# Approach in the development and evaluation of DNA molecular markers in *Brassica* species.



## MSc. Minor Thesis Report

**Jorge Alemán Báez**

M.Sc. student Plant Sciences
Plant Breeding and Genetic Resources
Wageningen University & Research

August 2018

# Approach in the development and evaluation of DNA molecular markers in *Brassica* species

**Jorge Alemán Báez**

**M.Sc. Minor Thesis**

Plant Sciences Group
Laboratory of Plant Breeding
Growth & Development Group

WAGENINGEN UNIVERSITY & RESEARCH

**Supervisors:**
dr.ir. A. B. Bonnema
ing. J. Bucher

August 2018

## Acknowledgements

Working in the growth and development group has been a beautiful and enriching experience. During these 4 months, I had the opportunity in reaffirming my passion for plant sciences. This minor thesis was my last step in the pursuit of a Plant Science Master degree. I feel happy since from my experiences in the last year I decided to continue my studies and start a PhD in this amazing university (Wageningen University and Research). All that remains to do, is to thanks the people who took part in this last project.

First of all, I would like to thank my supervisor Guusje Bonnema for guiding me during these four months. It was a very enriching experience every time you spend time explaining, suggesting and advising my concerns. Thank you very much for making Brassica a very interesting species to me. I really enjoyed working with you.

A special thanks to Johan Bucher my daily supervisor. Thank you for improving my skills and learnings. Moreover, thank you for breaking the barrier between supervisor and student. I really appreciated.

Finally, thanks to Consejo Nacional de Ciencia y Tecnologia (CONACyT) for giving me the opportunity to do a MSC degree in this lovely university.

# Table of Contents


Acknowledgements ..... 3

Table of Contents ..... 4

Abbreviations ..... 6

Abstract ..... 7

Introduction ..... 8

Project 1 Development of new RCBr strains with defined marker genotypes ..... 10

Project 2 Validation of leaf morphology candidate genes in *B. oleracea* ..... 12

Objectives ..... 18

Project 1: Development of new RCBr strains with defined marker genotypes ..... 18

Project 2: Validation of candidate genes associated to genetic studies to leaf morphology in *B. oleracea* morphotypes. ..... 18

Material and Methods ..... 19

Development of new RCBr strains with defined marker genotypes ..... 19

Plant Material ..... 19

Growth conditions, phenotyping and sampling ..... 21

DNA Isolation ..... 21

Primers and validation of primers genome position ..... 21

Polymerase chain reaction (PCR) and gel electrophoresis. ..... 21

Validation of candidate genes associated to genetic studies to leaf morphology in *B. oleracea* morphotypes. ..... 22

B. oleracea genes of interest and Brassica data base ..... 22

Primer Design ..... 22

PCR and gel electrophoresis to amplify genes ..... 22

DNA extraction and purification from agarose gel ..... 24

NanoDrop spectrophotometer ..... 24

Sequencing ..... 24

Software's ..... 24

Enzymatic digestion ..... 25

Results ..... 26

Development of new RCBr strains with defined marker genotypes ..... 26

Validation of candidate genes associated to genetic studies to leaf morphology in *B. oleracea* morphotypes. ..... 32

BoGATA19 (GATA transcription factor 19) (Bol028986) ..... 32

BoUBP15 (Ubiquitin carboxyl-terminal hydrolase 15) (Bol038288) ..... 32

BoSPR1 (SPIRAL 1) (Bol032553) ..... 37
BoNGTPT (Non-green Triose phosphate/phosphate translocator) (Bol027034) ..... 38
Discussion ..... 43
Conclusion ..... 46
Recommendations ..... 47
References ..... 48
Appendix ..... 51

## Abbreviations

| | |
|---|---|
| A | Adenine |
| aa | Amino Acid |
| *anl* | *Anthocyaninless* |
| bp | Base pair |
| BRAD | *Brassica* Database |
| C | Cytosine |
| CAPS | Cleaved Amplified Polymorphic Sequence |
| CDS – | Coding Sequence |
| CK | Cytokinin |
| DAS | Days after sowed |
| DNA | Deoxyribonucleic acid |
| dNTPs | Deoxyribonucleotide triphosphate |
| G | Guanine |
| gDNA | Genomic DNA |
| GWAS | Genome Wide Association Study |
| His | Histidine |
| Indel | Insertion/Deletion |
| LOD | Logarithm of odds |
| mRNA | Messenger RNA |
| MT | Cortical Microtubules |
| ng | Nanograms |
| NGTPT | Non-green Triose phosphate/phosphate translocator |
| PCR | Polymerase Chain Reaction |
| QTL | Quantitative Trait Loci |
| RNA | Ribonucleic acid |
| RCBr | Rapid Cycle Brassica rapa |
| SAM | Shoot Apical Meristem |
| SI | Self-incompatibility |
| SNP | Single Nucleotide Polymorphism |
| SPR1 | Spiral 1 |
| T | Thiamine |
| TAE | Tris-acetate EDTA |
| TBE | Tris-boreta EDTA |
| UBP15 | Ubiquitin Carboxyl-terminal hydrolase 15 |
| µl | Microliter |
| µM | Micromole |
| VNTR | Variable Number Tandem Repeat |
| WGT | Whole Genome Triplication |
| Ygr | Yellow-green |

**Abstract**

*Brassica oleracea* and *Brassica rapa* are two *Brassica* species of large agricultural importance. Both *B. oleracea* and *B. rapa* crops are constantly demanded to adjust their genotype for specific agronomic traits. Despite the optimism for conventional breeding, new technologies, as DNA markers, will be needed to maximize the probability of success. Nowadays, DNA molecular markers increase breeding precision and efficiency by mapping agronomically important genes and assisting in the selection procedure. Therefore, our research will focus on assessing and understanding the growing importance of molecular markers in *Brassica* species. Two projects involving the development and use of molecular markers were conducted. The first searched for the development of Rapid Cycle *Brassica rapa* (RCBr) strains with defined marker genotypes for teaching purposes. Two variable number tandem repeat (VNTR)-type markers developed by Slankster *et al.*, 2012 were evaluated for linkage to the anthocyaninless locus. Both markers showed a high degree polymorphism useful for marker trait association. However, the development of fixed RCBr strains still remain to be completed. The second project intended the validation *of BoGATA19* (Bol028986), *BoUBP15* (Bol038288), *BoSPR1* (Bol032553), and *BoNGTPT* (Bol027034) genes linked to markers associated with leaf morphology in *B. oleracea*. The first step was to sequence these genes in selected *B. oleracea* accessions representing heading cabbage, cauliflower and kohlrabi crops. However, *BoGATA19* was not PCR amplified and *BoSPR1,* although amplified by PCR, sequencing was not successful. *BoUBP15* and *BoNGTPT* were PCR amplified and sequenced. Sequence alignment identified the presence of two paralogue *BoUBP15* genes product from a Single Nucleotide Polymorphism (SNP) and three paralogues *BoNGTPT* genes result from two polymorphisms (one SNP and one Insertion/Deletion (Indel)). Three cleaved amplified polymorphic sequence (CAPS) markers, one for each polymorphism, were developed and evaluated. Unfortunately, the developed markers could not be validated for leaf morphology. Therefore, linkage of *BoGATA19*, *BoUBP15*, *BoSPR1*, and *BoNGTPT* to leaf traits still remains to be validated.

**Keywords:** RCBr, marker trait association, *anthocyaninless*, *Brassica oleracea*, leaf morphology, DNA molecular markers.

## Introduction

*Brassica* species are grouped into the plant genus *Brassica*, which belongs to the tribe *Brassiceae*. Additionally, *Brassiceae* tribe appertain to the plant family *Brassicaceae*. In total, *Brassicaceae* family is composed by 3709 species and 338 genera (Warwick *et al*., 2006). Among the abundant *Brassica* species, six of them are of large agricultural importance including important vegetables (*B. oleracea* (broccoli, cabbage and cauliflower) and *B. rapa* (Chinese cabbage, pak choi and turnip)) as well as oilseed crops (*B. napus*, *B. rapa*, *B. juncea* and *B. carinata*) and condiments (*B. nigra*, *B. juncea* and *B. carinata*) (Wang *et al*., 2011). Three of these species are diploid (*B. rapa*, n=10; *B. nigra*, n=8; and *B. oleracea*, n=9), whereas, the other three are allotetraploids (*B. juncea*, n=18; *B. napus*, n=19; *B. carinata*, n=17) derived from each pair of the three diploid species. These six widely cultivated *Brassica* species are part of the so-called triangle of U. The 'U's triangle visualizes the heritage of these six *Brassica* species (**Figure 1**) (Nagaharu, 1935).

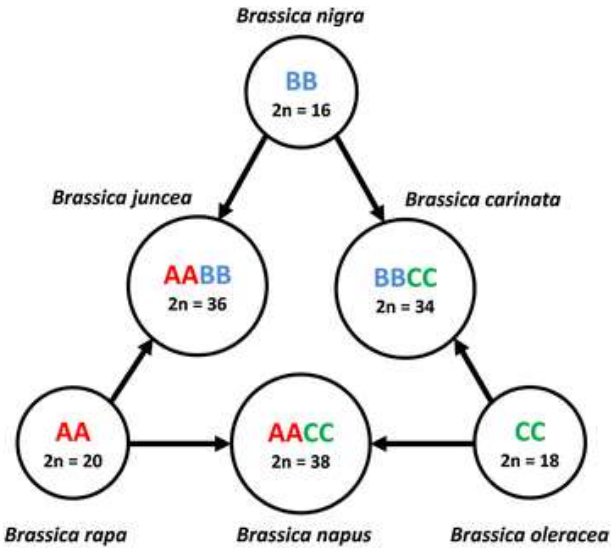



**Figure 1**. U's triangle. Heritage of six *Brassica* Species. The three diploid species A, B and C (*B. rapa*, *B. nigra* and *B. oleracea*) are placed on the three vertices of the triangle. Meanwhile, the three allotetraploids (*B. juncea*, *B. napus* and *B. carinata*), which are hybrids of AB, AC, and BC, respectively, are placed in the middle of the three edges of the triangle.

The concept of the U's triangle has promoted the genetic study of these species thus, has been successfully applied to aid in revealing and understanding the relationship among the *Brassica* crops (Cheng et al. 2014). Varieties, from each *Brassica* species, hold a rich diversity in speciation

and morphotypes. Cheng, Wu, and Wang (2014) award the speciation and expansion of morphotypes, the very diverse crops within each *Brassica* species, to an event so-called the Whole Genome Triplication (WGT). The WGT describes the triplication, genomic rearrangement and gene evolution of ancestral genomes that promoted the appearance of the varieties among the *Brassica* species. Based on comparative subgenome analysis, a two-step theory of polyploidization is suggested to explain the WGT event in *Brassica* species (Cheng *et al*., 2014). In the first step, two ancestral genomes merged together. Subsequently, a round of genomic reshuffling and gene fractionation produced a new diploid.

This first duplication is likely to have been an allotetraploidization however, an autotetraploidization cannot be excluded. In the second step, the emerged diploid combined with a third ancestral genome resulting in a hexaploid, which again stabilized into a diploid genome. After the WGT event, genes fractioned and further multicopy gene subfunctionalization or neofunctionalization promoted the parallel evolution of different morphotypes among the Brassica species (Cheng *et al*., 2014). The origin and variation of *Brassica* morphotypes have become of special interest for breeders in order to understand and employ the genetic puzzle surrounding these species.

Work for breeders is constantly respond to the necessity of developing genotypes with specific agronomic traits, fit to target environments, pathogen resistance, altered growth conditions and consumer preferences. Therefore, breeders face the endless task of continually developing new crop varieties (Evans, 1997). Despite the optimism for conventional breeding, new technologies will be needed to maximize the probability of success. Here it is where a biotechnological approach comes into operation. Biotechnology includes a broad area of science involving living systems and organisms to develop new technologies used in research, agriculture, industry and the clinic. Inside the biotechnology one area, DNA marker technology offers great opportunities for plant breeding. Based on the genetic linkage, DNA markers can be used for the detection of the allelic variation of genes underlying a trait of interest. Thus, the assistance of DNA markers can considerably increase breeding precision and efficiency (Collard and Mackill, 2008).

Nowadays, molecular markers have been widely used in mapping agronomically important genes and assisting breeding and selection procedures in *Brassica* genomes. The complete sequencing of the model plant *Arabidopsis thaliana* (The *Arabidopsis* Initiative 2000), to which *Brassica*

species represent the closest crop relatives, together with new marker technologies has opened the possibility for the investigation of the complex structure of *Brassica* genomes (Snowdon and Friedt 2003).

Focus of the project

In this thesis, we will focus our study on assessing and understanding the growing importance of molecular markers in *Brassica* species. Two different and connected research lines were followed however. The first aims at the development of new Rapid Cycle *Brassica rapa* (RCBr) strains with defined marker genotypes for teaching purposes. The second will attempt to validate of four candidate genes associated with leaf morphology in three *B. oleracea* morphotypes (cauliflower, heading cabbage and kohlrabi´s).

Project 1 Development of new RCBr strains with defined marker genotypes

Rapid cycling *Brassica rapa*, also known as Fast Plants, is a model plant widely used for teaching purposes. Fast plants were developed by selection of *B. rapa* for the traits of lack of seed dormancy, rapid seed maturation, petite growth habit, short time to flowering and high female fertility (Williams and Hill, 1986). The RCBr present an entire life cycle from sowing to the harvest of the produced seed (seed to seed) of 35-40 days. RCBr flowering stage can be visualized after only 12-14 days after sowed (DAS). Almost all brassica species, including RCBr, demonstrate self-incompatibility (SI). Self-incompatibility is a phenomenon were self versus non-self-pollen is recognized and it is exploited by the plant to prevent self-fertilization and thereby to increase genetic variation (Cheung *et al*., 1996). However, SI can be overcome by bud-pollination. Bud-pollination indicates that in buds where the stigma factors concerning with self-incompatibility response have not yet accumulated in sufficient quantities to inhibit self-pollen germination and tube growth (Shivanna, 1978). The addition of all these characteristics became the rapid-cycling lines an excellent organism for teaching purposes (Slankster *et al*., 2012).

One trait of special interest for teaching purposes in genetic using RCBr is the anthocyaninless (anl) trait, also known as the non-purple stem. Anthocyanins are flavonoid pigments responsible for purple/red coloration in stem and leaves where *anthocyaninless* (*anl*) mutants fail to produce anthocyanin pigments. Anthocyaninless trait is widely used as a model recessive trait since the phenotype is clearly visible at naked eye and at very early seedling stages. Thus, lessons in

Mendelian inheritance can be easily performed using RCBr model plant (Burdzinski and Wendell, 2007).

