

Evaluation of Changes in Glucosinolates Profile and Biosynthetic Genes Expression in Turnip (*Brassica rapa* L.)

Supervisors

Dr. ir. Guusje Bonnema

Dr. Jules Beekwilder

Dr.ir. Chris Maliepaard

Ir. Johan Bucher



March 26, 2015

Student: David Lagarrigue

Student number: 821209495080

Table of Contents

Acknowledgements.....	4
Abstract.....	5
Introduction	5
Glucosinolates.....	5
<i>Brassica rapa</i>	6
Turnip	6
Glucosinolates biosynthetic pathway, genes and transcription factors involved	7
Side chain elongation	7
Core structure formation	7
Secondary modifications.....	8
Transcription factors	9
Transport, synthesis and storage of GLS.....	10
Aim of the MSc thesis project	10
Research questions	11
Materials and methods.....	11
Project background	11
LC-MS data	13
Candidate gene selection criteria	15
qRT-PCR of Br-GTR paralogs	17
Primers design.....	17
Primers list	17
Protocols	17
Data analysis	18
Data pre-processing.....	18
Principal components analysis.....	19
Hierarchical clustering	19
Self-organizing maps	19
Correlation matrix.....	19
Other statistical tests	20
Relative gene expression (RGE)	20
Results.....	21
Background	21
GLS data evaluation	21

GLS profiles	22
Principal components analysis (PCA)	23
PCA for GLS	23
PCA for genes	26
Correlation matrix.....	28
Hierarchical clustering	31
Glucosinolates.....	31
Biosynthetic genes and paralogs	32
Self-Organizing map (SOM).....	49
Discussion.....	51
GLS distribution.....	51
Transcription factors as genes of major influence in GLS biosynthetic pathways	52
MYB28 – MYB29.....	52
MYB34 - MYB51	53
GTR genes activity	54
MAM genes role in AGLS profile	54
AOP2 genes are mostly expressed in leaves and correlations with other AGLS biosynthetic genes.....	56
Conclusions	57
Recommendations	57
References	58
Appendix 1: LC-MS data.....	65
Appendix 2: qRT-PCR data	74
Gene expression data as $-\Delta Ct$	74
Gene expression data as $2^{-\Delta\Delta Ct}$	77
Appendix 3: Wilcoxon and t-tests.....	80
T-test.....	80
Normality verifications.....	80
Equal variances verification	81
T-tests.....	82
Wilcoxon tests on GLS.....	83
Normality verifications.....	83
Wilcoxon rank-sum tests on all GLS.....	84
Wilcoxon rank-sum tests on GLS with significant differences between accessions.....	87
Appendix 4: SOM maps not shown in the report	90

Acknowledgements

I want to thank my supervisors Guusje Bonnema, Jules Beekwilder and Chris Maliepaard for giving me the opportunity to do my thesis in this interesting subject in the Growth and Development group, as well as for their constant help during this period. I am also very thankful for having the opportunity to assist to the Glucosinolates and Beyond conference in 2014. Johan Bucher also has all my gratitude for his great help with the execution of the qRT-PCR experiments on the glucosinolate transporter genes, as well as for his advice on primers design. All of this has contributed enormously to my learning experience during this process.

I also want to thank Shuhang Wang and Dr. Jun Gu Lee for generating the gene expression data and the LC-MS data, respectively.

I am also grateful of Dr. Hussam Hassan Nour-Eldin from the University of Copenhagen for providing the primers for the GTR2 gene paralogs in *B. rapa*.

My friends Jose, George, Fokion and Carola have also helped to make this period in Wageningen unforgettable with their company, support and friendship.

Marcela was also a formidable support with her love and wisdom during this process, especially in a period when I needed it the most. She was definitely a source of inspiration during this process.

And last, but not least, I thank my family for being always with me, despite the physical distance.

Abstract

Glucosinolates are secondary metabolites that have a role in plant defense in the *Brassicaceae* family. In addition, there is ongoing research on their cancer chemopreventive properties. A study aiming to correlate glucosinolates profile and gene expression has been conducted in two accessions of turnip (*Brassica rapa* ssp. *rapa*), measuring gene expression by means of qRT-PCR and glucosinolates profile (relative peak area) by means of LC-MS in several tissues and developmental stages. Data analysis was done by Principal Components Analysis (PCA), hierarchical clustering and self-organizing maps (SOM). Tissue was the main factor behind glucosinolates profile variation. Accession played a secondary role in glucosinolates profile variation. In terms of gene expression, tissue was also the main factor behind gene expression differences. Overall, correlation between biosynthetic genes and glucosinolates was not high, which means that there might be other factors that also influence the quantity of glucosinolates, such as transport processes, translational efficiency, protein stability and differential splicing. There is also a paralog of Methylthioalkylmalate synthase 3 (MAM3) that appears to make a difference in the side chain length (4C to 5C) of aliphatic glucosinolates between accessions.

Introduction

Glucosinolates

Glucosinolates (GLS) are beta-thioglucoside N-hydroxysulfate secondary metabolites that have a side chain (R) and a sulphur-linked beta-D-glucopyranose moiety, as shown in Figure 1 (Fahey *et al*, 2001; Mithen, 2001). These compounds can mainly be found in members of the *Brassicaceae* family, and in a limited number of other plant families, such as the *Caricaceae* and *Capparaceae* families (Fahey *et al*, 2001; Mithen, 2001)

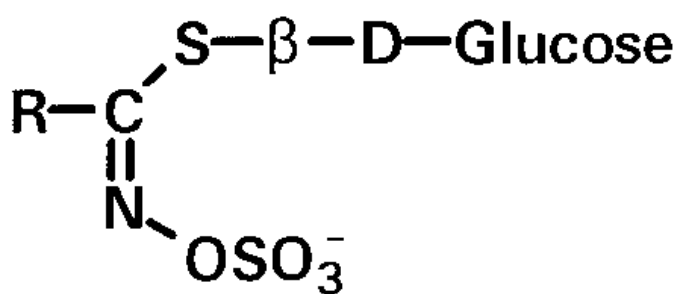


Figure 1. Glucosinolate core structure. Source: Mithen (2001)

According to the precursor amino acid and the side chain modifications, these compounds can be classified as aliphatic (AGLS, aliphatic glucosinolates), which are derived from alanine, leucine, isoleucine, methionine and valine, indolic (IGLS, indolic glucosinolates), which are derived from tryptophane, and aromatic (derived from phenylalanine or tyrosine) (Halkier and Gershenzon, 2006). These compounds have been associated with protections against herbivores in plants, and damage to plant tissues by herbivory will result in the hydrolysis of GLS to isothiocyanates, nitriles and thiocyanates by the action of myrosinases. These hydrolysis products are the bioactive compounds that confer protection against herbivores, as well as positive and negative nutritional effects

(Mithen, 2001). GLS and myrosinases coexist in the plant but they are separated from each other. Upon plant tissue disruption, such as leaves, GLS that are stored in cell vacuoles of GLS-rich cells called S-cells come into contact with myrosinases that are stored in immediately adjacent cells (Mithen, 2001; Koroleva *et al*, 2000).

Although GLS have been long known due to their presence in species from the *Brassicaceae* family, it is in the last 40 years that these compounds have gained higher relevance in agriculture. This is due to the increasing importance of rapeseeds (*Brassica rapa*) as oil and animal feed crops (Halkier and Gershenzon, 2006). GLS as a family of secondary metabolites have both nutritional and health-detrimental properties (Mithen, 2001). An example of a health-detrimental glucosinolate is progoitrin (2-hydroxy-3-butenyl glucosinolate), which is known to cause goiter. Therefore, plant breeders have reduced the content of the glucosinolate progoitrin because of its goiterogenic effects in the meal of oilseeds to allow the residue after crushing to be sold as animal feed supplement (Halkier & Gershenzon, 2006). However, there are GLS such as glucoraphanin (4-methylsulfinylbutyl glucosinolate) that, when hydrolyzed to sulforaphane by the action of myrosinase, have an effect upon the induction of Phase II enzymes. These enzymes have a chemoprotective effect, since they block chemical carcinogenesis, thus helping to prevent the onset of cancers (Shapiro *et al*, 1998; Talalay, 1999, Keum *et al*, 2004). Because of these reasons, there is renewed interest in recent years on research on biosynthesis and genetic regulation of GLS, even though the effect of GLS intake in food (Brassica vegetables) on cancer prevention might need to be further tested, since medical studies in this aspect do not always find an association between brassica vegetables consumption and cancer prevention, despite the chemopreventive effects of GLS (Murillo and Metha, 2001; Terry *et al*, 2001; Giovanucci *et al*, 2003; Yang *et al*, 2010) .

Brassica rapa

The wide growing range of *Brassica rapa* L. from the Mediterranean region to Central Asia probably helped it to become the first domesticated *Brassica* species as a multipurpose crop (turnip roots, seeds, young flowering shoots and leaves of heading and non-heading Chinese cabbage) (Gomez-Campo and Prakash, 1999).

Brassica rapa (A genome) is one of three main diploid species that form the U's triangle, alongside *Brassica nigra* (B genome) and *Brassica oleracea* (C genome) (U, 1935). Park *et al* (2005) demonstrated genome triplication in both the A and C genomes relative to the *A. thaliana* genome in *Brassica napus* through a genetic map based on markers derived from a bacterial artificial chromosome library from *A. thaliana*. Wang *et al* (2011) assembled the approximately 283.8 Mb genome, excluding repetitive sequences of *Brassica rapa* ssp. *pekinensis* (Chinese cabbage) line Chiifu-401-42 and confirmed the genome triplication event in *Brassica rapa*. Because of this, many genes in *Brassica rapa* appear in triplicate.

Turnip

Brassica rapa var *rapa* (turnip) is cultivated worldwide both as a vegetable and as animal fodder (Rakow, 2004). Lee *et al* (2013) evaluated the variation of GLS content in turnip tubers of forty-eight accessions, representing Asian, European, fodder and vegetable turnips. Eleven different GLS were

identified and monitored in turnip tubers by means of desulfo and liquid chromatography-mass spectrometry (LC-MS) for intact GLS analysis. Based on GLS profiles, the accessions were clustered, and four clusters of accessions were clearly defined based on their glucosinolate composition.

Glucosinolates biosynthetic pathway, genes and transcription factors involved

GLS biosynthesis is divided in three stages: First, side chains of aliphatic and aromatic amino acids such as methionine or phenylalanine are elongated by the addition of methyl groups. Second, a reconfiguration of the amino acid moiety itself gives rise to the core structure of GLS. Third, the formed GLS undergoes secondary modifications (Halkier and Gershenzon, 2006). These stages are described below, as well as in Figure 2.

Side chain elongation

The GLS biosynthesis process begins with the deamination of methionine or phenylalanine by the action of BCAT4 branched-chain amino-acid aminotransferase (Schuster, 2006), resulting in a 2-oxo acid. This oxo-acid enters an elongation cycle made of three consecutive transformations: Condensation with acetyl-CoA by a methylthioalkylsynthase (MAM), forming a substituted 2-malate derivative. Next, an isomerization by an isopropylmalate isomerase (IPMI) yields a 3-malate derivative which undergoes oxidative decarboxilation by the action of a isopropylmalate dehydrogenase (IPM-DH). This results in a 2-oxo-acid that is elongated by the addition of a methyl group (Sønderby *et al*, 2010a; Halkier and Gershenzon, 2006). At the end of the elongation cycle, the 2-oxo-acid can either undergo one or more rounds of elongation or be transaminated by BCAT3 branched-chain amino-acid aminotransferase to form the corresponding amino acid and then enter the core structure formation round of the pathway (Knill *et al*, 2008; Halkier and Gershenzon, 2006)

The cloning of the GS-ELONG-QTL by Kroymann *et al* (2001), which controls in the side chain length of aliphatic GLS, allowed the identification of genes MAM1, MAM2 and MAM3. MAM1 catalyzes the condensations in the first three elongation cycles, MAM2 only in the first one, and MAM3 in the production of all aliphatic GLS (Kroymann *et al*, 2001; Textor *et al*, 2007).

Core structure formation

Precursor amino acids with their elongated side chains are converted to aldoximes by cytochromes P450 of the CYP79 family (Sønderby *et al*, 2010a). CYP79B2 and CYP79B3 metabolize tryptophan, while CYP79A2 works with phenylalanine as substrate. CYP79F1 converts all chain-elongated Met derivatives and CYP79F2 converts only the long chained Met derivatives (Mikkelsen *et al*, 2000; Hull *et al*, 2000; Wittstock and Halkier, 2000 Hansen *et al*, 2001; Chen *et al*, 2003). Then, aldoximes are

oxidized to activated compounds by cytochromes P450 of the CYP83 family (Sønderby *et al*, 2010a). CYP83A1 converts aliphatic aldoximes whereas both tryptophan and phenylalanine aldoxime derivatives are converted by CYP83B1 (Naur *et al*, 2003, Bak *et al*, 2001; Bak and Feyereisen, 2001).

Once the activated aldoximes are conjugated to a sulfur donor, the resulting S-alkylthiohydroximates are converted to thiohydroximates by the C-S lyase SUR1 (Mikkelsen *et al*, 2004). Next, the thiohydroximates are S-glucosylated by glucosyltransferases of the UGT74 family (Sønderby *et al*, 2010a). UGT74B1 glucosylates aromatic thiohydroximates and UGT74C1 has an apparent role on the glucosylation of aliphatic GLS (Grubb *et al*, 2004; Grubb *et al*, 2014). After the glucosylation, the desulfo GLS are sulfated by the sulfotransferases SOT16 (Indolic and aromatic GLS), SOT17 and SOT 18 (aliphatic substrates) (Klein and Papenbrock, 2009).

Secondary modifications

The side chain of newly formed glucosinolate can be modified in many ways. Aliphatic GLS can be oxygenated, hydroxylated, alkylated and benzoylated, whereas indolic GLS can be methoxylated and hydroxylated (Halkier and Gershenzon, 2006, Sønderby *et al*, 2010a).

Kliebenstein *et al* (2001) identified GS-ELONG, GS-OX, GS-AOP and GS-OH as the responsible QTLs for the diversity of side chains of aliphatic GLS in Arabidopsis. S-oxygenation is most likely carried out by flavin monooxygenase FMO_{GS-OX1}, originated in the GS-OX locus (Hansen *et al*, 2007).

AOP2 and AOP3 genes are located within the GS-AOP QTL. AOP2 converts S-oxygenated GLS to alkenyl GLS, whereas AOP3 transforms S-oxygenated GLS to hydroxyalkyl GLS (Kliebenstein *et al*, 2001). Progoitrin, a glucosinolate known to have goiterogenic effects has its origin in the GS-OH locus. Its production is dependent on the presence of both AOP2 and MAM1 genes (Hansen *et al*, 2008, Sønderby *et al*, 2010a).

For indolic GLS, CYP81F2 is the gene associated to the 4-hydroxylation of I3M (indolyl-3-methyl glucosinolate), which is the precursor of the most commonly found indolic GLS in Arabidopsis (Pfalz *et al*, 2009)

mention that MYB76 plays a role in the spatial distribution of GLS in the leaf, thus having a possible transport role. MYB28 is the strongest regulator in the biosynthesis of aliphatic GLS, whereas MYB29 is an accessory part of this pathway (Hirai *et al*, 2007; Sønderby *et al*, 2007).

In the biosynthesis of IGLS, MYB51 has a central role in the regulation of this kind of GLS in leaves, where IGLS are mostly synthesized and accumulated. It is also the lead regulator of indolic GLS in shoots (Brown *et al*, 2003; Gigolashvili *et al*, 2008; Frerigmann and Gigolashvili, 2014). MYB34 is decisive for the biosynthesis of indolic GLS in roots, contributing in a lesser degree to the production of indolic GLS in shoots (Frerigmann and Gigolashvili, 2014). Although MYB122 does not significantly contribute to the production of indolic GLS, it has an accessory role in the regulation of indolic GLS upon environmental challenges, helping to lessen the decrease of indolic GLS upon these circumstances (Frerigmann and Gigolashvili, 2014).

Transport, synthesis and storage of GLS

The AGLS transporter genes GTR1 and GTR2 were identified and characterized in *Arabidopsis thaliana* in 2012 by Nour-Eldin *et al* as AT3G47960 (GTR1) and AT5G62680 (GTR2). Roots are the main organ of GLS accumulation until bolting (Andersen and Halkier, 2014; Petersen *et al*, 2002). Andersen and Halkier (2014) demonstrated that upon bolting, AGLS from both roots and leaves are transported to the seeds, and that the inflorescence represents a larger sink than roots at this developmental stage. Also, Andersen *et al* (2013), establish that AGLS that differ in chain length also differ in site of synthesis and storage. Short-chain AGLS are mainly synthesized and stored in the rosette, whereas long-chained aliphatic GLS are synthesized in both above and below organs, but are mainly retained in roots. With regard to IGLS, Andersen *et al* (2013) also establish that this type of GLS is synthesized in rosette and roots, using xylem and phloem for bidirectional distribution across the plant. GTR3 (AT1G18880) has been identified and characterized by Jorgensen *et al* (2014, unpublished) as an IGLS transporter.

Aim of the MSc thesis project

The objective of this project is to evaluate both the variation of GLS and gene expression profiles in different tissues and during development from seedling to flowering plant in two fodder turnip *B. rapa* ssp *rapa* accessions, as well as to correlate data for the relative abundance of GLS and activity of GLS biosynthetic genes. At the moment of starting the project, most of the data was already available, such as glucosinolate profiles of different tissues and developmental stages and qRT-PCR for the selected candidate biosynthesis and regulatory genes, but not for the GTRs. This is because GTR1, GTR2, and GTR3 transporter genes were identified after the data were generated. During this thesis project, the gene expression data for these genes was generated for at the same tissues and developmental stages as the rest of the gene expression data that are already available.

In order to fulfill the objective of this project, the following research questions need to be addressed:

Research questions

1. How do GLS profiles vary with tissue, developmental stage and between two different *B. rapa* accessions?
2. Can major regulatory genes for GLS profiles be identified?
3. Do these genes and paralogs act in a tissue specific way? Do these genes and paralogs act throughout all the developmental stages? Or is their action limited to a certain stage?

Materials and methods

Project background

In preparation for this project, tissue samples of two different turnip accessions (FT-004 and FT-086) that greatly differ in GLS profile according to the study of Lee *et al* (2013) were collected at seven time points during plant development between July 2010 and June 2011. Information about these accessions is shown in Table 1.

There were three blocks with 25 – 30 plants per accession and block. The seeds were sown on the 7th of July, 2010. The first transplant to jiffy pots was on the 12th of July, 2010 and the second transplant to bigger pots was on the 28th of July, 2010. On each sampling date for GLS, 1 plant per block was randomly selected yielding a total of three biological replicates per measurement. Sampling dates for each tissue are detailed in Table 2. Depictions of some developmental stages are shown in Figure 3.

Table 1. *B. rapa* accessions used in the project (Zhang *et al*, 2014)

Accession	Cultivar name	Collection	Genebank ID	Country of origin
FT-004	Lange Gele Bortfelder	Wageningen University	CGN06678	Denmark
FT-086		Wageningen University	CGN07223	Pakistan

The tissue samples collected included seedling, young leaves, mature leaves, inner turnip (inner tissue), outer turnip (outermost turnip layer) and in the later stages, tissue samples of stem, flower and seeds. Both GLS profiles and quantity and the expression of a selection of the genes that act on the GLS biosynthesis pathway was measured by means of LC-MS and quantitative real time polymerase chain reaction (qRT-PCR), respectively.



Figure 3. Some of the developmental stages of the *B. rapa* accessions in which gene expression and GLS profile studies were carried out. T1 (26-Jul-10): Seedling. T2 (16-Aug-10): Turnip formation. T3(30-Aug-10): Turnip formation. T4 (16-Sep-10): Not pictured. T5 (29-Sep-10) : Not pictured. T6 (5-Apr-11): Fully developed plant. This stage is the most interesting point for evaluation of gene expression and GLS content, since all tissues are fully developed. T7 (21-Jun-11): Turnip evaluation only. Source: Wageningen University Plant Breeding Brassica Group.

Table 2. Tissues and time points sampled for GLS presence in this project. Selected samples for gene expression are marked with X.

Accession	Time point	Tissue	Sampling date	Gene expression
FT-004	T1	seedling	26-Jul-10	X
FT-004	T2	young leaves	16-Aug-10	X
FT-004	T2	mature leaves	16-Aug-10	
FT-004	T2	inner turnip	16-Aug-10	X
FT-004	T2	outer turnip	16-Aug-10	
FT-004	T3	young leaves	30-Aug-10	
FT-004	T3	mature leaves	30-Aug-10	
FT-004	T3	inner turnip	30-Aug-10	X
FT-004	T3	outer turnip	30-Aug-10	
FT-004	T4	young leaves	16-Sep-10	
FT-004	T4	mature leaves	16-Sep-10	X
FT-004	T4	inner turnip	16-Sep-10	
FT-004	T4	outer turnip	16-Sep-10	
FT-004	T5	young leaves	29-Sep-10	
FT-004	T5	mature leaves	29-Sep-10	
FT-004	T5	inner turnip	29-Sep-10	
FT-004	T5	outer turnip	29-Sep-10	
FT-004	T6	young leaves	05-Apr-11	X
FT-004	T6	mature leaves	05-Apr-11	X
FT-004	T6	stem	05-Apr-11	X
FT-004	T6	flower	05-Apr-11	X
FT-004	T6	seed	05-Apr-11	X

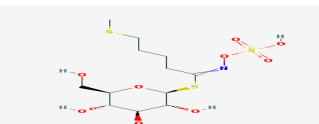
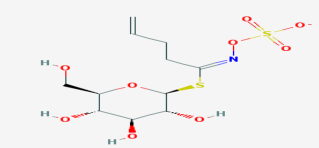
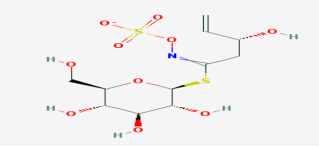
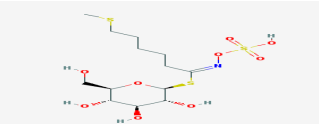
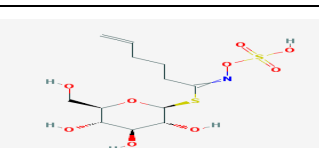
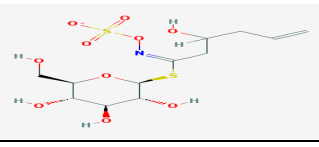
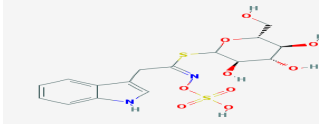
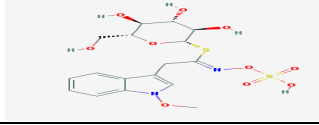
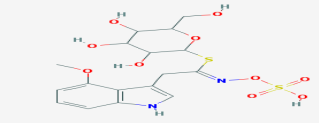
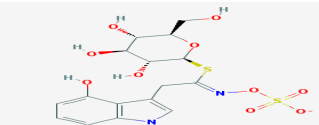
FT-004	T6	inner turnip	05-Apr-11	X
FT-004	T6	outer turnip	05-Apr-11	
FT-004	T7	inner turnip	21-Jun-11	
FT-004	T7	outer turnip	21-Jun-11	
FT-086	T1	seedling	26-Jul-10	X
FT-086	T2	young leaves	16-Aug-10	X
FT-086	T2	mature leaves	16-Aug-10	
FT-086	T2	inner turnip	16-Aug-10	X
FT-086	T2	outer turnip	16-Aug-10	
FT-086	T3	young leaves	30-Aug-10	
FT-086	T3	mature leaves	30-Aug-10	
FT-086	T3	inner turnip	30-Aug-10	X
FT-086	T3	outer turnip	30-Aug-10	
FT-086	T4	young leaves	16-Sep-10	
FT-086	T4	mature leaves	16-Sep-10	X
FT-086	T4	inner turnip	16-Sep-10	
FT-086	T4	outer turnip	16-Sep-10	
FT-086	T5	young leaves	29-Sep-10	
FT-086	T5	mature leaves	29-Sep-10	
FT-086	T5	inner turnip	29-Sep-10	
FT-086	T5	outer turnip	29-Sep-10	
FT-086	T6	young leaves	05-Apr-11	X
FT-086	T6	mature leaves	05-Apr-11	X
FT-086	T6	stem	05-Apr-11	X
FT-086	T6	flower	05-Apr-11	X
FT-086	T6	seed	05-Apr-11	X
FT-086	T6	inner turnip	05-Apr-11	X
FT-086	T6	outer turnip	05-Apr-11	
FT-086	T7	inner turnip	21-Jun-11	
FT-086	T7	outer turnip	21-Jun-11	

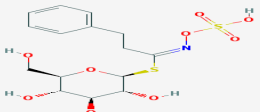
LC-MS data

The data for the relative peak area (RPA) for the GLS outlined in

Table 3 were collected by visiting scientist Dr. Jun Gu Lee in 2011. Table 3 presents the detected GLS by means of LC-MS. These GLS make up the GLS profile that will be discussed in this project. The information on the relative peak area for the detected GLS was gathered from the tissues detailed in Table 2. From each tissue, three biological replicates were taken, and the resulting relative peak area value is the average of these replicates

Table 3. GLS in study, group and molecular structure. Molecular structure image source: PubChem (2015)

Trivial name	Name	Abbreviation	Group	Structure
Glucoerucin	4-methylthiobutyl glucosinolate	ERU	Aliphatic	
Gluconapin	3-butenyl glucosinolate	NAP	Aliphatic	
Progoitrin	2-Hydroxy-3-butenyl glucosinolate	PRO	Aliphatic	
Glucoberteroin	5-Methylthiopentyl glucosinolate	BER	Aliphatic	
Glucobrassicinapin	pent-4-enylglucosinolate	CAN	Aliphatic	
Gluconapoleiferin	2-hydroxy-4-pentenylglucosinolate	NAPOL	Aliphatic	
Glucobrassicin	3-Indolylmethyl glucosinolate	BRA	Indolic	
Neoglucobrassicin	1-Methoxy-3-indolylmethyl glucosinolate	NBRA	Indolic	
4-methoxyglucobrassicin	4-methoxyglucobrassicin	4MBRA	Indolic	
4-hydroxyglucobrassicin	4-Hydroxy-3-indolylmethyl glucosinolate	4HBRA	Indolic	

Gluconasturtiin	Phenethylglucosinolate	NAS	Aromatic	
-----------------	------------------------	-----	----------	--

Candidate gene selection criteria

Previously, the quantitative expression of selected genes in the GLS biosynthetic pathway was determined in a selected subset of the samples that were taken for GLS determination as shown in Table 4.

The selection of candidate genes was as follows: the gene must act in a metabolite-type specific way. That is to say, the gene is related to a specific type of metabolite (aliphatic, indolic or phenyl), or also to a certain group of individual GLS. Another important criterion is that a gene should be key for a particular stage in the biosynthetic process (side chain elongation, core structure formation and secondary modifications). Another criterion is that some of these genes are identified as candidate genes in QTL papers. Transcription factors were also selected. In addition, GTR1, GTR2, and GTR3 transporter genes have been considered for the gene expression study, given their role in the bidirectional transport of aliphatic GLS through xylem and phloem. The gene expression data was generated by another student (S. Wang) in 2012 by means of quantitative real-time polymerase chain reaction (qRT-PCR). cDNA was extracted from the tissues marked as selected for gene expression in Table 2. Three technical replicates were made for each accession-tissue-timepoint combination so that relative gene expression could be measured. In the case of GTR genes, two technical replicates were considered.

At the beginning of the qRT-PCR experiments, AOP3 was also characterized as BRA018521. However, Wang et al (2011), found that AOP3 is not expressed in *B. rapa*. Therefore, the data for this paralog was put under the code name AOP2_2_1, and was not further considered into the study.

Table 4. Previously selected candidate genes and selection criteria

Gene	Role	Pathway	<i>B. rapa</i> paralog	Paralog code	Pathway stage
AOP2	Catalyses the conversion of S-oxygenated GLS to alkenyl GLS	Aliphatic	BRA000848 BRA018521 BRA034180	AOP2_1 AOP2_2 AOP2_3	Secondary modifications
CYP79A2	Conversion to aldoximes of phenyl compounds	Phenyl	BRA028764	CYP79A2	Core structure formation
CYP83A1	Aliphatic aldoxymes conversion	Aliphatic	BRA016908 BRA032734	CYP83A1_1 CYP83A1_2	Core structure formation
CYP83B1	Metabolizes both indolic and phenyl acetaldoximes	Indolic	BRA034941	CYP83B1	Core structure formation
GS-OH	Responsible for progoitrin biosynthesis	Aliphatic	BRA022920 BRA021670 BRA021671	GS-OH_1 GS-OH_2 GS-OH_3	Secondary modifications
MAM1	Catalyses condensation of 3 elongation cycles	Aliphatic	BRA029355 BRA013007 BRA018524	MAM1_1 MAM1_2 MAM1_3	Side chain elongation
MAM3	Catalyses condensation of 6 elongation cycles	Aliphatic	BRA013011 BRA013009 BRA021947 BRA029356	MAM3_1 MAM3_2 MAM3_3 MAM3_4	Side chain elongation
MYB28	Transcription factor	Aliphatic	BRA012961 BRA035929	MYB28_1 MYB28_2	
MYB29	Transcription factor	Aliphatic	BRA005949	MYB29	
MYB34	Transcription factor	Indolic	BRA013000 BRA029349 BRA029350 BRA035954	MYB34_1 MYB34_2 MYB34_3 MYB34_4	
MYB51	Transcription factor	Indolic	BRA016553 BRA025666 BRA031035	MYB51_1 MYB51_2 MYB51_3	
GTR1	Aliphatic GLS transporter	Aliphatic	BRA018096 BRA033782 BRA019521	GTR1_1 GTR1_2 GTR1_3	
GTR2	Aliphatic GLS transporter	Aliphatic	BRA010111 BRA029248 BRA035885 BRA035886	GTR2_1 GTR2_2 GTR2_3 GTR2_4	
GTR3	Indolic GLS transporter	Indolic	BRA025695 BRA016534 BRA031054	GTR3_1 GTR3_2 GTR3_3	

qRT-PCR of Br-GTR paralogs

Primers design

The coding sequences of the paralogs of GTR1, GTR2, and GTR3 were downloaded in FASTA format from the *B.rapa* database BRAD website using the *B. rapa* genome version 1.5 (Cheng *et al*, 2011). Then, the primers were designed in the Primer3Plus website (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>), selecting qPCR in special settings. The main selection criteria for forward and reverse primer combinations were, besides uniqueness in the genome, that the difference of melting temperature (T_m) was no larger than 0.8 Celsius degrees between forward and reverse primer. The targeted T_m was of 60 Celsius degrees. Another selection criterion was to choose primer combinations with a product size between 80 and 191 bp. Next, the primers were blasted to the *B. rapa* genome using the BLAST tool available at the website so the uniqueness of the primers sequences in the genome could be checked. The obtained primers are presented in Table 5.

The primers for all GTR2 paralogs for *B. Rapa* were kindly provided by Dr. Hussam Hassan Nour-Eldin (University of Copenhagen, Denmark) (2014, unpublished).

Once the primers were checked, these were synthesized (Biolegio BV, Nijmegen, The Netherlands) and then diluted in a 100µM stock solution. Working solutions had a concentration of 10 µM.

Primers list

Table 5. List of primers used in the qRT-PCR experiment

Genes	Forward primer	Reverse primer
GTR1_BRA018096	5'-GCAAAACAGACTCTAGGGATGG-3'	5'-CTTCAGCTATCCCTGCAAGAAC-3'
GTR1_BRA033782	5'-TCCCGAGAGTAGGTTTGTGAG-3'	5'-ATTCTACGACCGTGCCTTGTC-3'
GTR1_BRA019521	5'-TCCGACTTTACCTAAGCTGGTC-3'	5'-GAAAGCCTCTGGTAGTCCATTG-3'
GTR2_BRA029248	5'-ACGTGGGTAAAGGCTGTGAAGC-3'	5'-AACCTGAACTGGTCGGTGTACTTG-3'
GTR2_BRA010111	5'-ATGTGGCTGATCCCAACTCG-3'	5'-TGCCTCGTCCTGTGAACCATTG-3'
GTR2_BRA035885	5'-AGGTTAGAGCCAGTGGAACAG-3'	5'-TGTCGAGAAACCTGAATTGTTGGG-3'
GTR2_BRA035886	5'-CGAGGAGCGTAGGAGAACTTTAGC-3'	5'-ATCTGCCCAATAGCTCCAATGCG-3'
GTR3_BRA025695	5'-CATAACACGCGTCATCGTAGTC-3'	5'-GTATGGCTCAGTTTCGAGTTCC-3'
GTR3_BRA016534	5'-GAGGAGGTTTCGGACTTAAGGTTAG-3'	5'-CTCTATCAACGGTTACGTCACTAGC-3'
GTR3_BRA031054	5'-CCCATCAGCAGTAAGTGAAGTATC-3'	5'-ACACAGAAGAACGGTAGCTCTCAC-3'

Protocols

qRT-PCR experiments were carried out for GTR1, GTR2, and GTR3 in 96-well plates. One technical replicate was done per paralog. In each well, 8 µl of the following master mix were pipeted:

- 5 µl SYBR mix
- 2.4 µl of Milli-Q water
- 0.3 µl of forward primer
- 0.3 µl of reverse primer

Then, 2 µl of the cDNA corresponding for a given accession-tissue-time point combination were added to the well.

