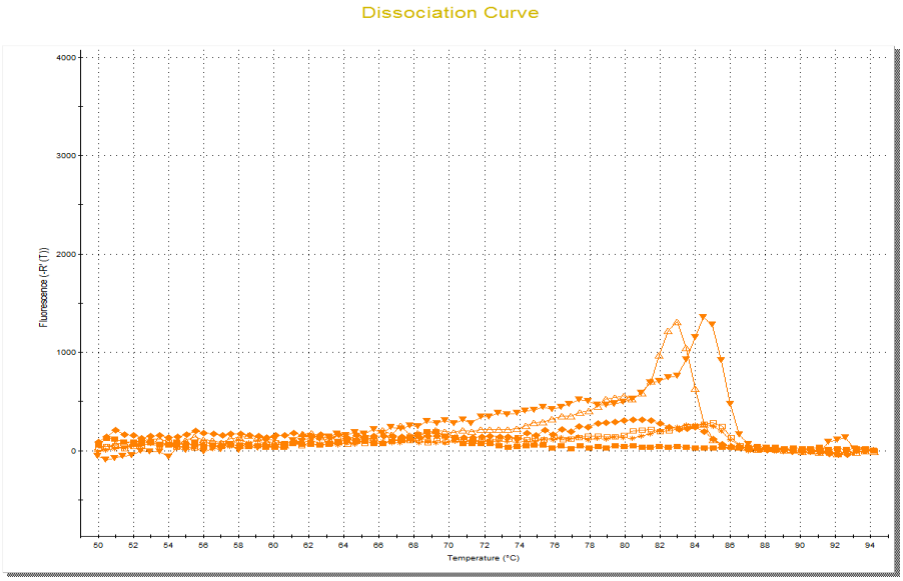
**Primers F93-F108 and F109-F120**

***Vicia faba***




❑: P93-94
♦ : P95-96
• : P97-98
✕: P99-100
°: P101-102
▲ : P103-104
≤ (reversed empty triangle): P101-102
▼ : P103-104