Slankster and team (2012) developed fourteen DNA-based genetic markers (**Table 1**) for RCBr, designed for use in the teaching laboratory environment. These markers were selected for robust amplification in PCR, polymorphism in RCBr strains, and alleles which can be easily resolved in simple agarose slab gels. Among these markers, three are expected to be found on chromosome A09 (D9*Brapa*S1, D9*Brapa*S4, and *Park9-HaeIII*), the chromosome which Burdzinski and Wendell (2007) reported holding the anthocyaninless locus. Markers D9BrapaS1 and D9BrapaS4 contain a longer repeated element like variable number tandem repeats (VNTR), whereas Park9-HaeIII belongs to a single nucleotide polymorphism (SNP) based marker (Slankster *et al*., 2012). The use of VNTR markers relies on the nucleotide length between alleles.

| Variable number tandem repeat-type markers for rapid cycling ***Brassica rapa***. | | | |
|---|---|---|---|
| Name | Genome position1 | Forward primer sequence | Reverse primer sequence |
| *D1BrapaS1* | A01:2129419‥2130033 | GGAGGAGCAAGCAGGACCAGGA | ACGCTGTGATTTGTTGCTTCCGA |
| *D1BrapaS2* | A01:1997477‥1998260 | GCGATGCGTATTGGTGGCCG | CCGTCGCCGGTTCACAAACCA |
| *D1BrapaS3* | A01:2092628‥2093030 | AAGCAAAGCCAGCGGCGGAT | GGCTGGTCACCCACAGGCAC |
| *D1BrapaS4* | A01:2125066‥2125999 | TGGGCGTTGTTCTCATGTTCGGT | ACCCGCCATTCTCCCCACCT |
| *D1BrapaS5* | A01:1975693‥1976021 | TCCAAAGTTCTGCTCGAGGGTCC | CCAGGAATGGCAGCATTAGACCGA |
| *D1BrapaS6* | A01:2129419‥2130034 | GGAGGAGCAAGCAGGACCAGGA | ACGCTGTGATTTGTTGCTTCCGA |
| *D1BrapaS7* | A01:4323134‥4323402 | TGGTCCCTGGATGCGCGGAA | GTGGCTGGTCACCGGTGTTGT |
| *D2BrapaS1* | A02:2005018‥2005276 | CCGAACCGTCTCTAACCGAATCGC | GGAGCTAGCATCGCTCGCGG |
| *D2BrapaS2* | A02:21600350‥21601504 | ACTTTGTGAGCTTTGGCTGTTGGT | AGCCCAGAAAGCGGTTCAGGA |
| *D3BrapaS1* | A03:25334204‥25334706 | ATGTGGCGCGCTGCCATTGA | CGCTAAGCATCCTTAACATTTTCGTGC |
| *D9BrapaS1* | A09:7345386‥7345818 | CCAGCCAAATCGTCACTCATGCGA | TGCATGCCTAAGAGTTTGGAGTAACAC |
| *D9BrapaS3* | A09:1547528‥1548201 | TCGTGGGACGGTCCTCTTGCT | ACCAAACTCTCCACCTCGGACA |
| *D9BrapaS4* | A09:28258896‥28259404 | AGCGATGTAGCACCCGAGTCCA | TCGAGCTGAGAGGGAAGCTGTGA |
| *D9BrapaS5* | A09:34072636‥34072917 | CCTTGGCTGCATCAGGCGCA | TCCAAAAGTGAGGCTGCCTTAGTGA |

**Table 1**. Primer sequences and genome location by Slankster and team (2012)

Slankster and team (2012) report the development of new homozygous RCBr strains with optimized genotype for the use of D9*Brapa*S1, D9*Brapa*S4 and *Park9-HaeIII* markers linked to the *anthocyaninless* locus (**Figure 2**) for the use in an instructional lab.

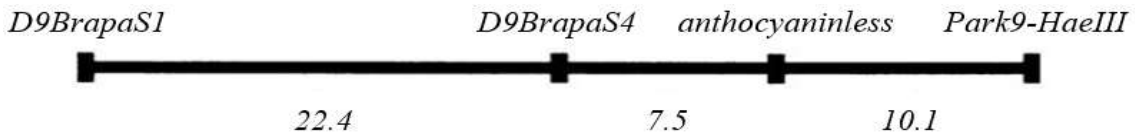



**Figure 2.** Genetic map, in fixed RCBr strains, of the *anthocyaninless* locus relative to markers D9*Brapa*S1, D9*Brapa*S4 and *Park9-HaeIII* (Slankster et al., 2012). Map distances are presented in Kosambi centimorgans.

This project will focus on the evaluation and validation of Slankster and colleagues (2012) RCBr markers D9*Brapa*S1 and D9*Brapa*S4, and on the development of RCBr strains with genotypes optimized for its use.

Project 2 Validation of leaf morphology candidate genes in *B. oleracea*

*B. oleracea* species contains the greatest variation among the *Brassica* crops (Cartea *et al*., 2011). Within each crop, a great variability of morphologies, colours, productive periods and environmental adaptabilities has been documented (Branca *et al*., 2013). Consequently, *B. oleracea* is the economically most important within the *Brassica* species (Cartea *et al*., 2011). Nowadays, some examples of widely cultivated *B. oleracea* crops include: cabbages (red, white, savoy, conical), broccoli and Romanesco, cauliflowers, kales, Brussels sprouts, collard greens, kohlrabi´s and ornamental subspecies (Bonnema *et al.,* 2011). The large variation inside the *B. oleracea* species is mainly addressed to its domestication through the pass of time by humans. Each crop type is distinguished by domestication traits that are not commonly found among the wild populations. For example, cabbages form a head that is made of leaves surrounding the terminal bud. Differently, Brussels sprouts are smaller heading structures, similar to cabbages, developed from the leaf axils. Kale was selected for its long marrow-stem meanwhile, kohlrabi for a short and thickened in a bulb-like shape stem. Finally, broccoli and cauliflowers are mainly characterized by an arrest in the switch from vegetative growth to flowering stage (Smith and King, 2000).

In the last decades, breeding of *B. oleracea* has mainly concerned crops as cabbages, cauliflower, broccoli and Brussels sprouts. *B. oleracea* breeding projects resulted in vigorous, high yielding and uniform cultivars. Biotic/a-biotic resistance has been an important breeding objective. Additionally, quality (colour, shape, taste and nutritional qualities) have been pursued to address the consumer's preferences. One trait that aligns both breeding objectives (resistance and quality)

is leaf morphology. Leaves are evolved organs, from branches (Bar and Ori, 2014), specialized in capturing the light energy. Via photosynthesis, leaves transform light energy into chemical energy. In most plants, leaves conform the major part above the ground surface. Therefore, leaves play a key role in the plant interaction with the above-ground environment. Cultivated *Brassicas* possess broadly expanded thin leaves containing a well-supplied chlorophyll content. High chlorophyll content in leaves increases the absorbance and utilization of solar radiation, water, and fertilizers. Additionally, breeding for leaf morphology increases the level of acceptance in *Brassica* crops were leaves traits are the anatomical parts consumed as food. Therefore, is of crucial importance to understand the process of leave development in *B. oleracea* crops.

Nowadays, genetic and molecular studies have contributed on unmask the process involved on leaf formation however, many aspects still remain unexplained (Tsukaya, 2013). Leaf morphology can be divided into a wide variety of traits from size, shape, and colour to cellular traits such as stomatal density (Chitwood *et al*., 2013). The focus of these project will be delimited to leaf and petiole size/shape.

The beginning of leaf development starts in the Shoot Apical Meristem (SAM). SAM is the source of all cells that ultimately will form the shoot, including the subset that ends up building leaves (Veit, 2004). Generally, cells in the central zone of the SAM remains in an undifferentiated and low division state. Is in the peripheral zone where leaf development starts with the formation of the leaf primordia.

Class 1 *KNOTTED-like homeobox* (*KNOX1*) transcription factors are required to maintain undifferentiated cells in the SAM by increasing the biosynthesis of the plant phytohormone cytokinin (CK) (Scofield and Murray, 2006). Moving away from the SAM central zone into the peripheral zone, auxin level increases. High concentration of auxin downregulates KNOX1 transcription factors. Thus, the function of KNOX1 becomes compromised as it moves away from the central zone. Additionally, auxin levels in the peripheral zone trigger ASYMMETRIC LEAF1/ROUGH (ARP) proteins. ARP proteins will initiate cell differentiation leading to leaf primordium development. Once the leaf primordia outgrowth apart the SAM, abaxial/adaxial polarity will be established.

The abaxial cell fate is promoted by members of the *YABBY* gene family. The expression of *YABBY* genes is positively regulated by the transcription factor KANADI (KAN) and ARF 3/4. Meanwhile, *KAN* and *ARF3/4* are activated by auxin concentration. Delimitation of the abaxial cell fate is negatively regulated by the expression *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*), and *REVOLUTA* (*REV*) genes in the adaxial domain. At the same time, *PHB*, *PHV* and *REV* genes encode for class III homeodomain-leucine zipper (HD-ZIPIII) proteins which promote the adaxial cell fate. Finally, HD-ZIPIII proteins are negatively regulated by miRNA 165/166, founded in the abaxial domain (Kalve *et al*., 2014). This way the abaxial-adaxial domains are internally regulated.

After the establishment of abaxial/adaxial polarity, leaf outgrowth starts. Leaf outgrowth is a combined result of cell division (proliferation) and cell expansion. Leaf cell division rate and duration are regulated by multiple redundant pathways. Transcription factors members of the *GROWTH-REGULATING FACTOR* (*GRF*) and *GRF-INTERACTING FACTOR* (*GIF*) families play an important role in promoting proliferation. Conversely, *CIN* type *TEOSINTE-BRANCHED1/CYCLOIDEA/PCF* (*CIN-TCP*) transcription factor family negatively regulates cell proliferation (Hepworth & Lenhard, 2014). Additionally to transcription factors, genes involved in protein degradation, like E3 ubiquitin ligases (Gonzalez, Vanhaeren, & Inzé, 2012) and genes involved in ribosomal processes and plant hormones (cytokinin, auxin, brassinosteroids and gibberellins) influence the rate and duration of cell proliferation (Fujikura, Horiguchi, Ponce, Micol, & Tsukaya, 2009). After cell proliferation, auxins, brassinosteroids, and gibberellins also play an important role in cell expansion. Cell expansion is principally the result of vacuole or turgor driven expansion. Endoreduplication, the process whereby DNA replicates repeatedly without alternating divisions through mitosis, is also involved in cell expansion (Kalve *et al*., 2014). The combined cell division and cell expansion not only affect the final size of the leave, but also the shape of the leaf.

The final stage in leaf development is the differentiation stage where cells differentiate into cells with a specialized function as trichomes, guard cells, and vascular tissue cells. The differentiation from a basic leaf cell to a specific cell type is regulated by pathways specific to each differentiation process (Kalve *et al*., 2014).

Most research in leaf development has been conducted in *A. thaliana* and *B. rapa*, little is known in *B. oleracea* crops. However, thanks to their close phylogenetic relationship, leaf development

studies could be conducted in *A. thaliana* as model plant and *B. rapa* species could serve as a base for *B. oleracea* varieties research. Xiao and team (2014) focus on the genetic dissection of leaf development in *B. rapa* by a genetical genomics approach. In this research, twenty-five colocalized phenotypic quantitative trait loci (QTLs) contributing to natural variation for leaf morphological traits, leaf number, plant architecture, and flowering time were identified. After several Genome Wide Association Studies (GWAS) performed in a collection of *B. oleracea* accessions, as part of a TKI project. A total of 1000 accessions, representing modern hybrids, gene bank accessions and wild species were collected and genotyped using GBS (genotyping by sequencing) resulting in about 18,000 SNPs. MSc students Fabian Topper, Floris Slob, Twan Groot, Josephine van Eggelen and Harm Brouwer, from the Wageningen University and Research Plant Breeding department, performed GWAS on variation in leaf traits (from diverse field experiments) and the initial candidate gene list proposed by Xiao and group (2014) was adapted. In a simplified way, GWAS analysis test for the association between genetic markers and phenotypes. Each marker hold a LOD (logarithm of odds) score. Markers with a LOD over the calculated threshold become candidate markers. Candidate markers possess the probability of holding, in a region around (for *Brassica* species 50 kb to either direction), a gene with an effect in the phenotype of interest. From this Xiao adapted list (**Appendix 1**) four candidate genes, product of a marker LOD score, were selected for this project with the intention of validating them for leaf morphology. The selected genes are *BoGATA19* with a *B. oleracea* code (Bocode) Bol028986, a LOD score of 17.40 and linked to leaf width trait; *BoUBP15* with a Bocode Bol038288, a LOD score of 5.28 and linked to stem length trait; *BoSPR1* with a Bocode Bol032553, a LOD score of 1.80 and linked to stem length trait; and finally *BoNGTPT* with a Bocode Bol027034, a LOD score of 7.07 and linked to leaf length trait.

Candidate Genes

*BoGATA19 (GATA transcription factor 19) (Bol028986)*

GATA is a family of transcription factors holding a single zinc finger domain. In Arabidopsis, *GATA* family include 30 members encoding for GATA3-type transcription factors (Zhao *et al*., 2004). In animals, GATA factors show participation in cell development, differentiation, and proliferation. In plants, GATA factors have been related to regulating developmental processes and responses to environmental stimuli as light signaling, photoperiod control of flowering, seed germination, lateral root founder cell specification, and stress responses. Zhao and team (2004)

suggested that members of the GATA family as HANABA TARANU (HAN) transcription factor are involved in controlling cell proliferation and differentiation since the ectopic expression of *HAN* causes growth retardation, aberrant cell division, and loss of meristem activity. *GNC* (*GATA* factor, nitrate-inducible, carbon metabolism-involved), another member of the *Arabidopsis GATA* family, is suggested to play an essential role in chlorophyll and glucose synthesis (Bi *et al*., 2005). Despite the large available information about *GATA* as a family of transcription factor, little is known about the in vivo functions of most of these factors.

*BoUBP15 (Ubiquitin carboxyl-terminal hydrolase 15) (Bol038288)*

*Ubiquitin-specific proteases* (*UBP*s) are a highly conserved family in eukaryotes involved in protein de-ubiquitination. Ubiquitin covalent modification of proteins plays a central role in diverse cellular pathways such as cell-cycle progression, signal transduction, DNA repair, endocytosis, and apoptosis. In *Arabidopsis*, the *UBP* family is integrated by at least 27 *UBP* genes (Amerik and Hochstrasser, 2004). All *UBP* genes encode for proteins containing a UBP domain with conserved motifs as cysteine (Cys) and histidine (His) boxes (Amerik and Hochstrasser, 2004). In *Arabidopsis*, *UBP15* mutants exposed narrower, serrated and flat rosette leaves, partially due to a defect in cell proliferation. Conversely, the overexpression of *UBP15* showed opposite phenotypes. (Liu *et al*., 2008). Lui and collaborators (2008) research suggest the participation of *UBP15* in leaf development by controlling cell proliferation. Although some members of the *UBP* family such as *UBP15* have been the focus of study, much remains to scrutinize on the function of the rest of the members.

*BoSPR1 (SPIRAL 1) (Bol032553)*

SPIRAL 1 protein is localized to cortical microtubules (MT) according to the results from genetic complementation of the SPR1 protein with green fluorescent protein (Nakajima *et al*., 2004). Nakajima and team (2004) suggested that *SPR1* is essential for anisotropic growth in rapidly elongating cells after the observations, in *Arabidopsis*, of an increased hypocotyl elongation in plants overexpressing *SPR1* and a right–handed helical growth in roots and etiolated hypocotyls in *spr1* mutants. Few studies have been addressed to understand completely the function of *SPIRAL 1* however, the existing ones give a clear impression of its participation for directional control in cell elongation (Furutani *et al*., 2000).