The total volume of reaction mix was 10 µl added on each well. Before running the plates in the qRT-PCR machine, the plates were sealed and centrifuged at 600 rpm in a Heraeus Multifuge 3 s-r centrifuge. After being centrifuged, the qRT-PCR reaction was carried out in a combined Bio-Rad CFX 96 Real-Time System and C1000 Thermal Cycler (Bio-Rad, CA, USA) using the standard qRT-PCR amplification protocol available in this machine, which is presented in Figure 4.

1	2	3	4	5	6	7	
95.0 °C 03:00	95.0 °C 00:15	60.0 °C 01:00	G O T O	95.0 °C 00:10	65.0 °C 00:05	10.0 °C Forever	E N D
	←		2 39 X				

Figure 4. qRT-PCR amplification protocol.

Actin was used as reference gene for the qRT-PCR experiments. The data output for the gene expression of each paralog was in the format of the average of the $-\Delta Ct$ value for all technical replicates. That is to say, it is the negative of the difference between the expressions of the gene of interest versus the expression of the reference gene Actin.

$$\Delta Ct = Ct \text{ Target gene} - Ct \text{ Reference gene}$$

The PCR plate layout is detailed in the Appendix 2

Data analysis

Data pre-processing

Once the datasets for both GLS RPA and RGE ($-\Delta Ct$) were generated, both datasets were transformed by means of centering around the data series mean and range scaling according to van den Berg *et al* (2006) using Microsoft Excel 2010.

The definition of centering in combination with range scaling is as follows:

$$\sim X_{ij} = \frac{X_{ij} - \overline{X_i}}{(X_{i_{max}} - X_{i_{min}})}$$

Where:

$\sim X_{ij}$ = range-scaling transformed data point

X_{ij} = data point

$\overline{X_i}$ = mean of all data points in the series

$X_{i_{max}}$ = highest value in the data series

$X_{i_{min}}$ = lowest value in the data series

In order to work with metabolite intensities, Smilde *et al* (2005) define range scaling as dividing each measurement by the range of those measurements over all samples. This allows comparing metabolites relative to the biological response range (van den Berg, 2006). According to Smilde (2005) this method will allow to remove instrumental response factors from the data, obtaining

relative concentration for each variable. An important property of range scaling is the equal treatment of all levels of metabolite variation and scaling is related to biology (van den Berg *et al*, 2006). A possible disadvantage is that the range is determined by only two individual values (minimum and maximum among all) which are also the ones influenced most by outliers in the data.

Upon range-scaling transformation, both GLS and gene expression datasets were imported into GeneMaths XT version 2.12, build 2 (Applied Maths, Sint-Martens-Latem, Belgium) for Principal Components analysis (PCA), hierarchical clustering analysis (HCA) and self-organizing maps (SOM) cluster analysis.

Principal components analysis

Once the transformed datasets were imported into GeneMaths, a Principal Components Analysis (PCA) was carried out in order to get an idea on which factors are driving the variation in GLS and gene expression profiles. The accession-tissue-timepoint combinations were grouped by accession, tissue and timepoint. The PCA analyses were based upon accession-tissue-time point combinations of the samples.

For the PCA of GLS, every point in the score plot represents the average of three biological replicates and each point in the loading plot represents the individual contribution of a particular GLS to the principal components. In the case of the GLS biosynthetic genes PCA, every point in the score plot is the mean of the relative gene expression ($-\Delta Ct$) of three technical replicates and every point in the loading plot is the contribution of the relative gene expression of each GLS biosynthetic gene paralog to the score.

Hierarchical clustering

To detect clusters of high or low gene activity per accession, as well as tissue and time point, the data was clustered in both rows and columns using the dendrogram option available in GeneMaths. No data filters were used. The similarity criteria used was Pearson correlation and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) was used as the method to summarize clustered entries. The distance between clusters is measured as the distance between the averages over the elements of each cluster.

Self-organizing maps

In order to evaluate the similarity of paralogs expression, as well as to get an overview of possible co-regulatory schemes of GLS biosynthetic genes and paralogs, self-organizing maps (SOMs) were generated for both gene expression and GLS relative peak area datasets using the available option in GeneMaths. No data filters were used. The dissimilarity criteria used was linearly scaled Euclidean distance. Three grids were used for each dataset. The larger the grid, the more clusters will be formed, and these clusters will tend to be smaller. The chosen grid sizes for the clusters were 3 x 3, 4 x 4, and 5 x 5 for the gene expression data, and of 2 x 2, 3 x 3, and 4 x 4 for the GLS data.

Correlation matrix

With the purpose of correlating gene expression to GLS relative presence, a correlation matrix was generated in GenStat 16th Edition (VSN International, Hemel Hempstead, UK) using the BICORRELATE procedure included in the Biometris GenStat Procedure Library Edition 16 (Wageningen, The

Netherlands). This procedure calculates Pearson correlation coefficients between two variables including as many values as possible, leaving out only those samples for which there is a missing value for one of the two included variables (but does not leave out samples that have missing values for other variables than these two). The significance (using a two-sided t-test) for each correlation coefficient was also obtained.. All significant correlations taken into consideration in the Results section are at the 0.05 level. Only significant correlations with $r > 0.399$ were considered for the Results section.

Other statistical tests

Any other statistical tests, such as t-test and Wilcoxon rank-sum test were carried upon the raw data to detect differences between accessions in terms of GLS presence or gene expression in GenStat 16th Edition. As the LC-MS data is made of averages of biological replicates, all replicates were considered as independent measurements in order to have more statistical power, having a total of $n = 156$ samples ($n = 78$ samples per accession). For gene expression, the total number of accession-tissue-timepoint combinations was 22 (average of three technical replicates per accession-tissue-timepoint combination) differences are deemed as significant at $p = 0.05$. Assumption verifications for t-test and were performed by means of Shapiro-Wilk test for normality and homogeneity of variances was verified by means of Bartlett's test. In the case of non-normality or unequal variances, Wilcoxon rank sum test was used. This test had to be used to detect significant differences between accessions in terms of GLS presence, since the distribution for several GLS was non-normal. The information related to these tests can be found in Appendix 3.

Relative gene expression (RGE)

In order to visualize the fold change in gene expression across tissues and time points, that is to say, the relative gene expression (RGE) the $2^{-\Delta\Delta Ct}$ method developed by Livak and Schmittgen (2001) was applied to the dataset.

$$RGE = 2^{-((Ct_{Target} - Ct_{ActinT}) - (Ct_{Control} - Ct_{ActinC}))}$$

Where:

CtTarget: Number of DNA doubling cycles required for the target gene to pass the gene expression detection threshold

CtActinT: Number of DNA doubling cycles required for the actin gene to pass the gene expression detection threshold in the same sample as the target gene

CtControl: Number of DNA doubling cycles required for the same gene in the control sample to pass the gene expression detection threshold in the same sample as the target gene

CtActinC: Number of DNA doubling cycles required for the actin gene to pass the gene expression detection threshold in the control sample

In this case, the selected control for each paralog was the gene expression in seedlings of FT-004 (T1-seedling-FT-004). This information is displayed next to each hierarchical clustering heat map as complementary information.

Results

Background

To improve readability of the Result and Discussion sections, Figure 5 shows an overview of the GLS and genes involved in selected steps of the GLS biosynthetic pathway.

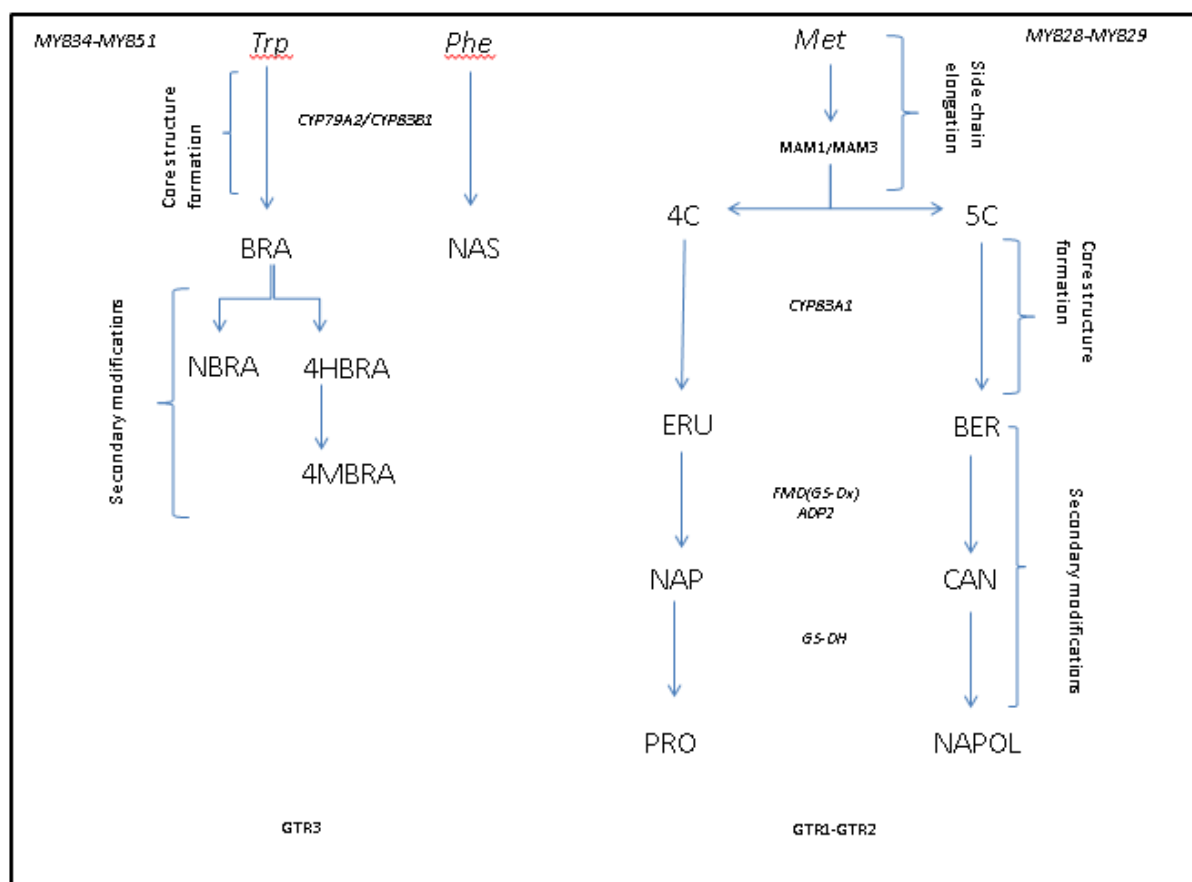


Figure 5. Simplification of the GLS biosynthetic pathway with GLS and genes evaluated in this study. In uppercase: GLS. In Italic: Genes involved in each step. Bottom genes: GLS transporter genes.

GLS data evaluation

An evaluation on the consistency of the data was performed by calculating the Coefficient of Variation (CV) of the three biological replicates that make each measurement of RPA for each GLS. In 52 accession-tissue-timepoint combinations, each made of 3 biological repeats per tissue, 11 GLS were measured, yielding a total of 572 total measurements of averages. From these 572 measurements, only 59 (10.3%) had a CV over 1, which speaks about a reasonable reproducibility of the results, thus reassuring the good quality of the data.

GLS profiles

Figures 6 to 9 show the profile per GLS. GLS were grouped by profile similarity per accession, tissue, and timepoint. In Figure 6, it is shown that the AGLS ERU, BER and NAS, are mainly present in turnip tissues, and absent (ERU and BER) or low in leaves and flowers and seeds. IGLS are more present in young leaves and outer turnips.

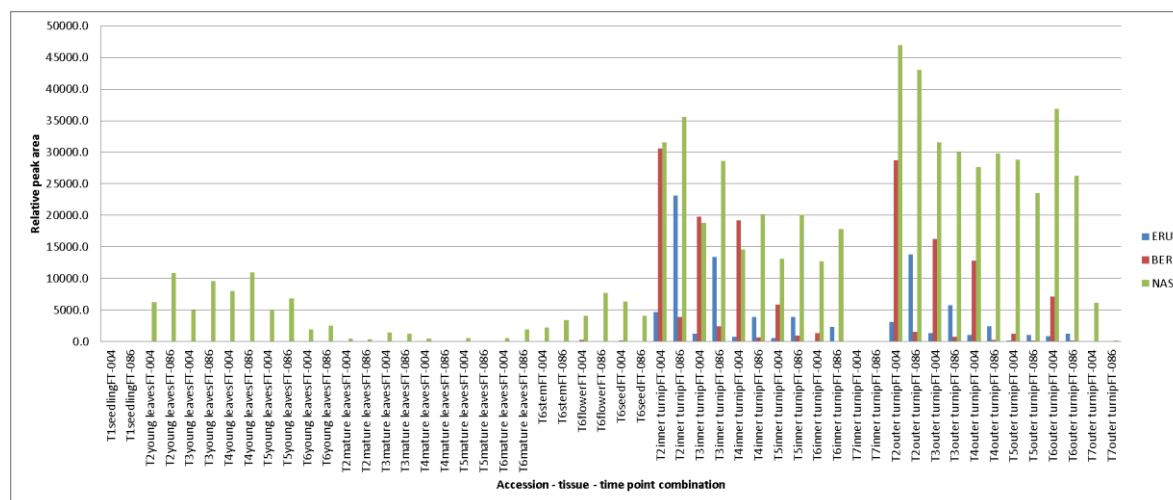


Figure 6. Distribution profile for ERU, BER, and NAS across accessions, tissues, and time points.

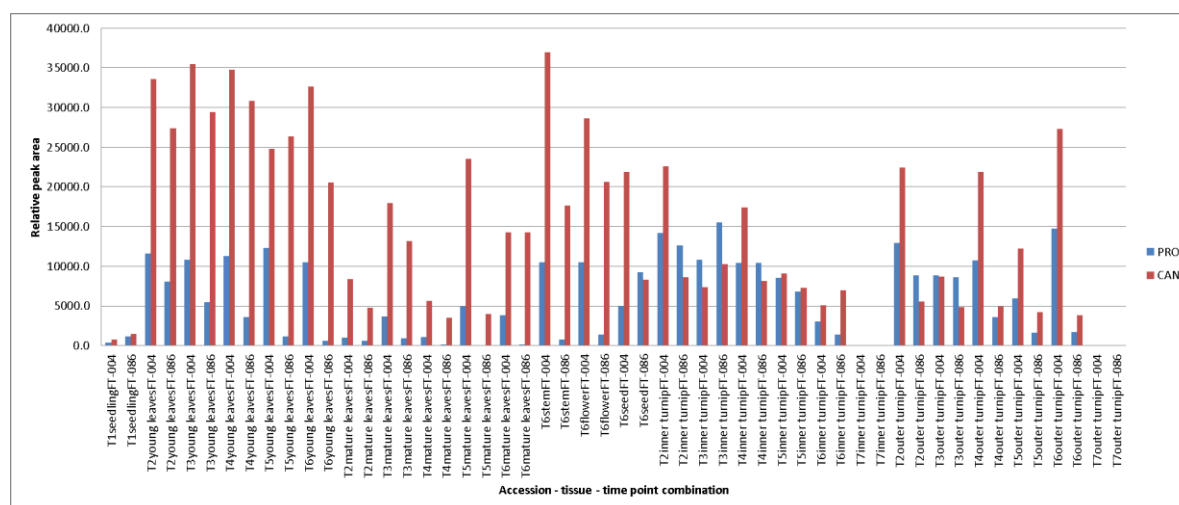


Figure 7. Distribution profile for PRO and CAN across accessions, tissues, and time points.

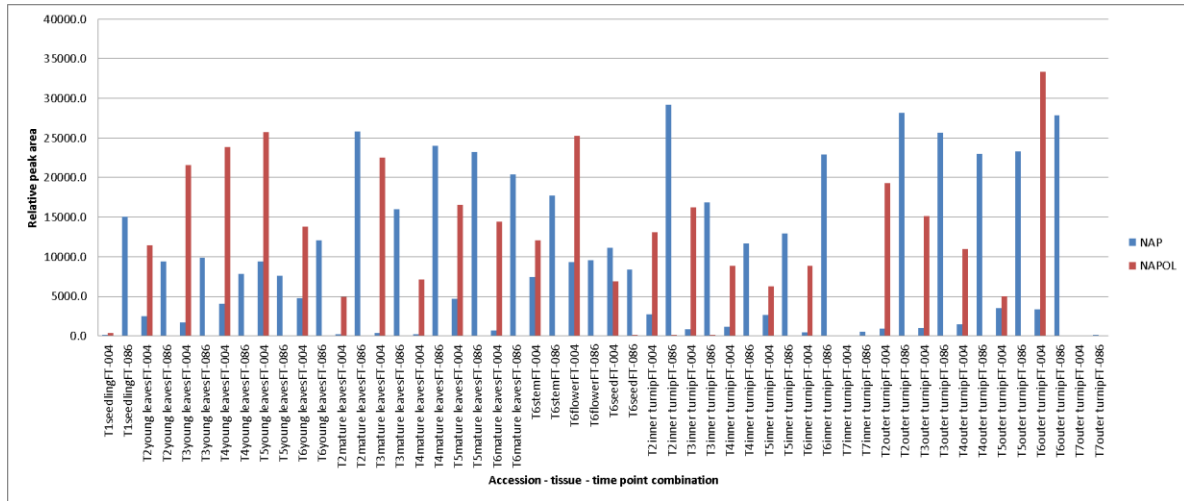


Figure 8. Distribution profile for NAP and NAPOL across accessions, tissues, and time points.

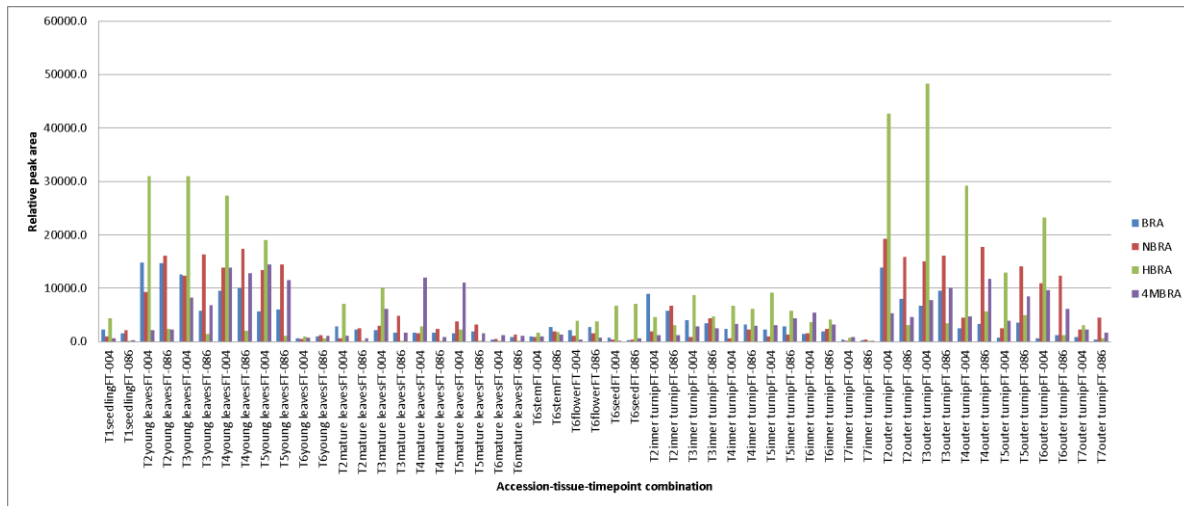


Figure 9. Distribution profile for BRA, NBRA, 4HBRA and 4MBRA across accessions, tissues, and time points.

Principal components analysis (PCA)

PCA for GLS

A PCA for the GLS was generated from the range-scaled data. Apparently, the variation along the first principal component PC1 (explaining 46% of the variation) is driven mainly by tissue. There is also a dispersion of the accessions along the second principal component axis (PC2), which explains 19.8% of the variation. In the direction of PC1, the turnip tissues belonging to T7 do not cluster together with the rest of the tissues (Figure 10). T7 is the harvested turnip after bolting and seed set, so it is also possible that the lower biological activity related to this developmental stage is driving the variation. To see whether tissue is the main factor behind changes in the GLS profile, another PCA analysis without inner and outer turnip from T7 was done (Figure 11). In Figure 11, it can be seen that although PC1 explains slightly less variation (44 %), tissue still clearly varies along this

principal component. Nevertheless, In Figure 11 PC2 explains more variation (24.8 %), which separates samples in terms of accession. However, this should be expected, given that the two *B. rapa* accessions for this study were selected based on differences in their GLS profiles. Table 6 shows GLS that are significantly more abundant in one accession than in the other.

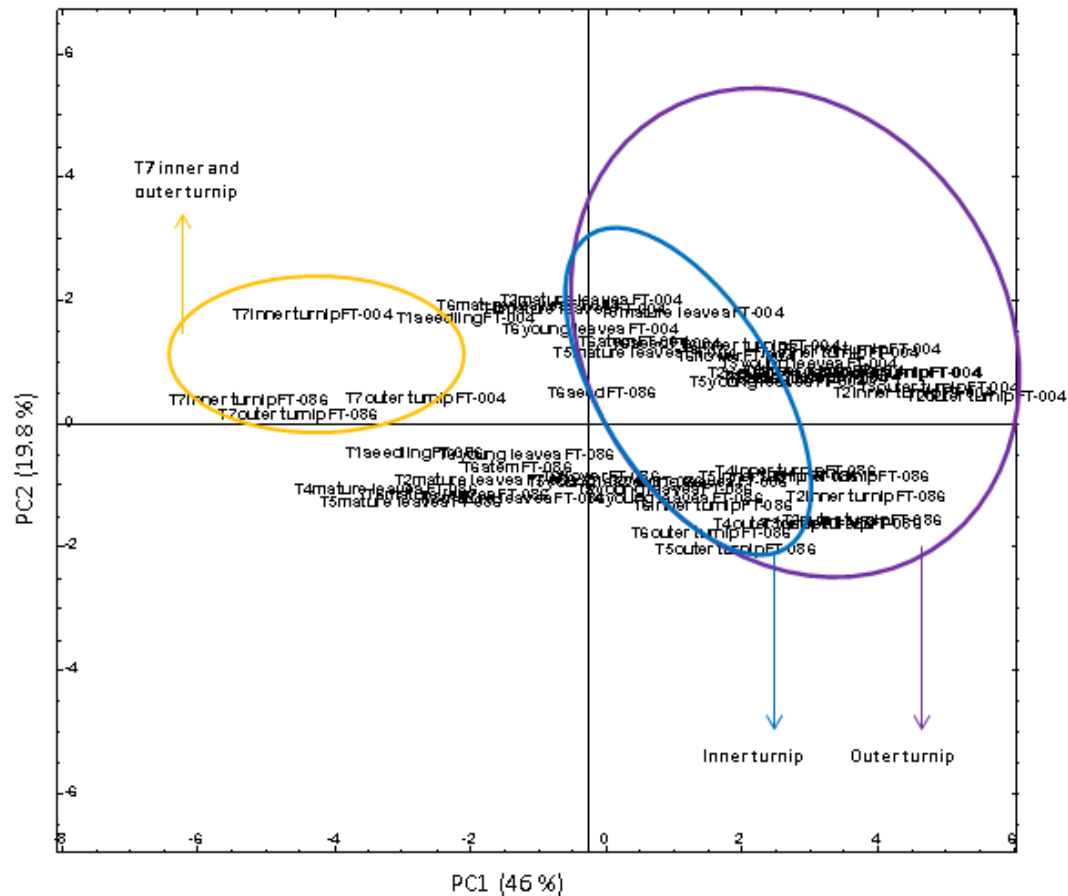


Figure 10. PCA score plot of GLS including turnips in T7. T7 clusters separately from the rest of the tissues.

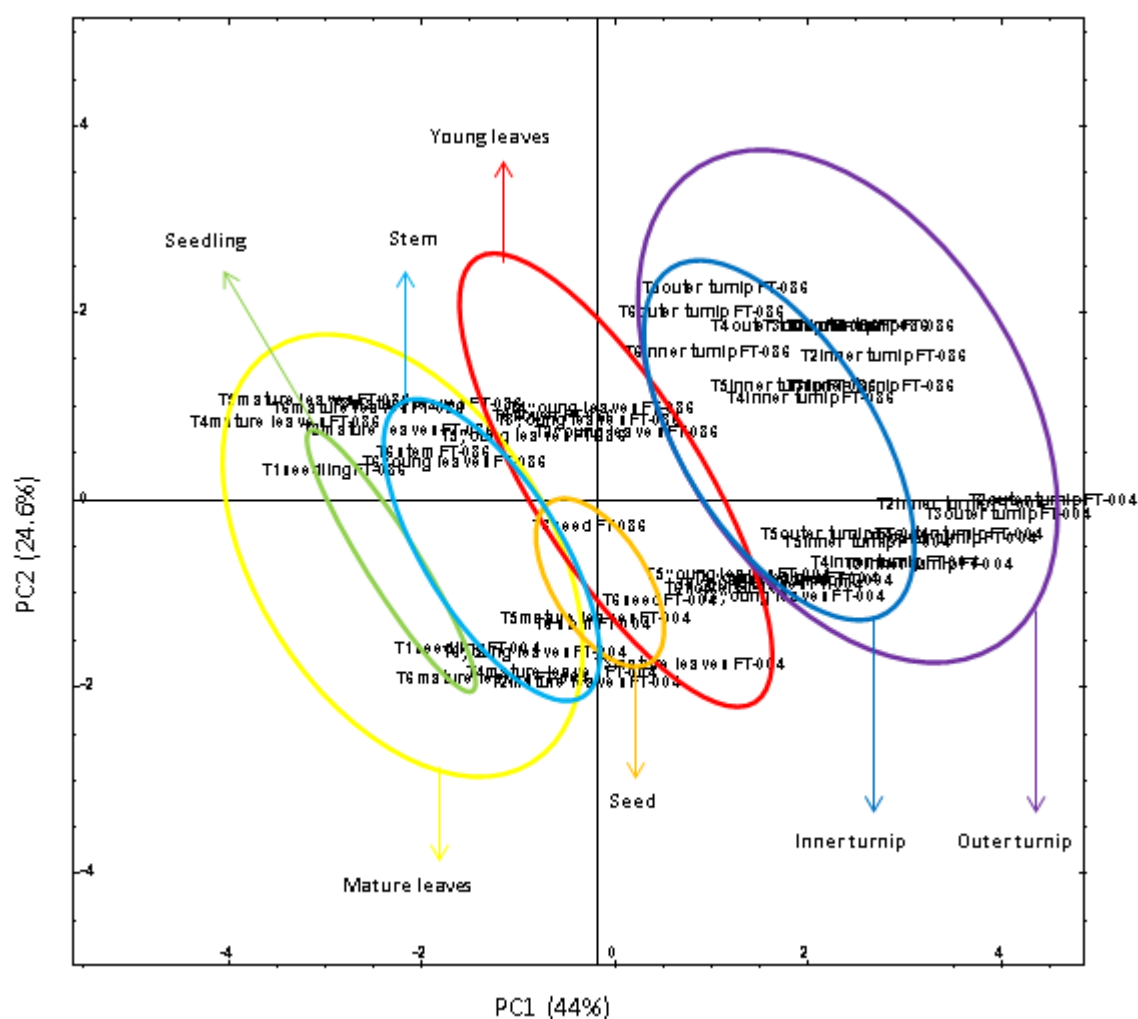


Figure 11. PCA score plot for GLS without T7 turnip tissues. Data points distribute along the main principal component are distributed in a tissue-wise manner.

Looking at the PCA plot from a broader grouping perspective, that is to say, viewing young and mature leaves together as leaves and inner and outer turnip as turnip, it is even more evident that there is a difference in GLS profile between the turnips and the above ground part of the plant (Figure 12). ERU, BER, NAS are grouped in the turnip area, 4HBRA is in the middle, and the AGLS NAP, NAPOL, CAN, PRO, as well as the IGLS BRA, NBRA, tend to be more related to the above ground part of the plant.

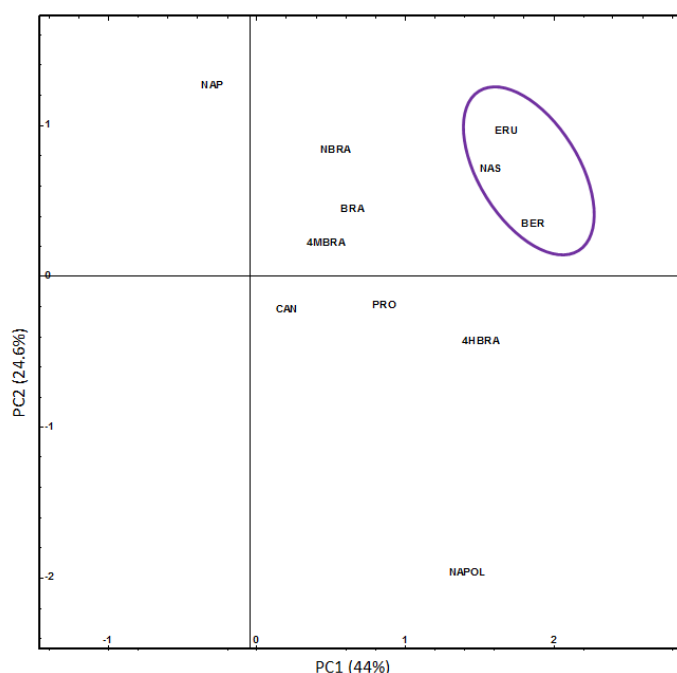


Figure 12. PCA loading plot with GLS distribution across the PCA plot. Encircled are ERU, NAS, and BER, which seem to be related to turnip.

Table 6. Differences in relative presence of GLS between accessions FT-004 and FT-086 (Wilcoxon Rank-Sum Test, n=156)

GLS	Accession with higher overall relative presence	Significance (p-value)
4HBRA	FT-004	<0.001
BER	FT-004	<0.001
CAN	FT-004	<0.001
NAP	FT-086	<0.001
NAPOL	FT-004	<0.001
NBRA	FT-086	<0.001
PRO	FT-004	<0.001

PCA for genes

A second PCA was carried out for the gene expression dataset on a subset of the samples which also was used to measure GLS (Figures 13 and 14). PC1 explains 40% of the total variation, whereas PC2 accounts for 15.5% of the variation. It can be seen in Figure 13 that the variation in gene expression profile is driven by tissue across PC1, although this variation is not as marked as in the PCA for GLS. Interestingly, it can be seen that the seed tissues for both accessions are clustered together within the inner turnip cluster accession. It is possible that their condition of sink tissues is playing a role in their gene expression profile. It is also observed that there is dispersion in terms of time points along the Y axis. Here, all tissues belonging to T6 but seeds are above the Y axis, whereas tissues belonging to other time points are below the X axis. This suggests there might be a change in the gene expression profiles once when this developmental stage takes place. Unlike in the GLS PCA analysis, no separation between accessions was observed, even when looking at the third and subsequent principal components.

Figure 14 shows the loading plot of the PCA analysis for genes. Here, genes are tightly clustered together in the plot area homologous to the area where leaves are clustered in the score plot. Likewise, MAM1, MAM3, and CYP83A1 paralogs, all of them related with the biosynthesis of AGLS, are also clustered in that area. MYB28 paralogs are clustered together, whereas MYB29 is separate in the plot, expressed in an area occupied by turnip tissues in the PCA plot. In this area, there are mostly GTR, GS-OH, and MYB34 paralogs. MYB34 paralogs, with the exception of MYB34-3 are clustered together in the area related to turnip tissues.