***Pisum sativum***

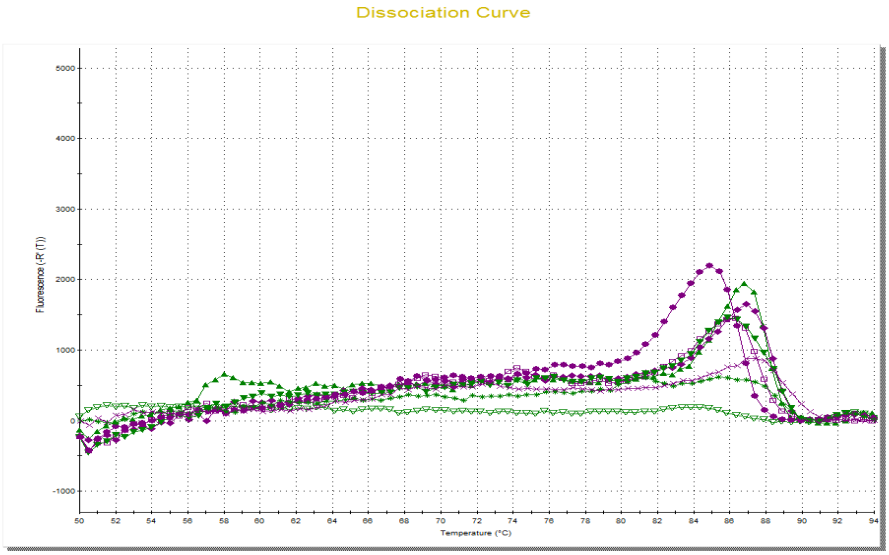



▼ : F109-110
✱ : F111-112
■ : F113-114
Δ : F115-116
❑: F117-118
♦ : F119-120

## Appendix C : Melting curves barcoding PCR

**Primers F17-F32 and P17-32**

***Vicia faba***




▲ : F17-18

❑: F19-20

≤ (reversed empty triangle): F21-22

♦ : F23-24

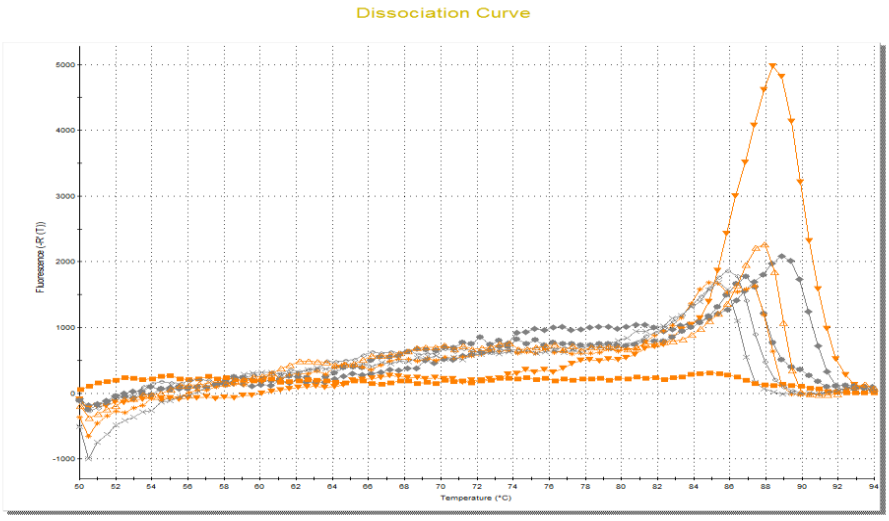
▼ : F25-26

• : F27-28

✱ : F29-30

✕: F31-32

***Pisum sativum***




♦ : P17-18

▼ : P19-20

• : P21-22

✱ P23-24

✕: P25-26

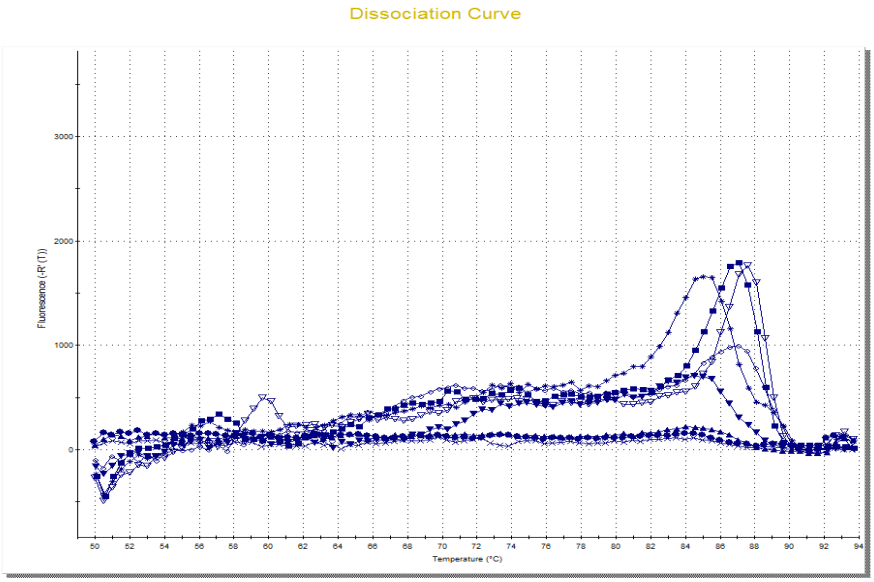
■ : P27-28

°: P29-30

Δ : P31-32

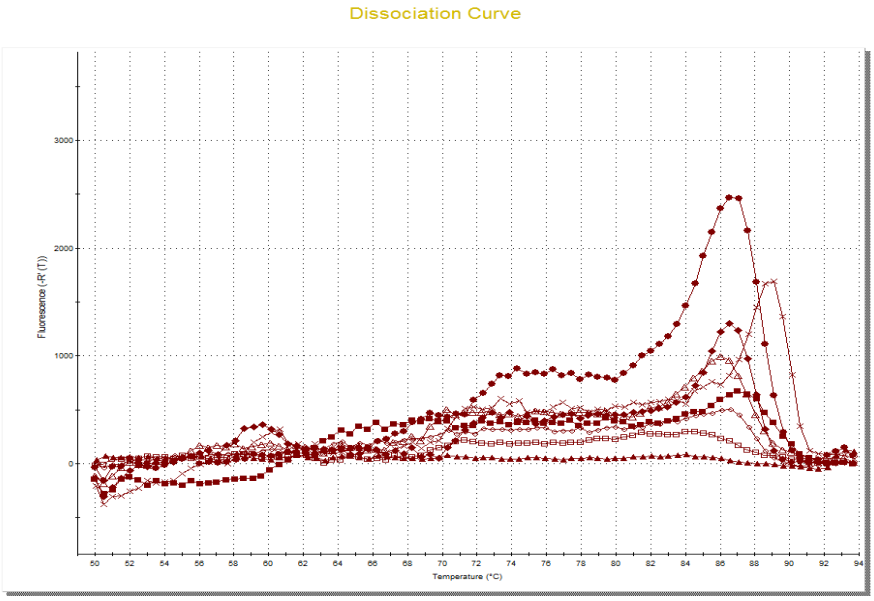
**Primers F33-F48 and P33-P48**

***Vicia faba***




• : F33-34
✕: F35-36
°: F37-38
▲ : F39-40
≤ (reversed empty triangle): F41-42
▼ : F43-44
✱ : F45-46
■ : F47-48