*BoNGTPT (Non-green Triose phosphate/phosphate translocator) (Bol027034)*

NGTPT proteins mediate the export of fixed carbons from the chloroplast into the cytosol in form of triose phosphates. In *Arabidopsis*, *NGTPT* genes are expressed in photosynthetically active tissues. Studies in the *cTPT* (chloroplast triose phosphate/phosphate translocator) member of the *NGTPT* family, reported that *cTPT* is almost exclusively expressed in photosynthetic tissues (leaves) and also in cauliflower inflorescences and the expression of the *cTPT* genes is linked to the photosynthetic carbon metabolism (Fischer et al., 1997). However, studies of *NGTPT* genes in the participation in leaf morphology has not be conducted. Nevertheless, the expression of *NGTPT* genes in photosynthetic tissues opens the possibility of interacting in the leaf development process.

**Objectives**

Project 1: Development of new RCBr strains with defined marker genotypes

1. Confirm the genetic position of Slankster *et al*. 2012 markers by a Blast search in the *B. rapa* genome sequence database (BRAD).
2. Evaluate the DNA polymorphism degree and allele frequencies responsible for markers D9*Brapa*S1 and D9*Brapa*S4 RCBr obtained from Wisconsin Fast Plants and Carolina Biological Supply.
3. Develop two RCBr parental lines with genotypes, to easy score in an F2 population the Mendelian loci *anthocyaninless*, optimized for the use of D9*Brapa*S1 and D9*Brapa*S4 markers.

Project 2: Validation of candidate genes associated to genetic studies to leaf morphology in *B. oleracea* morphotypes.

1. Generate primers specific for each candidate gene and amplify the genes from heading cabbage, cauliflower and kohlrabi accessions.
2. Sequence the amplified products.
3. Align sequences for polymorphism detection.
4. Develop molecular markers and genotype this over a subset of the collection.
5. Calculate association between the developed marker and the leaf traits.

## Material and Methods

Development of new RCBr strains with defined marker genotypes

*Plant Material*

Seven *B. rapa* RCBr strains were acquired from two different distributors. Four strains from Wisconsin Fast Plants and three different from Carolina Biological Supply.

**Wisconsin Fast Plants**

RC1 - Purple Stem (*ANL/ANL*) (Purple Hypocotyl and Green Cotyledon)

Purple Stem, Hairy variety of Wisconsin Fast Plants were selectively bred for strong expression of a purple pigment, anthocyanin, visible throughout the stems, leaves, and bud tips. The ability to express anthocyanin (purple phenotype) is dominant, and multiple genes control the purple intensity.

RC2 - Standard

The Standard variety of the Wisconsin Fast Plants was selectively bred to have a uniform, short flowering time (12-14 days) under recommended growing conditions of continuous light, water, and nutrition. Shows a phenotype of green hypocotyl and green cotyledon.

RC3 - Non-Purple Stem (*anl/anl*) (Green Hypocotyl and Green Cotyledon)

Non-Purple Stem variety of Wisconsin Fast Plants do not produce anthocyanin (recessive trait), a purple pigment found in other Fast Plant varieties. The lack of the purple pigment based on a single mutation causes the plants to appear purely green.

RC4 - Non-Purple Stem, Yellow-Green Leaf (*anl/anl; ygr/ygr*) (Green Hypocotyl and Yellow-Green Cotyledon)

The Non-Purple Stem, Yellow-Green Leaf variety of Wisconsin Fast Plants® is homozygous for two recessive traits: non-purple stem (*anthocyaninless, anl*) and yellow-green leaf *(ygr*). Plants of this variety do not express anthocyanin, and shows pale yellow green leafs when compared to other Fast Plants varieties. The Non-Purple Stem, Yellow-Green traits are the result of an allelic variation of single genes.

F2 Population (Purple/Non-purple stem; Green/Yellow-Green Leaf)

F2 of Wisconsin Fast Plants is the offspring generation of intermating the Purple Stem/Green Leaf with the Non-purple Stem/Yellow-Green Leaf. This stock demonstrates a two-gene inheritance model representative of Mendel's findings. F2 population will segregate on average in the following ratio:

- 9: purple stem (*ANL/ANL* or *ANL/anl*); dark green leaf (*YGR/YGR* or *YGR/ygr*)
- 3: non-purple stem (*anl/anl*); dark green leaf (*YGR/YGR* or *YGR/ygr*)
- 3: purple stem (*ANL/ANL* or *ANL/anl*); yellow-Green leaf (*ygr/ygr*)
- 1: non-purple stem (*anl/anl*); yellow-green leaf (*ygr/ygr*)

**Carolina Biological Supply**

Carolina strains are highly similar to Wisconsin strains. Both strains are selected for the same phenotypic traits as hypocotyl and cotyledon colour. Carolina strains were derived from the selection for high purple stem colour (RCBr1-119) and for from a non-purple stem population (RCBr1-104). However, the different between Wisconsin and Carolina strains relies on genotypic differences. The following Carolina strains were evaluated:

RCBr1-104: Green hypocotyl (*anl/anl*); yellow-green cotyledon (*ygr/ygr*) and hairless.
RCBr1-119: Purple hypocotyl (*ANL/ANL* or *ANL/anl*); yellow-green cotyledon (*ygr/ygr*), hairy.
RCBr1-033 (Standard): Green hypocotyl (*anl/anl*); green cotyledon *YGR/YGR* or *YGR/ygr*)

**Table 2** summarizes the Rapid cycle *Brassica* plants strains from Wisconsin Fast Plants and Carolina Biological Supply showing phenotype and genotype for hypocotyl and cotyledon colour.

| | Strain | Phenotype | | Genotype | |
|---|---|---|---|---|---|
| | | Hypocotyl | Cotyledon | Hypocotyl | Cotyledon |
| Wisconsin Fast Plants | Standard | Green | Green | - | - |
| | Purple Stem | Purple | Green | *ANL/ANL* | *YGR/YGR* |
| | Non-Purple Stem | Green | Green | *anl/anl* | *YGR/YGR* |
| | Non-Purple Stem, Yellow-Green Leaf | Green | Yellow-Green | *anl/anl* | *ygr/ygr* |
| | F2 Population | - | - | - | - |
| Carolina Biological Supply | PRCBr1-104 | Green | Yellow-Green | *anl/anl* | *ygr/ygr* |
| | RCBr1-119 | Purple | Yellow-Green | *ANL/ANL* | *ygr/ygr* |
| | RCBr1-033 (Standard) | Green | Green | - | - |

*Growth conditions, phenotyping and sampling*

Rapid cycle *B. rapa* seeds were sown on trays with potting mix and covered with vermiculite. Trays were placed under a high intensity light proportioned by fluorescent lights and at greenhouse temperature (20-25 °C). After 4-10 days, cotyledon and hypocotyl colour was scored. Additionally at the same time point of phenotyping, a small part of the leaf, from each plant, was excised for DNA extraction.

*DNA Isolation*

DNA extraction was performed on young leaves using the Qiagen DNeasy® plant mini kit following the manufacturer's protocol available in www.qiagen.com/handbooks. DNeasy® extraction kit relies on the cell membrane lysis and the digestion of the contaminating RNA in the sample. After several centrifugation steps at 13000rpm, extracted DNA possessed a high purity and quality. Extracted DNA samples of 50 µl were stored in the freezer at -20°C.

*Primers and validation of primers genome position*

Primers sequences from markers D9*Brapa*S1 and D9*Brapa*S4 were obtained from Slankster *et al*., 2012 article. Table 1 shows the primer sequence of each marker. To validate the genomic position and uniqueness of the primers their genome position was determined by a BLAST search on the *B. rapa* genome sequence version 1.1 using the *Brassica* Database (BRAD).

*Polymerase chain reaction (PCR) and gel electrophoresis.*

PCR was performed in a 10 µl reaction volume. Each sample mixture contained: 0.1 µl of Dream Taq polymerase, 1.0 µl of 10X deTaq buffer, 1.0 µl (5 µM) of each primer, 0.4 µl of dNTPs (5 pmol/l), 1 µl DNA extraction or DNA sample, and 5.5 µl Milli-Q. Configuration of PCR steps were: Firstly denaturation at 94°C for 5 min, then 28-32 cycles of 94°C for 30 seconds for the denaturing stage, 55° for 30 seconds for the annealing stage and 72°C for 1 minute for the extending stage, followed by 72°C for 10 min and finally 10°C for unspecified time. Finally, PCR products were separated by electrophoresis in a 1.5% TAE agarose gel at 80 V for 4 hrs. Bands were detected by ethidium bromide staining.

Validation of candidate genes associated to genetic studies to leaf morphology in *B. oleracea* morphotypes.

*B. oleracea genes of interest and Brassica data base*

The four *B. oleracea* genes of interest were selected from Xiao *et al*., 2014 adapted list (**Appendix 1**) of candidate list for leaf morphology. Selection criteria of genes was: genes result from markers with a high LOD score and with a short distance between marker and gene. Genomic (gDNA) and messenger RNA (mRNA) sequences of selected genes were obtained from BRAD data base source *B. oleracea* (chromosome v1.0) 2013-06-08.

*Primer Design*

Sequences from the different genes of interest were obtained from the *Brassica* (BRAD) web-based database and tested for the existence of orthologous genes. No orthologous genes were identified for any gene of interest. Candidate primers were obtained by the online platform Primer3 (version 4.0) and Primer Quest Tool from Integrated DNA Technologies. Software Oligo 6 was used for testing the interaction between primers for dimerization analysis, hairpin formation and melting temperature (Tm). Primers (**Table 3**) were selected with a Tm above 60°C (65-68°C); at least 20 base pair (bp) length and g/c content over 50%. All primer were blasted in BRAD database for genome position confirmation. Primers were synthesized by Biolegio BV Laboratories.

*PCR and gel electrophoresis to amplify genes*

PCR was performed in a 50 µl reaction volume. Each sample mixture contained: 0.5 µl of Dream Taq polymerase, 5.0 µl of 10X deTaq buffer, 5.0 µl (5 µM) of each primer, 2.0 µl of dNTPs (5 pmol/l), 5 µl DNA extraction or DNA sample, and 27.5 µl Milli-Q. Configuration of PCR steps were: Firstly denaturation at 94°C for 5 min, then 28-32 cycles of 94°C for 30 seconds for the denaturing stage, 53-60°C for 30 seconds for the annealing stage and 72°C for 1 minute for each kilobase of the expected product for the extending stage, followed by 72°C for 10 min and finally 10°C for unspecified time. PCR products were separated by electrophoresis in a 1.5% TAE agarose gel at 80 V for 4 hrs. Bands were detected by ethidium bromide staining.

**Table 3.** List of relevant primers used during the project for amplification purposes.

| Gene | Lenght | Primer Name | Sense | Primer | Start | Stop |
|---|---|---|---|---|---|---|
| GATA19 | 967 | PrJA001F | F | CATCATCACTCCTCGCCTTATG | 31 | 53 |
| GATA19 | 967 | PrJA002R | R | AGAAGGGATCGGTCGTGATA | 908 | 888 |
| GATA19 | 967 | PrJA003F | F | CTTCTGTTGACTGCACTCTCTC | 53 | 75 |
| GATA19 | 967 | PrJA004R | R | CGTGATATCATGGCTCTCTCTG | 894 | 873 |
| UBP12 | 1449 | PrJA011F | F | CTACCTCCCATCATGACACTTC | 713 | 735 |
| UBP12 | 1449 | PrJA012R | R | GGAGGATTTCTGAGTCCATGAG | 75 | 53 |
| UBP12 | 1449 | PrJA013F | F | TGTGGACGCAGAGAACAAA | 186 | 205 |
| UBP12 | 1449 | PrJA014F | R | GGTGGTTCCTATCACGACAAA | 509 | 488 |
| UBP12 | 1449 | PrUBP1501F | F | CGAACCCAAGACATATCCTCTAAC | 323 | 347 |
| UBP12 | 1449 | PrUBP1502R | R | TGGATCTACTACCTCCCATCAT | 743 | 719 |
| UBP12 | 1449 | PrUBP1503F | F | TCGAAGTTGTAATGATTCTCTGGA | 350 | 374 |
| UBP12 | 1449 | PrUBP1504R | R | CTACCTCCCATCATGACACTTC | 735 | 713 |
| UBP12 | 1449 | Pr UBP1506R | R | ACACTTCGGCCTCCTCT | 719 | 702 |
| SPR1 | 412 | PrJA021F | F | GCTGTGGTGGAGGTCAAA | 16 | 34 |
| SPR1 | 412 | PrJA022R | R | CAGGAGCTGCATGGACTTT | 369 | 350 |
| SPR1 | 412 | PrJA023F | F | GGAGGTCAAAGCTCACTCAA | 25 | 45 |
| SPR1 | 412 | PrJA024R | R | CATGGACTTTGGTCGAAGGA | 360 | 340 |
| NGTPT | 2100 | PrJA031F | F | TCTCCCTCGCTTCCTCTC | 27 | 45 |
| NGTPT | 2100 | PrJA032R | R | GGCCATCAGAAAGAGAGACATC | 1505 | 1483 |
| NGTPT | 2100 | PrJA033F | F | TCCAATCGTTGGTGGAGTTG | 1145 | 1165 |
| NGTPT | 2100 | PrJA034R | R | GGCTTGATACGCTTCACTCT | 2081 | 2061 |
| NGTPT | 2100 | PrJA035F | F | CCTCGGATCTCAACGTCTCT | 89 | 109 |
| NGTPT | 2100 | PrJA036R | R | CTTCGGAGAAGAAAGTCACAGG | 1527 | 1505 |
| NGTPT | 2100 | PrJA16F | F | GACAGAGGTCTCATTCAACTGG | 1178 | 1199 |
| NGTPT | 2100 | PrJA17R | R | CCACAAGACGTTTGAGCAATG | 1389 | 1369 |
| NGTPT | 2100 | PrJA18F | F | CTTCAGTGACAGAGGTCTCATT | 1171 | 1192 |
| NGTPT | 2100 | PrJA19R | R | TGAGCAATGTACAGGGAACTTA | 1377 | 1355 |
| NGTPT | 2100 | PrNGTPT01F | F | TCCAATCGTTGGTGGAGTTG | 1145 | 1164 |
| NGTPT | 2100 | PrNGTPT02R | R | CACACGCTTCACACAGTTTC | 1871 | 1852 |
| NGTPT | 2100 | PrNGTPT03F | F | CTTGCTTCAGTGACAGAGGT | 1166 | 1686 |
| NGTPT | 2100 | PrNGTPT04R | R | AAGCATTAACAGGCGAGACC | 1929 | 1909 |
| NGTPT | 2100 | PrNGTPT06R | R | TCTTAGGCTTTGGCTTGATAC | 2091 | 2071 |
| NGTPT | 2100 | PrJA041F | F | CGGTAGTGTCCTCATTACCTTC | 713 | 735 |
| NGTPT | 2100 | PrJA042R | R | CCCGAGGCTCATGTTAGTAAAG | 948 | 926 |
| NGTPT | 2100 | PrJA043F | F | CTGTGACTTTCTTCTCCGAAGG | 1528 | 1506 |
| NGTPT | 2100 | PrJA044R | R | GAAGCAGAGTGCAGCAATAAG | 1689 | 1668 |

*DNA extraction and purification from agarose gel*

The product from a PCR was loaded on a TAE 1.0% agarose gel. Using a sterile blade, the fragment of gel containing the amplified product with the expected length was carefully excised. The solid gel was then dissolved by heat and product purification was realized using MinElute® Gel Extraction Kit from Qiagen TM following the manufacturer's protocol available in www.qiagen.com/handbooks.