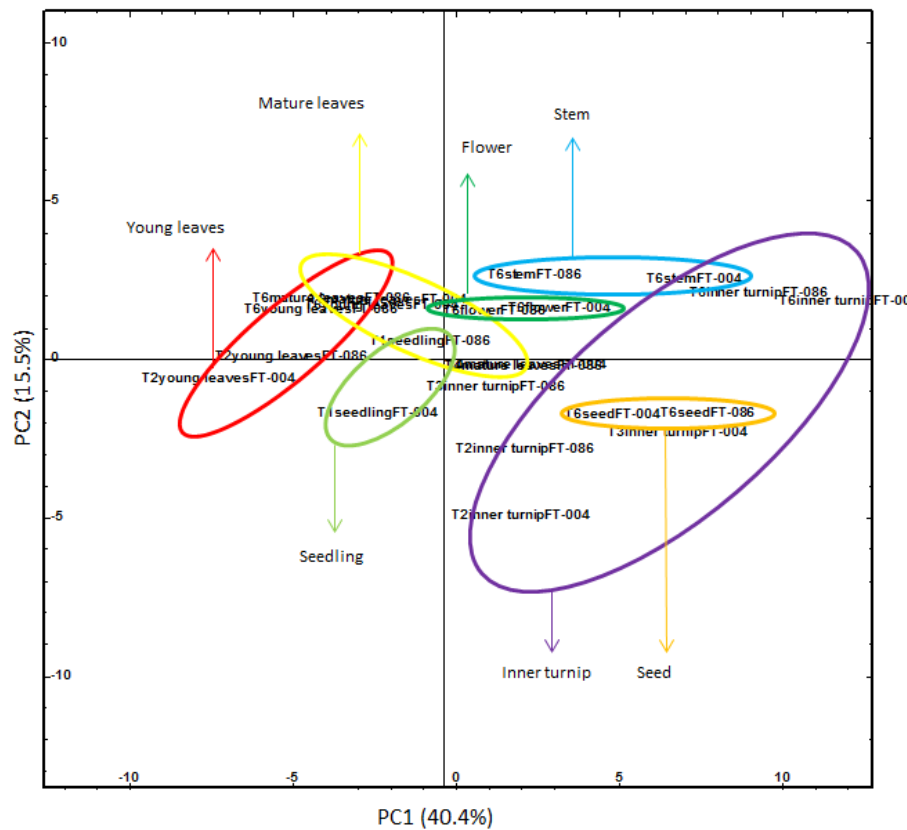


Figure 13. PCA score plot for GLS biosynthetic genes obtained by analysing the relative gene expression ($-\Delta Ct$) of the average of three technical replicates of every sample mentioned in Table 4 .

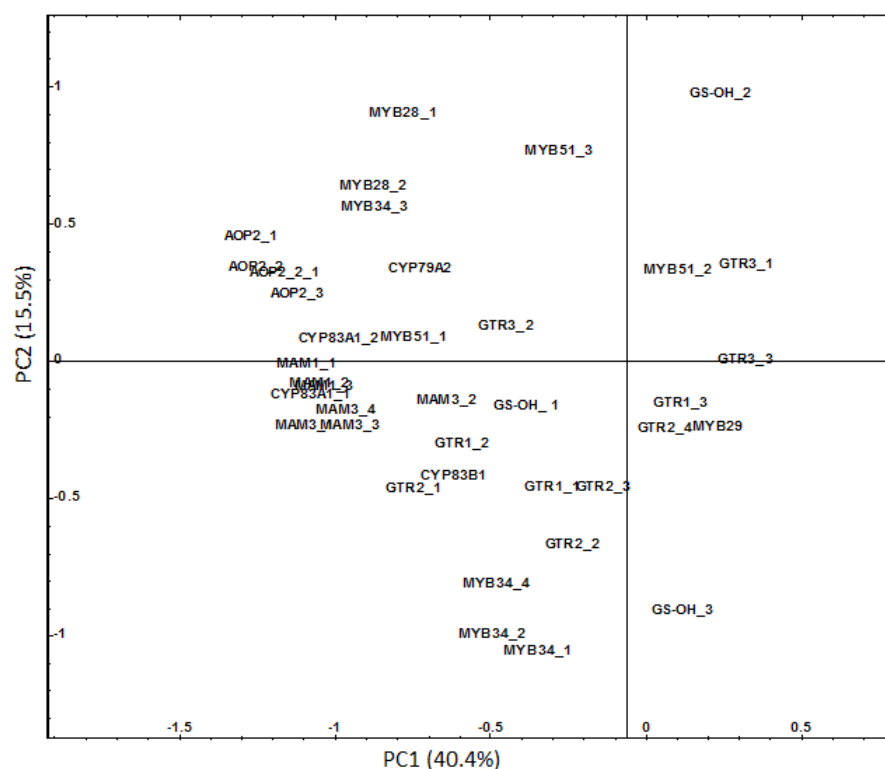


Figure 14. PCA loading plot for GLS biosynthetic genes expression.

Correlation matrix

After executing the BICORRELATE procedure in GenStat, a correlation matrix was obtained (Table 7). This correlation matrix features the correlations between genes and GLS. Correlations over 0.4 and below -0.4 are color-coded according to an interval in order to detect strong positive and negative correlation hotspots. The P-values of these correlations are presented in Table 8.

Two main high correlation hotspots can be distinguished. AOP2 genes are strongly correlated to each other, as well as to MYB28 transcription factors, CYP83A1 paralogs, MAM1 and also MAM3 paralogs. The second high correlation hotspot is between CYP83A1 paralogs and both MAM1 and MAM3 paralogs.

Table 7. Correlation matrix for GLS and genes. The two main high correlation hotspots are highlighted. In blue: AOP2 -MAM-MYB28-CYP83A1. In black: Correlations between CYP83A1 and MAM genes. Correlations highlighted in green: $r = 0.4 - 0.599$. Correlations highlighted in red: $r > 0.599$. Correlations highlighted in yellow: $r = -0.4 - -0.599$. Correlations highlighted in orange: $r < -0.599$

[illegible]

Hierarchical clustering

Glucosinolates

After hierarchical clustering was performed based on Pearson correlations, a heat map was obtained (Figure 15). A distinctive separation of tissues and accessions can be seen in the plot, which is in accordance to what is presented in Figure 11 (PCA analysis). Again, it can be seen that for the accession-tissue-time point combinations, the two main branches of the dendrogram tend to correspond to one accession each. Two main cluster groups arise from each main branch. For both accessions, one of these groups belong to the above ground tissues (leaves and seedling), whereas the other group belongs to the below ground tissues (inner and outer turnip). Turnip tissue seems to be an area of high relative presence of GLS. However, T7 is a developmental stage with very low relative GLS content (Figure 15, highlighted in blue). On the contrary, inner turnip in previous developmental stages appears to be a GLS-rich tissue (Figure 15, highlighted in orange).

There is also differentiated relative presence of compounds per accession and tissue. NAS, ERU, and BER seem to be strongly related to turnip tissue (Figure 15, highlighted in yellow, being relatively low or undetectable in other tissues (Appendix 1) . NAPOL appears to be highly abundant in FT-004 tissues only (Figure 15, highlighted in white), whereas NAP appears to be most abundant in FT-086 (Figure 15, highlighted in purple). IGLS with the exception of 4HBRA cluster together, showing higher abundance in turnips and young leaves Differences in RPA in terms of accession can be observed in Table 6, considering all time points and tissues sampled.



Figure 15. Hierarchical clustering heat map for GLS profile. Green areas indicate lower relative presence of GLS. Red squares indicate high relative presence of GLS. Highlighted in yellow: Profiles of NAS, ERU, and BER across accessions, tissues and time points. Highlighted in orange: Indication of high relative GLS presence in turnip tissues. Highlighted in white: differences in relative presence of NAPOL between accessions FT-004 and FT-086. Highlighted in dark blue: Relative lack of GLS in T7. Highlighted in purple: differences in relative presence of NAP between accessions.

It is also notable that, despite the separation by accession in the heat map, seeds of both accessions are clustered together. Therefore, there is a similarity in GLS profile of seeds despite the differences in overall GLS profile between both accessions.

Table 9 shows the correlations between GLS and their associated p-values. In the AGLS, ERU is significantly correlated to NAP and PRO ($r=0.46$ and 0.41 , respectively), which are subsequent steps in the side chain elongation phase in the AGLS pathway. It is also correlated to NAS ($r=0.61$). Although these GLS have different precursors and therefore do not share any pathway step, these are turnip-specific GLS, so it is possible that the high correlation coefficient is given by the tissue factor. However, turnip specific ERU (4C) is not correlated to the turnip specific BER (5C), its side chain elongated equivalent. In the IGLS group, BRA is significantly correlated to NBRA, 4HBRA and 4MBRA ($r = 0.7$, 0.51 and 0.34 , respectively)

Table 9. A: Pearson correlation matrix for GLS. B: Probabilities for each correlation coefficient (n=52)

A											
ERU											
NAP	0.46										
PRO	0.41	-0.16									
BER	0.14	-0.3	0.5								
CAN	-0.19	-0.27	0.5	0.09							
NAPOL	-0.18	-0.49	0.58	0.34	0.57						
NAS	0.61	0.23	0.56	0.55	-0.05	0.1					
BRA	0.2	-0.02	0.46	0.3	0.45	0.19	0.36				
NBRA	0.17	0.22	0.25	0.09	0.27	0.11	0.5	0.7			
4HBRA	-0.04	-0.37	0.53	0.47	0.34	0.59	0.42	0.51	0.42		
4MBRA	-0.05	-0.04	0.18	-0.02	0.26	0.35	0.25	0.34	0.68	0.34	
	ERU	NAP	PRO	BER	CAN	NAPOL	NAS	BRA	NBRA	4HBRA	4MBRA
B											
ERU											
NAP	0.001										
PRO	0.002	0.257									
BER	0.309	0.033	0								
CAN	0.187	0.049	0	0.539							
NAPOL	0.202	0	0	0.014	0						
NAS	0	0.096	0	0	0.702	0.485					
BRA	0.155	0.878	0.001	0.032	0.001	0.185	0.01				
NBRA	0.228	0.112	0.069	0.521	0.05	0.434	0	0			
4HBRA	0.784	0.007	0	0	0.013	0	0.002	0	0.002		
4MBRA	0.711	0.779	0.208	0.894	0.061	0.012	0.072	0.013	0	0.015	
	ERU	NAP	PRO	BER	CAN	NAPOL	NAS	BRA	NBRA	4HBRA	4MBRA

Biosynthetic genes and paralogs

Overview

The hierarchical clustering performed on the genes dataset yields a heat map (Figure 16, where there is a distinctive cluster of high RGE for AGLS biosynthetic paralogs in tissues that mostly belong to leaves, though stem, flower, and inner turnip of FT-086 are also present (Figure 16, highlighted in orange). Looking at the dendrogram structure per accession-tissue-timepoint combination, the

upper group of branches matches the high RGE cluster seen in the heat map (Figure 16, highlighted in blue).

The heat map also shows differences in overall GLS profile between tissue-timepoint combinations. Turnip in T6, when plants are already flowering and setting seeds is an area of low overall RGE (Figure 16, highlighted in yellow). In Figure 16, the GLS profile at mature leaves in T4 is highlighted in white given the similarity of gene expression profile at this developmental stage and tissue between both accessions and, at the same time, is in a separate branch group in the dendrogram. This shows that this tissue-timepoint combination is very different from mature leaves at T6, which cluster together with young leaves of T6 and T2. The lack of samples of mature leaves at other developmental stages remains an issue when it comes to see whether the GLS profile of this tissue-timepoint is very different from other tissue-timepoint combinations, especially in leaf tissue.

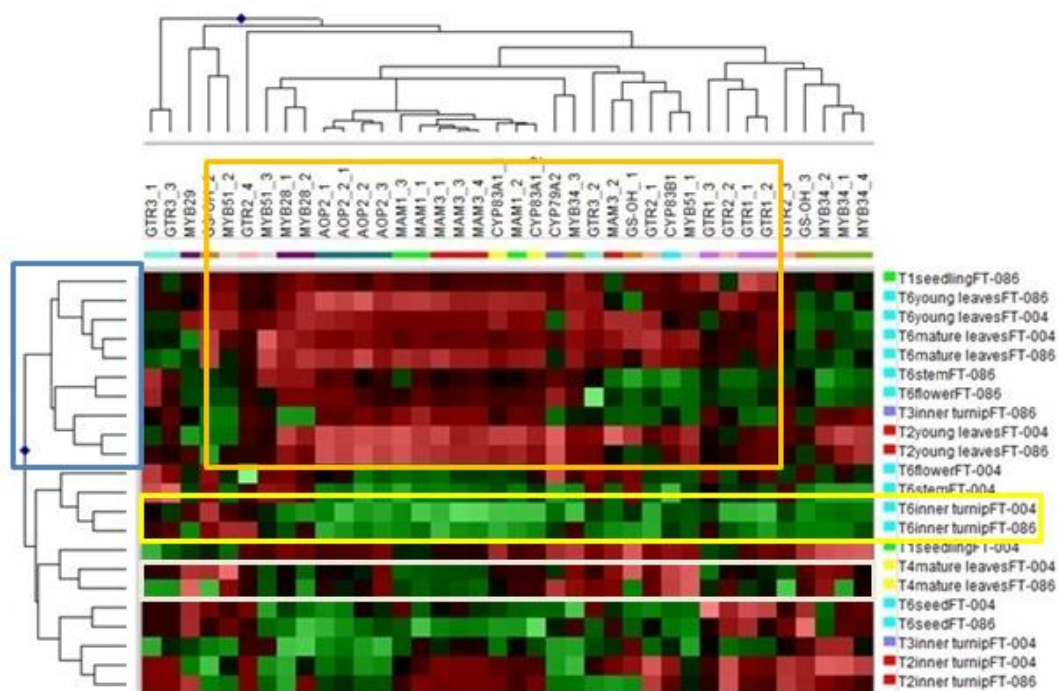


Figure 16. Overview of hierarchical clustering heat map for GLS biosynthetic paralogs. Red patches indicate high RGE. Green patches indicate low RGE. Highlighted in orange: High R cluster of aliphatic GLS genes. Highlighted in yellow: low RGE in T6 turnips. Highlighted in white: Different RGE profile at mature leaves in T4. Highlighted in blue: Dendrogram branch group related to a high RGE cluster of AGLS.

MAM1

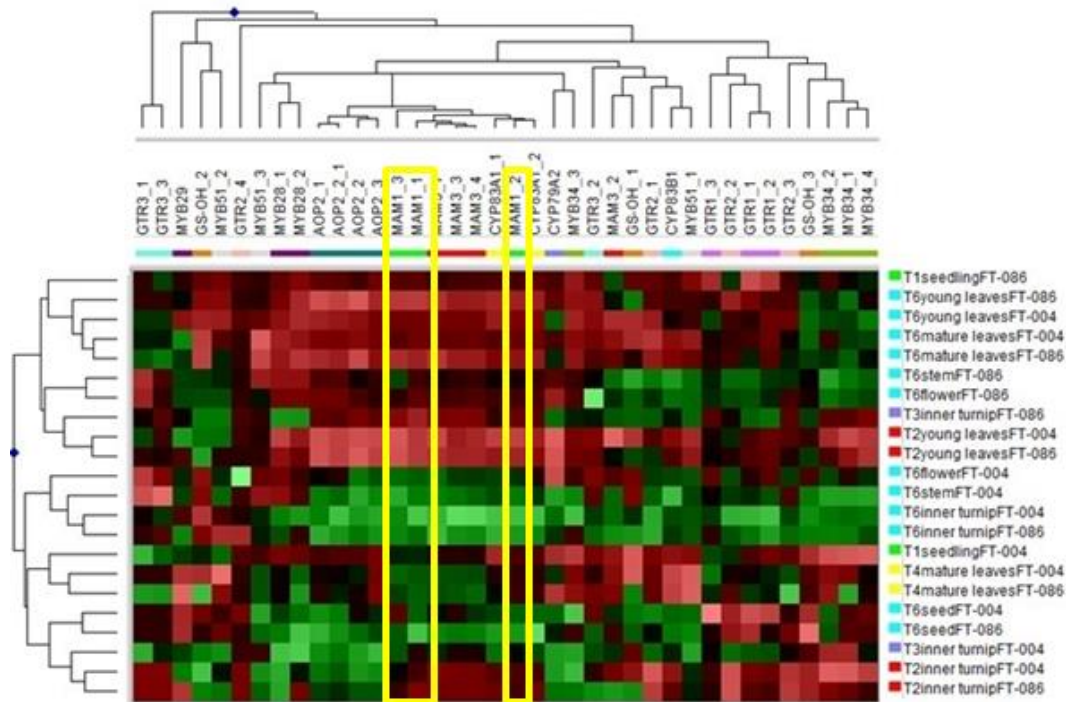


Figure 17. Hierarchical clustering RGE heat map of GLS biosynthetic genes. MAM1 paralogs are highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

There is a high correlation between the gene expression profiles of MAM1_1 and MAM1_3 as seen in the heat map (Figure 17), since these paralogs are closely situated to each other, having very similar expression patterns. Although MAM1_2 is at a distance in the heat map from MAM1_1 and MAM1_3, its expression pattern is quite similar to these paralogs across tissues and time points. The correlations among MAM1 paralogs are all statistically significant and in the range of $r=0.86-0.92$. Figure 18 shows that the highest expression for all MAM1 paralogs is in young leaves, and that RGE for mature leaves is at its lowest in T4 for both accessions. However, gene expression is 10-fold higher in mature leaves during T6 compared to T4. Relative gene expression is lower in stem, flowers and seeds of both accessions. Although in inner turnip at T2 gene expression is 100-fold lower than in T1 seedling, gene expression increases in turnips with the developmental stage reaching between 2 and 25 times the level of RGE in seedlings. However, RGE decreases again in inner turnip tissues at T6, being up to 100-fold lower than in FT-004 at the seedling stage.

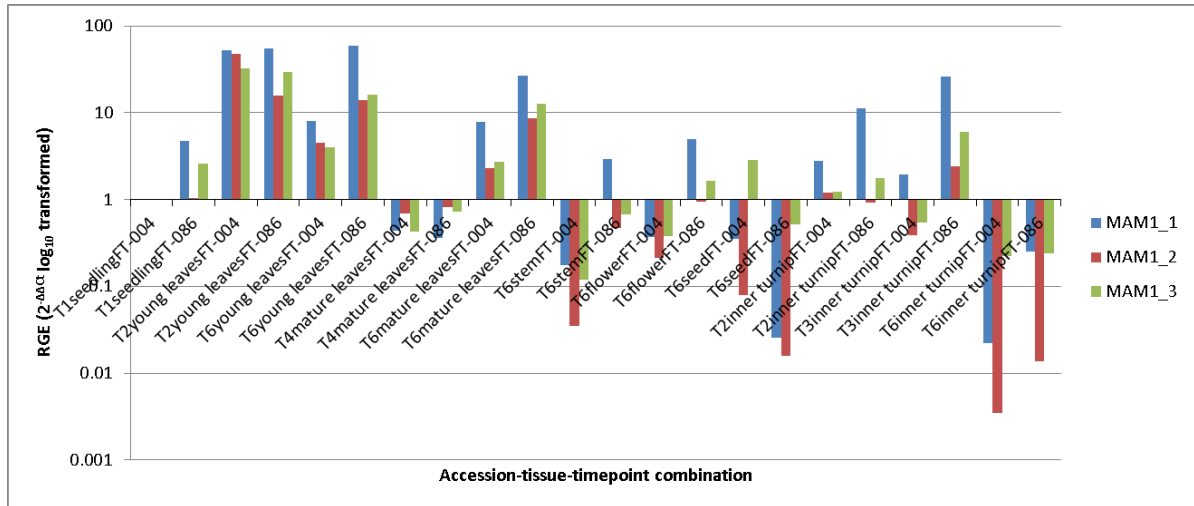


Figure 18. RGE pattern for MAM1 paralogs. The Y axis shows the fold-change in expression compared to seedling of accession FT-004.

MAM3

In Figure 19 and Figure 20, highlighting in the hierarchical clustering MAM3 paralogs and showing RGE pattern for MAM3 paralogs, respectively, it is seen that MAM3_1, MAM3_3, and MAM3_4 are clustered together and have a very similar RGE pattern. However, MAM3_2 is separated from this cluster, being significantly correlated with GS-OH_1 ($r=0.69$). It is also noticeable that MAM3_2 and GS_OH 1 tend to be higher expressed in tissues of FT-004. This trend is confirmed by means of a t-test performed for both paralogs with respect to the accession factor. Both paralogs are significantly higher expressed in FT-004 than in FT-086 (t-test). When looking at Figures 6 to 8, the only AGLS that follows such trend is NAPOL (5C), which is the product of conversion of CAN by GS-OH. MAM3_2 is negatively correlated to ERU and NAP ($r=-0.41$ and -0.51 , respectively), though the negative correlation between MAM3_2 and ERU falls short out of the significance threshold of significance ($p=0.06$). Also, GS-OH_1 is positively correlated to NAPOL ($r=0.56$), and it is also negatively correlated to NAP (4C) ($r=-0.74$, $p<0.001$). This correlation makes sense, since GS-OH hydroxylates CAN to produce NAPOL.

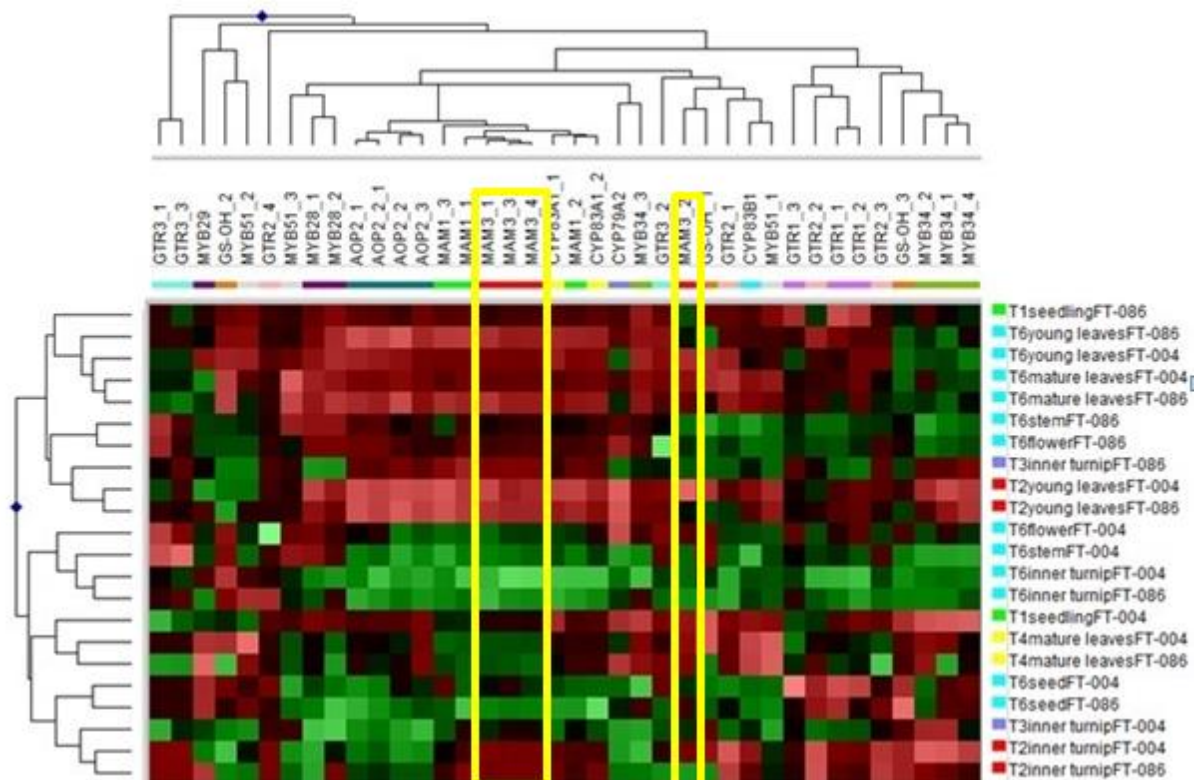


Figure 19. Hierarchical clustering RGE heat map of GLS biosynthetic genes with MAM3 paralogs highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

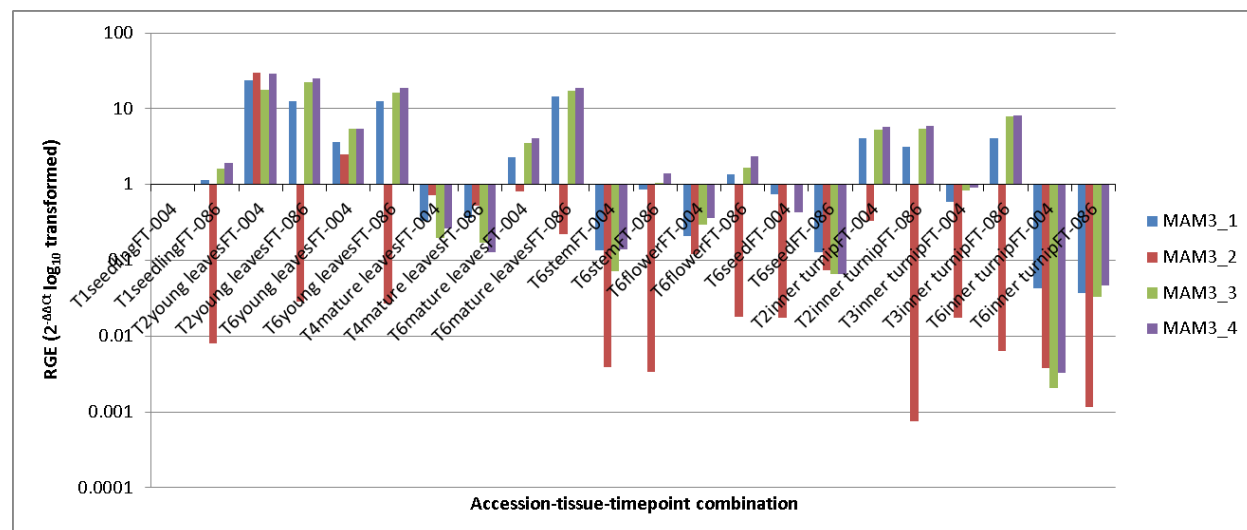


Figure 20. RGE pattern for MAM3 paralogs .The Y axis shows the fold-change in expression compared to seedling of accession FT-004.

AOP2

All AOP2 paralogs are closely correlated to each other, as seen in the correlation matrix, and in Figures 21 and 22. Figure 21 also shows that all AOP paralogs are higher expressed in leaves, young and mature. The only exception to this is mature leaves in T4, where gene expression is 10- fold lower compared to leaves in other time points, but still higher than in stem, turnip tissues. Comparing MAM1, MAM3, and AOP paralogs RGE patterns, these are closely correlated, as seen in Figures 23 and in the correlation matrix. Only negative correlations are observed between AOP2 paralogs and AGLS. The only significant correlation is a negative correlation between AOP2_3 and BER, which could be expected from the conversion of BER to CAN by AOP2, but there is no significant positive correlation between AOP2 and NAP or CAN.

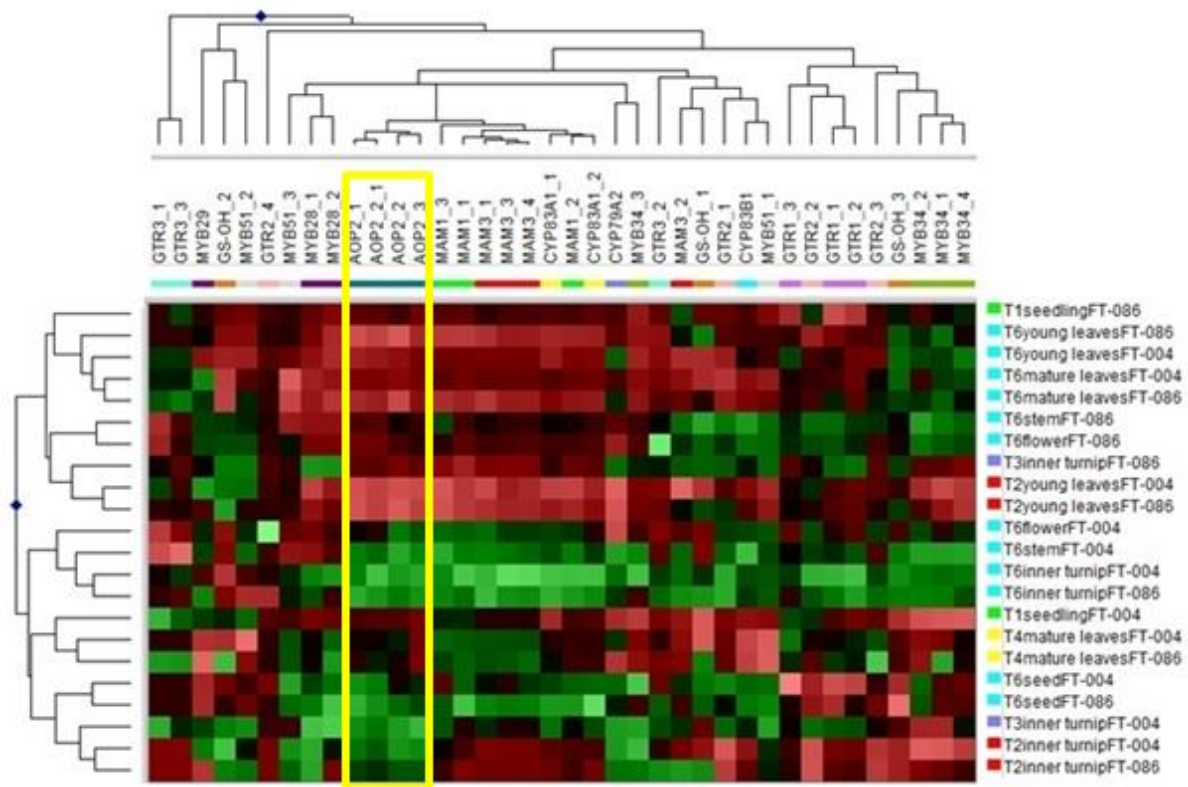


Figure 21. Hierarchical clustering RGE heat map of GLS biosynthetic genes with AOP2 paralogs highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

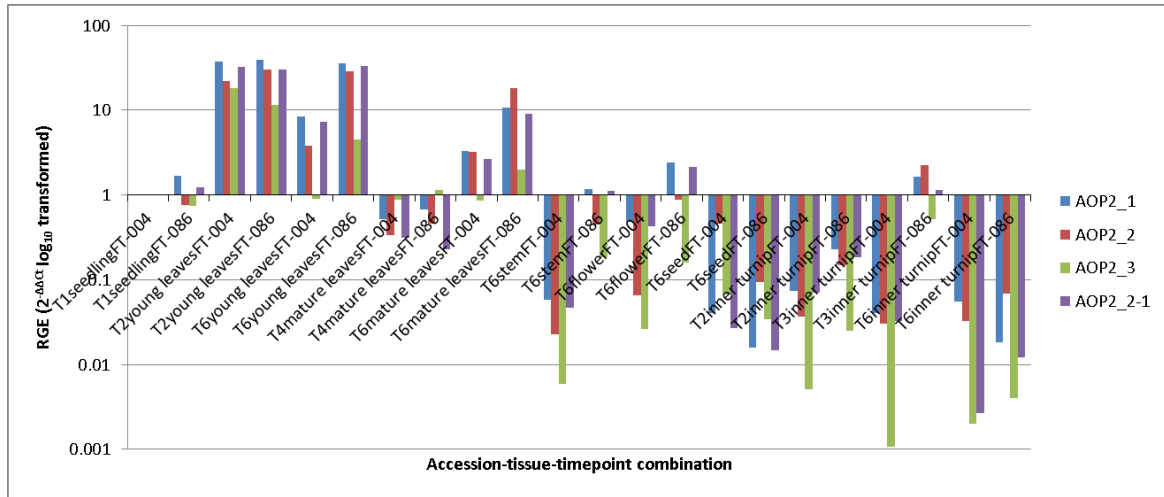


Figure 22. RGE pattern for AOP2 paralogs.