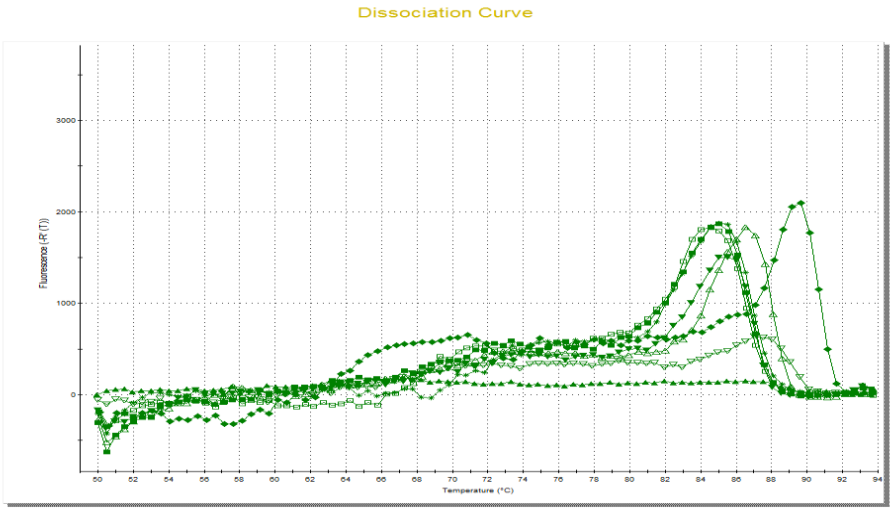
***Pisum sativum***




■ : P33-34
Δ : P35-36
❑: P37-38
♦ : P39-40
• : P41-42
✕: P43-44
°: P45-46
▲ : P47-48

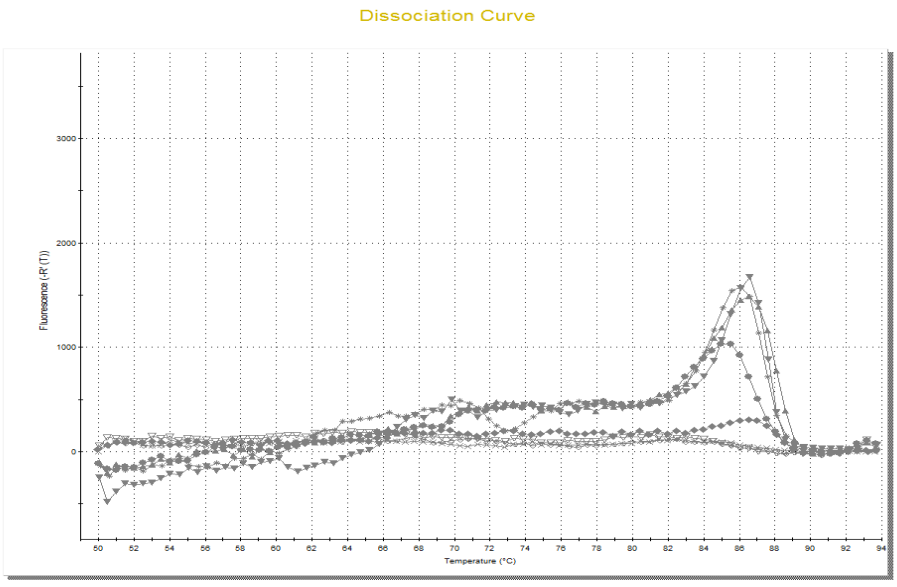
**Primers F49-F64 and P49-P64**

***Vicia faba***




▲ : F49-50
≤ (reversed empty triangle): F51-52
▼ : F53-54
✱ : F55-56
■ : F57-58
Δ : F59-60
❑: F61-62
♦ : F63-64