*NanoDrop spectrophotometer*

NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) was used for measuring concentration and purity of nucleic acid samples. For this step, 1.5 µl of the genetic sample was pipetted onto the end of a fiber optic cable (receiving fiber) and then closed with a second fiber optic cable (the source fiber). A PC based software provides the report of the concentration and purity of the different samples.

*Sequencing*

For this step, 400-500 ng of pDNA is used, plus 2.5 µl of one primer (25 pmol) and fully filled with Milli-Q water for a final volume of 10 µl. Samples were sequenced by GATC BIOTECH, an external sequencing service.

*Software's*

**MEGA 6**

MEGA 6 is an integrated tool for conducting automatic and manual sequence alignment. With the help of MEGA 6 obtained gene sequences from the different morphotypes were aligned by ClustalW alignment. Sequence alignment allowed the identification of gene polymorphisms. Additionally, nucleotide sequences were translated into amino acids sequences.

**ApE**

ApE is a proprietary sequence editor using a modified GenBank format to save sequences. ApE allows the recognition of restriction enzymes and virtual digestion of the proposed sequence. ApE

software was specifically used for the development of CAPS markers. ApE helped with the recognition of restriction enzymes with restriction sites located in each proposed polymorphism.

*Enzymatic digestion*

Three different CAPS markers were designed and SACI, HpyCH4V and FspBI restriction enzymes were employed. The enzymatic digestion with each restriction enzyme was performed following the manufactures recommended protocol for digestion of PCR products directly after amplification. All restriction enzymes used followed the same recommended protocol. Each sample unit contained: 10 µl of PCR reaction mixture, 2 µl of 10X Buffer R and 0.1-0.2 µl of restriction enzyme. Reaction mix was incubated at 37 °C for 16 hrs. For enzyme inactivation, samples were incubated at 65°C for 20 min. After inactivation, digested PCR product was loaded into TAE 1.0% agarose gels and separated by electrophoresis at 80 V for 4 hrs. Bands were detected by ethidium bromide staining.

## Results

Development of new RCBr strains with defined marker genotypes

### Genome position validation of Slankster primers

Our first assay aimed at the validation of the genome positions of the primers, developed by Slankster *et al*., 2012 (henceforth referred only as Slankster). Genome position was determined by a BLAST search of the *B. rapa* genome sequence version 1.1 using the *Brassica* Database (BRAD) as Slankster specifies in his article. Results, from blasting the primer sequences and compare them with the reported by Slankster indicate a several mismatches in the genome position. However, discrepancies did not compromise further assays.

### Phenotype and validation of markers D9*Brapa*S1 and D9*Brapa*S4 in Wisconsin Fast Plants parental lines and F2 population

Ten plants from each Wisconsin Fast Plants variety Purple Stem and Green Leaf (RC1), Standard (RC2), Non-Purple Stem and Green Leaf (RC3) and Non-purple Stem/Yellow-Green Leaf (RC4) (henceforth referred as parental lines) and ninety F2 (Purple/Non-purple stem; Green/Yellow-Green Leaf) were sown. Ten DAS, all plants were scored for purple/non-purple stems and green/yellow-green traits. Additionally, DNA was isolated from all 130 plants.

All parental lines phenotype, for the stem (hypocotyl) and leaf (cotyledon) traits, matched with the proposed by Wisconsin Fast Plants. Meanwhile, the F2 phenotyping results for the monohybrid segregation of hypocotyl and cotyledon colour, 52/90 plants score a purple and 38/90 present a green colour for the hypocotyl meanwhile, for the cotyledon 66/88 score a green and 22/88 shows a yellow-green colour (**Table 4**). Moreover, results for the dihybrid segregation of hypocotyl and cotyledon colour in the F2 population are: 39/88 plants score a purple stem and green leaves, 29/88 plants show a green stem and green leaves, 13/88 plants score a purple stem and yellow-green leaves and only 7/88 show a green stem and yellow-green leaves (**Table 4**). Two plants present a green stem however, leaf colour was unable to be determined.

Next, the observed values from phenotyping the F2 population were evaluated for their fit for Mendelian inheritance of a monohybrid segregation ratio of 3:1 and in a dihybrid segregation ratio of 9:3:3:1 (**Table 4**).

F2 Population

| | Observed (o) | Expected (e) | $((o-e)^2/e)$ |
|---|---|---|---|
| Hypocotyl (Stem) | | | |
| Purple | 52 | 67.5 | 3.56 |
| Green | 38 | 22.5 | 10.67 |
| Total | 90 | 90 | 14.23 |
| Cotyledon (Leaf) | | | |
| Green | 66 | 68 | 0.06 |
| Yellow | 22 | 20 | 0.18 |
| Total | 88 | 88 | 0.24 |

| Phenotype (Stem-Leaf) | Observed (o) | Expected (e) | $((o-e)^2/e)$ |
|---|---|---|---|
| Purple-Green | 39 | 49.5 | 2.22 |
| Purple-Yellow | 13 | 16.5 | 0.74 |
| Green-Green | 29 | 16.5 | 9.46 |
| Green-Yellow | 7 | 5.5 | 0.4 |
| Total | 88 | 88 | 12.82 |

**Table 4.** On the left, F2 population phenotypic score and calculation of chi-square values for monohybrid segregation of hypocotyl and cotyledon colour separately; on the right F2 population phenotypic score and calculation of chi-square for independent dihybrid segregation of hypocotyl and cotyledon colour.

Using a chi-square test a chi-square value ($x^2$) of 14.23, 0.24 and 12.82 was calculated for the hypocotyl, cotyledon and (hypocotyl $^{x}$ cotyledon), respectively. The calculated $x^2$ was compared with $x^2 = 3.84$ value from a 95% confident interval and one degree of freedom. Only the observed values from cotyledon colour fit in a monohybrid segregation ratio of 3:1 whereas hypocotyl colour not. The dihybrid segregation of hypocotyl and cotyledon colour is significant different to the expected segregation of 9:3:3:1.

Next, the segregation of alleles from markers D9BrapaS4 and D9BrapaS4 for parental lines and F2 population were evaluated. Both markers were selected since, according to Slankster, are the closest to the *anthocyaninless* locus (**Figure 2**). Segregation of alleles from selected marker for parental lines, clearly indicates that none of the parental lines are homozygous (**Table 5**). Therefore, a chi-square test in parental lines was not performed. Meanwhile, for the F2 population the allele segregation from D9SrapaS1 and D9BrapaS4 was evaluated by a chi-square for its fit in a 1:2:1 segregation. For D9BrapaS1 a $x^2$ of 62.47 and for D9BrapaS4 a $x^2$ of 149.6 was calculated. Both chi-squares values clearly indicates that segregation of alleles from D9BrapaS1 and D9Brapas4 for the F2 population behave in alternative way to 1:2:1 segregation (**Table 5**).

| | Parental Population | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Genotype | | | | Marker | | | | | |
| | ANL | | YGR | | D9*Brapa* S1 | | | D9*Brapa* S4 | | |
| Strain | Anl/. | anl/anl | YGR/. | ygr/ygr | aa | ab | bb | aa | ab | bb |
| RC1 | 9 | | 9 | | 7 | 2 | - | 8 | - | - |
| RC2 | | 9 | 9 | | 6 | 3 | - | - | - | - |
| RC3 | | 9 | 9 | | 1 | 4 | 4 | 8 | - | 1 |
| RC4 | | 10 | | 10 | 7 | 2 | 1 | 9 | 1 | - |

| F2 Population | | | | | | |
|---|---|---|---|---|---|---|
| | D9*Brapa* S1 | | | D9*Brapa* S4 | | |
| Alleles | Observed | Expected | $((o-e)^2/e)$ | Observed | Expected | $((o-e)^2/e)$ |
| aa | 42 | 22.5 | 16.9 | 71 | 21.75 | 111.52 |
| ab | 42 | 45 | 0.2 | 13 | 43.5 | 21.38 |
| bb | 6 | 22.5 | 45.37 | 3 | 21.75 | 16.16 |
| Total | 90 | 90 | 62.47 | 87 | 90 | 149.6 |

**Table 5**. On top, allele score from markers D9BrapaS1 and D9BrapaS4 for Wisconsin Fast Plants parental lines. On the bottom, allele score and calculation of chi-square values from markers D9BrapaS1 and D9BrapaS4 for Wisconsin F2 population.

Slankster reported the allele frequencies of three RCBr varieties, with a similar phenotype to Wisconsin Fast plants, using 20 samples per strain from Carolina Biological Supply. We compared Slankster allele frequencies with our obtained Wisconsin allele frequencies (**Table 6**). For marker, D9*Brapa*S1 showed a significant difference in allele frequencies for the purple (0.89 observed vs 0.53 expected for allele one and 0.11 observed vs 0.47 expected for allele two) and non-purple stem (0.80 observed vs 0.53 expected for allele 1 and 0.20 observed vs 0.47 expected for allele two) strains. Moreover, both Standard strains (Wisconsin and Carolina) showed no significant difference in allele frequencies between the observed (0.17 for allele one and 0.83 for allele 2) and the reported by Slankster (0.21 for allele 1 and 0.79 for allele two). With regard to marker D9*Brapa*S4, the three strains from Wisconsin plants allele frequencies show a minor deviation to the ones obtained by Slankster. It is important to mention that three alleles were identified with D9*Brapa*S4 in both standard strains as reported by Slankster.

| Marker | Allele | Strain and Hypocotyl (Stem) Colour | | | | | |
|---|---|---|---|---|---|---|---|
| | | Purple Stem | | Non-Purple Stem | | Standard | |
| | | Scored | Expected | Scored | Expected | Scored | Expected |
| D9S1 | 1 | 0.89 | 0.53 | 0.80 | 0.53 | 0.17 | 0.21 |
| | 2 | 0.11 | 0.47 | 0.20 | 0.47 | 0.83 | 0.79 |
| D9S4 | 1 | 0.00 | 0.00 | 0.05 | 0.03 | 0.22 | 0.15 |
| | 2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.16 | 0.06 |
| | 3 | 1.00 | 1.00 | 0.95 | 0.97 | 0.61 | 0.79 |

**Table 6.** Marker D9*Brapa*S1 and D9*Brapa*S4 allele frequencies in Wisconsin Fast Plants parental lines and compared to allele frequencies reported by Slankster.

## Phenotype and validation of markers D9*Brapa*S1 and D9*Brapa*S4 in Carolina Biological Supply parental lines.

After comparing the alleles frequencies strains from Wisconsin Fast Plants, we requested and acquire the same Carolina RCBr strains used by Slankster. Three different strains were employed: RCBr-104 (Non-purple Stem, Yellow-Green Leafs), RCBr-119 (Purple Stem, Yellow-Green Leafs) and Standard RCBr 033 (Green Stem, Green Leafs). Ten plants, from each variety, were sowed in soil pots. Fourteen DAS stem and leaf colour were scored (**Figure 3**). All thirty plants show the expected phenotype.




**Figure 3.** Rapid Cycle *Brassica rapa* from Carolina Biological Supply (14 days after sowing).

DNA was extracted from all plants after scoring for stem and leaf colour. Markers D9*Brapa*S1 and D9*Brapa*S4 were evaluated using this DNA. For marker D9*Brapa*S1, two different alleles were observed in all strains (**Figure 4**), indicating that these strains are not homozygous. For strain RCBr-104 plant number 6 shows both alleles (heterozygous), for the rest of the samples only one

allele was identified. For RCBr-119 only the first two plants show and homozygous genotype, the rest of the samples present both alleles. Finally, for the strain RCBr-033, both homozygous (dominant and recessive) genotypes are present and plant 8 shows a heterozygous genotype.

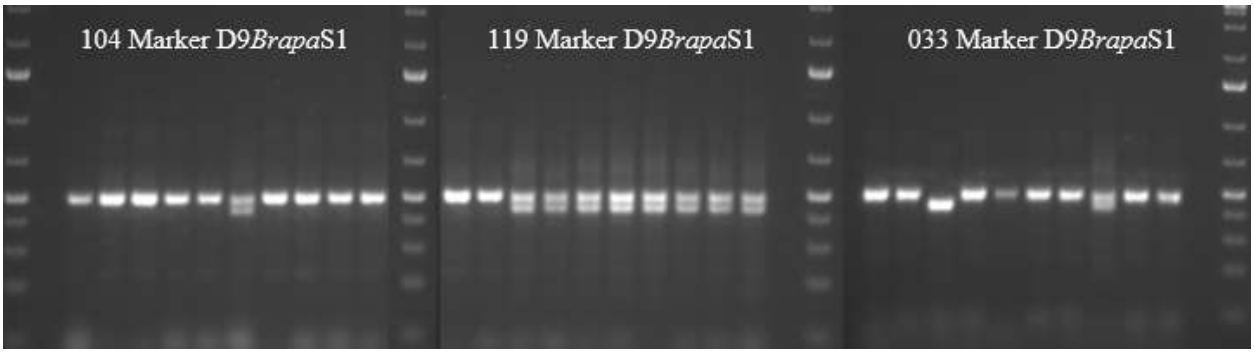



**Figure 4.** Allele resolution of marker D9*Brapa*S1 for Carolina plants.

Regarding marker D9S4, the resolution of alleles for RCBr-104 showed two alleles for plants 4 and 8. While the rest of the samples showed only the upper allele, for RCBr-119 only one allele was observed whereas for RCBr-033 three alleles were identified (**Figure 5**).

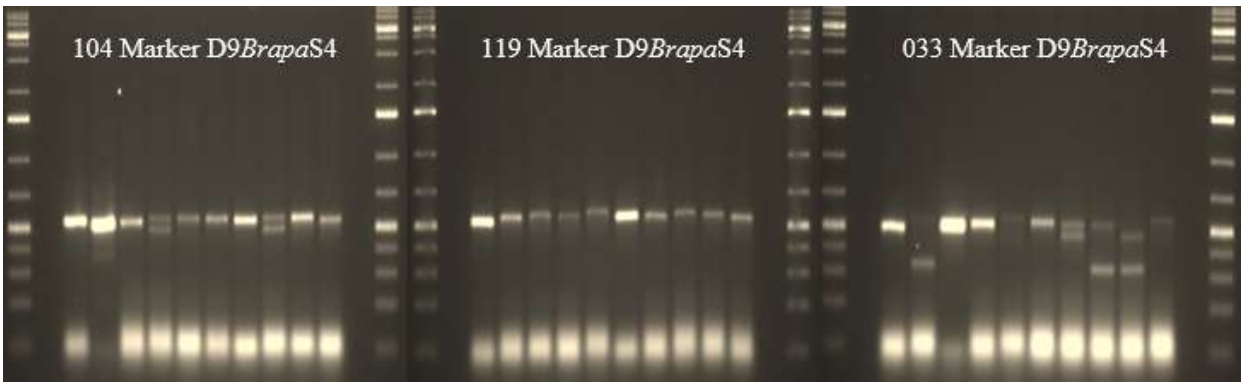



**Figure 5.** Allele resolution of marker D9*Brapa*S4.

Next, the first five plants from RCBR-104 and -119 were compared "side by side" (**Figure 6**). It was clearly observable that the unique allele, produced by D9*Brapa*S4, in strain 119 showed a similar length sequence to the upper allele in strain 104.