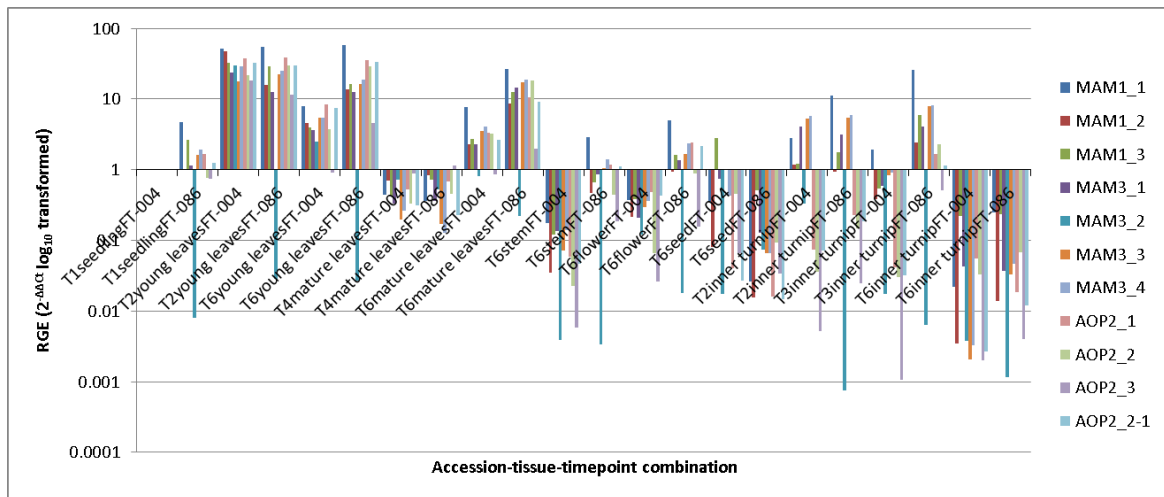


Figure 23. Comparison of RGE patterns among MAM1, MAM3 and AOP paralogs

MYB28 and MYB29 aliphatic pathway transcription factors

Both MYB28 paralogs, MYB28_1 and MYB28_2, are expressed in a similar way, as illustrated in Figure 24 where they are clustered together in the dendrogram. This similarity in RGE patterns is also shown in Figure 25, and by their strong correlation ($r=0.76$).

On the contrary, the MYB29 RGE tends to be higher when and where MYB28 paralogs RGE is lower. (Figure 26 and 27). This is also reflected in the negative correlation between MYB29 with MYB28_1 and MYB28_2 ($r=-0.27$ and -0.4 , respectively), though these correlation coefficients are not significantly different from zero.

A positive effect of the expression of both MYB28 paralogs in the expression of MAM, AOP and CYP83A1 is inferred from Table 7 (High correlation cluster, highlighted in blue). Both MYB28_1 and

MYB28_2 are positively correlated to all MAM, AOP2 and CYP83A1 paralogs. Though not all correlations are statistically significant, the strongest correlations are found with AOP2 and CYP83A1 paralogs

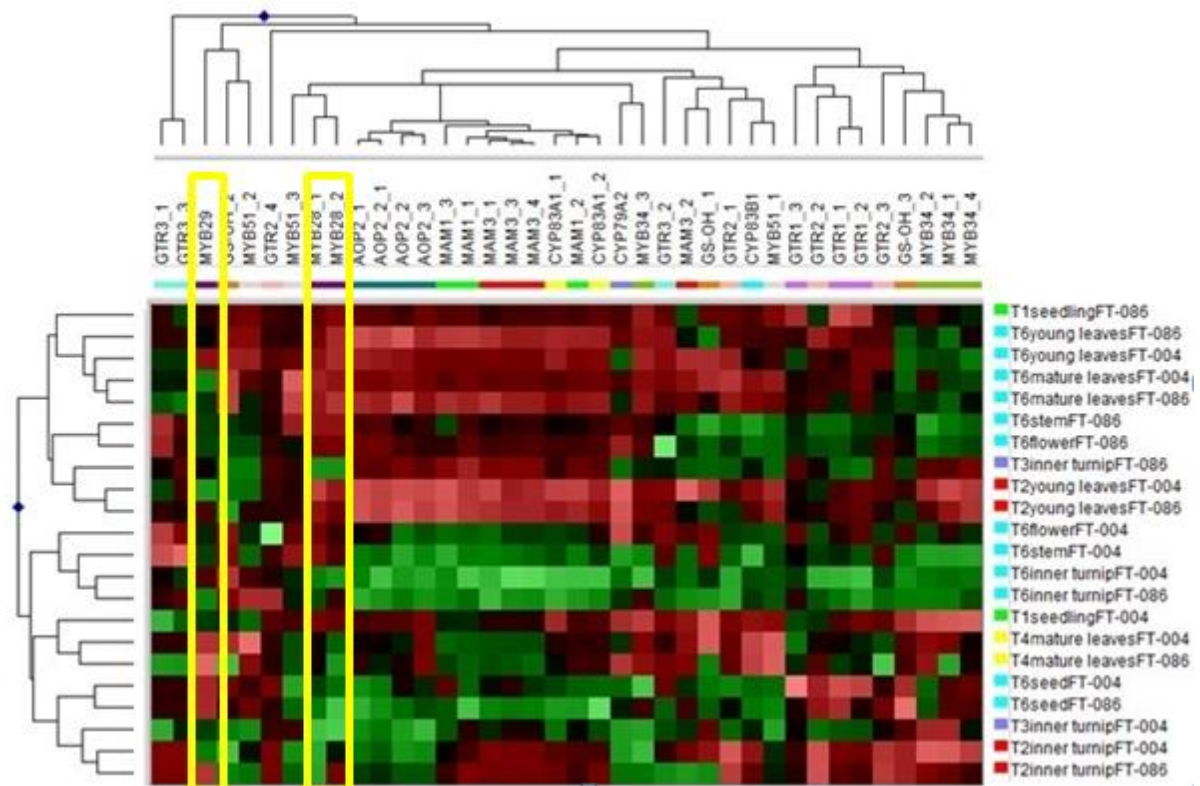


Figure 24. Hierarchical clustering RGE heat map of GLS biosynthetic genes with MYB28 and MYB29 paralogs highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

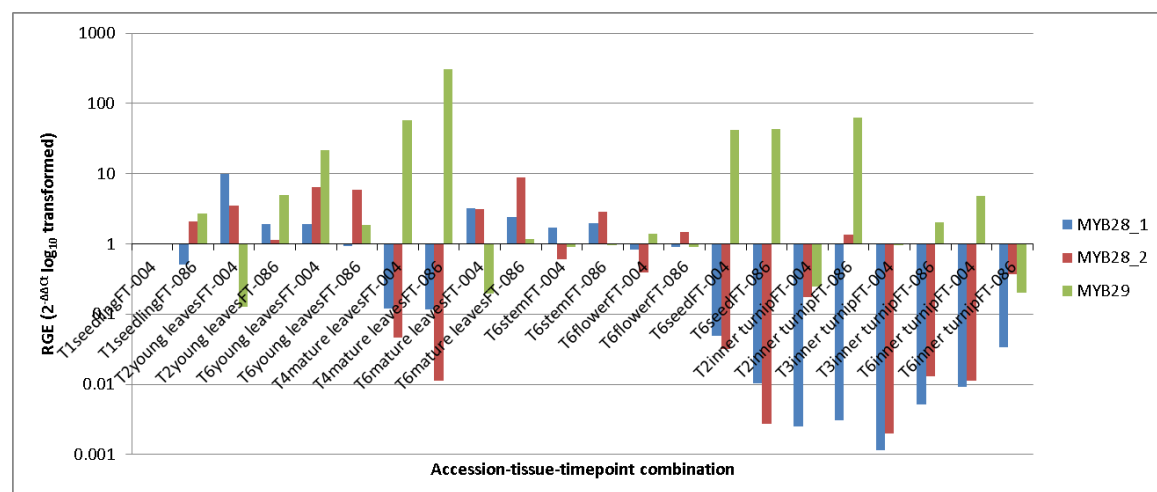


Figure 25. RGE pattern for MYB28 and MYB29 transcription factor paralogs.

GS-OH

The GS-OH paralogs are far from each other in the hierarchical clustering heat map (Figure 26). This fact is also evident in Table 7, showing no correlation among GS-OH_1, GS-OH_2, and GS-OH_3. Figure 27 also shows the differences in RGE pattern among the three GS-OH paralogs. It is also interesting to see that GS-OH_2 has a similar expression pattern with MYB51_2, since these genes belong to the AGLS and IGLS biosynthetic pathways respectively. The correlation between their RGE patterns is positive ($r=0.44$). The same phenomenon is observed in the correlations between GS-OH_3 and the IGLS TFs MYB34_1, MYB34_2, and MYB34_4 which are also moderately strong ($r=0.58$, $r=0.43$, and $r=0.56$, respectively). Contrary to what could have been expected, according to the correlation matrix in Table none of these paralogs show any significant correlation to PRO. For NAPOL, product after conversion CAN by GSOH, the only significant correlation is with GS-OH_1 ($r = 0.56$, $p = 0.007$). The only significant correlation between GS-OH and MAM paralogs is between MAM3_2 and GS-OH_1 ($r = 0.69$, $p < 0.001$). Likewise, there are no significant correlations between AOP2 and GS-OH paralogs.

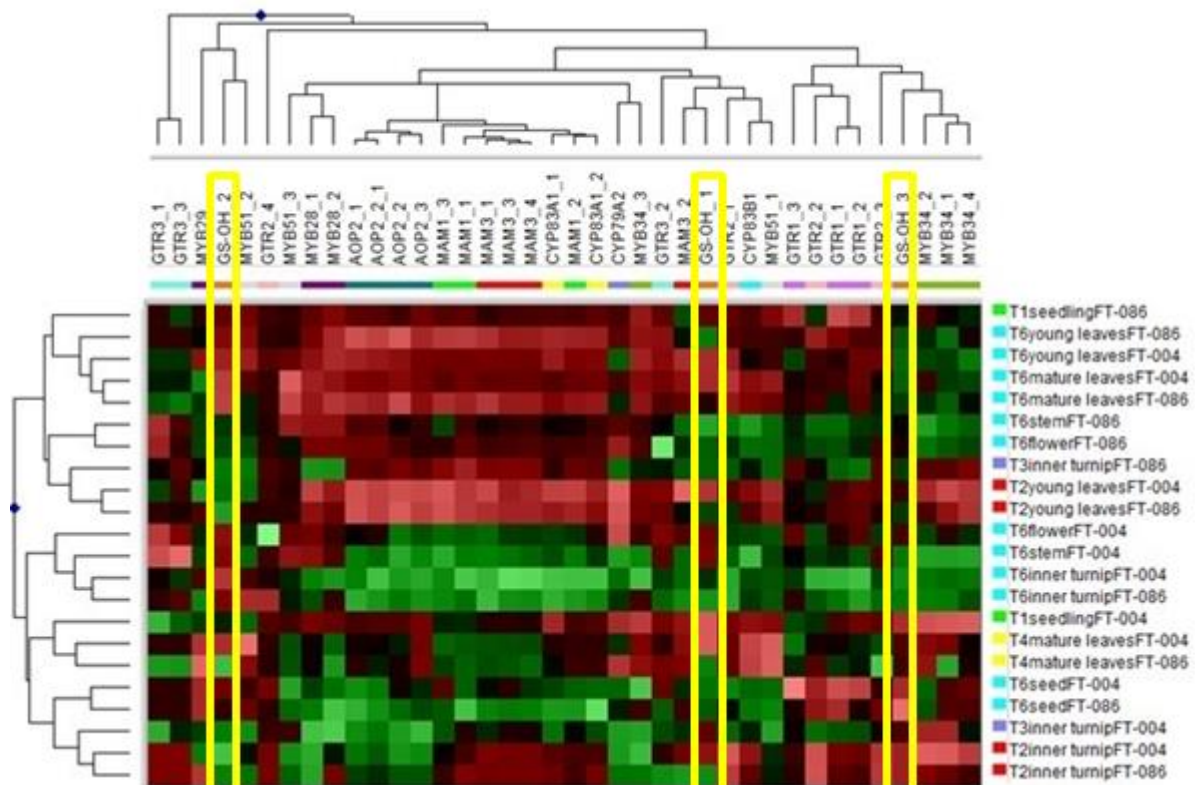


Figure 26. Hierarchical clustering RGE heat map of GLS biosynthetic genes with GS-OH paralogs highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

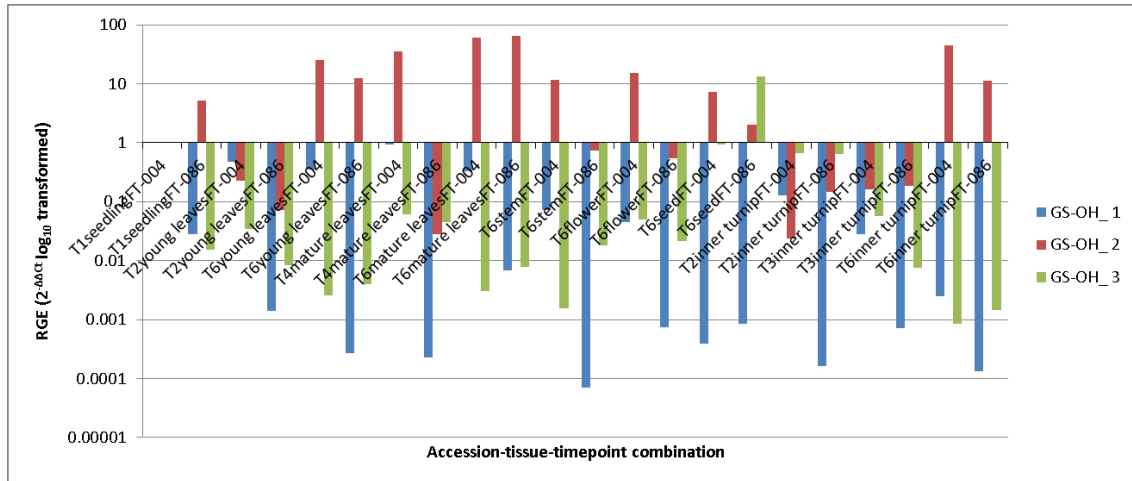


Figure 27. RGE pattern for GS-OH gene paralogs.

CYP83A1

In the core structure formation stage of the AGLS biosynthetic pathway, CYP83A1 plays a key role by metabolizing aliphatic aldoximes (Naur, 2003). CYP83A1 paralogs CYP83A1_1 and CYP83A1_2 are close to each other in the hierarchical clustering map (Figure 28). Their RGE patterns are similar, but CYP83A1_2 is more upregulated in all tissues and time points than CYP83A1_1 except in seeds of FT-086 (Figure 29). These genes are also present in the high RGE cluster mentioned in Figure 16 (highlighted in orange). These paralogs, and their expression is at its highest in leaves, and at its lowest in turnip, following a similar trend to the other genes belonging to the AGLS pathway. Although CYP83A1 plays a key role in the core structure formation of AGLS, there are no significant correlations with any of the AGLS measured in the same samples.

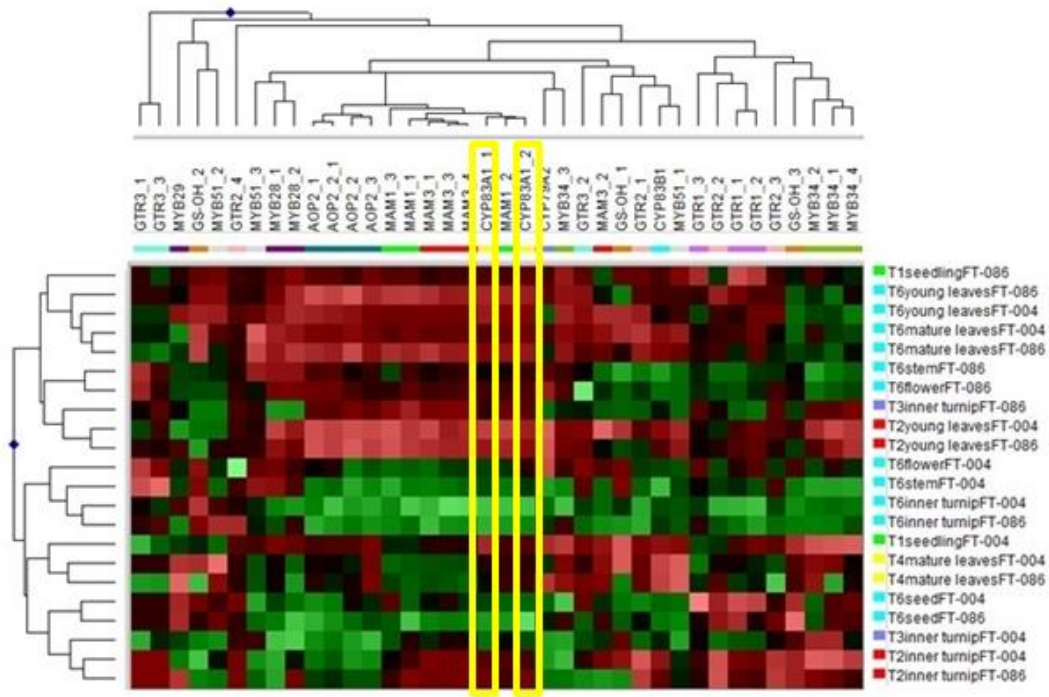


Figure 28. Hierarchical clustering RGE heat map of GLS biosynthetic genes with CYP83A1 paralogs highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

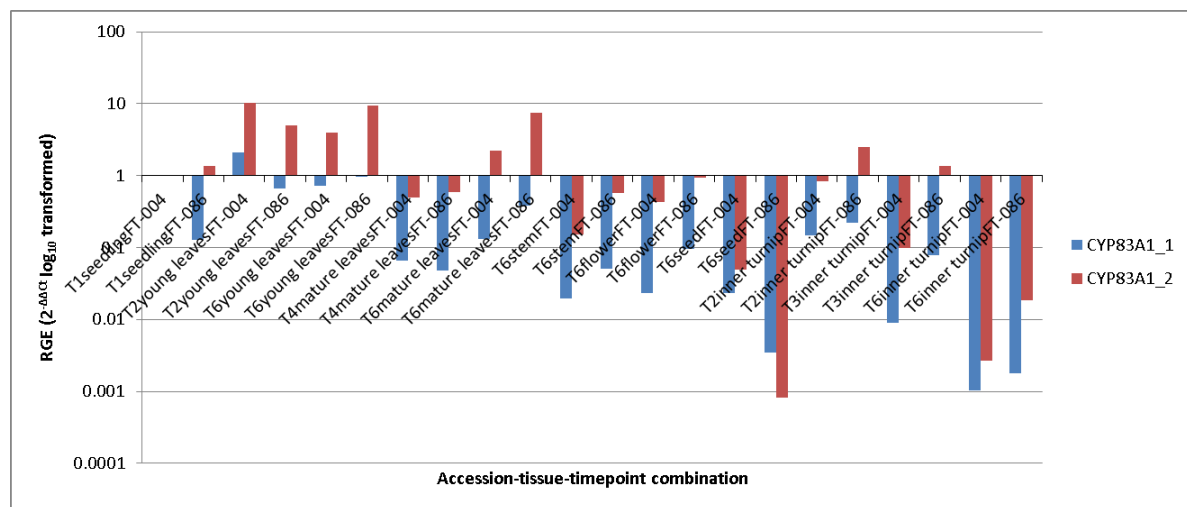


Figure 29. RGE profile for CYP83A1 gene paralogs.

GTR1

AGLS transporter gene paralogs GTR1_1, GTR1_2 are clustered together with GTR2_2, and have similar RGE pattern (Figure 30). However, GTR1_3 is the most upregulated paralog across all tissues and time points, with exception of T4 mature leaves in FT-086. In both GTR1_2 and GTR1_3 the highest expression point is in the seeds of FT-004, which is 1000-fold lower than seeds of FT-086, and in seedlings of FT-086 (Figure 31). In contrast, the highest expression point of GTR1_1 is at T2 of FT-004 inner turnip, with its RGE decreasing in turnip as time moves on.

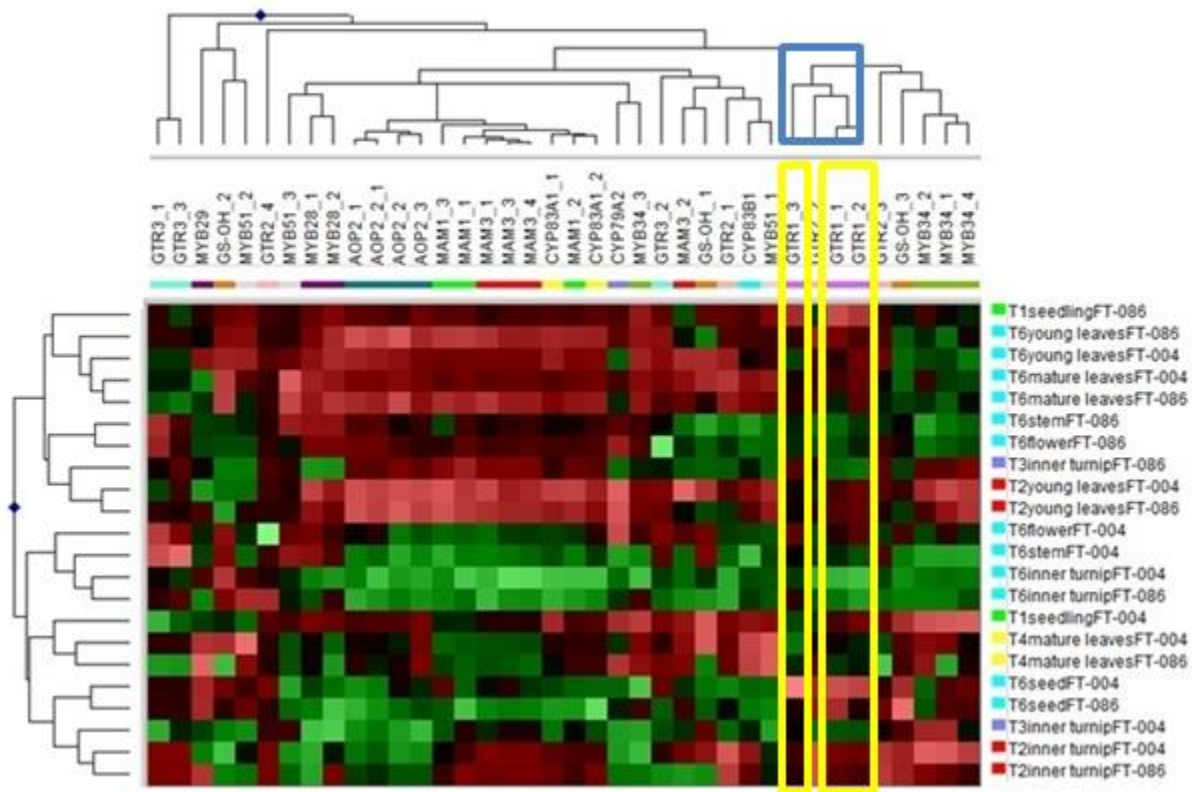


Figure 30. Hierarchical clustering heat map with GTR1 paralogs highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

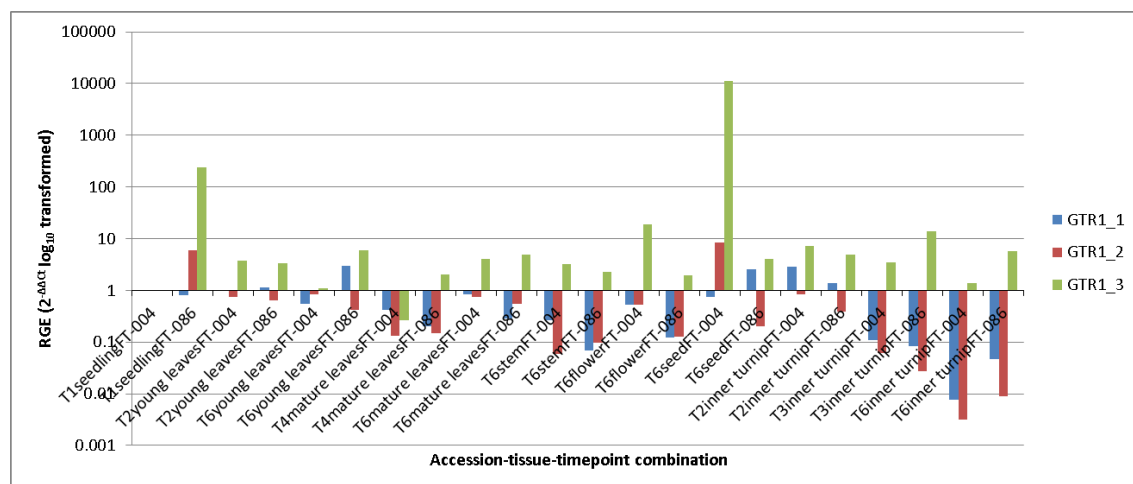


Figure 31. RGE profile for GTR1 gene paralogs

GTR2

It has to be mentioned that there were missing RGE values for some accession-tissue-timepoint combinations. RGE patterns for GTR2 paralogs are similar except for GTR2_4 (Figures 32 and 33). It is possible that it is grouped separately due to an outlier in the RGE of flower in FT-004. Here, the $-\Delta C_t$ value is -36.17 and the $2^{-\Delta\Delta C_t}$ value is 3.65×10^{-9} , so the gene may well be deemed as non-

expressed. Although 5 out of 22 data points for GTR2_4 are missing, it can be seen that in inner turnip it follows the same pattern as the other GTR2 paralogs..

Overall, GTR2 paralogs reach their highest expression in inner turnips (T2-T3), seeds, as well as young and mature leaves in T6.

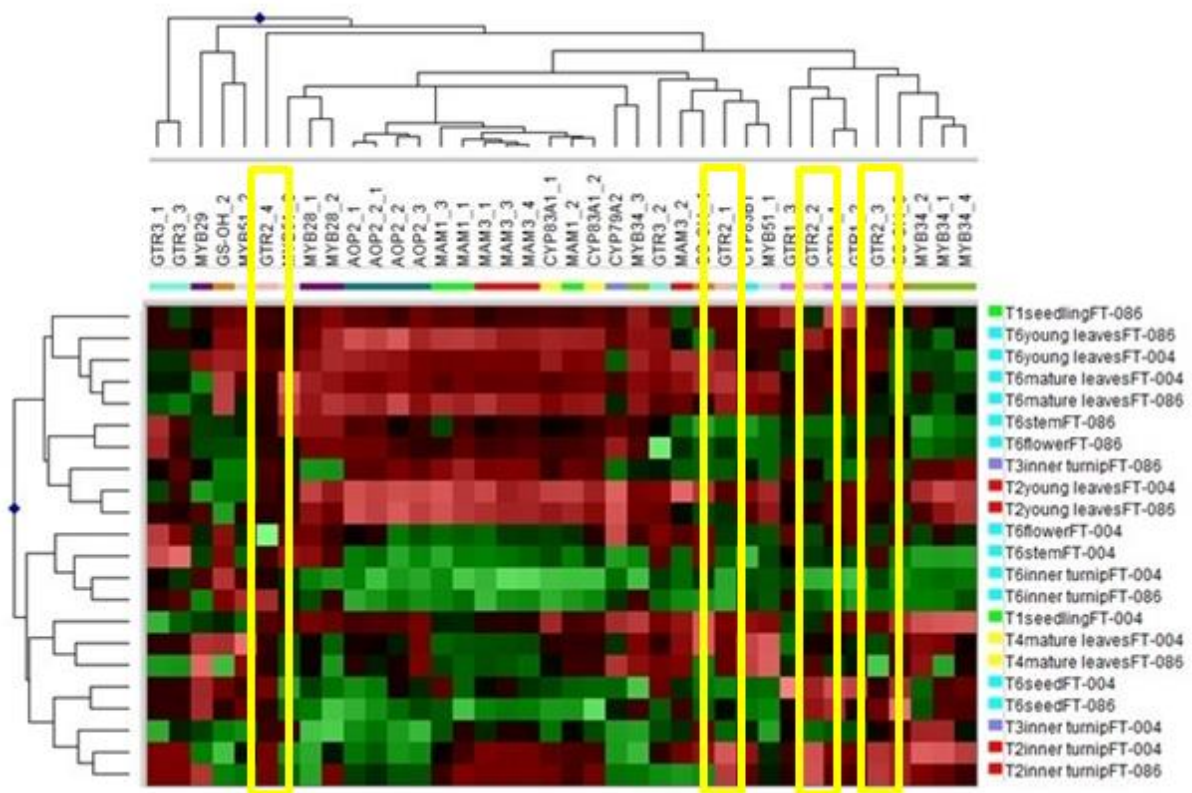


Figure 32. Hierarchical clustering RGE heat map of GLS biosynthetic genes with GTR2 paralogs highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

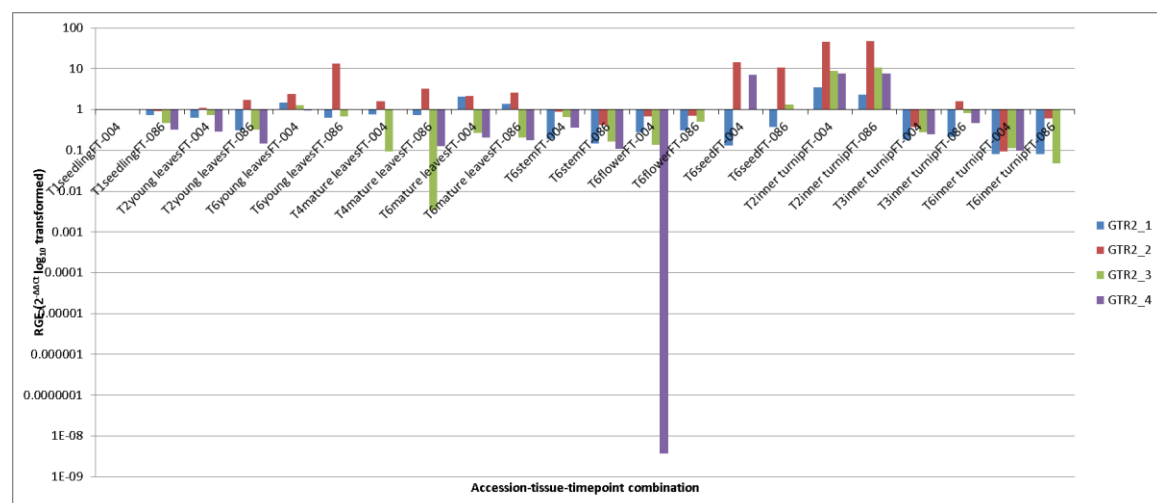


Figure 33. RGE profile for GTR2 gene paralogs

MYB34 and MYB51 indolic pathway transcription factors

Transcription factors related to the activation of the IGLS pathway, MYB34 paralogs MYB34_2, MYB34_1, and MYB34_4 are strongly correlated to each other, as seen in Figure 34 and 35, as well as in Table 7. However, MYB34_3 is in a different branching group in the dendrogram, close to and with a positive significant correlation to CYP79A2 ($r=0.64$), which belongs to the phenyl glucosinolates core structure formation pathway stage.

According to Figure 34, MYB34_2, MYB34_1, and MYB34_4 are more expressed in turnip tissues, except for T6. MYB34_3 is apparently more upregulated (up to 1000-fold RGE differences) in tissues in the aerial part of the plant. However, it is lower expressed in turnip tissues. It is also correlated to all AOP paralogs, MYB28_1, MYB28_2, CYP83A1_1, CYP83A1_2, MAM1_2, and MAM3_2. At the same time, it also has a positive correlation with CYP83B1, and a negative correlation with GTR3_3.

The expression pattern of MYB51 paralogs is somewhat different from MYB34. Only in the case of MYB51_1 a significant difference ($p < 0.001$) is observed in terms of the expression of the overall RGE per tissue type. Here, MYB51_1 is significantly more expressed in the seedling, young and mature leaves than in the turnip (Figure 36). As for MYB51_2 and MYB51_3, there was no significant difference in terms of their overall RGE per tissue type.