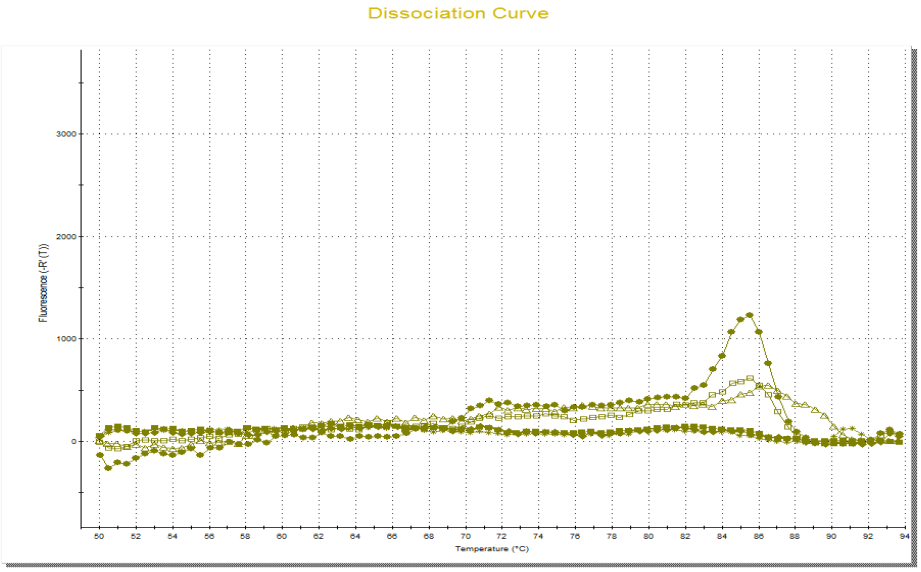
Pisum sativum




♦ : P49-50
• : P51-52
✕: P53-54
°: P55-56
▲ : P57-58
≤ (reversed empty triangle): P59-60
▼ : P61-62
✱ : P63-64

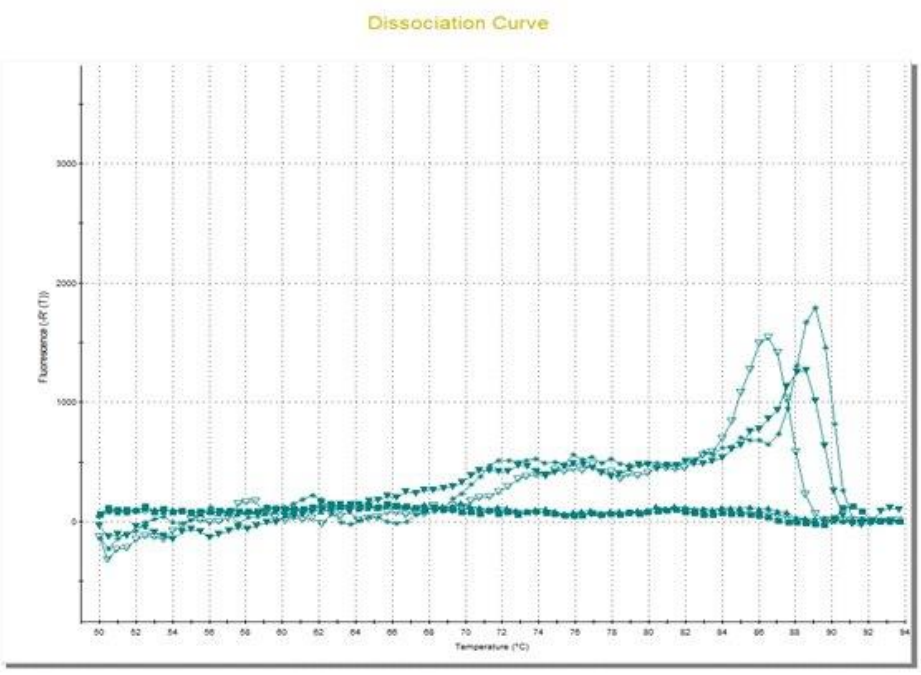
**Primers F65-76 and P65-P76**

***Vicia faba***




✱ : F65-66
■ : F67-68
Δ : F69-70
❑ : F71-72
♦ : F73-74
°: F75-76

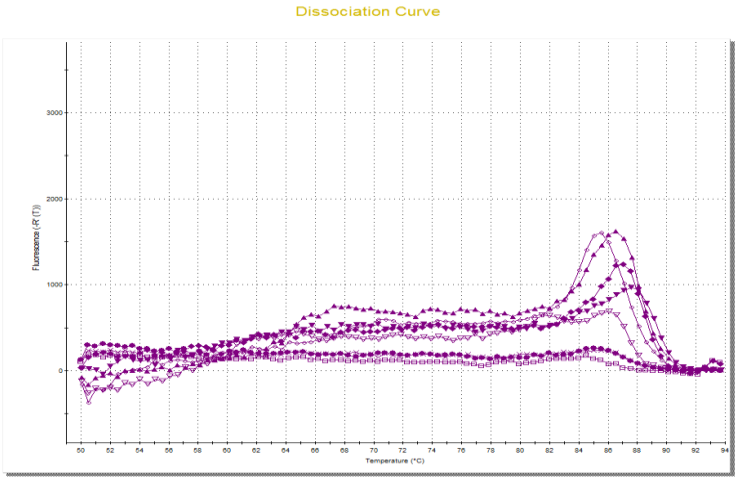
***Pisum sativum***




°:P65-66
▲ : P67-68
≤ (reversed empty triangle): P69-70
▼ : P71-72
✱ : P73-74
■ : P75-76