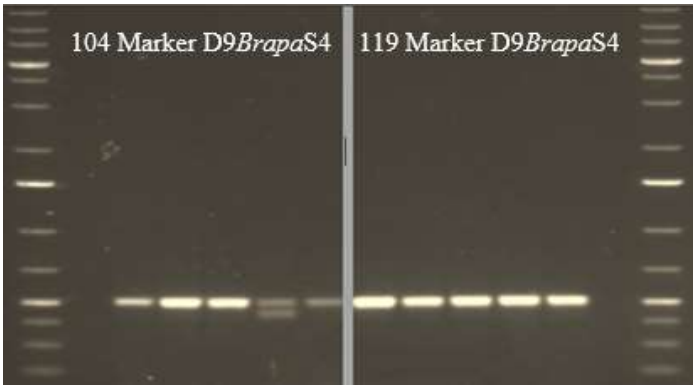



**Figure 6.** Allele resolution of marker D9*Brapa*S4. First block of five samples come from Carolina strain 104, second block from strain 119.

As done with Wisconsin Fast plants, Carolina plants allele frequencies were estimated and compared with the ones published by Slankster (also obtained from Carolina Biological Supply). Table 7 resumes the D9*Brapa*S1 and D9*Brapa*S4 allele frequencies in RCBr-104, -119 and -033 strains. For D9*Brapa*S1 similar range frequencies are reported between the score observed (ours) and the expected (Slankster). Thus, closer results and also higher degree of polymorphism was obtained for D9*Brapa*S1 with RCBr from Carolina Biological Supply than strains from Wisconsin Fast Plants. For D9*Brapa*S4, similar results to Wisconsin Fast plants were obtained.

| | | Strain and Hypocotyl (Stem) Colour | | | | | |
|---|---|---|---|---|---|---|---|
| Marker | Allele | 104 (Green) | | 119 (Purple) | | 033 (Standard) | |
| | | Scored | Expected | Scored | Expected | Scored | Expected |
| D9S1 | 1 | 0.05 | 0.15 | 0.60 | 0.53 | 0.15 | 0.21 |
| | 2 | 0.95 | 0.85 | 0.40 | 0.47 | 0.85 | 0.79 |
| D9S4 | 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.15 |
| | 2 | 0.10 | 0.13 | 0.00 | 0.00 | 0.10 | 0.06 |
| | 3 | 0.90 | 0.87 | 1.00 | 1.00 | 0.70 | 0.79 |

**Table 7.** Marker D9*Brapa*S1 and D9*Brapa*S4 allele frequencies in Carolina Biological Supply parental lines compared with the allele frequencies reported by Slankster.

Our next step was to develop new strains with defined marker genotypes by self-crossing Carolina plants showing heterozygous alleles. Unfortunately, plants were affected by thrips and the pest compromised further experiments.

Validation of candidate genes associated to genetic studies to leaf morphology in *B. oleracea* morphotypes.

Experiments with four candidate genes were performed to identify the allelic variation resulting in a function variance in *B. oleracea* leaf morphology and the development of molecular marker for mapping. Therefore, from the different polymorphisms identified by aligning gene sequences three were chosen randomly for further investigation. Results for *BoGATA19*, *BoUBP15*, *BoSPR1*, and *BoNGTPT* genes are presented.

*BoGATA19 (GATA transcription factor 19) (Bol028986)*

*BoGATA19* transcript sequence according to BRAD database presents a length of 660 bp (**Figure 7**). Figure 7 shows the gene structure of *BoGATA19* (Bol028986). *BoGATA19* present a total nucleotide length of 966 bp with one intron (295-601) of 306 bp. Amplification, using gradient a PCR assay, with primers PrJA001F and PrJA002R located in 31/966 and 908/966, respectively, yielded no bands in almost all accessions after ethidium bromide staining and electrophoresis. Then, a new set of primers PrJA003F and PrJA004R (53/966 and 894/966) were developed. After a PCR with all possible combination between the two sets of primers, still, no band was observed after ethidium bromide staining and electrophoresis. Consequently, following steps with the sequence strategy were not performed.

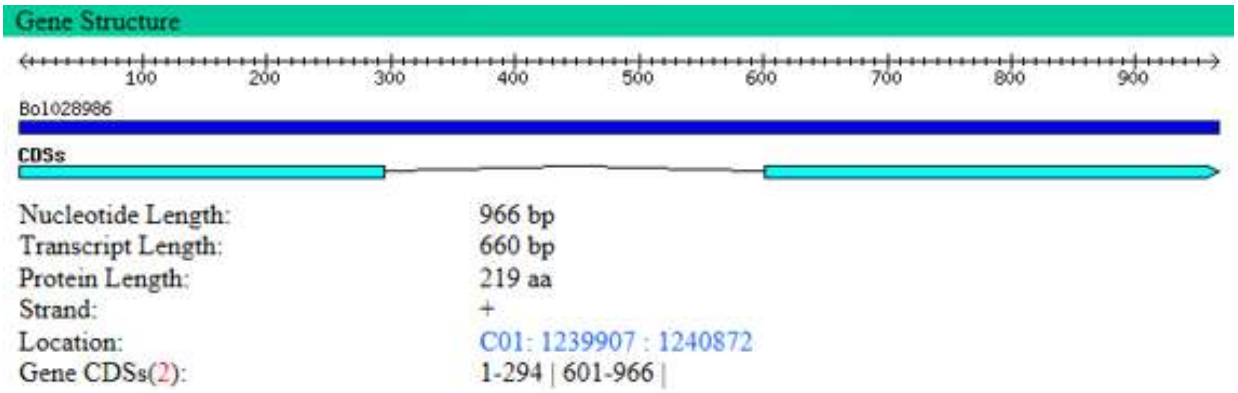



**Figure 7.** *BoGATA19* (Bol028986) gene structure obtained from BolBase.

*BoUBP15 (Ubiquitin carboxyl-terminal hydrolase 15) (Bol038288)*

*BoUBP15* sequence obtained from BRAD database for *B. oleracea* exhibit a gene fragment of 1448 bp. Figure 8 shows the gene structure of *BoUBP15*, an intron (767-1421) of 654 bp is

observed. Primers PrJA011F and PrJA012R with position 53/1448 and 894/1448, respectively, were used for a PCR assay.

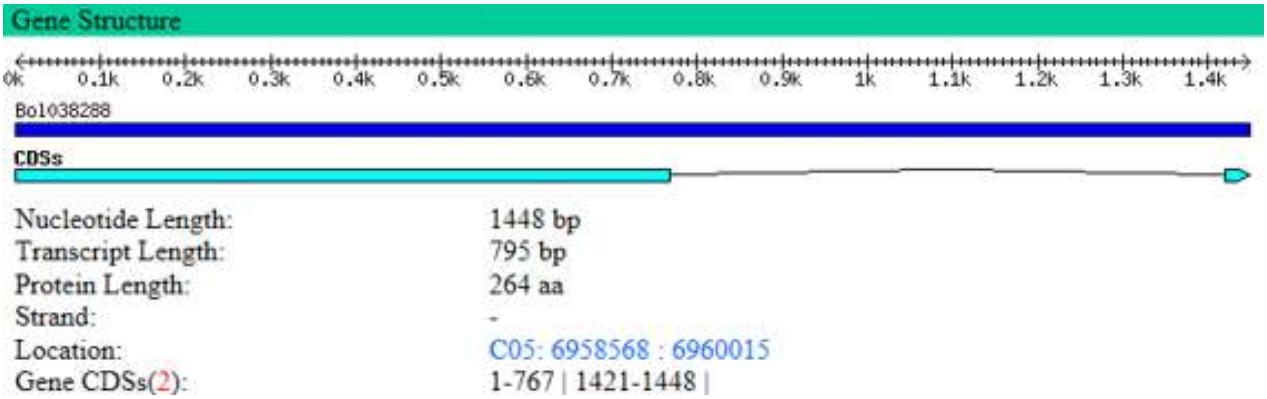



**Figure 8.** *BoUBP15* (Bol038288) gene structure obtained from BolBase.

Thirteen, from eighteen, accessions yielded, after electrophoresis, a clear band with the expected gene length (**Appendix 2**). PCR product after a successful amplification was then cut directly from TAE 1.0% agarose gel and purified using the MinElute Gel Extraction Kit, to get rid of primer dimers and other unwanted amplified sequences contaminating the sample. After DNA purification clear and single bands were obtained (not shown).

Purified single DNA fragments from heading cabbage (795, 718, 160 and 026), cauliflower (069, 511 and 297) and kohlrabi (317, 096, 145, 581 and 699) accessions were sequenced in separate samples using forward and reverse primers (PrJA011F and PrJA012R). Obtained sequences, from GATC laboratories, were then tested for clear and single peaks in the electropherogram record and only the sequences matching this criteria were used in the search of polymorphisms. With the help of MEGA 6 software, heading cabbage (795 and 026), cauliflower (511 and 287) and kohlrabi (317, 096 and 581) sequences were aligned and three polymorphisms were identified (**Figure 9**, **Figure 11**, and **Figure 13**).

The first polymorphism located in 388/966 bp is a SNP (**Figure 9**) were kohlrabi accession 317 and heading cabbage accession 026 contains an adenine nucleotide, while the rest of the accessions, gDNA, and mRNA contained a cytosine nucleotide. This SNP (SNP1) is located inside the coding sequence (CDS). After translating the nucleotide sequence, containing the SNP1, into

an amino acid (aa) sequence, was observed that the SNP1 performs as a silent mutation since all sequences translate for the same aa (**Figure 10**).

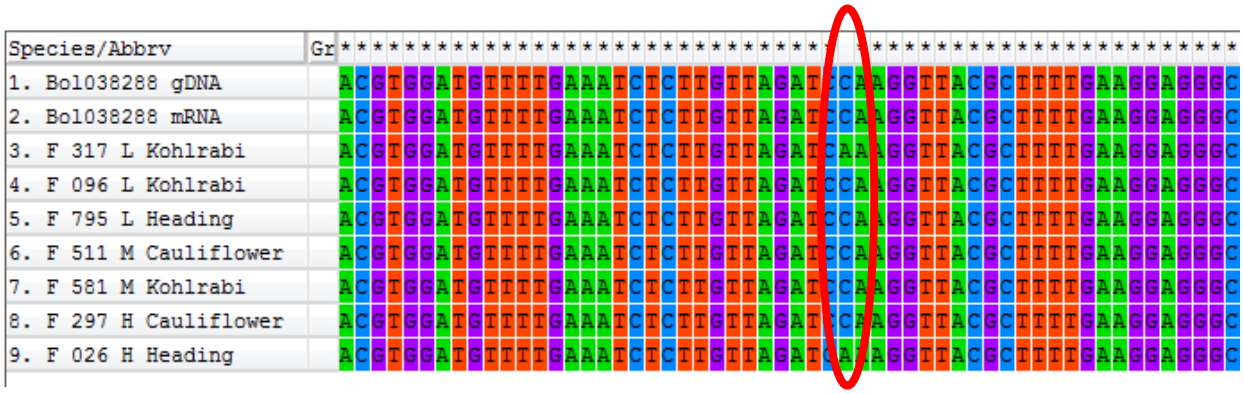



**Figure 9.** Nucleotide alignment of gDNA, mRNA and several morphotypes sequences for the *BoUBP15* (Bol038288) gene. The first polymorphism, a SNP of A/C, is circled in red.

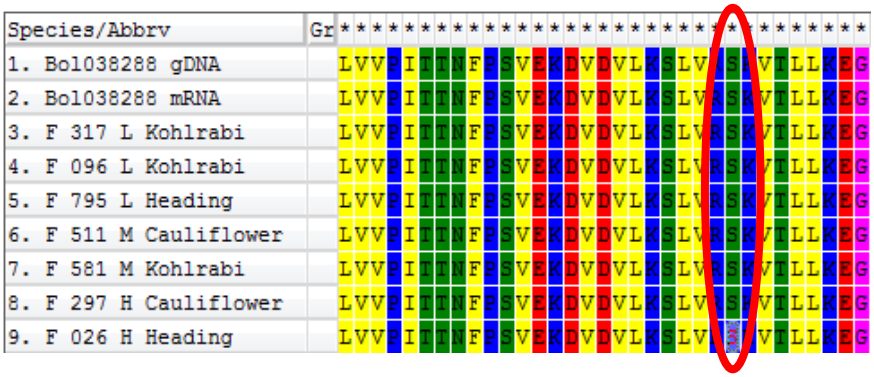



**Figure 10.** Amino acid alignment of gDNA, mRNA and several morphotypes sequences for the *BoUBP15* (Bol038288) gene. The first polymorphism, a silent SNP, is circled in red.

The second polymorphism was founded in position 302/966 bp as an SNP (SNP2) where accession 511 from cauliflower, 581 from kohlrabi and 026 from heading cabbage showed an adenine nucleotide meanwhile, the rest of accessions including the gDNA and mRNA reference sequence contained a guanine nucleotide (**Figure 11**) Once again the SNP was located inside the CDS and after translating the nucleotide sequence into an amino acid sequence a silent mutation was identified (**Figure 12**).

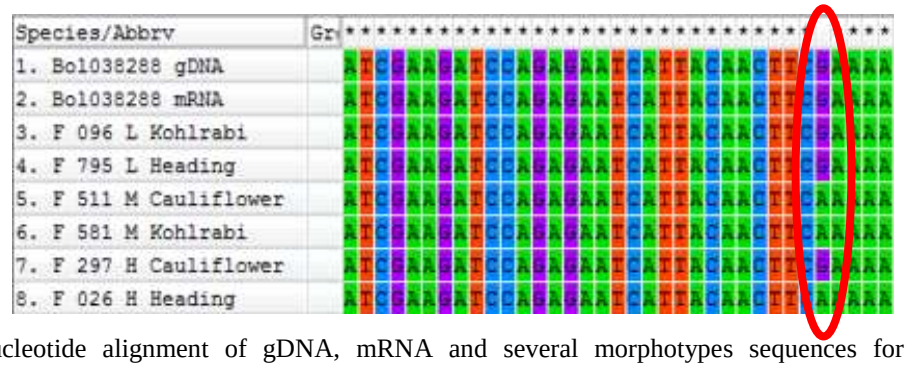



**Figure 11.** Nucleotide alignment of gDNA, mRNA and several morphotypes sequences for the *BoUBP15* (Bol038288) gene. The second polymorphism, a SNP of A/C, is circled in red.

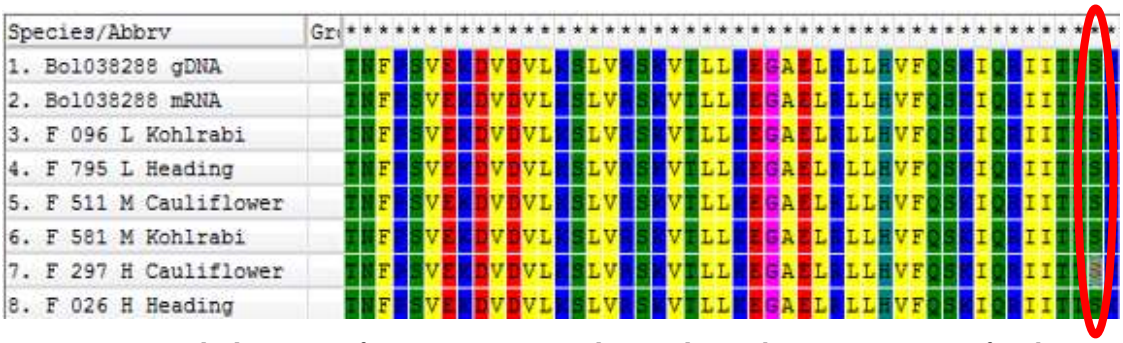



**Figure 12.** Amino acid alignment of gDNA, mRNA and several morphotypes sequences for the *BoUBP15* (Bol038288) gene. The second polymorphism, a silent SNP, is circled in red.