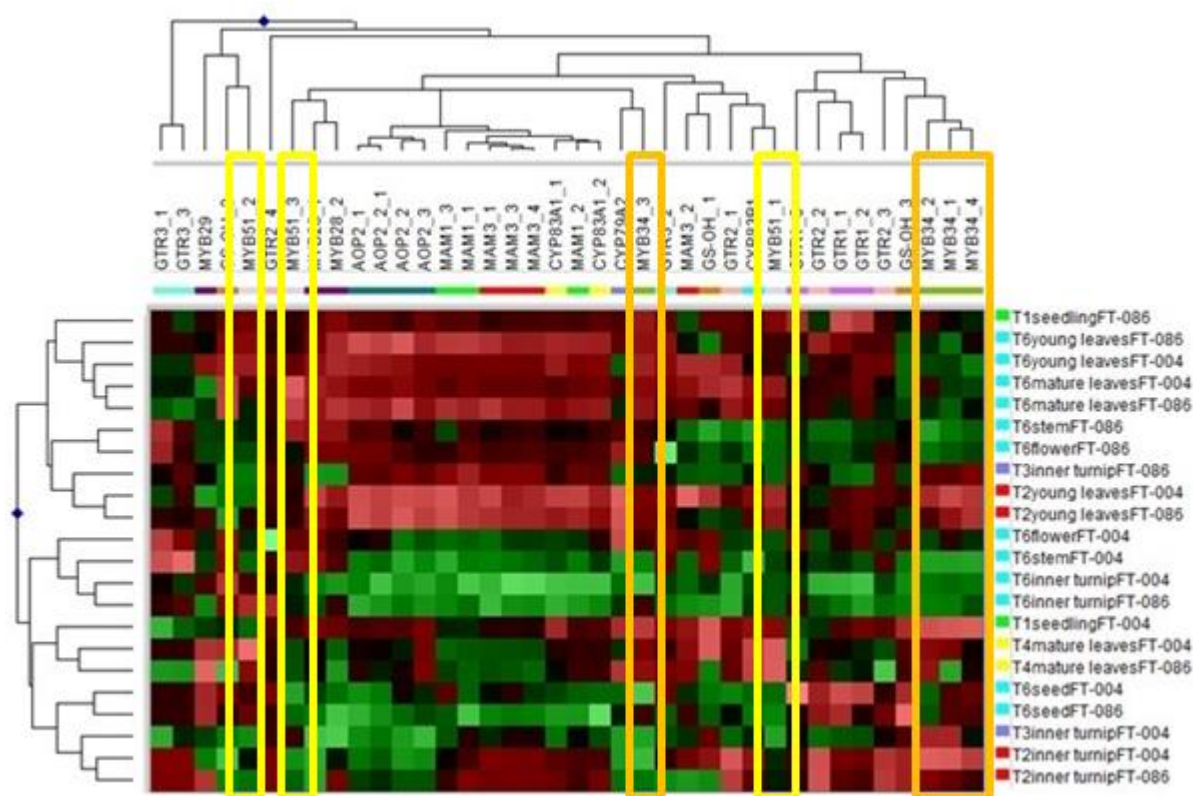


Figure 34. Hierarchical clustering RGE heat map of GLS biosynthetic genes with MYB34 paralogs highlighted in orange. MYB51 paralogs are highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

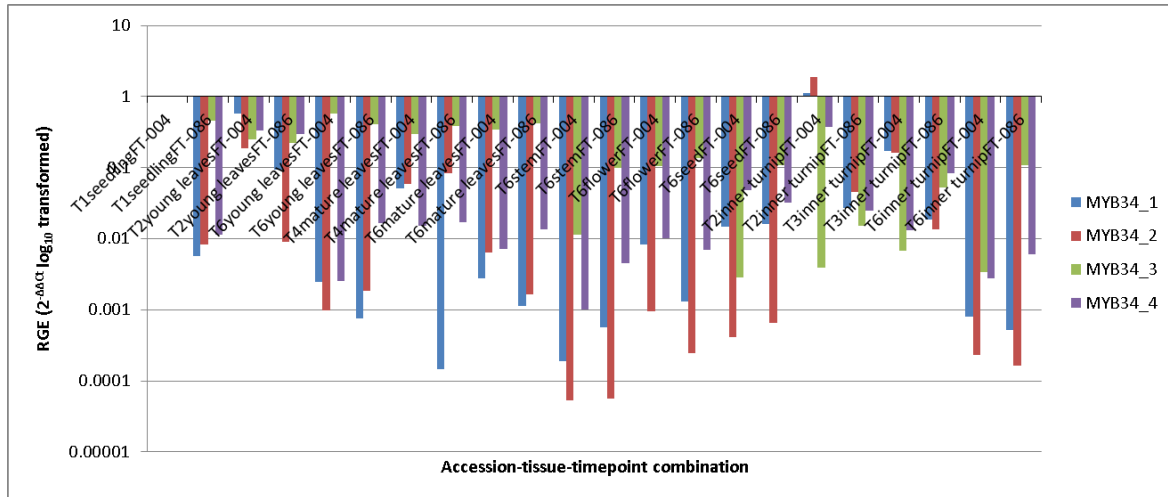


Figure 35. RGE profile for MYB34 gene paralogs

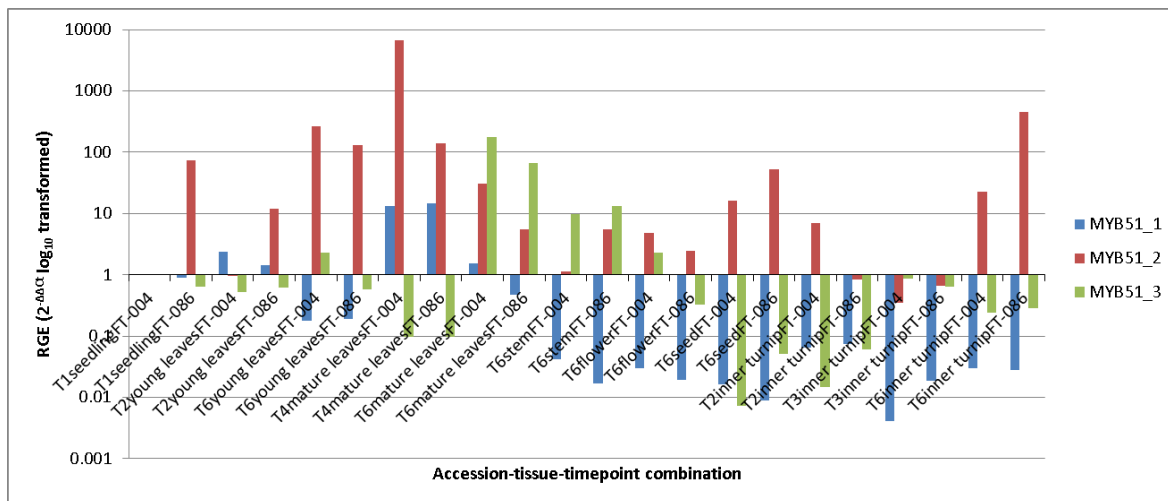


Figure 36. RGE profile for MYB51 gene paralogs

CYP79A2 and CYP83B1

CYP79A2 is related to the core structure formation stage in the phenyl GLS pathway. Besides being significantly correlated to MYB51_3, it also has a positive correlation to MYB51_1 ($r=0.49$). CYP83B1 also has strong correlations with several indolic pathway transcription factors, such as MYB34_2, MYB34_3, MYB34_4, and MYB51_1. Although the gene product of CYP83B1 plays a key role in the IGLS pathway, there are no significant correlations between it and the IGLS reviewed in this project. At the same time, there is no significant correlation between CYP79A2 and NAS.

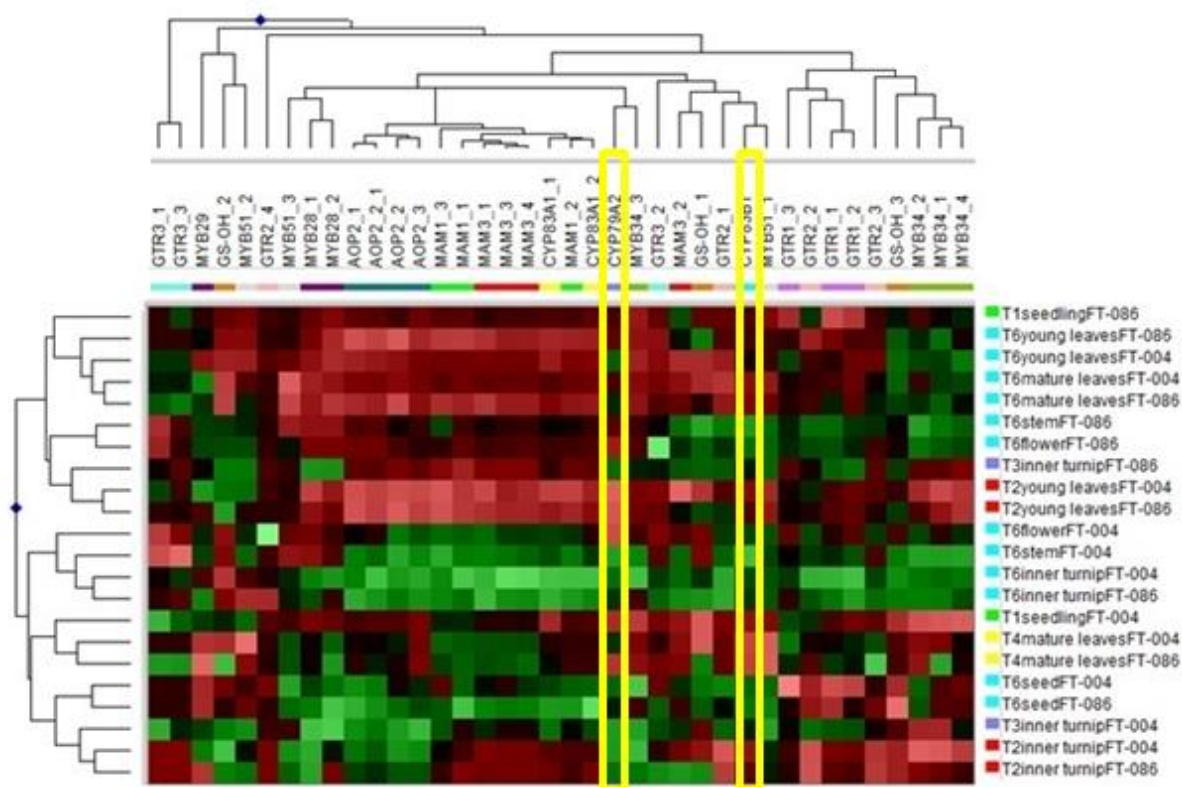


Figure 37. Hierarchical clustering RGE heat map of GLS biosynthetic genes with CYP79A2 and CYP83B1 paralog highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

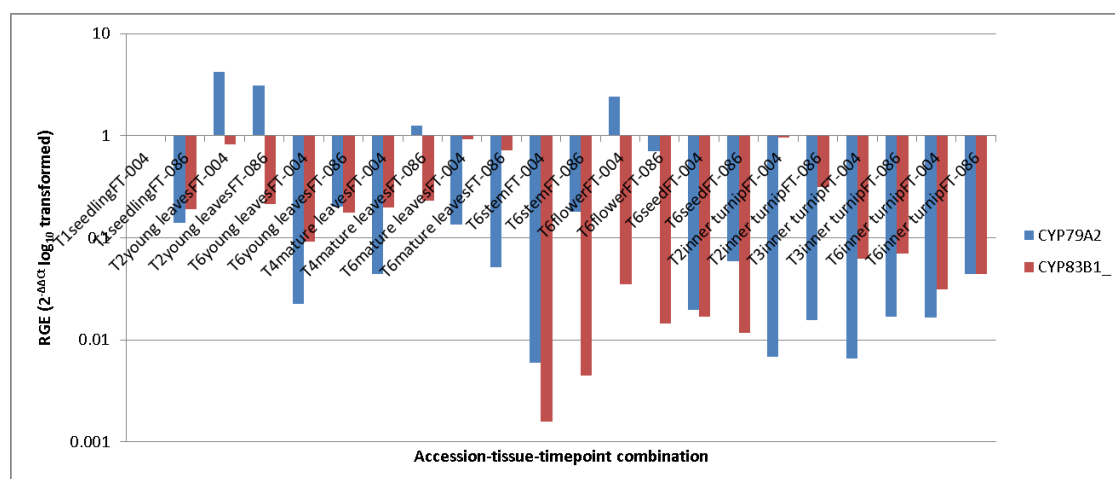


Figure 38. RGE profile for CYP79A2 and CYP83B1 genes

GTR3, indolic GLS transporter

The hierarchical clustering heat map in Figure 39 shows that GTR3_1 and GTR3_3 belong to a branching group of their own in the heat map, whereas GTR3_2 is clustered mostly with biosynthetic genes belonging to the aliphatic pathway. As seen in Tables 3 and 6 of Appendix 2, there are several missing data points for GTR3_1 and GTR3_3, so a possible reason for their position in the dendrogram is because the missing data values were imputed by GeneMaths from the mean of both row and column values. As GTR3_2 is the paralog with the least missing values, it clusters together

with other genes, though not of the same biosynthetic pathway. Figure 40 shows the RGE pattern of GTR3 genes.

Due to the fact that there are many missing data points, it is not possible to make solid conclusions about the RGE patterns of these paralogs, especially for GTR3_1 and GTR3_3

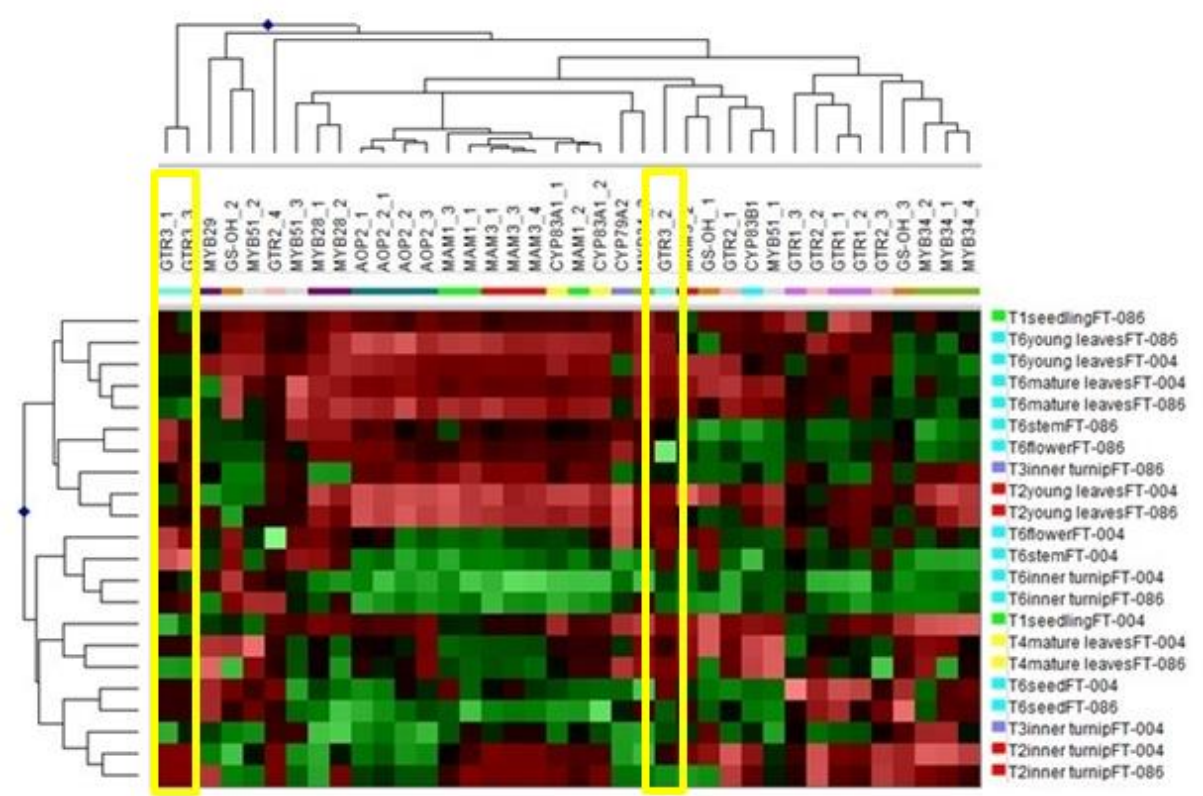


Figure 39. Hierarchical clustering RGE heat map of GLS biosynthetic genes with GTR3 paralogs highlighted in yellow. Red patches indicate high RGE. Green patches indicate low RGE.

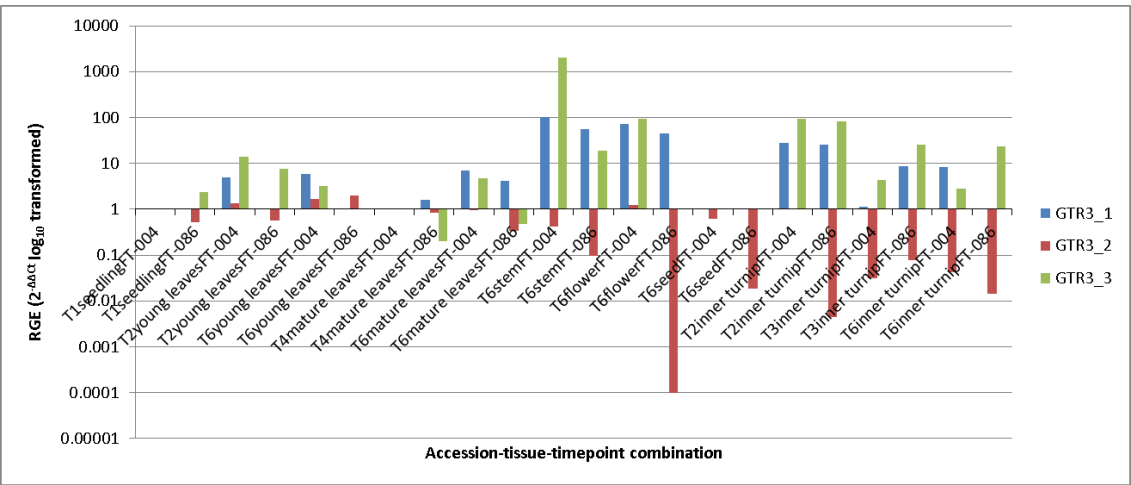


Figure 40. RGE profile for GTR3 gene paralogs

Self-Organizing map (SOM)

Biosynthetic genes and paralogs

With the purpose of evaluating strength of relationships between GLS biosynthetic paralogs in terms of their RGE, as well as to evaluate co-regulatory patterns, three SOMs were generated with different grid sizes. In this section, SOMs for 3 x 3 and 5 x 5 for GLS biosynthetic genes and 2 x 2 SOM for GLS are displayed. The 4 x 4 SOM for biosynthetic genes and 3 x 3 and 4 x 4 SOM for GLS are displayed in Appendix 4. Small grid sizes allow visualizing larger cluster of genes that are similarly expressed, thus visualizing co-regulatory schemes. Larger grid sizes allow visualizing stronger relationships between genes, pointing to a possible common pathway. In the 3 x 3 SOM grid, all MAM1 and MAM3 paralogs, except for MAM3_2 are clustered together with both CYP83A1 paralogs, having the genes that participate in the side chain elongation and core structure formation together. All AOP2 paralogs are also clustered together with MYB28_2. Thus, in the upper part it is possible to see most of the paralogs that act in the AGLS biosynthetic pathway. However, both MYB28 paralogs are not clustered together in this grid, but they join each other in the 4 x 4 and 5 x 5 SOM grids. MYB29 is always clustered apart from both MYB28 paralogs.

GTR1_1 and GTR1_2 are not clustered together in the 3 x 3 SOM grid, but they do cluster together in the 4 x 4 and 5 x 5 SOM grids. The same occurs with GTR3_1 and GTR3_3. However, given their unique patterns that were imputed by GeneMaths due to their missing values, it is difficult to say whether the associations are real or not.

The overall trend is the same as in the hierarchical clustering heat map, where the clusters of AGLS-related paralogs tend to group together as the grid size increases. However, there are also several clusters formed by paralogs that belong to different GLS pathways

As the size of the SOM grid increases, the number of clusters also increases. At the 5 x 5 SOM, the following paralog associations remain throughout all the grid size increases:

- All AOP2 paralogs
- MYB34_1, MYB34_2, and MYB34_4.
- MAM3_2 and GS_OH_1
- MYB28_1 and MYB34_3
- MYB51_2 and GS-OH_2
- GTR3_1 and GTR3_3

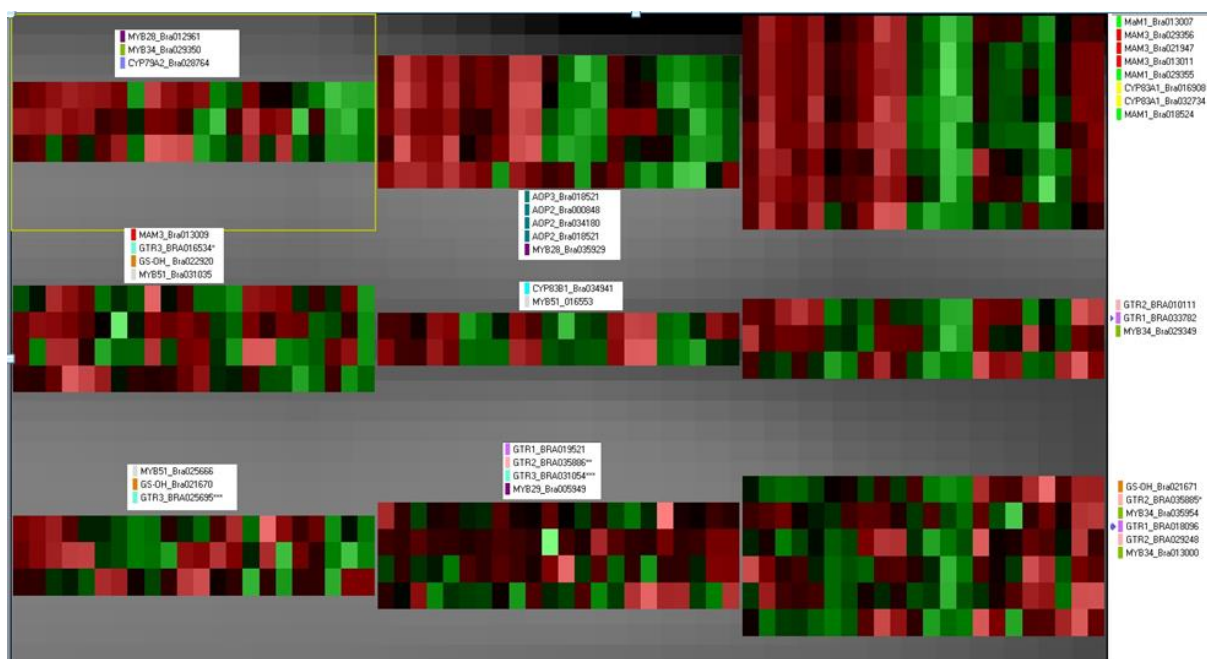


Figure 41. 3 x 3 SOM for GLS biosynthetic paralogs

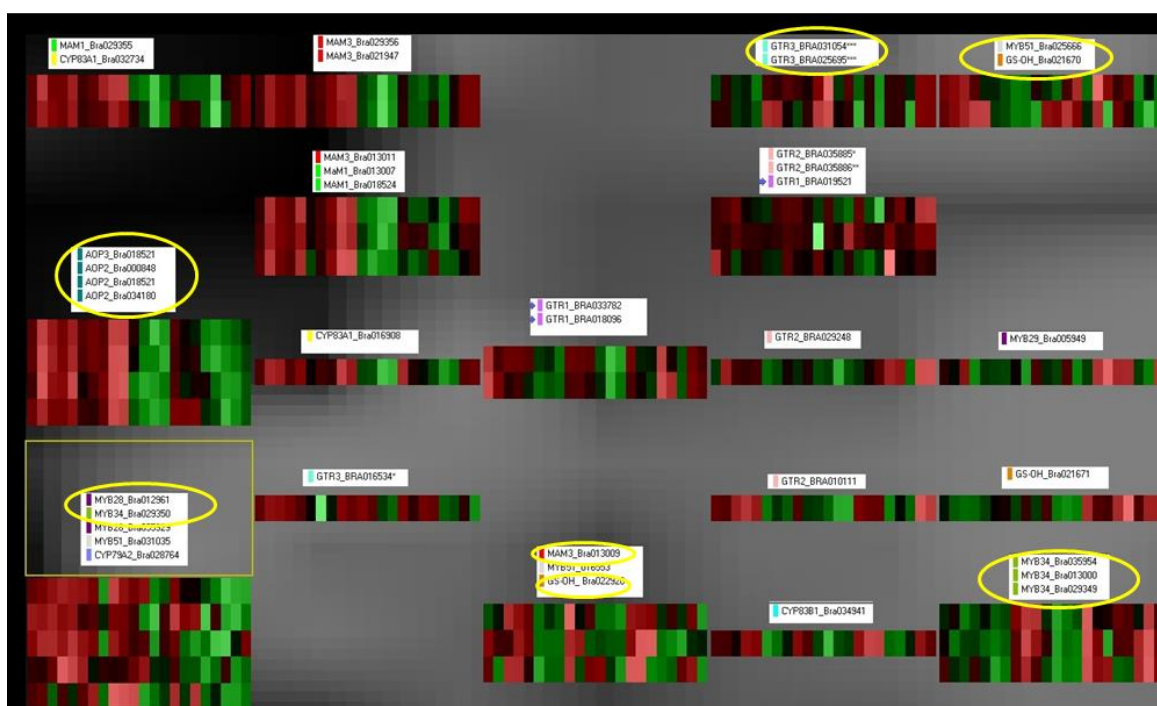


Figure 42. 5 x 5 SOM for GLS biosynthetic paralogs

Glucosinolates

Given the high tissue specificity of BER and ERU, this is the only cluster that remains intact after grid size increases. In the 2 x 2 SOM grid, GLS tend to cluster according to relative presence in tissues. BER and ERU form a single cluster. Below this cluster, NAP and NAS form another single cluster.

These GLS are mostly present, but not exclusively, in turnips. On the left side of the grid, there are two clusters with GLS that are more distributed throughout the plant. In one cluster, PRO, 4HBRA, and NAPOL are found. Below this cluster, BRA, 4HBRA, NBRA and CAN are together.

As the grid increases in size, NAS does not cluster with any other GLS and the first clusters are broken down into smaller clusters.

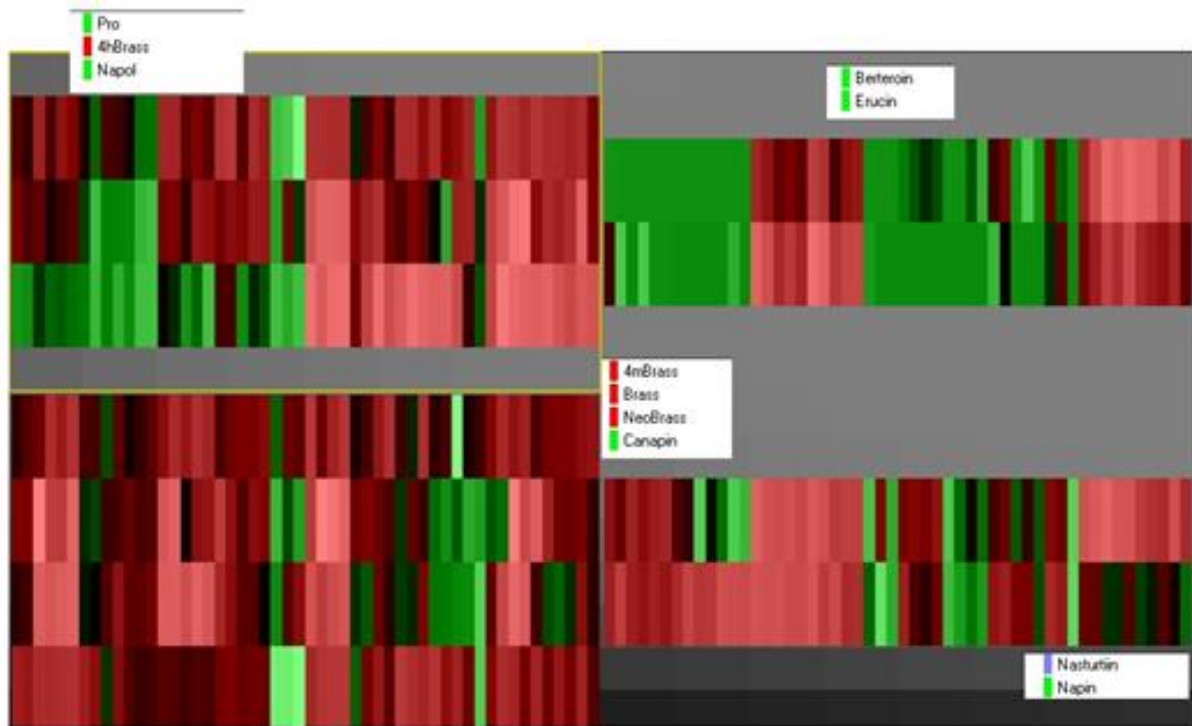


Figure 43. 2 x 2 SOM for GLS profile

Discussion

GLS distribution

The GLS distribution across the plant is mainly determined by the plant organ (Figure 6). ERU, BER and to a lesser degree, NAS, are turnip-specific GLS, showing very little presence in other plant tissues, regardless of the developmental stage of the plant. Given the GLS distribution profiles, the main difference in terms of tissue distribution seems to be between above-ground tissues and below-ground tissues for both AGLS and IGLS. In Figure 17, seed tissues are clustered together with turnip tissues of FT-004 in terms of GLS profile. In terms of gene expression, Figure 18 shows both seed tissue samples clustered together with turnip in T2 and T3. This suggests that seeds also become GLS sink tissues like turnips in T2 and T3. This is in accordance with Andersen and Halkier (2014), who demonstrated in *Arabidopsis* that, upon bolting, seeds become the sink tissue instead of the roots. Even though turnips are not roots, these are reserve organs.

Another factor that determines variation in GLS is the accession. It is seen that some GLS are significantly more present in one accession than in the other in Table 6. According to the current dataset, this should not be surprising, given the fact that FT-004 and FT-086 were chosen for this project because of the differences of their GLS profiles (Lee *et al*, 2013). Differences between GLS profiles from diverse genotypes have been reported in previous studies in Pak Choi (Wiesner *et al*, 2013), in *Arabidopsis* (Kliebenstein *et al*, 2001), and in *B. rapa* by Lou *et al* (2008). In the study carried out by Wiesner *et al* (2013), the differences in GLS profile were due to variations in the aliphatic glucosinolates profile, especially in differences of levels of PRO, NAP and the presence or absence of NAPOL. In Figure 12, NAP and NAPOL are at opposite extremes of the second principal component, driving the variation across this component. Therefore, it seems that NAP and NAPOL make the difference in the GLS profile of each accession. Kliebenstein *et al* (2001) analysed the GLS profiles of 39 *Arabidopsis* ecotypes from diverse environments, finding a large variability of GLS profiles among these genotypes. In their results, the variability of GLS profiles is originated in three loci: GS-*Elong*, GS-AOP and GS-OH, with more than 60% of the variation due to the GS-OH locus. This can be related with the high positive correlation between GS-OH_1 and NAPOL ($r = 0.56$, $p = 0.007$), as well as the negative correlation between NAP and GS-OH_1 ($r = -0.74$, $p < 0.001$), as NAP is the precursor of PRO. Kliebenstein *et al* (2001), also hypothesize that the wide diversity of GLS profiles among the *Arabidopsis* accessions is in response to the pressure of different herbivore agents present in each environment. Lou *et al* (2008), identified several QTLs for accumulation of GLS in leaves in two double haploid (DH) populations in two different growth seasons.

Transcription factors as genes of major influence in GLS biosynthetic pathways

MYB28 – MYB29

MYB28 transcription factor paralogs are significantly correlated with the major AGLS biosynthetic genes. Such are the positive correlations of MYB28_2 with MAM paralogs ($r = 0.48 - 0.63$), except with MAM3_2. However, the only significant correlations of MYB28_1 with MAM paralogs are with MAM3_2 and MAM1_2 ($r=0.44$ and 0.48 , respectively). Both MYB28 paralogs have significant positive correlations with all AOP2 paralogs, ranging between $r=0.52$ and $r=0.72$, as well as with CYP83_A1 paralogs (r between 0.54 and 0.72). In a previous study, the importance of MYB28 and MYB29 had been proved when, upon the construction of a *myb28myb29* double mutant in *Arabidopsis*, the plant was devoid of AGLS (Beekwilder *et al*, 2008). The similarity of the expression patterns of MYB28, MAM1, MAM3 and AOP2 is illustrated in Figure 16, where these paralogs are grouped together in the high RGE hotspot highlighted.

However, there are no significant correlations between MYB28 or MYB29 paralogs with any GS-OH paralog. It is also important to point that no significant correlations were found between MYB28 paralogs and AGLS. MYB29 shows no significant correlations to any AGLS whatsoever. Even though the expression of MYB29 is negatively correlated to both MYB28 paralogs, no significantly negative correlations between MYB29 and any AGLS biosynthetic gene are observed. Likewise, only 11 out of a total of 174 correlations between AGLS and AGLS biosynthetic genes are statistically significant. It is also interesting to observe in Figure 44 that while MYB28 paralogs are at their lowest expression in turnips, ERU and BER at their highest relative peak area. This phenomenon of no correlation between the expression of transcription factors and biosynthetic genes and GLS has been observed

by S nderby *et al* (2010b), who suggested that the metabolic chemotype is decoupled from the level of biosynthetic transcripts. They also point to the fact that there may be other mechanisms besides direct transcriptional regulation that have an influence on the AGLS profile. Possible mechanisms could be transport to other organs of the plant, as well as translational efficiency, protein stability and alternative splicing. Further studies on GLS biosynthesis and profile should consider these factors.