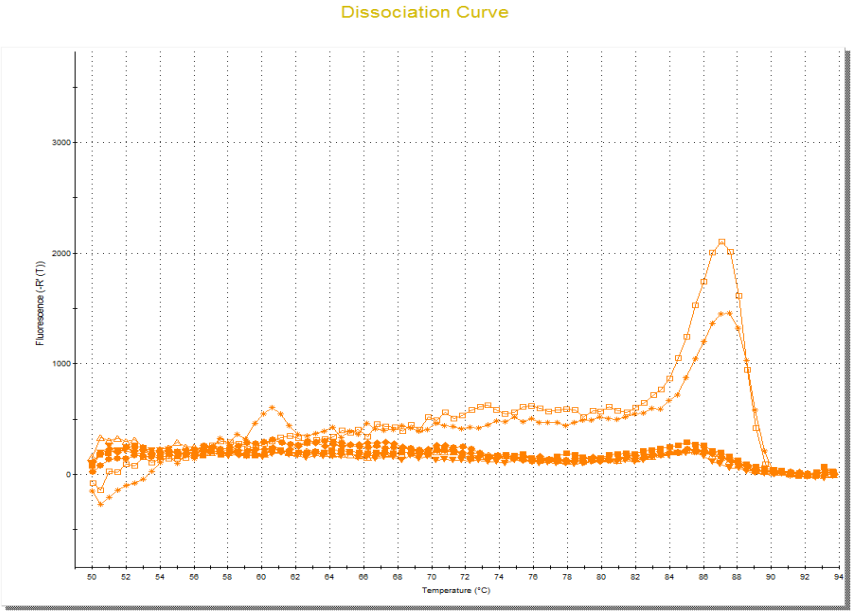
**Primers F77-F92 and P77-P90**

***Vicia faba***




❑: F77-78
♦ : F79-80
°: F81-82
✕: F83-84
°: F85-86
▲ : F87-88
≤ (reversed empty triangle): F89-90
▼ : F91-92

***Pisum sativum***




▼ : P77-78
✱ : P79-80
■ : P81-82
Δ : P83-84
❑: P85-86
♦ : P87-88
°: P89-90

## Appendix D: Primers that worked in genomic PCR and barcoding PCR

| *Vicia faba* | *Pisum sativum* |
|---|---|
| F03-04 | P03-04 |
| F05-06 | P05-06 |
| F07-08 | P07-08 |
| F09-10 | P09-10 |
| F11-12 | P11-12 |
| F13-14 | P13-14 |
| F15-16 | P15-16 |
| F17-18 | P17-18 |
| F19-20 | P19-20 |
| F23-24 | P21-22 |
| F25-26 | P23-24 |
| F27-28 | P25-26 |
| F31-32 | P29-30 |
| F37-38 | P31-32 |
| F41-42 | P33-34 |
| F43-44 | P35-36 |
| F45-46 | P39-40 |
| F47-48 | P41-42 |
| F51-52 | P43-44 |
| F53-54 | P45-46 |
| F55-56 | P55-56 |
| F57-58 | P57-58 |
| F59-60 | P61-62 |
| F61-62 | P63-64 |
| F63-64 | P69-70 |
| F69-70 | P71-72 |
| F71-72 | P73-74 |
| F75-76 | P79-80 |
| F79-80 | P85-86 |
| F85-86 | |
| F87-88 | |
| F89-90 | |
| F91-92 | |
| F93-94 | |
| F95-96 | |
| F97-98 | |
| F99-100 | |
| F101-102 | |
| F103-104 | |
| F105-106 | |
| F107-108 | |
| F109-110 | |
| F111-112 | |

| | |
|---|---|
| F115-116 | |
| F117-118 | |
| F119-120 | |

# Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
**Development of a high-throughput PCR & sequencing approach for SNP identification in Leguminosae species**

Richting: **master in de industriële wetenschappen: biochemie**
Jaar: **2015**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

**Maes, Ine**

Datum: **10/06/2015**