The third and last polymorphism harboured two adenine insertion/deletion (Indel) located in76/966 bp (Indel1) and 88/966 bp (Indel2) within the CDS (**Figure 13**). Figure 13 shows that accessions 317 from kohlrabi and 297 from cauliflower contained both (Indel1 and Indel2) whereas, accession 581 from kohlrabi presented only Indel2. Once we took the nucleotide sequence from the three accessions (317, 297 and 581), after translation into amino acid sequence and comparison with the gDNA and mRNA translated reference sequence it was confirmed, that Indel1 and Indel2 translated for a different amino acid sequence (**Figure 14**).

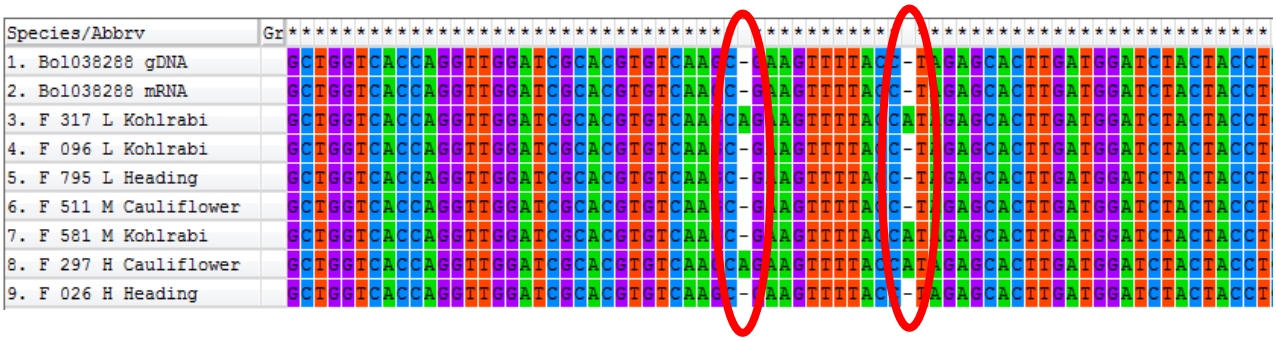



**Figure 13.** Nucleotide alignment of gDNA, mRNA and several morphotypes sequences for the *BoUBP15* (Bol038288) gene. The third polymorphism, two adenine INDEL, is circled in red.

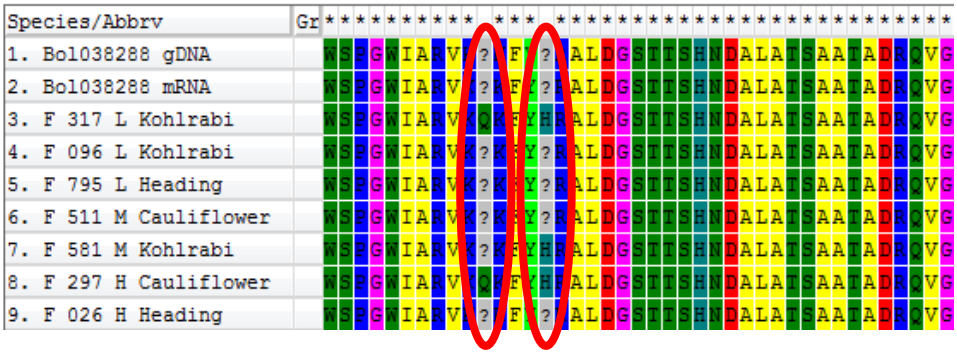



**Figure 14.** Amino acid alignment of gDNA, mRNA and several morphotypes sequences for the *BoUBP15* (Bol038288) gene. The third polymorphism, two INDEL, is circled in red.

Our next step was to develop CAPS makers, using the ApE software, with a restriction site to digest at the exact position where our Indel´s were located. For Indel1 no restriction enzyme was available for the desired position. However, for Indel2 two restriction enzymes (MaeI and FspBI) were found to digest at the exact position where Indel2 was located (**Figure 15**). MAE1 or FspBI will theoretically cut only when a *CTAG* nucleotide arrangement is present. Thus, MAE1/FspBI will cut only when Indel2 (adenine) is present.

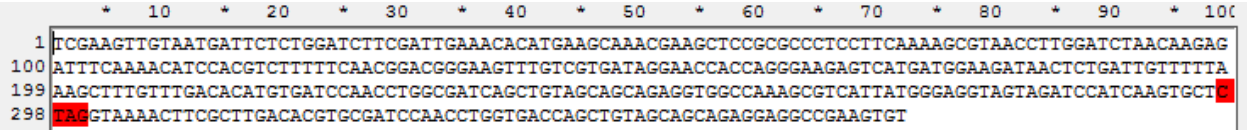



**Figure 15.** Sequence amplified by primers *UBP15*-03F/06R. Red box show the restriction site for MaeI/FspBI restriction enzyme.

Results after enzymatic digestion by MaeI/FspBI showed a partial digestion in all employed samples (**Figure 16**). Samples 4, 5 and 8 were supposed to remain undigested however, all samples exhibit a similar outcome.

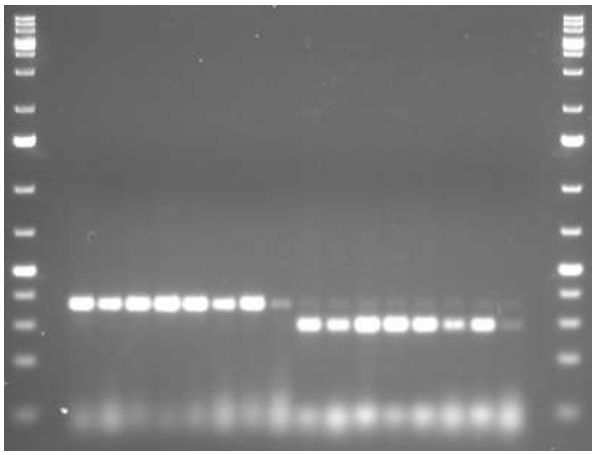


**Figure 16.** Resolution of alleles of *BoUBP15* (Bol038288) digested by MaeI/FspBI, first 8 samples contain the undigested PCR product of primers *UBP15*-03F/06R meanwhile, the last 8 samples contains the same PCR product after MaeI/FspBI digestion.

*BoSPR1 (SPIRAL 1) (Bol032553)*

*BoSPR1* gene structure (BRAD database) exhibit a total length of 412 bp with a 67 bp intron (**Figure 17**). After a PCR with primers PrJA021F (16/412) and PrJA022R (369/412), ten of eighteen accessions (**Appendix 2**) yielded a clear band with the expected gene length (not shown). PCR product was then cut directly from TAE 1.0% agarose gel and purified.

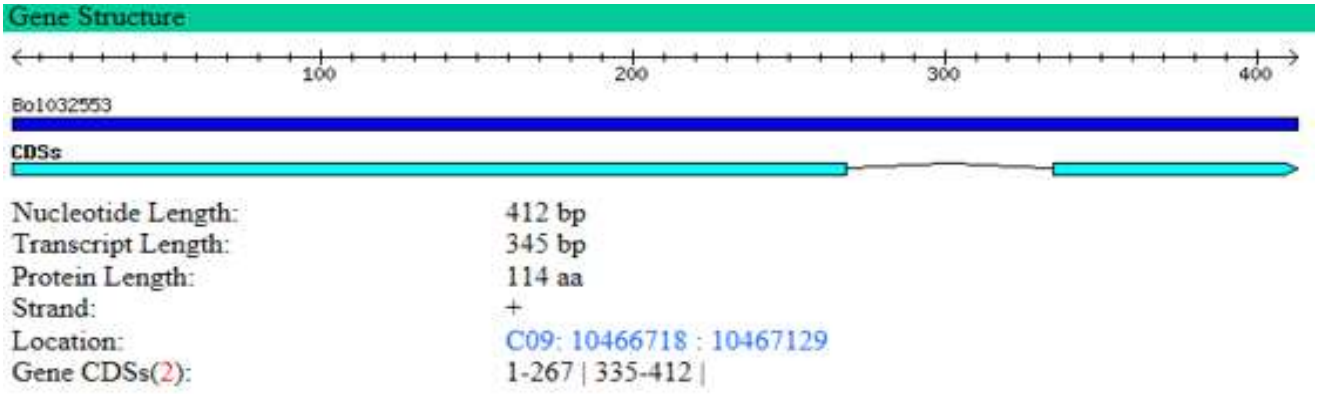



**Figure 17.** *BoSPR1* (Bol032553) gene structure obtained from BolBase.

Purified single DNA fragments from heading cabbage (795, 165 and 537), cauliflower (209, 069, 369 and 676) and kohlrabi (096 and 699) accessions were sequenced in separate samples using forward and reverse primers (PrJA021F and PrJA022R). The obtained sequences showed unclear and multiple peaks in the electropherogram record rendering the procedure for sequence alignment

difficult (not shown). A new set of primers (PrJA023F and PrJA024R) were developed and better results than using primers PrJA021F and PrJA022R in amplification were obtained (**Figure 18**). However, sequencing assay was not performed for time reasons.

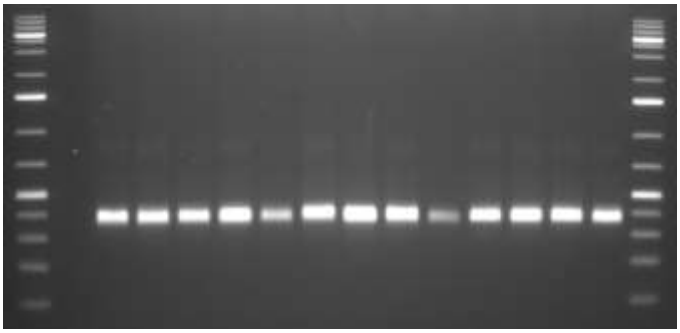


**Figure 18**. Resolution of *BoSPR1* (Bol032553) gene amplified bands by primers PrJA023F and PrJA024R.

*BoNGTPT (Non-green Triose phosphate/phosphate translocator) (Bol027034)*

*BoNGTPT* gene structure (BRAD database) shows a considerably long gene with a total length of 2099 bp (**Figure 19**) with eight introns of different length. For sequencing purposes, *BoNGTPT* gene was divided into two segments by developing two sets of primers with an overlapping region. Primers PrJA031F and PrJA032R were intended to amplify from bp 27 to 1505 (segment1) meanwhile PrJA033F and PrJA034R from bp 1145 to 2081 (segment2).

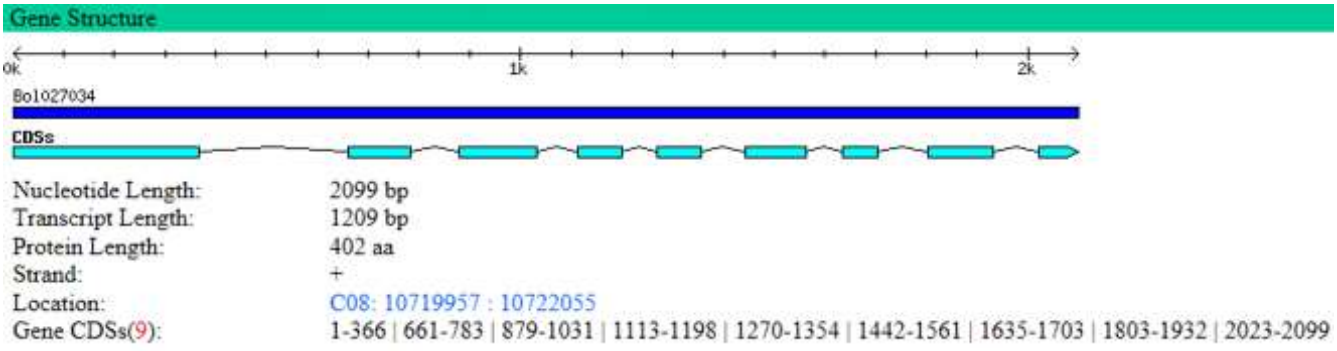



**Figure 19.** *BoNGTPT* (Bol027034) gene structure obtained from BolBase.

Better amplifying results were obtained with the second fragment (PrJA033F and PrJA034R) since 8 of 9 accessions were amplified whereas for the first segment (PrJA031F and PrJA032R) only 5 of 9 were successful (not shown). Amplified PCR products from both segments were directly cut from the gel plate, purified and tested for DNA purity (not shown). Purified single DNA fragments

from segment1 (heading cabbage (451 and 541), cauliflower (027 and 020) and Kohlrabi (181)) and from segment2 (heading cabbage (451, 541 and 168), cauliflower (027 and 020) and kohlrabi (803, 084 and 181)) were sent for sequencing in separate samples using their respective primers. Only the obtained sequencing results showing clear and single peaks in the electropherogram (not shown) record were pooled for the identification of polymorphisms. After the alignment of heading cabbage (168, 451 and 541), cauliflower (027 and 020), kohlrabi (803,084 and 181) sequences, by MEGA6 software, no polymorphisms were identified in segment1 however, two polymorphisms were identified in segment2.

The first polymorphism founded inside the *BoNGTPT* gene was an Indel of 7 base pairs located in 1239-1245/2099 and by aligning the reference gDNA with the mRNA it was found that the Indel was part of an intron (**Figure 20**). Next step was to develop a CAPS marker which digests exactly where the Indel is located.




**Figure 20.** On top, nucleotide alignment of gDNA and several morphotypes sequences for the *BoNGTPT* (Bol027034) gene. The first polymorphism, an adenine INDEL, is circulated in red. On the bottom, nucleotide alignment of gDNA and mRNA for the NGTPT gene.

By using the ApE software, the HpyCH4V (TGCA) restriction enzyme with a restriction site located on the adenine Indel was identified. However, HpyCH4V digest 5 times in the amplified sequence by primers PrJA033F and PrJA034R **(Figure 21)**. Thus, new primers PrJA16F (1178/2099 bp) and PrJA19R (1377/2099 bp bp) were developed for amplify a shorter fragment and ensure that HpyCH4V will digest only once in the amplified sequence from *BoNGTPT* gene (**Figure 21**)

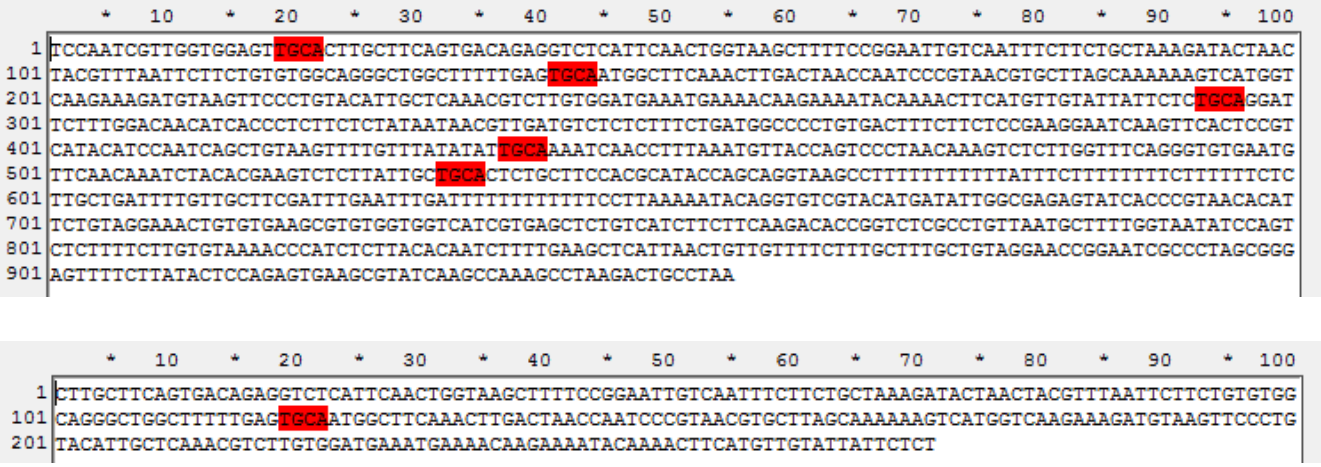



**Figure 21.** Red box shows the restriction sites for HpyCH4V restriction enzyme. On top, *BoNGTPT* (Bol027034) sequence amplified by primers PrJA033F and PrJA034R shows 5 restriction sites for HpyCH4V. On the bottom, *BoNGTPT* (Bol027034) sequence amplified by primers PrJA16F and PrJA19R shows a single restriction site for HpyCH4V.