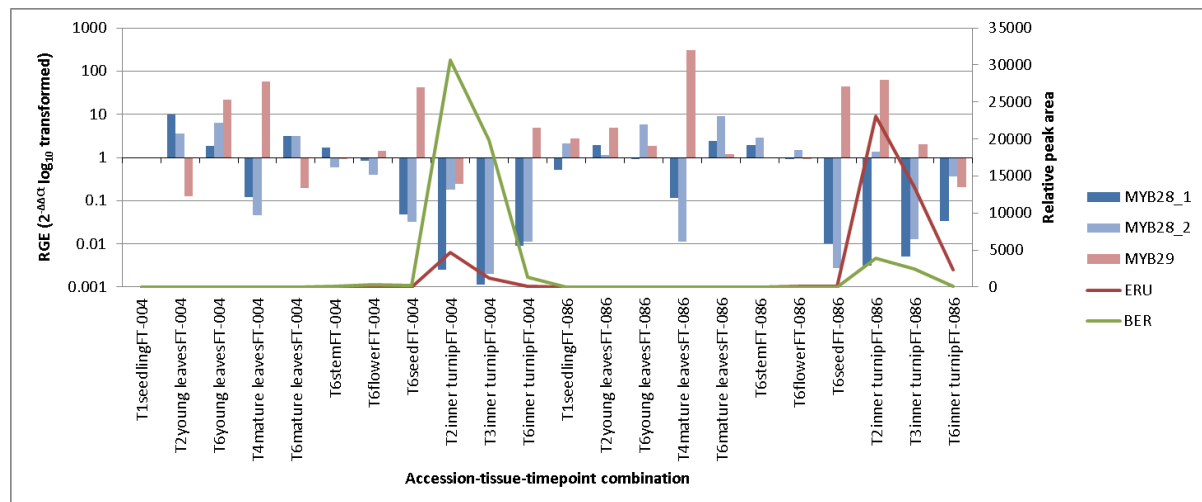


Figure 44. GLS profile versus gene expression. MYB28 and MYB29 transcription factors vs ERU and BER.

Besides these factors, there can be other gene expression factors that can explain the observed lack of correlation between MYB transcription factors and AGLS presence. An example of this is that there are recently identified transcription factors that are described as having a role in the GLS biosynthetic gene expression. Such is the case of the basic Helix-Loop-Helix (bHLH) transcription factors MYC2, MYC3, and MYC4 (Schweizer *et al*, 2013). It is suggested that these transcription factors form complexes with every known GLS-biosynthesis related transcription factor (MYB), thus playing a role in GLS biosynthesis regulation. Schweizer *et al* (2013) also demonstrated that MYC2, MYC3 and MYC4 can also play a role in the expression of GLS biosynthetic genes by binding to their promoter at a specific motif. However, the number of genes regulated by MYC2, MYC3 and MYC4 is unknown.

MYB34 - MYB51

CYP83B1, which metabolized both indolic and aromatic oximes at the core structure formation phase (Naur *et al*, 2003) is positively correlated to MYB34_2, MYB34_3, MYB34_4, and MYB51_1 ($r=0.79, 0.43, 0.49, 0.81$, respectively). In the case of CYP79B2, it is only positively correlated to MYB34_3 and MYB51_1 ($r = 0.63$ and 0.49 , respectively). Unlike in the case of the AGLS being studied in this project, there is positive correlation between transcription factors and IGLS. BRA is correlated to MYB34_1, MYB34_2, and MYB34_4 ($r = 0.6, 0.46$ and 0.63 , respectively). NBRA is only correlated to MYB34_4 ($r = 0.48$), 4HBRA is correlated to both MYB34_1 and MYB34_4 ($r = 0.52$ and 0.43), and 4MBRA is only correlated to MYB51_2 ($r=0.47$). However, these correlation coefficients are not so high, taking into account the role of MYB34 and MYB51 in the activation of the IGLS pathway, so there might be other factors that have an impact in the amount of IGLS. Triple mutants of *myb34myb51myb122* in *Arabidopsis* are devoid of IGLS (Frerigmann and Gigolashvili, 2014), showing the importance of these transcriptions factors in the activation of the IGLS pathway. Just as

in the case of AGLS, MYC2, MYC3, and MYC4 transcriptions factors can also be involved in the IGLS biosynthetic genes expression. Frerigmann (2014), identified bHLH05 as a transcription factor that interacts with MYB51, as well as bHLH04, bHLH05 and bHLH06/MYC2 as regulators of GLS biosynthesis in *Arabidopsis*. As in the AGLS biosynthesis, factors such as transport processes, translational efficiency, protein stability and differential splicing may play a role in the amount of synthesized AGLS.

GTR genes activity

GTR1 and GTR2, which are involved in the bidirectional transport of AGLS, tend to be highly expressed in seeds and in early developmental stages in turnips (T2). GTR1_3 is up to 10,000-fold higher expressed in the seed stage than in the seedling stage of FT-004. GTR2 paralogs, especially GTR2_2 and GTR2_4 are also highly active in the seed tissue. However, all GTR2 paralogs reach their peak of activity in the turnip at T2 in both accessions. In turnips, GTR1_1 and GTR1_2 are highly expressed at T2, and then their expression decreases in time in this tissue. The fact that GTR2 paralogs follow a relatively similar expression pattern in inner turnips (T2-T3), as well as in young and mature leaves in T6 and seeds in T6, points at the possibility that AGLS transport processes from turnips to leaves and seeds occur at these developmental stages. Although there is a possibility that gene expression of GTR2_4 in flowers of FT-004 is lower than the detection threshold (more than 36 cycles), it would be good to carry out the experiment again. In this case, it was not possible due to time constraints.

The high expression of GTR1_3 in seed tissue is in accordance with the role of GTR1 and GTR2 as glucosinolate importers (Nour-Eldin *et al*, 2012). This GTR1 and GTR2-mediated transport of GLS to the seeds occurs upon bolting (Andersen and Halkier, 2014). Unlike another studies carried out in *Brassica* species, the seeds were not the most GLS-rich tissues in this study (Brown *et al*, 2003; Kim *et al*, 2013). However the fact that GTR transporters could have other roles besides GLS transport also has to be considered. An example of this is that Saito *et al* (2015) reported that GTR1 may also be involved in the transport of hormones such as gibberellin in *Arabidopsis*.

Given the missing data points in the GTR3 paralogs RGE, it is difficult to make any conclusions about their expression profile and relationship with another genes and GLS. Definitely, further studies will be needed in order to assess more in depth the behaviour of GTR3 paralogs in *B. rapa*.

MAM genes role in AGLS profile.

In the two accessions being assessed in this project, NAPOL, the 5C GLS that is a product of GSOH conversion of CAN, is in higher amounts in FT-004. At the same time, MAM3_2 and GS-OH_1 are higher expressed in FT-004 than in FT-086. As there is a significant strong positive correlation between GS-OH_1 and MAM3_2 ($r=0.69$, $p < 0.001$), and between NAPOL and GS-OH_1 ($r=0.56$, $p = 0.007$), The fact that a similar correlation type is not observed with PRO and GS-OH is unexpected since GS-OH in *Arabidopsis* is responsible for the biosynthesis of PRO by means of the hydroxylation of 3-butenyl-glucosinolate to form 2-hydroxybut-3-enyl glucosinolate (PRO) (Hansen *et al*, 2008). Therefore, MAM3_1 and GS-OH_1 might be key regulators of NAPOL biosynthesis. Figure 45 shows the RGE of MAM3_2, GS-OH_1 and the quantitative variation in NAPOL. The peaks of presence of NAPOL are related with higher RGE in FT-004 tissues.

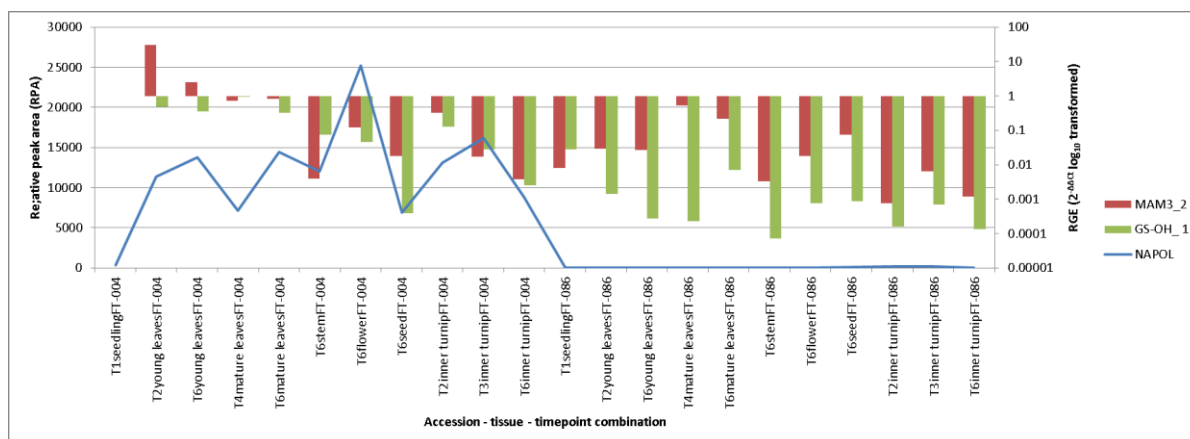


Figure 45. Relationship between RGE of MAM3_2 and GS-OH_1 with relative presence of NAPOL.

All 5C AGLS in this study are significantly higher expressed overall in FT-004 compared to FT086 (BER, CAN and NAPOL, Table 6). Interestingly, neither BER nor CAN show any significant correlation to any MAM gene, which have a major role in chain elongation..

In fact, in a previous study (Hirani *et al*, 2013) where homologous GSL-ELONG genes were replaced in *B. rapa* by the *B. oleracea* non-functional GSL-ELONG gene, the SCAR marker BraMAM1-1 was used. This marker is related to the presence of 5C AGLS. BraMAM1-1 amplifies a region that is within MAM3_2 gene (as per primer BLAST analysis and review in the *B. rapa* genome database BRAD, version 1.5). The physical proximity of MAM3_2 to MAM1_2 and MAM3_1 may lead to think that these genes tend to be coexpressed. However, in this study this does not seem to be the case, since MAM3_2 has a different expression profile across both accessions than the other MAM genes, especially in leaves (Figure 46). Furthermore, MAM3_2 has no correlations with most MAM genes, except for MAM1_2 and MAM3_1, though the correlation coefficients are not remarkably high (0.53, $p = 0.012$ and 0.4, $p = 0.066$, respectively). Further studies with more genotypes and a larger number of samples with qRT-PCR expression will be required in order to assess the role of MAM3_2 as the single responsible gene related to the 5C AGLS presence in *B. rapa*, as well as to assess the substrate-specificity of GS-OH_1 for CAN. These studies, besides considering several 4C and 5C AGLS-rich accessions, should focus mainly on leaf tissue, given the higher activity in leaves of MAM genes.

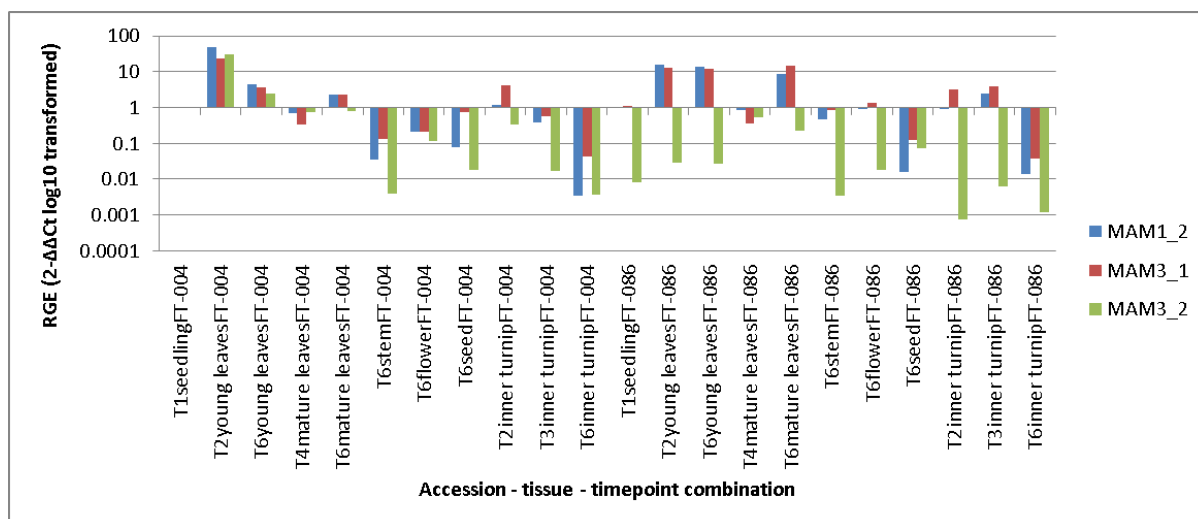


Figure 46. RGE patterns of MAM3_2, MAM3_1 and MAM1_2. MAM3_2 (in green) is more expressed in leaves of FT-004 than in FT-086.

AOP2 genes are mostly expressed in leaves and correlations with other AGLS biosynthetic genes

As it can be seen in Figures 21 and 22, AOP2 genes are mostly upregulated in leaves. This is in accordance with the characterization of the AOP2 gene in *Arabidopsis* by Neal *et al* (2010). Here, AOP2 was found to be most abundant in leaf and stem tissues. In addition, that study also found that the AOP2 expression is light dependent, since there are light regulatory elements that were identified in the promoter region of the gene. They also found that the level of transcript of AOP2 decreases rapidly with the quantity of light. This is further proof that there are other factors besides biosynthetic gene expression influencing GLS profile.

The high correlations between AOP genes, MAM, CYP83A1 and MYB28 shown in the correlation matrix in Table 7 can also be seen in the PCA analysis for gene expression in Figures 13 and 14, where in the loading plot these genes sit together in an area that is homologous to where leaf tissues sit together in the score plot. The high RGE hotspot in the hierarchical clustering heat map in Figure 15 is mostly comprised of leaf tissue. However, GS-OH genes are not strongly correlated to the rest of the genes in the AGLS pathway in this study, as they are placed in the loading plot in Figure 14 in an area that is homologous to where turnip tissues sit together in the score plot. This suggests that PRO and NAPOL can be similarly synthesized in both leaves and turnips, whereas ERU and BER are mostly synthesized in leaves and then transported to the turnips where import is carried out by GTR1 and GTR2. NAP and CAN may also be more synthesized in leaves, as suggested by the PCA analysis, and then be translocated to the turnips for further transformation into PRO and NAPOL, respectively.

Conclusions

- Tissue seems to be the most important factor related to both the GLS profile and the gene expression profile.
- The genotype is also an important factor in the GLS profile, though in a lesser degree than the type of tissue where GLS are present.
- GLS relative presence and gene expression are most often not correlated. There might be other factors that play a role in the amount of biosynthesized glucosinolates, such as transport processes, translational efficiency, protein stability and differential splicing. There are also other transcription factors recently discovered that also have an influence in the activation of the biosynthetic pathways and further studies will be required in order to fully elucidate their role.
- Leaves are tissues with high AGLS biosynthetic gene activity.
- Turnips are tissues with a very specific GLS profile. ERU, BER and NAS to a lesser degree are turnip specific compounds. This profile could be explained due to the action of GTR genes that import ERU and BER into turnips
- The action of MAM3_2 (BRA013009) may have effects on the GLS profile by the addition of a 5th C to the side chain of deaminated Met.

Recommendations

- In order to fully validate the GTR gene expression study, as well as for the rest of the GLS biosynthetic genes paralogs, it would be advisable to sequence the amplicons and then blast them to the *B. rapa* genome. This will be a sound proof that the correct genes were evaluated.
- Given the fact that there were missing data points for GTR2_4, GTR3_1, GTR3_2, GTR3_3, it is suggested to repeat the qRT-PCR experiment in order to better evaluate their RGE profile. It is also advisable to have more gene expression samples, especially from T3 and T5, since it would have allowed to evaluate the RGE patterns of side chain elongation, core structure formation and secondary motivation with better precision and to study the GTR genes behaviour in depth. It will also allow seeing if the developmental stage does not really have a

role in the GLS profile since in this case, T6 is overrepresented if compared to the other time points. It is also be advisable to redefine sampling time points and set them to more regular intervals in order to have a better view of how both GLS profile and gene expression evolve over time, since in this case, there are almost 7 months between T5 and T6.

- Although the AGLS biosynthetic gene selection for this study covers all stages of the AGLS biosynthetic pathway, the IGLS gene selection could be covered in more detail. It would be advisable to perform RGE studies on genes that are involved in secondary modifications of IGLS after BRA is synthesized, such as CYP81F2, CYP81F3, CYP81F4, which are involved in the conversion of BRA to 4HBRA or NBRA, as well as IGMT1 and IGMT2, which convert 4HBRA to 4MBRA (Pfalz *et al*, 2009; Pfalz *et al*, 2011). This will allow having a better view of where IGLS are biosynthesized.
- It would be interesting to evaluate the role of MAM3_2 (BRA013009) in the change from a 4C dominant to a 5C dominant AGLS profile. This could possibly be carried out with a mutant of MAM3_2 in FT-004, or any other accession rich in 5C AGLS compared to a wild type control. Then, LC-MS or another suitable technique can be applied on leaf and turnip tissues before bolting, since GLS will tend to go to the tissues after bolting.

References

- Andersen, T. G. and B. A. Halkier (2014). "Upon bolting the GTR1 and GTR2 transporters mediate transport of glucosinolates to the inflorescence rather than roots." *Plant Signal Behavior* 9(1).
- Andersen, T. G., et al. (2013). "Integration of biosynthesis and long-distance transport establish organ-specific glucosinolate profiles in vegetative Arabidopsis." *Plant Cell* 25(8): 3133-3145.
- Bak, S. and R. Feyereisen (2001). "The Involvement of Two P450 Enzymes, CYP83B1 and CYP83A1, in Auxin Homeostasis and Glucosinolate Biosynthesis." *Plant Physiol* 127(1): 108-118.
- Bak, S., et al. (2001). "CYP83B1, a Cytochrome P450 at the Metabolic Branch Point in Auxin and Indole Glucosinolate Biosynthesis in Arabidopsis." *The Plant Cell Online* 13(1): 101-111.
- Beekwilder, J., et al. (2008). "The impact of the absence of aliphatic glucosinolates on insect herbivory in Arabidopsis." *PLoS ONE* 3(4): e2068.

- Celenza, J. L., et al. (2005). "The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate homeostasis." *Plant Physiol* 137(1): 253-262.

- Chen, S., et al. (2003). "CYP79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in Arabidopsis." *The Plant Journal* 33(5): 923-937.

- Cheng, F., et al. (2011). "BRAD, the genetics and genomics database for Brassica plants." *BMC Plant Biol* 11: 136.

- Dubos, C., et al. (2010). "MYB transcription factors in Arabidopsis." *Trends Plant Sci* 15(10): 573-581.

- Fahey, J. W., et al. (2001). "The chemical diversity and distribution of glucosinolates and isothiocyanates among plants." *Phytochemistry* 56(1): 5-51.

- Fang, J., et al. (2012). "Tissue-specific distribution of secondary metabolites in rapeseed (*Brassica napus* L.)." *PLoS ONE* 7(10): e48006.

- Frerigmann, H. and T. Gigolashvili (2014). "MYB34, MYB51 and MYB122 Distinctly Regulate Indolic Glucosinolate Biosynthesis in Arabidopsis thaliana." *Molecular Plant*.

- Gigolashvili, T., et al. (2007)a. "The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in Arabidopsis thaliana." *Plant J* 51(2): 247-261.

- Gigolashvili, T., et al. (2007)b. "The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana." *Plant J* 50(5): 886-901.

- Gigolashvili, T., et al. (2008). "Specific and coordinated control of indolic and aliphatic glucosinolate biosynthesis by R2R3-MYB transcription factors in Arabidopsis thaliana." *Phytochemistry Reviews* 8(1): 3-13.

- Giovannucci, E., et al. (2003). "A Prospective Study of Cruciferous Vegetables and Prostate Cancer." *Cancer Epidemiology Biomarkers & Prevention* 12(12): 1403-1409.

- Gómez-Campo, C. and S. Prakash (1999). "2 Origin and domestication." *Developments in plant genetics and breeding* 4: 33-58.

- Grubb, C. D., et al. (2004). "Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis." *Plant J* 40(6): 893-908.
- Grubb, C. D., et al. (2014). "Comparative analysis of Arabidopsis UGT74 glucosyltransferases reveals a special role of UGT74C1 in glucosinolate biosynthesis." *Plant J* 79(1): 92-105.
- Halkier, B. A. and J. Gershenzon (2006). "BIOLOGY AND BIOCHEMISTRY OF GLUCOSINOLATES." *Annual Review of Plant Biology* 57(1): 303-333.
- Hansen, B. G., et al. (2007). "Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in Arabidopsis." *Plant J* 50(5): 902-910.
- Hansen, B. G., et al. (2008). "A novel 2-oxoacid-dependent dioxygenase involved in the formation of the goiterogenic 2-hydroxybut-3-enyl glucosinolate and generalist insect resistance in Arabidopsis." *Plant Physiol* 148(4): 2096-2108.
- Hansen, C. H., et al. (2001). "Cytochrome p450 CYP79F1 from arabidopsis catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates." *J Biol Chem* 276(14): 11078-11085.
- Hirai, M. Y., et al. (2007). "Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis." *Proc Natl Acad Sci U S A* 104(15): 6478-6483.
- Hull, A. K., et al. (2000). "Arabidopsis cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis." *Proc Natl Acad Sci U S A* 97(5): 2379-2384.
- Keum, Y. S., et al. (2004). "Chemoprevention by isothiocyanates and their underlying molecular signaling mechanisms." *Mutat Res* 555(1-2): 191-202.
- Klein, M. and J. Papenbrock (2009). "Kinetics and substrate specificities of desulfo-glucosinolate sulfotransferases in Arabidopsis thaliana." *Physiol Plant* 135(2): 140-149.
- Kliebenstein, D. J., et al. (2001). "Genetic control of natural variation in Arabidopsis glucosinolate accumulation." *Plant Physiol* 126(2): 811-825.
- Knill, T., et al. (2008). "Arabidopsis branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis." *Plant Physiol* 146(3): 1028-1039.

- Knill, T., et al. (2009). "Arabidopsis thaliana encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation." *Plant Mol Biol* 71(3): 227-239.

- Koroleva, O. A., et al. (2000). "Identification of a New Glucosinolate-Rich Cell Type in Arabidopsis Flower Stalk." *Plant Physiol* 124(2): 599-608.

- Kroymann, J., et al. (2001). "A Gene Controlling Variation in Arabidopsis Glucosinolate Composition Is Part of the Methionine Chain Elongation Pathway." *Plant Physiol* 127(3): 1077-1088.

- Lee, J. G., et al. (2013). "Evaluation of glucosinolate variation in a collection of turnip (*Brassica rapa*) germplasm by the analysis of intact and desulfo glucosinolates." *J Agric Food Chem* 61(16): 3984-3993.

- Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻($\Delta\Delta C_T$) Method." *Methods* 25(4): 402-408.

- Lou, P., et al. (2008). "Quantitative trait loci for glucosinolate accumulation in *Brassica rapa* leaves." *New Phytol* 179(4): 1017-1032.

- Mikkelsen, M. D., et al. (2000). "Cytochrome P450 CYP79B2 from Arabidopsis catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid." *J Biol Chem* 275(43): 33712-33717.

- Mikkelsen, M. D., et al. (2004). "Arabidopsis mutants in the C–S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis." *The Plant Journal* 37(5): 770-777.

- Mithen, R. (2001). "Glucosinolates – biochemistry, genetics and biological activity." *Plant Growth Regulation* 34(1): 91-103.

- Murillo, G. and R. G. Mehta (2001). "Cruciferous Vegetables and Cancer Prevention." *Nutrition and Cancer* 41(1-2): 17-28.

- Naur, P. (2003). "CYP83A1 and CYP83B1, Two Nonredundant Cytochrome P450 Enzymes Metabolizing Oximes in the Biosynthesis of Glucosinolates in Arabidopsis." *Plant Physiol* 133(1): 63-72.

- Neal, C. S., et al. (2010). "The characterisation of AOP2: a gene associated with the biosynthesis of aliphatic alkenyl glucosinolates in *Arabidopsis thaliana*." *BMC Plant Biol* 10: 170.
- Nour-Eldin, H. H., et al. (2012). "NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds." *Nature* 488(7412): 531-534.
- Park, J. Y., et al. (2005). "Physical mapping and microsynteny of *Brassica rapa* ssp. *pekinensis* genome corresponding to a 222 kbp gene-rich region of *Arabidopsis* chromosome 4 and partially duplicated on chromosome 5." *Mol Genet Genomics* 274(6): 579-588.
- Petersen, B., et al. (2002). "Composition and content of glucosinolates in developing *Arabidopsis thaliana*." *Planta* 214(4): 562-571.
- Pfalz, M., et al. (2009). "The gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in *Arabidopsis*." *Plant Cell* 21(3): 985-999.
- Pfalz, M., et al. (2011). "Metabolic engineering in *Nicotiana benthamiana* reveals key enzyme functions in *Arabidopsis* indole glucosinolate modification." *Plant Cell* 23(2): 716-729.
- Rakow, G. (2004). Species origin and economic importance of *Brassica*. *Brassica*, Springer: 3-11.
- Saito, H., et al. (2015). "The jasmonate-responsive GTR1 transporter is required for gibberellin-mediated stamen development in *Arabidopsis*." *Nat Commun* 6: 6095.
- Sawada, Y., et al. (2009). "Omics-based approaches to methionine side chain elongation in *Arabidopsis*: characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase." *Plant Cell Physiol* 50(7): 1181-1190.
- Schuster, J., et al. (2006). "Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*." *Plant Cell* 18(10): 2664-2679.
- Schweizer, F., et al. (2013). "Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior." *Plant Cell* 25(8): 3117-3132.

- Shapiro, T. A., et al. (1998). "Human metabolism and excretion of cancer chemoprotective glucosinolates and isothiocyanates of cruciferous vegetables." *Cancer Epidemiology Biomarkers & Prevention* 7(12): 1091-1100.
- Smilde, A. K., et al. (2005). "Fusion of Mass Spectrometry-Based Metabolomics Data." *Analytical Chemistry* 77(20): 6729-6736.
- Sønderby, I. E., et al. (2007). "A Systems Biology Approach Identifies a R2R3 MYB Gene Subfamily with Distinct and Overlapping Functions in Regulation of Aliphatic Glucosinolates." *PLoS ONE* 2(12): e1322.
- Sønderby, I. E., et al. (2010)a. "Biosynthesis of glucosinolates--gene discovery and beyond." *Trends Plant Sci* 15(5): 283-290.
- Sønderby, I. E., et al. (2010)b. "A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis." *Plant Physiol* 153(1): 348-363.
- Talalay, P. (1999). "The War against Cancer: New Hope." *Proceedings of the American Philosophical Society* 143(1): 52-72.
- Terry, P., et al. (2001). "Brassica vegetables and breast cancer risk." *JAMA* 285(23): 2975-2977.
- Textor, S., et al. (2007). "MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in Arabidopsis." *Plant Physiol* 144(1): 60-71.
- U, N. 1935. Genomic analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* 7: 389–452
- van den Berg, R. A., et al. (2006). "Centering, scaling, and transformations: improving the biological information content of metabolomics data." *BMC Genomics* 7: 142.
- Wang, X., et al. (2011). "The genome of the mesopolyploid crop species *Brassica rapa*." *Nat Genet* 43(10): 1035-1039.
- Wang, H., et al. (2011). "Glucosinolate biosynthetic genes in *Brassica rapa*." *Gene* 487(2): 135-142.

- Wiesner, M., et al. (2013). "Genotypic variation of the glucosinolate profile in pak choi (*Brassica rapa* ssp. *chinensis*).
J Agric Food Chem 61(8): 1943-1953.
- Wittstock, U. and B. A. Halkier (2002). "Glucosinolate research in the *Arabidopsis* era."
Trends Plant Sci 7(6): 263-270.
- Yang, G., et al. (2010). "Isothiocyanate exposure, glutathione S-transferase polymorphisms, and colorectal cancer risk." The American Journal of Clinical Nutrition 91(3): 704-711.
- Zhang, N., et al. (2014). "Morphology, carbohydrate composition and vernalization response in a genetically diverse collection of Asian and European turnips (*Brassica rapa* subsp. *rapa*)."
PLoS ONE 9(12): e114241.