Results after enzymatic digestion by HpyCH4V shows a complete digestion in all employed samples (**Figure 22**). At least, sample 3 was supposed to remain undigested. However, all samples exhibit a similar outcome.

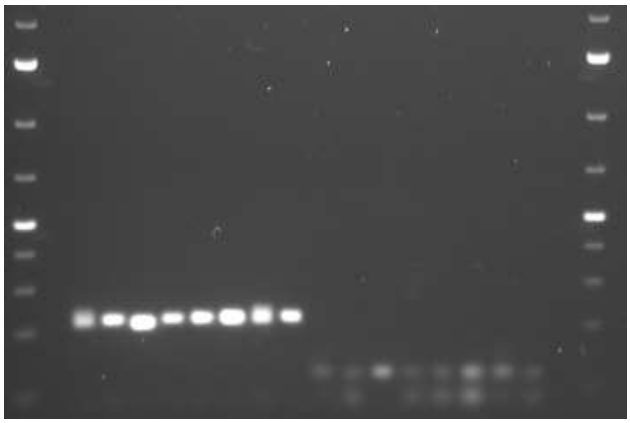


**Figure 22.** Resolution of alleles for *BoNGTPT* (Bol027034) digested by HpyCH4V, first 8 samples contain the undigested PCR product of primers PrJA16R and PrJA19R meanwhile, the last 8 samples contains the same PCR product after HpyCH4V digestion.

The second and last polymorphism in the *BoNGTPT* gene is an SNP located in 1883/2099 bp were kohlrabi accession 803 and heading cabbage 541 accession contains a guanine base, meanwhile

the rest of the accessions, the reference gDNA and mRNA contain a thymine base (**Figure 23**). This SNP is located inside the CDS, and after translation of the nucleotide sequence into the amino acid sequence, it was observed that SNP acts as a silent SNP (**Figure 24**). Next we continue to develop CAPS marker for analysing this polymorphism.

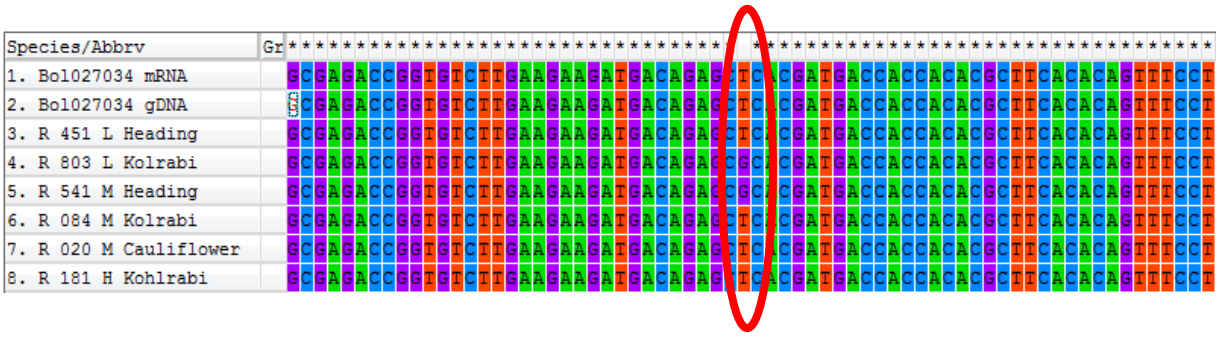



**Figure 23.** Nucleotide alignment of gDNA, mRNA and several morphotypes sequences for the *BoNGTPT* (Bol027034) *gene*. The polymorphism, a SNP of T/G, is circulated in red.

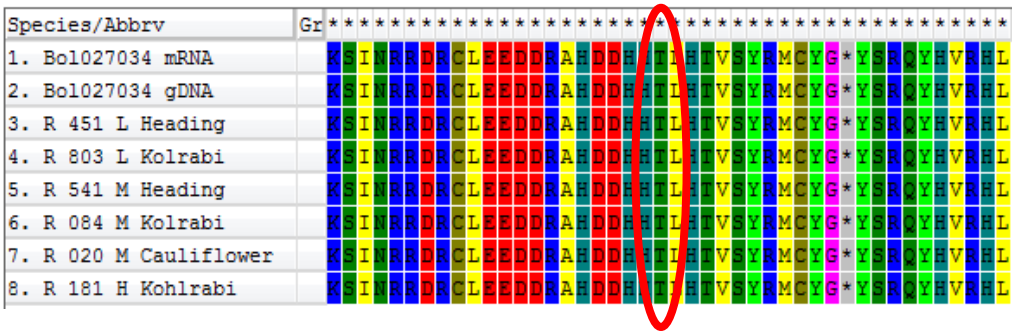



**Figure 24.** Amino acid alignment of gDNA, mRNA and several morphotypes sequences for the *BoNGTPT* (Bol027034) gene. The polymorphism, a silent SNP, is circulated in red.

Using ApE software, the restriction enzyme SACI (*GAGCTC*) was identified to digest exactly where the SNP was located. New primers (*NGTPT*01F, *NGTPT*02R, *NGTPT*03F, *NGTPT*04R, and *NGTPT*06R) were developed since primers PrJA033F and PrJA034R were not specific enough (not shown). Primer combination *NGTPT*03F and *NGTPT*06R, amplifying a segment of 973 bp, yielded the clearest and almost single band of all possible combinations (not shown). PCR product from this primer set was then digested with SacI restriction enzyme. Figure 25 shows ten PCR products, from primers *NGTPT*03F/06R (1166-2091 bp), before and after SacI enzymatic digestion. Samples 2 (not observable) and 7 were expected to be undigested. However, all samples were at least partially digested. Thus, no further experiments were performed.

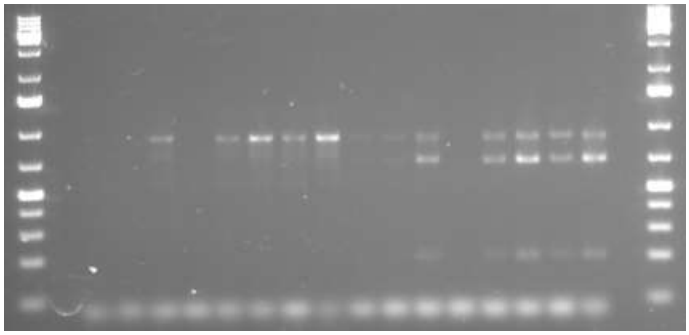


**Figure 25.** Resolution of alleles for *BoNGTPT* (Bol027034) digested by SacI, first 8 samples contain the undigested PCR product of primers *NGTPT*-03F/06R meanwhile, the last 8 samples contains the same PCR product after SacI digestion.

## Discussion

Development of new RCBr strains with defined marker genotypes to the *anthocyaninless* locus.

The goal of this project was to achieve RCBr strains, for teaching purposes, with defined marker genotypes closely linked to an easy observable trait. Four parental strains from Wisconsin Fast plants and one F2 population were evaluated first. All phenotypes observed in the parental lines correlated with the expected outcomes. However, based on a chi-square test the phenotypic segregation of purple/green hypocotyl and green/yellow-green cotyledon in the F2 population a difference segregation to the expected Mendelian segregation in a dihybrid cross of 9:3:3:1 was observed. Additionally, F2 monohybrid segregation of hypocotyl colour differed to the Mendelian segregation of 3:1. One possible explanation is the erroneous score for the hypocotyl colour since the intensity of anthocyanin pigmentation is a polygenic trait and can also be affected by the environment (Goldman, 1999), since low anthocyanin levels producing a light, almost indistinguishable, purple colour can be easily confused with a green colour.

Markers developed by Slankster and team (2012) were the starting point in developing teaching material for marker-trait association analysis. The first step was the validation of the genomic location of Slankster makers. Few incongruities were found between the marker position reported by Slankster and the position proportioned by BRAD database. However, marker location minimal discrepancies did not affect significantly the project outcome.

The next step was the evaluation of markers D9*Brapa*S1 and D9*Brapa*S4 due to their close location to the *anthocyaninless* locus (Slankster *et al*., 2012). Both selected markers were evaluated on RCBr strains acquired from Wisconsin Fast Plants and Carolina Biological Supply. Is clearly observable that parental lines from Wisconsin and Carolina strains do not present homozygous alleles for D9*Brapa*S1 and D9*Brapa*S4. Therefore, is difficult to calculate the genetic distance or recombination rate between markers and the *anl* locus.

Obtained allele frequencies highly resemble to the results reported by Slankster and team (2012) however, an insignificant difference can be observed between the allele frequencies obtained in this project and the ones reported by Slankster. This slightly variation in allele frequencies can be explained by the number of samples employed. According to Slankster article (2012) the reported allele frequencies are product of 20 samples per strain meanwhile, this project allele frequencies

are the outcome of only 10 samples per strain. Slankster and colleagues (2012) concluded that strains acquired from Carolina Biological Supply are better option, over Wisconsin Fast Plants, for the use of marker D9*Brapa*S1. Marker D9*Brapa*S1 yield a higher degree of polymorphism in all Carolina strains evaluated by Slankster and results from this project support the same conclusion. Therefore, D9*Brapa*S1 and D9*Brapa*S4 high level of polymorphism in Carolina strains enabled the development of RCBr strains with genotypes optimized for the use of these markers. For developing new strains of RCBr with defined marker genotypes it was proposed to choose the plants expressing heterozygous alleles from each marker and by bud-pollination (for avoiding self-incompatibility) self-cross the heterozygous plant and select in the progeny the desired genotype. Unfortunately, RCBr plants from this project were damaged by Thrips compromising further experiments.

Validation of candidate genes associated to genetic studies to leaf morphology in *B. oleracea* morphotypes.

The first step for the validation of candidate genes associated to genetic studies to leaf morphology was to amplify the gene sequence underlying each gene. Almost all primers yielded bands after gel-electrophoresis with the exception of *BoGATA19* related primers. Different approaches were employed to overcome this problem as the evaluation of DNA material, the use of 4 different primer combination, gradient PCR with 6 different annealing temperatures, the use of two different polymerases (Taq and Phusion). Nevertheless, it was not possible to amplify the desired sequence. *BoGATA19* primers, as all primers genes, were blasted using BRAD database to confirm the sequence location in the desired chromosome. However still not results were obtained. Failure in this step can be caused by primer dimerization, hair spin formation or the presence of polymorphisms in the annealing primer location.

After successful amplification of *BoUBP15*, *BoSPR1* and *BoNGTPT* genes, samples were sent for sequencing. Results from this sequences were tested and confirmed for its use. Obtained electropherogram records for each morphotype (heading cabbage, cauliflower and kohlrabi´s) were tested to confirm a trustable sequencing assay. Electropherogram records from *BoUBP15* and *BoNGTPT* genes showed clear and single peaks when sequenced using forward and reverse primers. However, gene sequence form *SPR1* shows multiple peaks in the electropherogram chart making it unreliable for the search of polymorphisms. New primers were developed for amplifying

again *SPR1* and obtained sequences were sent for sequencing. However, the new sequencing results showed the same problem. After two times failure in sequencing *BoSPR1* further experiments was ceased. This can be explained by the use of unspecific primers, annealing and amplifying more than one sequence. Therefore, the sequencing results showed not only one sequence read.

*BoUBP15* and *BoNGTPT* sequences from heading cabbage, cauliflower and kohlrabi were aligned separately using the MEGA6 software. Acquired sequences were aligned together with the genomic sequence (gDNA) and the coding sequence (mRNA) obtained from the BRAD database. The use of the gDNA and mRNA allows discerning if a polymorphism is located in the coding sequence or forms part of an intron region. From the five polymorphisms identified, two were located in *BoUBP15* gene and three inside the *BoNGTPT* gene. By random selection one Indel in the *BoUBP15* gene and two polymorphism (one SNP and one Indel) in the *BoNGTPT* gene were selected for further investigation since time for research was limited.

A CAPS marker was then developed for each polymorphism. Using the ApE software, a restriction enzyme digesting for each specific polymorphism was identified. New primers following two criteria were developed. In a first scenario, primers will amplify a fragment length between 300-900 bp. Alternatively only one restriction site (for a specific restriction enzyme) will be located inside the amplified fragment. With these conditions, CAPS markers will be easily applied and the differentiation between digested/un-digested fragments will be easily noticed. Unfortunately, the CAPS marker approach did not produce the expected results since, partial degradation can be visualised on all three polymorphisms. Three theories can explain these results. Firstly, the sequence amplified by the new primers does not contain the polymorphism identified by the old primers. Since time was limited, the amplified sequence from the new primers was not sequenced to confirm the presence of the polymorphism identified with the old primers. Two, false positive polymorphisms product of erroneous reading during the sequencing. Finally, it can be due to the presence of heterozygous alleles where only one allele is digested resulting in an allele resolution of a mix of digested and undigested alleles. An alternative reason can be due to suboptimal conditions in the use of the restriction enzymes causing the cleaving in sequences which present a similar, however not identical, restriction site (star activity).

## Conclusion

This thesis aimed at studying the use and development of DNA genetic markers in two *Brassica* species. Two projects were intended: 1) the development of new Rapid Cycle *Brassica rapa* strains with genotypes designed for the use of Slankster *et al*., 2012 markers D9*Brapa*S1 and D9*Brapa*S4 selecting for the *anthocyaninless* locus; 2) the validation of *BoGATA19*, *BoUBP15*, *BoSPR1* and *BoNGTPT* as genes participating in leaf morphology in *B. oleracea* crops. For the RCBr project, Slankster markers D9*Brapa*S1 and D9*Brapa*S4 showed a higher polymorphism degree in strains obtained from Carolina Biological Supply than Wisconsin Fast Plants. Additionally, obtained allele frequencies from D9*Brapa*S1 and D9*Brapa*S4 markers resemble the reported allele frequencies from Slankster *et al*., 2012. Results from this project indicate that either in Wisconsin or in Carolina RCBr strains D9*Brapa*S1 and D9*Brapa*S4 marker alleles are not homozygous in parental lines. However, due to the irreversible plant damage caused from thrips, the development of new RCBr with defined marker genotype was not achievable. Nevertheless, the knowledge gained during this project shows the action plan needed for the development of new RCBr lines with fixed genotypes for teaching marker-trait association praxis. Regarding the project performed on *B. oleracea* morphotypes (heading cabbage, cauliflower and kohlrabi), *BoGATA19* gene was not amplified from the used DNA extracted samples. *BoSPR1* was fully amplified, however sequencing of the *SPR1* gene was unsuccessful. *BoUBP15* and *BoNGTPT* were fully or partially amplified and sequenced. Separate alignment of *BoUBP15* and *BoNGTPT* sequences from different morphotypes revealed three candidate polymorphisms, one in the *UBP15* and two in the *BoNGTPT* gene sequences. For each polymorphisms identified in *BoUBP15* and *BoNGTPT* gene sequences a specific CAPS marker was developed. Unfortunately, the three of them failed to discriminate if a *B. oleracea* accession present or is absent of the identified polymorphism. Therefore, further research in *BoGATA19*, *BoUBP15*, *BoSPR1* and *BoNGTPT* genes is needed to confirm or reject their participation on leaf morphology.

**Recommendations**

After four months of work in this thesis project, some recommendations can be suggested for improving the results quality. During the development of new RCBr with defined marker genotypes, parental lines ready for crosses and inter-crosses were insect damaged by thrips. Thus, for future practices is suggested to grow the RCBr plants under a free and controlled pest environment which safeguards the integrity of the project. Regarding to *BoGATA19*, *BoUBP15*, *BoSPR1* and *BoNGTPT* genes validation to leaf morphology in *B. oleracea* morphotypes is suggested to develop more specific primers and evaluate the new primers developed. Following this approach might increase the quantity and quality of gene sequences per gene. Thus, increasing the reliableness on the identified polymorphisms. Additionally, it is recommended to use principally the reverse primer for sequencing. Better results were obtained from using the reverse primer over the forward primer for sequencing. Moreover, it can be suggested to use different restriction enzymes, with the same restriction site, from different laboratories and optimize the reaction conditions for an optimal digestion.

## References


Amerik, A. Y., & Hochstrasser, M. (2004). Mechanism and function of deubiquitinating enzymes. Biochim Biophys Acta, 1695(1-3), 189-207. doi:10.1016/j.bbamcr.2004.10.003

Bar, M., & Ori, N. (2014). Leaf development and morphogenesis. Development, 141(22), 4219-4230. doi:10.1242/dev.106195

Bi, Y. M., Zhang, Y., Signorelli, T., Zhao, R., Zhu, T., & Rothstein, S. (2005). Genetic analysis of Arabidopsis GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity. Plant J, 44(4), 680-692. doi:10.1111/j.1365-313X.2005.02568.x

Bonnema, G., Del Carpio, D. P., & Zhao, J. (2011). Diversity analysis and molecular taxonomy of Brassica vegetable crops. Genetics, Genomics and Breeding of Vegetable Brassicas, 81-124.

Branca, F., Lucia, R., Alessandro, T., Poulsen, G., Maggioni, L., & Bothmer R, v. (2013). Diversity of Kale Growing in Europe as a Basis for Crop Improvement (Vol. 1005).

Burdzinski, C., & Wendell, D. L. (2007). Mapping the Anthocyaninless (anl) Locus in Rapid-Cycling *Brassica rapa* (RCBr) to Linkage Group R9. BMC Genetics, 8(1), 64. doi:10.1186/1471-2156-8-64

Cartea, M. E., Lema, M., Francisco, M., & Velasco, P. (2011). Basic information on vegetable Brassica crops. Genetics, Genomics and Breeding of Vegetable Brassicas, 1-33.

Cheng, F., Wu, J., & Wang, X. (2014). Genome triplication drove the diversification of Brassica plants. Horticulture Research, 1, 14024. doi:10.1038/hortres.2014.24

Cheung, A. Y. (1996). Pollen-pistil interactions during pollen-tube growth. Trends in Plant Science, 1(2), 45-51. doi:Doi 10.1016/S1360-1385(96)80028-8

Chitwood, D. H., Kumar, R., Headland, L. R., Ranjan, A., Covington, M. F., Ichihashi, Y. Sinha, N. R. (2013). A Quantitative Genetic Basis for Leaf Morphology in a Set of Precisely Defined Tomato Introgression Lines. Plant Cell, 25(7), 2465-2481. doi:10.1105/tpc.113.112391

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1491), 557-572. doi:10.1098/rstb.2007.2170

Evans, L. T. (1997). Adapting and improving crops: the endless task. Philosophical Transactions

of the Royal Society of London. Series B: Biological Sciences, 352(1356), 901-906. doi:10.1098/rstb.1997.0069

Fischer, K., Kammerer, B., Gutensohn, M., Arbinger, B., Weber, A., Häusler, R E., Flügge, U I., (1997). A new class of plastidic phosphate translocators: a putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. The Plant Cell Mar, 9 (3) 453-462; DOI: 10.1105/tpc.9.3.453

Fujikura, U., Horiguchi, G., Ponce, M. R., Micol, J. L., & Tsukaya, H. (2009). Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of Arabidopsis thaliana. Plant J, 59(3), 499-508. doi:10.1111/j.1365-313X.2009.03886.x

Furutani, I., Watanabe, Y., Prieto, R., Masukawa, M., Suzuki, K., Naoi, K., Hashimoto, T. (2000). The SPIRAL genes are required for directional control of cell elongation in Arabidopsis thaliana. Development, 127(20), 4443-4453.

Goldman I. L. (1999). Teaching recurrent selection in the classroom with Wisconsin Fast plants. Horttechnology 9, 579–584

Gonzalez, N., Vanhaeren, H., & Inze, D. (2012). Leaf size control: complex coordination of cell division and expansion. Trends Plant Sci, 17(6), 332-340. doi:10.1016/j.tplants.2012.02.003

Hepworth, J., & Lenhard, M. (2014). Regulation of plant lateral-organ growth by modulating cell number and size. Curr Opin Plant Biol, 17, 36-42. doi:10.1016/j.pbi.2013.11.005

Kalve, S., De Vos, D., & Beemster, G. T. (2014). Leaf development: a cellular perspective. Front Plant Sci, 5, 362. doi:10.3389/fpls.2014.00362

Liu, Y., Wang, F., Zhang, H., He, H., Ma, L., & Deng, X. W. (2008). Functional characterization of the Arabidopsis ubiquitin-specific protease gene family reveals specific role and redundancy of individual members in development. The Plant Journal, 55(5), 844-856.

Nagaharu, U. (1935). Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. Jap J Bot 7: 389-452.

Nakajima, K., Furutani, I., Tachimoto, H., Matsubara, H., & Hashimoto, T. (2004). SPIRAL1 Encodes a Plant-Specific Microtubule-Localized Protein Required for Directional Control of Rapidly Expanding Arabidopsis Cells. Plant Cell, 16(5), 1178-1190. doi:10.1105/tpc.017830

Scofield, S., & Murray, J. A. (2006). KNOX gene function in plant stem cell niches. Plant Mol

Biol, 60(6), 929-946. doi:10.1007/s11103-005-4478-y

Shivanna, K. R., Heslop-Harrison, Y., & Heslop-Harrison, J. (1978). THE POLLEN-STIGMA INTERACTION: BUD POLLINATION IN THE CRUCIFERAE. Acta Botanica Neerlandica, 27(2), 107-119. doi:doi:10.1111/j.1438-8677.1978.tb00265.x

Slankster, E. E., Chase, J. M., Jones, L. A., & Wendell, D. L. (2012). DNA-Based Genetic Markers for Rapid Cycling Brassica Rapa (Fast Plants Type) Designed for the Teaching Laboratory. Front Plant Sci, 3, 118. doi:10.3389/fpls.2012.00118

Smith, L. B., & King, G. J. (2000). The distribution of BoCAL-a alleles in Brassica oleracea is consistent with a genetic model for curd development and domestication of the cauliflower. Molecular Breeding, 6(6), 603-613. doi:10.1023/A:1011370525688

Snowdon, R. J., & Friedt, W. (2004). Molecular markers in Brassica oilseed breeding: current status and future possibilities. Plant Breeding, 123(1), 1-8. doi:doi:10.1111/j.1439-0523.2003.00968.x

Tsukaya, H. (2013). Leaf Development. The Arabidopsis Book / American Society of Plant Biologists, 11, e0163. doi:10.1199/tab.0163

Veit, B. (2004). Determination of cell fate in apical meristems. Curr Opin Plant Biol, 7(1), 57-64.

Warwick, S. I., Francis, A., & Al-Shehbaz, I. A. (2006). Brassicaceae: Species checklist and database on CD-Rom. Plant Systematics and Evolution, 259(2), 249-258. doi:10.1007/s00606-006-0422-0

Williams, P. H., & Hill, C. B. (1986). Rapid-cycling populations of Brassica. Science, 232(4756), 1385-1389. doi:10.1126/science.232.4756.1385

Xiao, D., Wang, H., Basnet, R. K., Zhao, J., Lin, K., Hou, X., & Bonnema, G. (2014). Genetic dissection of leaf development in Brassica rapa using a genetical genomics approach. Plant Physiology, 164(3), 1309-1325. doi:10.1104/pp.113.227348

Zhao, Y., Medrano, L., Ohashi, K., Fletcher, J. C., Yu, H., Sakai, H., & Meyerowitz, E. M. (2004). HANABA TARANU Is a GATA Transcription Factor That Regulates Shoot Apical Meristem and Flower Development in Arabidopsis. Plant Cell, 16(10), 2586-2600. doi:10.1105/tpc.104.024869

# Appendix

Appendix 1 - Xiao *et al*., 2014 adapted candidate list for leaf morphology. Candidates genes evaluated in this project are highlighted in green.

| | Trait | Gene of Interest in Region | Bol gene code | Distance (KB) Gene-Mark. | LOD |
|---|---|---|---|---|---|
| 1 | WidthLeaf | ROT3 | Bol028986 | 31404 | 17.4014 |
| 2 | WidthLeaf | ROT3 | Bol029002 | 46257 | 17.4014 |
| 3 | LengthLeaf | KAN2 | Bol014057 | 3233 | 2.79954 |
| 4 | AreaStem | IAA19 | | | 69.6152 |
| 5 | CoreLength | | Bol015765 | 5428 | |
| 6 | CoreLength | | Bol015767 | 21977 | |
| 7 | LengthLeaf | TCP12 | | | 1.47734 |
| 8 | AreaPetioleLargest | CLAVATA3 | Bol028637 | 14145 | 0.88582 |
| 9 | MeanWidthLeafCabbage | KNAT4 | | | 2.25604 |
| 10 | LengthStemLargest | KNAT4 | | | 3.76548 |
| 11 | LengthLeaf | INCURVATA2 | Bol015194 | 49956 | 5.76144 |
| 12 | LengthLeaf | SPR1 SPR2 | | | 0.85429 |
| 13 | TotalLength | | Bol025737 | 9953 | 9.02429 |
| 14 | LengthLeaf | LEP | | | 3.97249 |
| 15 | LengthLeaf | KAN1 | | | 3.5505 |
| 16 | AreaLeaf | MYB81 MYB104 | Bol027782 | 40775 | 10.4788 |
| 17 | LengthLeaf | LNG1 | Bol011079 | 42352 | 13.999 |
| 18 | LengthStemLargest | UBP15 | Bol038285 | 48526 | 5.28443 |
| 19 | LengthStemLargest | UBP15 | Bol038287 | 40250 | 5.28443 |
| 20 | LengthStemLargest | UBP15 | Bol038288 | 23793 | 5.28443 |
| 21 | LengthStemLargest | UBP15 | Bol038299 | 36515 | 5.28443 |
| 22 | LengthStemLargest | UBP15 | Bol038300 | 39436 | 5.28443 |
| 23 | MeanWidthLeaf | | Bol040785 | 41808 | 5.74454 |
| 24 | MeanWidthLeaf | | Bol040788 | 31404 | 5.74454 |
| 25 | MeanWidthLeaf | | Bol040789 | 27495 | 5.74454 |
| 26 | MeanWidthLeaf | | Bol040792 | 14626 | 5.74454 |
| 27 | MeanWidthLeaf | | Bol040798 | 17343 | 5.74454 |
| 28 | MeanWidthLeaf | | Bol040804 | 40508 | 5.74454 |
| 29 | AreaStem | | Bol039157 | 19665 | 16.1232 |
| 30 | AreaStem | | Bol039158 | 45933 | 16.1232 |
| 31 | CoreLength | | Bol026171 | 30480 | |
| 32 | CoreLength | | Bol026175 | 3951 | |
| 33 | AreaPetioleLargest | WOX3 | | | 0.00379 |
| 34 | LengthLeaf | NGTPT | Bol027034 | 34378 | 7.07629 |
| 35 | LengthLeaf | GTE4 CLO | | | 1.33261 |
| 36 | LengthStemLargest | SPR1 SPR2 | Bol032552 | 10449 | 1.80126 |
| 37 | LengthStemLargest | SPR1 SPR3 | Bol032553 | 24873 | 1.80126 |
| 38 | LengthStemCabbage | APRR3 TOE2 APUM19 | Bol017251 | 765 | 0.9396 |
| 39 | LengthStemCabbage | | Bol017253 | 32460 | 0.9396 |
| 40 | LengthLeaf | PI | | | 4.44391 |
| 41 | LengthStemCabbage | KAN1 | | | 9.93704 |
| 42 | LengthLeaf | | Bol019260 | 45232 | 9.24455 |
| 43 | LengthLeaf | | Bol019262 | 36960 | 9.24455 |

Appendix 2 – Accessions for each morphotype used in the B. oleracea project. In green are the accessions successfully amplified.

| Gene | Trait | Morphotype | Score Value | Accession |
|---|---|---|---|---|
| GATA19 | Leaf Lenght | Heading Cabbage | Low | 146 |
| | | | Medium | 790 |
| | | | High | 261 |
| | | Cauliflower | Low | 592 |
| | | | Medium | 61 |
| | | | High | 115 |
| | | Kohlrabi | Low | 181 |
| | | | Medium | 651 |
| | | | High | 112 |
| UBP15 | Petiole Lenght | Heading Cabbage | Low | 795 |
| | | | Low | 718 |
| | | | Medium | 165 |
| | | | Medium | 160 |
| | | | High | 537 |
| | | | High | 26 |
| | | Cauliflower | Low | 69 |
| | | | Low | 209 |
| | | | Medium | 511 |
| | | | Medium | 369 |
| | | | High | 297 |
| | | | High | 676 |
| | | Kohlrabi | Low | 317 |
| | | | Low | 96 |
| | | | Medium | 145 |
| | | | Medium | 581 |
| | | | High | 699 |
| | | | High | 699 |
| SPR1 | Petiole Area | Heading Cabbage | Low | 718 |
| | | | Low | 795 |
| | | | Medium | 160 |
| | | | Medium | 165 |
| | | | High | 26 |
| | | | High | 537 |
| | | Cauliflower | Low | 209 |
| | | | Low | 69 |
| | | | Medium | 369 |
| | | | Medium | 511 |
| | | | High | 676 |
| | | | High | 297 |
| | | Kohlrabi | Low | 96 |
| | | | Low | 317 |
| | | | Medium | 581 |
| | | | Medium | 145 |
| | | | High | 699 |
| | | | High | 699 |
| NGTPT | Leaf Width | Heading Cabbage | Low | 451 |
| | | | Medium | 541 |
| | | | High | 168 |
| | | Cauliflower | Low | 27 |
| | | | Medium | 20 |
| | | | High | 355 |
| | | Kohlrabi | Low | 803 |
| | | | Medium | 84 |
| | | | High | 181 |