Appendix 1: LC-MS data

Appendix 1, Table 1. LC-MS raw data. PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucoberteroin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Timep	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Eruc	Brass	Naturtiin	Bertero	4mBras	NeoBras
July 26, 2010	FT-004	T1	seedling	1*FT-004*TYL*1		522.7	348.2	273.5	4035.4	1195.2	0.0	1784.3	102.8	0.0	700.5	947.5
July 26, 2010	FT-004	T1	seedling	2*FT-004*TYL*2		214.7	250.1	0.2	3803.8	307.7	0.0	2288.0	30.0	0.0	401.8	1219.6
July 26, 2010	FT-004	T1	seedling	3*FT-004*TYL*3		438.6	543.8	26.1	5299.2	771.4	0.0	2777.3	83.5	0.0	688.4	863.1
July 26, 2010	FT-004	T1	seedling	FT-004*TYL*	AVG	392.0	380.7	99.9	4379.5	758.1	0.0	2283.2	72.1	0.0	596.9	1010.1
					SE	91.9	86.3	87.1	464.7	256.3	0.0	286.7	21.8	0.0	97.6	107.6
					CV	0.4	0.4	1.5	0.2	0.6	#DIV/0!	0.2	0.5	#DIV/0!	0.3	0.2
August 16, 2010	FT-004	T2	young leaves	4*FT-004*IL*1		11556.6	14024.0	1706.0	34702.2	35234.0	0.0	12854.8	9729.0	0.7	2390.9	5359.9
August 16, 2010	FT-004	T2	young leaves	5*FT-004*IL*2		14064.7	11473.6	5181.8	23628.8	41448.9	0.0	8182.0	4724.0	18.9	1500.8	9759.9
August 16, 2010	FT-004	T2	young leaves	6*FT-004*IL*3		9190.9	8731.7	620.3	34542.8	24084.4	0.0	23354.4	4421.0	0.4	2402.9	12882.3
August 16, 2010	FT-004	T2	young leaves	FT-004*IL*	AVG	11604.0	11409.7	2502.7	30957.9	33589.1	0.0	14797.0	6291.3	6.7	2098.2	9334.0
					SE	1407.2	1528.1	1375.7	3664.8	5079.7	0.0	4486.3	1721.1	6.1	298.7	2181.9
					CV	0.2	0.2	1.0	0.2	0.3	#DIV/0!	0.5	0.5	1.6	0.2	0.4
August 16, 2010	FT-004	T2	mature leaves	7*FT-004*OL*1		437.6	4889.4	101.4	8272.4	4361.3	0.0	4620.1	308.9	0.0	581.1	513.7
August 16, 2010	FT-004	T2	mature leaves	8*FT-004*OL*2		2149.4	6306.1	410.6	8061.5	18997.8	0.0	1897.3	793.8	0.0	760.8	745.4
August 16, 2010	FT-004	T2	mature leaves	9*FT-004*OL*3		516.1	3577.2	59.9	4828.9	1781.9	0.0	1995.7	263.7	0.0	1921.7	757.7
August 16, 2010	FT-004	T2	mature leaves	FT-004*OL*	AVG	1034.4	4924.2	190.6	7054.3	8380.3	0.0	2837.7	455.5	0.0	1087.8	672.3
					SE	558.0	788.0	110.7	1114.3	5360.7	0.0	891.7	169.7	0.0	420.1	79.3
					CV	0.9	0.3	1.0	0.3	1.1	#DIV/0!	0.5	0.6	#DIV/0!	0.7	0.2
August 16, 2010	FT-004	T2	innerturnip	10*FT-004*IR*1		14156.6	11918.9	1530.5	5308.8	19258.5	2588.3	11834.7	32002.7	32890.5	811.8	2507.3
August 16, 2010	FT-004	T2	innerturnip	11*FT-004*IR*2		15258.7	20136.0	1051.5	5029.1	23726.3	2268.6	6394.5	27446.9	28967.1	1443.7	1745.6
August 16, 2010	FT-004	T2	innerturnip	12*FT-004*IR*3		13209.5	7161.0	5686.5	3602.8	24786.4	9138.3	8509.5	35242.7	30033.9	1394.8	1320.2
August 16, 2010	FT-004	T2	innerturnip	FT-004*IR*	AVG	14208.3	13072.0	2756.2	4646.9	22590.4	4665.1	8912.9	31564.1	30630.5	1216.8	1857.7
					SE	592.1	3789.7	1471.7	528.2	1693.8	2238.5	1583.4	2261.1	1171.2	203.0	347.2
					CV	0.1	0.5	0.9	0.2	0.1	0.8	0.3	0.1	0.1	0.3	0.3
August 16, 2010	FT-004	T2	outerturnip	13*FT-004*OR*1		11146.9	12639.9	1121.5	41882.2	18083.9	2859.6	13218.5	47739.2	27878.9	4187.7	20009.8
August 16, 2010	FT-004	T2	outerturnip	14*FT-004*OR*2		13709.4	22401.4	841.7	43139.6	24434.2	3129.2	13873.0	45288.1	28601.6	5618.8	18287.0
August 16, 2010	FT-004	T2	outerturnip	15*FT-004*OR*3		14026.0	22803.6	886.5	43100.8	24850.1	3305.9	14564.7	47832.5	29716.5	6232.0	19574.0
August 16, 2010	FT-004	T2	outerturnip	FT-004*OR*	AVG	12960.8	19281.6	949.9	42707.5	22456.1	3098.2	13885.4	46953.3	28732.3	5346.2	19290.3
					SE	911.5	3322.9	86.8	412.8	2189.4	129.7	388.7	833.0	534.5	605.7	517.2
					CV	0.1	0.3	0.2	0.0	0.2	0.1	0.0	0.0	0.0	0.2	0.0
August 30, 2010	FT-004	T3	young leaves	16*FT-004*IL*1		9058.2	18788.2	3037.9	35021.3	31510.9	0.1	16412.1	4578.5	11.0	9178.5	12540.3
August 30, 2010	FT-004	T3	young leaves	17*FT-004*IL*2		11022.4	19419.5	1540.6	30261.3	40966.2	0.0	5298.8	5673.9	13.4	11045.3	8035.5
August 30, 2010	FT-004	T3	young leaves	18*FT-004*IL*3		12375.5	26370.6	565.4	27752.5	33913.3	0.0	15904.4	4674.0	25.1	4505.4	16456.0
August 30, 2010	FT-004	T3	young leaves	FT-004*IL*	AVG	10818.7	21526.1	1714.6	31011.7	35463.5	0.0	12538.4	4975.5	16.5	8243.1	12343.9

Appendix 1, Table 2. LC-MS data (Continued). PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucoberteroin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Time point	Tissue	qt of name	Peak Area	Pro	Napol	Napin	4hBras	Canapin	Erucin	Brass	Naturtiin	Berteroin	4mBras	NeoBras
					SE	963.0	2429.1	719.0	2131.6	2837.4	0.0	3622.8	350.3	4.4	1945.0	2432.8
					CV	0.2	0.2	0.7	0.1	0.1	1.7	0.5	0.1	0.5	0.4	0.3
August 30, 2010	FT-004	T3	mature leaves	19*FT-004*OL*1		5992.0	24383.8	320.7	13658.4	22738.6	0.0	1758.2	2359.9	18.3	6297.1	2305.3
August 30, 2010	FT-004	T3	mature leaves	20*FT-004*OL*2		1788.7	17058.4	560.1	5089.0	5495.6	0.0	1964.6	603.6	0.0	3632.7	5823.0
August 30, 2010	FT-004	T3	mature leaves	21*FT-004*OL*3		3334.1	25999.7	259.0	11129.0	25592.9	0.0	2837.2	1450.2	0.5	8322.9	825.3
August 30, 2010	FT-004	T3	mature leaves	FT-004*OL*	AVG	3704.9	22480.6	379.9	9958.8	17942.4	0.0	2186.7	1471.2	6.3	6084.2	2984.5
					SE	1227.5	2750.9	91.9	2542.0	6277.7	0.0	330.7	507.1	6.0	1358.1	1482.1
					CV	0.6	0.2	0.4	0.4	0.6	#DIV/0!	0.3	0.6	1.7	0.4	0.9
August 30, 2010	FT-004	T3	innerturnip	22*FT-004*IR*1		9968.0	13822.1	667.4	9690.4	6836.9	1302.1	4395.0	19736.0	20064.5	3097.6	879.2
August 30, 2010	FT-004	T3	innerturnip	23*FT-004*IR*2		11160.1	19104.4	622.5	7204.9	7958.3	895.8	3118.8	17900.4	19912.4	2651.1	662.2
August 30, 2010	FT-004	T3	innerturnip	24*FT-004*IR*3		11307.3	15675.9	1120.9	9177.3	7313.1	1418.9	4683.7	18925.8	19536.7	2752.2	868.6
August 30, 2010	FT-004	T3	innerturnip	FT-004*IR*	AVG	10811.8	16200.8	803.6	8690.9	7369.5	1205.6	4065.8	18854.1	19837.9	2833.6	803.3
					SE	424.0	1547.3	159.2	757.6	324.9	158.5	480.8	531.1	156.9	135.2	70.6
					CV	0.1	0.2	0.3	0.2	0.1	0.2	0.2	0.0	0.0	0.1	0.2
August 30, 2010	FT-004	T3	outerturnip	25*FT-004*OR*1		9035.3	15133.5	888.9	47061.3	8783.1	1420.6	5913.8	33584.4	17126.7	8086.7	14479.6
August 30, 2010	FT-004	T3	outerturnip	26*FT-004*OR*2		8687.0	14648.5	962.7	50081.2	8381.3	1227.2	6900.6	31506.4	15475.3	6808.2	15638.5
August 30, 2010	FT-004	T3	outerturnip	27*FT-004*OR*3		8919.1	15525.9	1111.9	47617.0	8902.0	1443.7	7366.2	29644.9	16060.9	8335.3	15055.9
August 30, 2010	FT-004	T3	outerturnip	FT-004*OR*	AVG	8880.5	15102.6	987.8	48253.1	8688.8	1363.8	6726.9	31578.6	16221.0	7743.4	15058.0
					SE	102.4	253.7	65.6	928.0	157.6	68.7	428.2	1137.8	483.4	473.1	334.5
					CV	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.0
September 16, 2010	FT-004	T4	young leaves	28*FT-004*IL*1		10966.2	27371.5	2841.9	28499.4	33800.0	0.1	5371.6	8211.5	10.1	12148.1	10269.4
September 16, 2010	FT-004	T4	young leaves	29*FT-004*IL*2		11694.3	23501.1	5901.9	29511.4	30653.1	0.0	16523.2	10175.5	9.9	15206.1	17732.8
September 16, 2010	FT-004	T4	young leaves	30*FT-004*IL*3		11212.8	20614.5	3457.3	24039.4	39706.3	0.0	6842.1	5560.9	1.3	14381.0	13630.8
September 16, 2010	FT-004	T4	young leaves	FT-004*IL*	AVG	11291.1	23829.0	4067.0	27350.1	34719.8	0.0	9579.0	7982.6	7.1	13911.7	13877.7
					SE	213.8	1957.5	934.5	1680.9	2653.6	0.0	3498.0	1337.0	2.9	913.4	2158.0
					CV	0.0	0.1	0.4	0.1	0.1	1.7	0.6	0.3	0.7	0.1	0.3
September 16, 2010	FT-004	T4	mature leaves	31*FT-004*OL*1		937.0	8303.1	3.9	2543.7	645.4	0.0	1383.6	168.8	0.9	10570.4	449.3
September 16, 2010	FT-004	T4	mature leaves	32*FT-004*OL*2		1257.4	3595.3	417.4	1699.9	12007.7	0.1	1775.8	495.9	0.0	13640.6	2774.9
September 16, 2010	FT-004	T4	mature leaves	33*FT-004*OL*3		988.5	9513.5	164.4	4399.2	4132.2	0.0	1893.7	583.4	0.0	11865.3	1384.1
September 16, 2010	FT-004	T4	mature leaves	FT-004*OL*	AVG	1061.0	7137.3	195.2	2880.9	5595.1	0.0	1684.4	416.0	0.3	12025.4	1536.1
					SE	99.3	1805.1	120.3	797.3	3360.6	0.0	154.2	126.2	0.3	889.9	675.6
					CV	0.2	0.4	1.1	0.5	1.0	1.7	0.2	0.5	1.7	0.1	0.8
September 16, 2010	FT-004	T4	innerturnip	34*FT-004*IR*1		10756.7	9906.9	1112.9	6569.7	18566.1	753.5	2407.7	13874.4	20226.6	3493.3	592.8
September 16, 2010	FT-004	T4	innerturnip	35*FT-004*IR*2		9610.0	7783.9	923.8	6217.7	16028.1	706.0	2308.2	14352.7	18943.7	3592.0	644.7

Appendix 1, Table 3. LC-MS data (Continued). PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucoberteroin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Time p	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Eruc	Bras	Naturti	Bertero	4mBras	NeoBras
September 16, 2010	FT-004	T4	innerturnip	36*FT-004*IR*3		10822.4	8761.5	1394.1	7463.0	17718.5	876.4	2427.5	15582.3	18488.4	2935.4	563.4
September 16, 2010	FT-004	T4	innerturnip	FT-004*IR*	AVG	10396.3	8817.4	1143.6	6750.2	17437.6	778.6	2381.1	14603.1	19219.6	3340.2	600.3
					SE	393.6	613.5	136.6	370.6	746.0	50.8	36.9	508.7	520.4	204.4	23.8
					CV	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.1	0.0	0.1	0.1
September 16, 2010	FT-004	T4	outerturnip	37*FT-004*OR*1		10960.0	11298.7	1609.6	32218.3	22698.5	1248.3	2812.3	27761.9	14022.5	3988.5	5255.1
September 16, 2010	FT-004	T4	outerturnip	38*FT-004*OR*2		10316.6	10431.9	1630.1	29421.1	22866.4	929.6	2502.8	27326.3	11141.2	4828.0	4899.0
September 16, 2010	FT-004	T4	outerturnip	39*FT-004*OR*3		10905.7	11081.2	1098.3	25935.5	20033.9	1059.6	2306.7	27711.3	13440.8	5474.8	3309.5
September 16, 2010	FT-004	T4	outerturnip	FT-004*OR*	AVG	10727.4	10937.3	1446.0	29191.6	21866.3	1079.2	2540.6	27599.8	12868.2	4763.8	4487.9
					SE	206.0	260.4	173.9	1817.3	917.4	92.5	147.2	137.6	879.7	430.2	598.1
					CV	0.0	0.0	0.2	0.1	0.1	0.1	0.1	0.0	0.1	0.2	0.2
September 29, 2010	FT-004	T5	young leaves	40*FT-004*IL*1		10428.2	29731.1	5019.3	17750.0	25909.5	0.0	7626.8	4711.2	3.1	14795.2	10299.2
September 29, 2010	FT-004	T5	young leaves	41*FT-004*IL*2		12598.7	18561.3	14760.9	15569.1	20402.1	0.0	2979.3	4351.4	1.8	15281.6	14684.7
September 29, 2010	FT-004	T5	young leaves	42*FT-004*IL*3		13864.4	28993.6	8380.6	23621.9	28001.2	0.1	6502.8	6177.0	1.0	13424.1	15131.6
September 29, 2010	FT-004	T5	young leaves	FT-004*IL*	AVG	12297.1	25762.0	9387.0	18980.3	24771.0	0.0	5703.0	5079.9	1.9	14500.3	13371.8
					SE	1003.3	3606.6	2856.8	2404.6	2266.3	0.0	1400.0	558.3	0.6	556.1	1541.7
					CV	0.1	0.2	0.5	0.2	0.2	1.7	0.4	0.2	0.6	0.1	0.2
September 29, 2010	FT-004	T5	mature leaves	43*FT-004*OL*1		3087.2	33133.9	96.6	4643.5	7268.6	0.0	2630.4	464.7	2.6	15526.0	8545.4
September 29, 2010	FT-004	T5	mature leaves	44*FT-004*OL*2		7146.5	3324.3	9574.1	501.2	34705.2	0.0	781.4	701.0	0.0	8835.8	1174.8
September 29, 2010	FT-004	T5	mature leaves	45*FT-004*OL*3		4436.9	13123.6	4271.5	1699.2	28627.3	0.0	1200.2	510.2	0.0	8783.8	1711.3
September 29, 2010	FT-004	T5	mature leaves	FT-004*OL*	AVG	4890.2	16527.3	4647.4	2281.3	23533.7	0.0	1537.4	558.6	0.9	11048.5	3810.5
					SE	1193.5	8772.0	2742.4	1230.7	8319.7	0.0	559.8	72.4	0.9	2238.8	2372.5
					CV	0.4	0.9	1.0	0.9	0.6	NDIV/0!	0.6	0.2	1.7	0.4	1.1
September 29, 2010	FT-004	T5	innerturnip	46*FT-004*IR*1		8005.6	5523.8	2756.0	8032.0	10085.5	492.6	1947.8	14138.8	4287.3	2695.4	901.5
September 29, 2010	FT-004	T5	innerturnip	47*FT-004*IR*2		8034.6	6430.1	2346.0	10725.3	7506.2	546.5	2619.6	12449.6	6583.1	3510.8	1255.2
September 29, 2010	FT-004	T5	innerturnip	48*FT-004*IR*3		9651.4	6720.2	2891.0	8806.2	9759.3	500.2	2267.4	12770.1	6657.6	2932.4	916.0
September 29, 2010	FT-004	T5	innerturnip	FT-004*IR*	AVG	8563.9	6224.7	2664.3	9187.8	9117.0	513.1	2278.3	13119.5	5842.7	3046.2	1024.2
					SE	543.8	360.3	163.9	800.6	810.9	16.9	194.0	518.0	778.0	242.2	115.6
					CV	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.2	0.1	0.2
September 29, 2010	FT-004	T5	outerturnip	49*FT-004*OR*1		4677.7	5198.5	2216.1	12127.0	7167.3	139.9	629.3	25089.5	1339.9	3449.3	2734.1
September 29, 2010	FT-004	T5	outerturnip	50*FT-004*OR*2		5521.2	5416.7	3509.0	14509.9	12288.9	182.6	840.4	28918.8	1326.6	4115.4	2897.9
September 29, 2010	FT-004	T5	outerturnip	51*FT-004*OR*3		7646.2	4320.1	4823.2	12190.2	17261.7	276.9	740.9	32457.0	1203.7	4158.3	1963.1
September 29, 2010	FT-004	T5	outerturnip	FT-004*OR*	AVG	5948.3	4978.4	3516.1	12942.4	12239.3	199.8	736.9	28821.7	1290.1	3907.7	2531.7
					SE	883.2	335.1	752.6	784.0	2914.1	40.5	61.0	2127.4	43.3	229.5	288.2
					CV	0.3	0.1	0.4	0.1	0.4	0.4	0.1	0.1	0.1	0.1	0.2

Appendix 1, Table 4. LC-MS data (continued). PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucoberberoin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Timepx	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Erucin	Brass	Naturtiin	Berteroin	4mBras	NeoBras
April 5, 2011	FT-004	T6	young leaves	52*FT-004*IL*1		10617.5	10181.2	5088.8	899.6	32372.0	0.0	303.3	2274.8	0.2	635.7	428.7
April 5, 2011	FT-004	T6	young leaves	53*FT-004*IL*2		7779.2	7800.7	4376.8	583.6	30088.1	0.0	332.6	1206.2	0.0	366.1	472.5
April 5, 2011	FT-004	T6	young leaves	54*FT-004*IL*3		13185.5	23289.8	4925.2	1389.3	35547.6	0.0	1156.1	2299.4	0.0	1219.1	537.0
April 5, 2011	FT-004	T6	young leaves	FT-004*IL*	AVG	10527.4	13757.3	4796.9	957.5	32669.2	0.0	597.3	1926.8	0.1	740.3	479.4
					SE	1561.3	4815.6	215.3	234.4	1583.0	0.0	279.5	360.4	0.1	251.7	31.4
					CV	0.3	0.6	0.1	0.4	0.1	#DIV/0!	0.8	0.3	1.7	0.6	0.1
April 5, 2011	FT-004	T6	mature leaves	55*FT-004*OL*1		4409.1	13222.0	919.6	85.9	21349.8	0.0	159.9	818.1	0.0	1204.2	542.0
April 5, 2011	FT-004	T6	mature leaves	56*FT-004*OL*2		2455.6	8272.2	613.1	108.5	7969.8	0.0	243.5	465.0	0.0	603.0	540.2
April 5, 2011	FT-004	T6	mature leaves	57*FT-004*OL*3		4490.1	21816.3	512.8	124.0	13510.8	0.0	867.0	486.6	0.0	1681.8	274.4
April 5, 2011	FT-004	T6	mature leaves	FT-004*OL*	AVG	3784.9	14436.8	681.8	106.2	14276.8	0.0	423.5	589.9	0.0	1163.0	452.2
					SE	665.1	3956.8	122.4	11.1	3881.4	0.0	223.1	114.3	0.0	312.1	88.9
					CV	0.3	0.5	0.3	0.2	0.5	#DIV/0!	0.9	0.3	#DIV/0!	0.5	0.3
April 5, 2011	FT-004	T6	innerturnip	58*FT-004*IR*1		2914.7	8433.3	446.2	3384.7	4817.6	104.4	1387.3	12438.8	1259.2	5112.0	1479.9
April 5, 2011	FT-004	T6	innerturnip	59*FT-004*IR*2		3070.5	8990.9	461.3	3701.7	5155.7	113.2	1451.4	12898.5	1363.3	5555.1	1582.4
April 5, 2011	FT-004	T6	innerturnip	60*FT-004*IR*3		3178.4	9032.7	462.1	3756.0	5261.8	117.1	1475.0	12903.3	1396.0	5675.3	1590.3
April 5, 2011	FT-004	T6	innerturnip	FT-004*IR*	AVG	3054.6	8819.0	456.5	3614.1	5078.3	111.6	1437.9	12746.9	1339.5	5447.5	1550.9
					SE	76.5	193.2	5.2	115.8	133.9	3.7	26.2	154.1	41.2	171.3	35.5
					CV	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.1	0.0
April 5, 2011	FT-004	T6	outerturnip	61*FT-004*OR*1		15649.5	35444.3	2004.2	30537.1	22912.0	643.8	706.2	33228.2	8922.9	10896.2	14607.3
April 5, 2011	FT-004	T6	outerturnip	62*FT-004*OR*2		15987.6	23664.4	5466.1	7810.8	32970.5	736.6	590.7	47731.9	4881.8	6197.5	5768.3
April 5, 2011	FT-004	T6	outerturnip	63*FT-004*OR*3		12625.8	40989.9	2685.1	31256.1	26091.6	1067.8	504.0	29694.9	7602.8	11937.7	12531.8
April 5, 2011	FT-004	T6	outerturnip	FT-004*OR*	AVG	14754.3	33366.2	3385.1	23201.3	27324.7	816.1	600.3	36885.0	7135.8	9677.1	10969.2
					SE	1068.7	5108.2	1058.9	7698.1	2968.4	128.7	58.6	5518.6	1189.7	1765.6	2668.5
					CV	0.1	0.3	0.5	0.6	0.2	0.3	0.2	0.3	0.3	0.3	0.4
April 5, 2011	FT-004	T6	stem	64*FT-004*Stem*1		10677.7	12242.1	7509.3	1687.6	36536.9	0.0	971.4	2108.7	55.7	966.9	926.9
April 5, 2011	FT-004	T6	stem	65*FT-004*Stem*2		10380.4	11983.2	7434.7	1691.4	38480.1	0.3	949.9	2223.6	68.0	956.7	912.1
April 5, 2011	FT-004	T6	stem	66*FT-004*Stem*3		10464.5	11882.8	7288.2	1680.3	35840.0	0.3	942.2	2473.6	63.7	951.2	900.1
April 5, 2011	FT-004	T6		FT-004*Stem*	AVG	10507.5	12036.0	7410.7	1686.5	36952.3	0.2	954.5	2268.6	62.5	958.3	913.0
					SE	88.5	107.0	65.0	3.3	789.9	0.1	8.7	107.7	3.6	4.6	7.7
					CV	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.1	0.1	0.0	0.0
April 5, 2011	FT-004	T6	flower	67*FT-004*FL*1		10814.7	27189.8	10674.3	3971.1	25339.8	33.5	1858.9	3833.6	323.3	438.3	1127.7
April 5, 2011	FT-004	T6	flower	68*FT-004*FL*2		10341.4	26055.7	9654.8	3542.8	28797.4	26.1	1001.7	3965.6	314.8	291.5	1215.2
April 5, 2011	FT-004	T6	flower	69*FT-004*FL*3		10295.7	22483.6	7525.3	4095.9	31825.7	7.3	3529.1	4470.1	230.5	399.1	838.4
April 5, 2011	FT-004	T6		FT-004*FL*	AVG	10483.9	25243.0	9284.8	3869.9	28654.3	22.3	2129.9	4089.8	289.5	376.3	1060.4

Appendix 1, Table 5. LC-MS data (Continued). PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucoberteroin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Timepx	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Eruc	Brass	Naturtii	Bertero	4mBras	NeoBras
					SE	165.9	1418.0	927.7	167.5	1873.7	7.8	742.1	194.0	29.6	43.9	113.8
					CV	0.0	0.1	0.2	0.1	0.1	0.6	0.6	0.1	0.2	0.2	0.2
April 5, 2011	FT-004	T6	seed	70*FT-004*Seed*1		8801.5	9576.5	14297.9	7152.7	21650.2	3.2	735.2	4448.9	315.8	5.7	105.7
April 5, 2011	FT-004	T6	seed	71*FT-004*Seed*2		0.0	4752.0	10430.0	6119.0	21678.4	5.4	492.4	5811.6	191.2	0.0	83.0
April 5, 2011	FT-004	T6	seed	72*FT-004*Seed*3		6313.9	6275.9	8554.5	6748.7	22299.5	29.9	887.4	8792.5	74.7	0.0	978.4
April 5, 2011	FT-004	T6		FT-004*Seed*	AVG	5038.5	6868.2	11094.1	6673.5	21876.0	12.8	705.0	6351.0	193.9	1.9	389.0
					SE	2619.6	1423.8	1690.9	300.8	211.9	8.5	115.0	1282.5	69.6	1.9	294.8
					CV	0.9	0.4	0.3	0.1	0.0	1.2	0.3	0.3	0.6	1.7	1.3
June 21, 2011	FT-004	T7	innerturnip	73*FT-004*IR*1		27.9	49.8	3.0	918.5	131.5	0.0	448.5	98.8	0.0	1228.6	172.7
June 21, 2011	FT-004	T7	innerturnip	74*FT-004*IR*2		40.5	13.3	49.0	267.0	51.4	0.0	195.4	18.5	0.1	83.8	21.5
June 21, 2011	FT-004	T7	innerturnip	75*FT-004*IR*3		33.0	1.0	101.3	924.9	57.0	0.0	448.5	51.2	0.0	1359.3	167.3
June 21, 2011	FT-004	T7	innerturnip	FT-004*IR*	AVG	33.8	21.4	51.1	703.5	80.0	0.0	364.2	56.2	0.0	890.6	120.5
					SE	3.7	14.6	28.4	218.2	25.8	0.0	84.4	23.3	0.0	405.1	49.5
					CV	0.2	1.2	1.0	0.5	0.6	#DIV/0!	0.4	0.7	1.7	0.8	0.7
June 21, 2011	FT-004	T7	outerturnip	76*FT-004*OR*1		23.7	6.3	121.1	6799.6	96.6	0.0	622.0	1030.6	0.0	2966.5	2257.6
June 21, 2011	FT-004	T7	outerturnip	77*FT-004*OR*2		3.1	0.0	0.0	1180.7	1.5	0.0	660.9	1712.8	0.0	639.8	2496.1
June 21, 2011	FT-004	T7	outerturnip	78*FT-004*OR*3		0.6	0.2	0.0	1286.2	1.8	0.0	1193.9	15638.5	0.0	3020.2	2056.0
June 21, 2011	FT-004	T7	outerturnip	FT-004*OR*	AVG	9.1	2.2	40.4	3088.8	33.3	0.0	825.6	6127.3	0.0	2208.8	2269.9
					SE	7.3	2.1	40.4	1855.7	31.6	0.0	184.5	4759.7	0.0	784.7	127.2
					CV	1.4	1.6	1.7	1.0	1.6	#DIV/0!	0.4	1.3	#DIV/0!	0.6	0.1
July 26, 2010	FT-086	T1	seedling	79*FT-086*TYL*1		1088.3	5.1	14877.1	134.2	1703.7	0.0	962.9	17.9	0.0	274.9	1778.7
July 26, 2010	FT-086	T1	seedling	80*FT-086*TYL*2		1127.1	4.7	15659.5	273.2	1391.6	0.0	1851.5	82.3	0.0	222.6	2660.7
July 26, 2010	FT-086	T1	seedling	81*FT-086*TYL*3		1304.8	10.0	14690.3	218.8	1317.2	0.0	1891.7	81.7	0.0	151.3	2170.3
July 26, 2010	FT-086	T1	seedling	FT-086*TYL*	AVG	1173.4	6.6	15075.6	208.7	1470.8	0.0	1568.7	60.6	0.0	216.3	2203.2
					SE	66.6	1.7	296.9	40.4	118.4	0.0	303.1	21.4	0.0	35.8	255.1
					CV	0.1	0.4	0.0	0.3	0.1	#DIV/0!	0.3	0.6	#DIV/0!	0.3	0.2
August 16, 2010	FT-086	T2	young leaves	82*FT-086*IL*1		4344.8	22.9	8713.7	1982.8	28862.8	0.0	19057.0	9828.9	0.0	2504.0	16459.1
August 16, 2010	FT-086	T2	young leaves	83*FT-086*IL*2		11935.6	64.3	9506.0	2158.5	26817.6	0.0	11339.9	13069.7	0.0	3160.4	16393.3
August 16, 2010	FT-086	T2	young leaves	84*FT-086*IL*3		7929.8	28.6	10035.5	2847.2	26389.1	0.0	13769.7	9667.9	0.0	1040.5	15396.3
August 16, 2010	FT-086	T2	young leaves	FT-086*IL*	AVG	8070.1	38.6	9418.4	2329.5	27356.5	0.0	14722.2	10855.5	0.0	2235.0	16082.9
					SE	2192.4	12.9	384.1	263.8	763.2	0.0	2278.1	1108.1	0.0	626.6	343.8
					CV	0.5	0.6	0.1	0.2	0.0	#DIV/0!	0.3	0.2	#DIV/0!	0.5	0.0
August 16, 2010	FT-086	T2	mature leaves	85*FT-086*OL*1		168.7	0.0	24893.8	183.0	4092.8	0.0	2399.0	555.5	0.0	868.6	3969.7
August 16, 2010	FT-086	T2	mature leaves	86*FT-086*OL*2		1194.3	13.8	25127.2	256.3	7164.1	0.0	2765.7	489.5	0.0	585.4	2255.9

Appendix 1, Table 6. LC-MS data (continued). PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucobertoin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Timep	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Eruc	Brass	Naturtiin	Berteroin	4mBras	NeoBras
August 16, 2010	FT-086	T2	mature leaves	87*FT-086*OL*3		380.6	0.6	27359.1	149.5	3144.0	0.0	1635.5	166.9	0.0	300.4	1248.3
August 16, 2010	FT-086	T2	mature leaves	FT-086*OL*	AVG	581.2	4.8	25793.4	196.3	4800.3	0.0	2266.7	404.0	0.0	584.8	2491.3
					SE	312.6	4.5	785.7	31.5	1213.2	0.0	332.9	120.1	0.0	164.0	794.4
					CV	0.9	1.6	0.1	0.3	0.4	#DIV/0!	0.3	0.5	#DIV/0!	0.5	0.6
August 16, 2010	FT-086	T2	innerturnip	88*FT-086*IR*1		15998.9	100.0	30681.0	4064.8	9185.5	34674.9	7618.3	33568.7	6928.6	880.6	2811.9
August 16, 2010	FT-086	T2	innerturnip	89*FT-086*IR*2		9467.4	325.1	25158.3	2735.8	10356.2	20488.6	4434.6	41209.9	3277.0	1273.7	9424.1
August 16, 2010	FT-086	T2	innerturnip	90*FT-086*IR*3		12427.7	59.0	31762.6	2501.7	6353.6	14205.4	5294.4	31929.4	1507.2	1363.6	7781.3
August 16, 2010	FT-086	T2	innerturnip	FT-086*IR*	AVG	12631.3	161.4	29200.6	3100.8	8631.8	23123.0	5782.4	35569.3	3904.3	1172.6	6672.4
					SE	1888.2	82.7	2045.2	486.8	1188.2	6054.1	950.9	2859.7	1596.2	148.3	1987.7
					CV	0.3	0.9	0.1	0.3	0.2	0.5	0.3	0.1	0.7	0.2	0.5
August 16, 2010	FT-086	T2	outerturnip	91*FT-086*OR*1		6745.1	53.9	27103.8	3554.1	5796.3	21766.1	6469.9	43220.3	2477.0	3269.0	16273.5
August 16, 2010	FT-086	T2	outerturnip	92*FT-086*OR*2		13835.6	146.5	28126.4	3638.0	6428.0	9661.2	5716.6	47264.3	1164.4	3608.9	16272.6
August 16, 2010	FT-086	T2	outerturnip	93*FT-086*OR*3		5950.5	18.0	29291.4	2040.3	4505.0	10003.7	11976.1	38758.1	870.8	6821.8	14999.3
August 16, 2010	FT-086	T2	outerturnip	FT-086*OR*	AVG	8843.7	72.8	28173.9	3077.5	5576.4	13810.3	8054.2	43080.9	1504.1	4566.5	15848.5
					SE	2506.4	38.3	632.0	519.1	565.9	3979.1	1973.0	2456.5	493.8	1131.9	424.6
					CV	0.5	0.9	0.0	0.3	0.2	0.5	0.4	0.1	0.6	0.4	0.0
August 30, 2010	FT-086	T3	young leaves	94*FT-086*IL*1		6364.2	13.4	8938.4	1706.4	31187.8	2.7	4826.9	10525.6	0.1	7316.1	19189.5
August 30, 2010	FT-086	T3	young leaves	95*FT-086*IL*2		4163.8	12.4	9422.0	1004.5	31609.8	0.0	7187.3	10596.1	0.0	7696.2	16478.6
August 30, 2010	FT-086	T3	young leaves	96*FT-086*IL*3		5873.1	26.8	11323.6	1516.1	25330.3	0.0	5389.4	7629.1	0.0	5419.1	13185.7
August 30, 2010	FT-086	T3	young leaves	FT-086*IL*	AVG	5467.0	17.5	9894.7	1409.0	29376.0	0.9	5801.2	9583.6	0.0	6810.5	16284.6
					SE	666.8	4.6	728.0	209.6	2026.5	0.9	711.8	977.5	0.0	704.3	1735.9
					CV	0.2	0.5	0.1	0.3	0.1	1.7	0.2	0.2	1.7	0.2	0.2
August 30, 2010	FT-086	T3	mature leaves	97*FT-086*OL*1		228.2	2.1	21723.0	66.7	5026.1	0.0	2330.1	80.8	0.0	1144.7	1297.8
August 30, 2010	FT-086	T3	mature leaves	98*FT-086*OL*2		1285.3	2.6	13498.9	272.7	18474.2	0.0	937.8	2282.6	0.0	1713.8	8281.0
August 30, 2010	FT-086	T3	mature leaves	99*FT-086*OL*3		1358.9	0.7	12825.5	280.3	16006.9	0.0	1667.0	1411.9	0.0	2263.2	4874.7
August 30, 2010	FT-086	T3	mature leaves	FT-086*OL*	AVG	957.5	1.8	16015.8	206.6	13169.0	0.0	1645.0	1258.4	0.0	1707.2	4817.8
					SE	365.3	0.6	2860.2	70.0	4133.3	0.0	402.1	640.2	0.0	322.9	2016.1
					CV	0.7	0.5	0.3	0.6	0.5	#DIV/0!	0.4	0.9	#DIV/0!	0.3	0.7
August 30, 2010	FT-086	T3	innerturnip	100*FT-086*IR*1		13843.4	214.2	17125.0	5183.5	11319.8	13578.9	3465.4	27499.6	2764.7	2464.0	3967.7
August 30, 2010	FT-086	T3	innerturnip	101*FT-086*IR*2		15719.3	124.5	17743.3	4564.4	10398.8	11413.2	3358.9	27430.8	2366.7	2557.9	3785.0
August 30, 2010	FT-086	T3	innerturnip	102*FT-086*IR*3		17067.2	163.2	15634.4	4518.3	9082.1	15313.1	3384.7	30997.3	2023.9	2503.7	5467.1
August 30, 2010	FT-086	T3	innerturnip	FT-086*IR*	AVG	15543.3	167.3	16834.2	4755.4	10266.9	13435.0	3403.0	28642.6	2385.1	2508.6	4406.6
					SE	934.8	26.0	625.9	214.4	649.3	1128.1	32.1	1177.5	214.0	27.2	532.9
					CV	0.1	0.3	0.1	0.1	0.1	0.1	0.0	0.1	0.2	0.0	0.2

Appendix 1, Table 7. LC-MS data (continued). PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucobertoin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Timep	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Eruc	Brass	Naturti	Bertero	4mBras	NeoBras
August 30, 2010	FT-086	T3	outerturnip	103*FT-086*OR*1		8337.6	44.2	24984.8	3388.0	4712.0	5780.7	10551.3	29621.7	722.1	10307.6	16052.9
August 30, 2010	FT-086	T3	outerturnip	104*FT-086*OR*2		8299.4	46.2	25233.4	3550.0	4722.3	5446.5	11131.1	28730.6	735.4	11215.1	15156.5
August 30, 2010	FT-086	T3	outerturnip	105*FT-086*OR*3		9106.4	43.1	26725.8	3496.3	5034.1	6196.2	6976.8	31769.9	819.4	8801.1	17113.1
August 30, 2010	FT-086	T3	outerturnip	FT-086*OR*	AVG	8581.1	44.5	25648.0	3478.1	4822.8	5807.8	9553.0	30040.8	759.0	10107.9	16107.5
					SE	262.8	0.9	543.7	47.6	105.7	216.8	1299.0	902.0	30.5	704.0	565.5
					CV	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.1	0.1	0.1	0.1
September 16, 2010	FT-086	T4	young leaves	106*FT-086*IL*1		2538.7	5.8	8239.4	2239.0	27982.9	2.7	7699.9	7042.9	0.1	10334.9	16295.2
September 16, 2010	FT-086	T4	young leaves	107*FT-086*IL*2		3083.8	8.3	8075.5	1128.6	33745.9	0.0	7697.8	11943.9	0.0	14571.0	16035.9
September 16, 2010	FT-086	T4	young leaves	108*FT-086*IL*3		5216.9	16.4	7074.7	2558.2	30783.3	0.0	14450.3	13992.1	0.0	13458.5	19781.1
September 16, 2010	FT-086	T4	young leaves	FT-086*IL*	AVG	3613.1	10.2	7796.5	1975.3	30837.4	0.9	9949.3	10992.9	0.0	12788.1	17370.7
					SE	817.2	3.2	364.0	433.3	1663.9	0.9	2250.5	2061.6	0.0	1267.9	1207.5
					CV	0.4	0.5	0.1	0.4	0.1	1.7	0.4	0.3	1.7	0.2	0.1
September 16, 2010	FT-086	T4	mature leaves	109*FT-086*OL*1		27.1	0.0	20517.8	71.1	2339.8	0.8	950.0	94.5	0.0	1090.1	4102.2
September 16, 2010	FT-086	T4	mature leaves	110*FT-086*OL*2		158.1	0.0	25431.6	75.5	3905.0	0.0	1096.9	68.6	0.0	577.7	1211.5
September 16, 2010	FT-086	T4	mature leaves	111*FT-086*OL*3		119.1	0.0	26044.4	90.0	4263.7	0.0	2958.9	34.5	0.0	766.6	1846.5
September 16, 2010	FT-086	T4	mature leaves	FT-086*OL*	AVG	101.4	0.0	23997.9	78.9	3502.9	0.3	1668.6	65.9	0.0	811.5	2386.7
					SE	38.8	0.0	1749.1	5.7	590.7	0.3	646.5	17.4	0.0	149.6	877.1
					CV	0.7	#DIV/0!	0.1	0.1	0.3	1.7	0.7	0.5	#DIV/0!	0.3	0.6
September 16, 2010	FT-086	T4	innerturnip	112*FT-086*IR*1		11452.4	67.9	10345.3	6305.4	9015.3	4175.9	3460.3	21869.9	758.3	3367.2	1930.5
September 16, 2010	FT-086	T4	innerturnip	113*FT-086*IR*2		9714.5	34.8	12615.7	6140.2	7600.5	3383.3	2493.2	17951.3	622.3	2630.0	1480.6
September 16, 2010	FT-086	T4	innerturnip	114*FT-086*IR*3		10084.4	44.0	12152.1	5994.2	7796.3	4075.9	3647.5	20608.1	644.3	2961.9	3222.7
September 16, 2010	FT-086	T4	innerturnip	FT-086*IR*	AVG	10417.1	48.9	11704.4	6146.6	8137.4	3878.4	3200.3	20143.1	675.0	2986.4	2211.3
					SE	528.6	9.9	692.6	89.9	442.6	249.2	357.7	1154.8	42.2	213.2	522.1
					CV	0.1	0.3	0.1	0.0	0.1	0.1	0.2	0.1	0.1	0.1	0.4
September 16, 2010	FT-086	T4	outerturnip	115*FT-086*OR*1		3623.3	11.6	22711.8	5126.2	4730.6	2290.1	3571.6	29524.9	264.4	11765.5	19471.3
September 16, 2010	FT-086	T4	outerturnip	116*FT-086*OR*2		3367.6	13.2	21644.1	5540.1	4393.4	2308.7	2344.9	28136.2	324.3	11155.5	17508.1
September 16, 2010	FT-086	T4	outerturnip	117*FT-086*OR*3		3806.8	15.6	24509.1	6251.1	5610.3	2787.1	3866.7	31741.4	351.4	12221.5	16199.0
September 16, 2010	FT-086	T4	outerturnip	FT-086*OR*	AVG	3599.2	13.5	22955.0	5639.1	4911.4	2462.0	3261.1	29800.8	313.4	11714.2	17726.2
					SE	127.4	1.2	835.9	328.5	362.8	162.7	465.9	1049.9	25.7	308.8	950.9
					CV	0.1	0.2	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.0	0.1
September 29, 2010	FT-086	T5	young leaves	118*FT-086*IL*1		867.9	1.0	8851.5	656.1	25498.2	0.2	5259.5	5677.6	0.0	11103.2	16501.1
September 29, 2010	FT-086	T5	young leaves	119*FT-086*IL*2		732.4	2.9	6758.1	494.8	25825.5	0.0	2984.0	4462.4	0.0	11006.4	11312.2
September 29, 2010	FT-086	T5	young leaves	120*FT-086*IL*3		1857.9	37.5	7088.9	1996.4	27757.7	0.0	9741.5	10420.3	0.0	12326.5	15574.5
September 29, 2010	FT-086	T5	young leaves	FT-086*IL*	AVG	1152.7	13.8	7566.2	1049.1	26360.5	0.1	5995.0	6853.4	0.0	11478.7	14462.6

Appendix 1, Table 8. LC-MS data (continued). PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucoberteroin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Timep	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Eruc	Brass	Naturti	Bertero	4mBras	NeoBras
					SE	354.7	11.9	649.7	475.9	705.0	0.1	1985.1	1817.6	0.0	424.8	1597.8
					CV	0.5	1.5	0.1	0.8	0.0	1.7	0.6	0.5	#DIV/0!	0.1	0.2
September 29, 2010	FT-086	T5	mature leaves	121*FT-086*OL*1		92.0	0.0	23147.1	53.1	3612.7	0.0	1937.1	66.5	0.0	1210.4	5586.4
September 29, 2010	FT-086	T5	mature leaves	122*FT-086*OL*2		34.6	0.0	20338.3	54.2	1906.9	0.0	1311.0	56.3	0.0	1890.4	2066.3
September 29, 2010	FT-086	T5	mature leaves	123*FT-086*OL*3		154.1	0.1	26168.7	90.5	6473.3	0.0	2335.3	198.7	0.0	1739.4	1978.1
September 29, 2010	FT-086	T5	mature leaves	FT-086*OL*	AVG	93.6	0.0	23218.0	65.9	3997.6	0.0	1861.1	107.2	0.0	1613.4	3210.3
					SE	34.5	0.0	1683.5	12.3	1332.2	0.0	298.1	45.9	0.0	206.1	1188.3
					CV	0.6	1.7	0.1	0.3	0.6	#DIV/0!	0.3	0.7	#DIV/0!	0.2	0.6
September 29, 2010	FT-086	T5	innerturnip	124*FT-086*IR*1		6284.3	18.2	13067.6	6004.0	7710.0	3534.1	2631.5	20780.0	881.4	4638.1	1448.1
September 29, 2010	FT-086	T5	innerturnip	125*FT-086*IR*2		6791.7	20.8	12110.4	5999.5	7667.0	4822.4	3075.9	20578.4	1046.4	4126.0	1228.9
September 29, 2010	FT-086	T5	innerturnip	126*FT-086*IR*3		7425.8	24.8	13590.3	5449.7	6384.1	3367.5	2968.2	18915.8	808.4	4177.2	1144.7
September 29, 2010	FT-086	T5	innerturnip	FT-086*IR*	AVG	6834.0	21.3	12922.8	5817.7	7253.7	3908.0	2891.9	20091.4	912.1	4313.8	1273.9
					SE	330.2	1.9	433.3	184.0	435.0	459.7	133.8	590.7	70.4	162.9	90.4
					CV	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1
September 29, 2010	FT-086	T5	outerturnip	127*FT-086*OR*1		1724.8	0.7	22014.1	5154.1	3339.0	869.7	3975.2	21065.6	134.6	10077.4	16793.8
September 29, 2010	FT-086	T5	outerturnip	128*FT-086*OR*2		1713.0	1.3	24339.7	4623.4	4183.0	1314.0	4575.3	25202.5	147.2	8695.1	14871.5
September 29, 2010	FT-086	T5	outerturnip	129*FT-086*OR*3		1377.9	1.0	23580.8	5074.8	5155.0	1098.3	2131.1	24428.4	181.3	6794.8	10486.2
September 29, 2010	FT-086	T5	outerturnip	FT-086*OR	AVG	1605.2	1.0	23311.5	4950.8	4225.6	1094.0	3560.5	23565.5	154.4	8522.4	14050.5
					SE	113.7	0.2	684.7	165.3	524.7	128.3	735.4	1269.8	14.0	951.5	1866.6
					CV	0.1	0.3	0.1	0.1	0.2	0.2	0.4	0.1	0.2	0.2	0.2
April 5, 2011	FT-086	T6	young leaves	130*FT-086*IL*1		527.0	0.0	14197.3	680.2	15529.8	0.0	912.0	2578.2	0.0	915.6	1025.1
April 5, 2011	FT-086	T6	young leaves	131*FT-086*IL*2		266.7	0.3	13324.5	402.7	25971.2	0.0	772.1	1857.5	0.0	540.6	577.1
April 5, 2011	FT-086	T6	young leaves	132*FT-086*IL*3		1032.7	22.3	8668.7	626.4	20015.4	0.0	1386.7	3164.2	0.0	1836.3	1889.5
April 5, 2011	FT-086	T6	young leaves	FT-086*IL*	AVG	608.8	7.5	12063.5	569.8	20505.5	0.0	1023.6	2533.3	0.0	1097.5	1163.9
					SE	224.9	7.4	1716.0	85.0	3024.1	0.0	186.0	377.9	0.0	384.9	385.2
					CV	0.6	1.7	0.2	0.3	0.3	#DIV/0!	0.3	0.3	#DIV/0!	0.6	0.6
April 5, 2011	FT-086	T6	mature leaves	133*FT-086*OL*1		79.7	3.9	23492.6	93.9	11018.9	0.0	1041.0	2681.4	0.0	745.9	1252.3
April 5, 2011	FT-086	T6	mature leaves	134*FT-086*OL*2		30.7	0.0	23016.0	57.7	16163.9	0.0	618.0	734.0	0.0	787.4	777.7
April 5, 2011	FT-086	T6	mature leaves	135*FT-086*OL*3		286.9	0.0	14711.2	64.3	15702.2	0.0	931.6	2471.1	0.0	1698.9	1801.2
April 5, 2011	FT-086	T6	mature leaves	FT-086*OL*	AVG	132.4	1.3	20406.6	72.0	14295.0	0.0	863.5	1962.2	0.0	1077.4	1277.1
					SE	78.5	1.3	2851.0	11.1	1643.5	0.0	126.8	617.1	0.0	311.0	295.7
					CV	1.0	1.7	0.2	0.3	0.2	#DIV/0!	0.3	0.5	#DIV/0!	0.5	0.4
April 5, 2011	FT-086	T6	innerturnip	136*FT-086*IR*1		1508.1	0.8	19622.0	6158.3	6062.7	6059.6	2169.9	15051.2	197.3	3274.6	1656.0
April 5, 2011	FT-086	T6	innerturnip	137*FT-086*IR*2		938.6	9.4	22452.6	3392.0	10416.7	619.8	1115.7	20637.7	28.7	3375.1	3320.8

Appendix 1, Table 9. LC-MS data (continued). PRO: Progoitrin, Napol: Gluconapoleiferin, Napin: Gluconapin, 4hBras: 4-hydroxyglucobrassicin, Canapin: Glucobrassicinapin, Erucin: Glucoerucin, Brass: Glucobrassicin, Naturtiin: Gluconasturtiin, Berteroin: Glucoberteroin, 4mBras: 4-methoxyglucobrassicin, Neobras: Neoglucobrassicin. AVG: Average peak area. SE: Standard error. CV: Coefficient of variation among 3 biological replicates.

Sampling date	Accession	Timepx	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Eruc	Brass	Naturti	Bertero	4mBras	NeoBras
April 5, 2011	FT-086	T6	innerturnip	138*FT-086*IR*3		1837.4	4.7	26629.3	3033.6	4330.2	182.4	2292.0	17860.7	15.2	2896.3	2114.4
April 5, 2011	FT-086	T6	innerturnip	FT-086*IR*	AVG	1428.0	5.0	22901.3	4194.6	6936.5	2287.3	1859.2	17849.9	80.4	3182.0	2363.7
					SE	262.5	2.5	2035.2	987.3	1810.5	1890.4	373.4	1612.7	58.6	145.8	496.5
					CV	0.3	0.9	0.2	0.4	0.5	1.4	0.3	0.2	1.3	0.1	0.4
April 5, 2011	FT-086	T6	outerturnip	139*FT-086*OR*1		1471.9	0.0	25899.0	1044.4	2921.9	1645.2	585.2	24515.0	93.9	4072.6	11706.8
April 5, 2011	FT-086	T6	outerturnip	140*FT-086*OR*2		1000.3	0.1	22651.0	1665.2	3369.8	1146.0	627.0	23847.1	83.3	7328.3	10674.9
April 5, 2011	FT-086	T6	outerturnip	141*FT-086*OR*3		2526.5	12.2	34963.5	1010.4	5127.7	921.0	2451.0	30569.8	349.0	6875.7	14797.7
April 5, 2011	FT-086	T6	outerturnip	FT-086*OR*	AVG	1666.2	4.1	27837.9	1240.0	3806.4	1237.4	1221.1	26310.6	175.4	6092.2	12393.2
					SE	451.2	4.0	3684.2	212.8	673.1	214.0	615.1	2138.3	86.9	1018.2	1238.6
					CV	0.5	1.7	0.2	0.3	0.3	0.3	0.9	0.1	0.9	0.3	0.2
April 5, 2011	FT-086	T6	stem	142*FT-086*Stem*1		812.3	1.3	16768.6	1850.5	10750.8	0.2	2006.5	3059.8	0.0	1224.9	1299.9
April 5, 2011	FT-086	T6	stem	143*FT-086*Stem*2		500.2	12.9	22151.9	1558.4	19284.8	0.0	1207.5	3413.9	0.0	626.9	1096.5
April 5, 2011	FT-086	T6	stem	144*FT-086*Stem*3		1045.2	0.0	14188.2	1816.9	22846.1	0.1	5136.5	3795.8	0.0	2282.6	3235.0
April 5, 2011	FT-086	T6		FT-086*Stem*	AVG	785.9	4.7	17702.9	1741.9	17627.3	0.1	2783.5	3423.1	0.0	1378.1	1877.1
					SE	157.9	4.1	2345.9	92.3	3588.6	0.1	1198.9	212.5	0.0	484.1	681.5
					CV	0.3	1.5	0.2	0.1	0.4	1.0	0.7	0.1	#DIV/0!	0.6	0.6
April 5, 2011	FT-086	T6	flower	145*FT-086*FL*1		1204.5	0.9	9922.0	4174.5	18317.7	107.0	3333.8	9901.0	1.9	837.7	1696.4
April 5, 2011	FT-086	T6	flower	146*FT-086*FL*2		819.6	9.1	9681.5	3216.9	25198.2	16.3	1668.2	5866.0	0.7	130.4	910.5
April 5, 2011	FT-086	T6	flower	147*FT-086*FL*3		2092.9	1.7	9131.2	3802.9	18434.2	44.3	3041.0	7409.1	0.1	1378.3	2135.7
April 5, 2011	FT-086	T6		FT-086*FL*	AVG	1372.3	3.9	9578.2	3731.4	20650.0	55.9	2681.0	7725.4	0.9	782.1	1580.9
					SE	377.0	2.6	234.1	278.7	2274.3	26.8	513.4	1175.5	0.5	361.3	358.4
					CV	0.5	1.2	0.0	0.1	0.2	0.8	0.3	0.3	1.0	0.8	0.4
April 5, 2011	FT-086	T6	seed	148*FT-086*Seed*1		9888.5	169.7	8423.6	5612.0	6473.8	79.5	162.2	2247.9	0.3	186.5	172.8
April 5, 2011	FT-086	T6	seed	149*FT-086*Seed*2		9518.2	80.9	8643.5	7469.1	3905.2	123.7	433.8	1073.6	5.3	379.3	275.6
April 5, 2011	FT-086	T6	seed	150*FT-086*Seed*3		8251.7	88.0	8083.9	7994.7	14397.4	99.9	345.1	8941.6	8.5	1352.7	641.7
April 5, 2011	FT-086	T6		FT-086*Seed*	AVG	9219.4	112.9	8383.6	7025.3	8258.8	101.0	313.7	4087.7	4.7	639.5	363.4
					SE	495.6	28.5	162.8	722.7	3157.6	12.8	80.0	2450.5	2.4	360.9	142.3
					CV	0.1	0.4	0.0	0.2	0.7	0.2	0.4	1.0	0.9	1.0	0.7
June 21, 2011	FT-086	T7	innerturnip	151*FT-086*IR*1		18.8	0.0	776.8	342.7	87.5	1.8	647.9	210.0	0.0	259.2	879.8
June 21, 2011	FT-086	T7	innerturnip	152*FT-086*IR*2		7.7	0.0	371.1	10.1	6.4	0.0	48.9	3.0	0.0	36.7	133.4
June 21, 2011	FT-086	T7	innerturnip	153*FT-086*IR*3		13.3	0.0	337.9	4.6	1.2	0.0	95.7	30.3	0.0	103.7	90.6
June 21, 2011	FT-086	T7	innerturnip	FT-086*IR*	AVG	13.3	0.0	495.3	119.1	31.7	0.6	264.1	81.1	0.0	133.2	367.9
					SE	3.2	0.0	141.1	111.8	27.9	0.6	192.3	64.9	0.0	65.9	256.2
					CV	0.4	#DIV/0!	0.5	1.6	1.5	1.7	1.3	1.4	#DIV/0!	0.9	1.2
Sampling date	Accession	Timepx	Tissue	qtof name	Peak Ar	Pro	Napol	Napin	4hBras	Canapin	Eruc	Brass	Naturti	Bertero	4mBras	NeoBras
June 21, 2011	FT-086	T7	outerturnip	154*FT-086*OR*1		5.5	0.1	397.7	239.7	67.5	0.0	472.5	312.3	0.0	217.6	247.8
June 21, 2011	FT-086	T7	outerturnip	155*FT-086*OR*2		0.0	0.0	2.0	248.3	0.0	0.0	61.1	71.3	0.0	166.9	1542.8
June 21, 2011	FT-086	T7	outerturnip	156*FT-086*OR*3		0.0	0.0	1.9	1375.5	0.0	0.0	519.4	22.6	0.0	4562.4	11674.1
June 21, 2011	FT-086	T7	outerturnip	FT-086*OR*	AVG	1.8	0.0	133.8	621.2	22.5	0.0	351.0	135.4	0.0	1649.0	4488.2
					SE	1.8	0.0	131.9	377.1	22.5	0.0	145.6	89.6	0.0	1456.8	3612.3
					CV	1.7	1.7	1.7	1.1	1.7	#DIV/0!	0.7	1.1	#DIV/0!	1.5	1.4

Appendix 2: qRT-PCR data

Gene expression data as $-\Delta Ct$

Appendix 2, Table 1. Gene expression data as $-\Delta Ct$

ID	MAM1_1	MAM1_2	MAM1_3	MAM3_1	MAM3_2	MAM3_3	MAM3_4	CYP79A2	CYP83A1_1	CYP83A1_2	CYP83B1	AOP2_1	AOP2_2	AOP2_3	AOP2_2-1	GS-OH_1	GS-OH_2
T1seedlingFT-004	-9.643	-7.930	-9.263	-7.607	-7.010	-7.927	-7.887	-11.573	-5.823	-5.383	0.027	-7.707	-12.240	-7.747	-7.010	0.680	-10.357
T2young leavesFT-004	-3.943	-2.350	-4.243	-3.033	-2.093	-3.760	-3.013	-9.510	-4.737	-2.017	-0.297	-2.493	-7.787	-3.573	-1.980	-1.740	-12.513
T2inner turnipFT-004	-8.167	-7.673	-8.967	-5.570	-8.610	-5.540	-5.350	-18.777	-8.567	-5.627	0.087	-11.457	-17.020	-15.347	-10.830	-3.647	-15.750
T3inner turnipFT-004	-8.700	-9.303	-10.137	-8.377	-12.857	-8.170	-8.023	-18.830	-12.640	-8.733	-4.030	-12.317	-17.273	-17.610	-11.967	-5.823	-12.990
T4mature leavesFT-004	-10.840	-8.460	-10.490	-9.180	-7.470	-10.250	-9.830	-16.070	-9.760	-6.390	2.360	-8.640	-13.820	-7.940	-8.680	0.770	-5.200
T6young leavesFT-004	-6.643	-5.753	-7.263	-5.760	-5.683	-5.487	-5.450	-17.053	-6.290	-3.380	-3.480	-4.620	-10.320	-7.900	-4.127	-2.193	-5.713
T6mature leavesFT-004	-6.687	-6.720	-7.823	-6.393	-7.303	-6.100	-5.873	-14.463	-8.767	-4.230	-0.133	-5.990	-10.557	-7.960	-5.593	-2.307	-4.447
T6inner turnipFT-004	-15.140	-16.090	-11.423	-12.160	-15.053	-16.847	-16.153	-17.477	-15.733	-13.943	-5.040	-11.900	-17.170	-16.707	-15.570	-9.324	-4.857
T6stemFT-004	-12.150	-12.760	-12.334	-10.506	-15.002	-11.726	-10.726	-18.956	-11.504	-8.092	-9.340	-11.810	-17.698	-15.158	-11.444	-4.457	-6.792
T6flowerFT-004	-11.063	-10.157	-10.660	-9.863	-10.070	-9.697	-9.347	-10.303	-11.230	-6.593	-4.873	-8.777	-16.167	-12.987	-8.240	-5.147	-6.447
T6seedFT-004	-11.133	-11.573	-7.757	-8.023	-12.833	-7.927	-9.103	-17.230	-11.250	-9.710	-5.913	-12.350	-13.370	-11.653	-12.210	-11.980	-7.487
T1seedlingFT-086	-7.400	-7.883	-7.877	-7.423	-13.967	-7.240	-6.927	-14.397	-8.783	-4.943	-2.413	-6.960	-12.637	-8.160	-6.703	-5.840	-7.970
T2young leavesFT-086	-3.867	-3.943	-4.397	-3.950	-12.137	-3.427	-3.240	-9.943	-6.413	-3.050	-2.240	-2.423	-7.330	-4.210	-2.110	-10.147	-14.150
T2inner turnipFT-086	-6.143	-8.037	-8.440	-5.947	-17.400	-5.497	-5.330	-17.587	-8.013	-4.083	-1.707	-9.823	-15.007	-13.073	-9.443	-13.260	-13.120
T3inner turnipFT-086	-4.947	-6.657	-6.687	-5.593	-14.293	-4.947	-4.877	-17.463	-9.513	-4.947	-3.857	-6.990	-11.063	-8.710	-6.803	-11.143	-12.803
T4mature leavesFT-086	-11.130	-8.210	-9.730	-9.080	-7.920	-10.470	-10.840	-11.250	-10.220	-6.160	2.130	-8.270	-13.370	-7.540	-9.140	-12.770	-15.530
T6young leavesFT-086	-3.767	-4.137	-5.243	-3.967	-12.230	-3.910	-3.633	-13.890	-5.873	-2.150	-2.537	-2.550	-7.393	-5.563	-1.960	-12.533	-6.727
T6mature leavesFT-086	-4.917	-4.803	-5.607	-3.763	-9.187	-3.820	-3.670	-15.843	-7.193	-2.467	-0.497	-4.290	-8.067	-6.750	-3.837	-7.850	-4.330
T6inner turnipFT-086	-11.647	-14.117	-11.340	-12.360	-16.743	-12.857	-12.317	-16.063	-14.940	-11.133	-4.527	-13.473	-16.103	-15.720	-13.383	-13.543	-6.880
T6stemFT-086	-8.107	-9.027	-9.847	-7.810	-15.233	-7.867	-7.400	-14.060	-10.123	-6.183	-7.843	-7.487	-13.407	-10.177	-6.863	-14.470	-10.787
T6flowerFT-086	-7.320	-8.020	-8.560	-7.143	-12.807	-7.177	-6.643	-12.070	-9.217	-5.473	-6.147	-6.433	-12.430	-10.410	-5.900	-11.060	-11.237
T6seedFT-086	-14.917	-13.920	-10.213	-10.587	-10.780	-11.850	-11.820	-15.653	-13.993	-15.653	-6.440	-13.680	-15.653	-12.610	-13.087	-10.860	-9.360

Appendix 2, Table 2. Gene expression data as $-\Delta Ct$ (Continued)

	GS-OH_2	GS-OH_3	MYB28_1	MYB28_2	MYB29	MYB34_1	MYB34_2	MYB34_3	MYB34_4	MYB51_1	MYB51_2	MYB51_3	GTR1_1	GTR1_2	GTR1_3	GTR2_1	GTR2_2
T1seedlingFT-004	-10.357	-7.297	-9.057	-7.157	-14.380	-4.900	-3.527	-3.083	-7.730	-4.683	-17.330	-11.873	-2.512	-7.204	-14.125	-2.640	-8.336
T2young leavesFT-004	-12.513	-12.163	-5.727	-5.327	-17.370	-5.680	-5.943	-5.077	-9.330	-3.433	-17.370	-12.820	-2.561	-7.635	-12.207	-3.310	-8.219
T2inner turnipFT-004	-15.750	-7.900	-17.677	-9.663	-16.417	-4.747	-2.610	-11.070	-9.133	-8.650	-14.540	-17.963	-0.978	-7.467	-11.271	-0.816	-2.812
T3inner turnipFT-004	-12.990	-11.443	-18.830	-16.123	-14.437	-7.440	-6.150	-10.300	-13.997	-12.600	-18.830	-12.087	-5.714	-11.228	-12.354	-5.082	-9.742
T4mature leavesFT-004	-5.200	-11.350	-12.120	-11.600	-8.510	-9.210	-7.620	-4.850	-13.780	-0.940	-4.600	-15.217	-3.769	-10.128	-16.072	-3.018	-7.672
T6young leavesFT-004	-5.713	-15.903	-8.140	-4.487	-9.937	-13.583	-13.540	-3.883	-16.367	-7.160	-9.287	-10.690	-3.391	-7.454	-13.993	-2.086	-7.058
T6mature leavesFT-004	-4.447	-15.640	-7.380	-5.513	-16.740	-13.417	-10.823	-4.607	-14.860	-4.083	-12.367	-4.400	-2.751	-7.619	-12.105	-1.564	-7.255
T6inner turnipFT-004	-4.857	-17.477	-15.830	-13.620	-12.100	-15.170	-15.607	-11.293	-16.223	-9.757	-12.850	-13.907	-9.523	-15.512	-13.661	-6.244	-11.730
T6stemFT-004	-6.792	-16.614	-8.276	-7.888	-14.510	-17.288	-17.718	-9.534	-17.684	-9.272	-17.126	-8.597	-4.416	-11.304	-12.436	-5.006	-8.511
T6flowerFT-004	-6.447	-11.637	-9.340	-8.520	-13.883	-11.813	-13.573	-6.317	-14.373	-9.747	-15.043	-10.693	-3.413	-8.109	-9.869	-4.472	-8.877
T6seedFT-004	-7.487	-7.420	-13.427	-12.123	-9.007	-10.967	-14.760	-11.533	-12.097	-10.637	-13.300	-18.956	2.938	-4.106	-0.646	-5.587	-4.501
T1seedlingFT-086	-7.970	-13.320	-10.017	-6.100	-12.930	-12.373	-10.447	-4.193	-14.200	-4.857	-11.150	-12.503	2.852	-4.625	-6.206	-3.080	-8.448
T2young leavesFT-086	-14.150	-14.183	-8.117	-6.960	-12.083	-8.310	-10.330	-5.253	-9.493	-4.180	-13.763	-12.560	-2.323	-7.852	-12.387	-4.348	-7.547
T2inner turnipFT-086	-13.120	-7.953	-17.397	-6.713	-8.403	-10.087	-8.003	-9.107	-13.043	-8.433	-17.603	-15.943	-2.029	-8.590	-11.816	-1.423	-2.758
T3inner turnipFT-086	-12.803	-14.350	-16.657	-13.423	-13.350	-10.657	-9.743	-7.353	-11.297	-10.407	-17.920	-12.507	-6.094	-12.414	-10.305	-4.894	-7.668
T4mature leavesFT-086	-15.530	-11.750	-12.150	-13.650	-6.130	-17.623	-7.090	-4.480	-13.600	-0.810	-10.210	-15.263	-4.841	-9.975	-13.114	-3.067	-6.627
T6young leavesFT-086	-6.727	-15.263	-9.167	-4.597	-13.483	-15.263	-12.620	-4.390	-13.657	-7.073	-10.287	-12.660	-0.945	-8.464	-11.526	-3.312	-4.610
T6mature leavesFT-086	-4.330	-14.313	-7.807	-4.000	-14.143	-14.667	-12.767	-4.330	-13.940	-5.747	-14.853	-5.840	-4.318	-8.095	-11.826	-2.154	-6.981
T6inner turnipFT-086	-6.880	-16.743	-13.943	-8.603	-16.697	-15.827	-16.083	-6.307	-15.117	-9.853	-8.520	-13.673	-6.907	-13.994	-11.629	-6.291	-9.044
T6stemFT-086	-10.787	-13.073	-8.097	-5.620	-14.400	-15.680	-17.643	-6.433	-15.520	-10.557	-14.870	-8.153	-6.384	-10.593	-12.920	-5.387	-9.620
T6flowerFT-086	-11.237	-12.817	-9.190	-6.583	-14.530	-14.483	-15.513	-5.980	-14.887	-10.363	-16.017	-13.510	-5.540	-10.203	-13.171	-4.345	-8.816
T6seedFT-086	-9.360	-3.547	-15.653	-15.653	-8.933	-10.863	-14.120	-6.303	-12.683	-11.477	-11.603	-16.163	1.172	-9.513	-12.094	-4.058	-4.889

Appendix 2, Table 3. Gene expression data as $-\Delta Ct$ (Continued)

ID	GTR2_3	GTR2_4	GTR3_1	GTR3_2	GTR3_3
T1seedlingFT-004	-9.291	-8.145	-15.866	-4.559	-11.698
T2young leavesFT-004	-9.728	-9.918	-13.541	-4.100	-7.922
T2inner turnipFT-004	-6.147	-5.182	-11.091		-5.165
T3inner turnipFT-004	-11.147	-10.170	-15.660	-9.572	-9.595
T4mature leavesFT-004	-12.679				
T6young leavesFT-004	-8.917	-8.217	-13.291	-3.820	-9.990
T6mature leavesFT-004	-11.170	-10.420	-13.032	-4.595	-9.443
T6inner turnipFT-004	-12.435	-11.518	-12.782	-9.098	-10.209
T6stemFT-004	-9.885	-9.598	-9.159	-5.787	0.729
T6flowerFT-004	-12.184	-36.173	-9.711	-4.224	-5.154
T6seedFT-004		-5.321		-5.282	
T1seedlingFT-086	-10.405	-9.808		-5.498	-10.467
T2young leavesFT-086	-10.919	-10.906		-5.404	-8.778
T2inner turnipFT-086	-5.880	-5.196	-11.181	-12.377	-5.327
T3inner turnipFT-086	-9.582	-9.243	-12.777	-8.252	-7.043
T4mature leavesFT-086	-17.424	-11.141	-15.199	-4.835	-14.038
T6young leavesFT-086	-9.867			-3.590	
T6mature leavesFT-086	-11.561	-10.627	-13.824	-6.122	-12.794
T6inner turnipFT-086	-13.662			-10.657	-7.119
T6stemFT-086	-11.904	-11.360	-10.069	-7.917	-7.446
T6flowerFT-086	-10.275		-10.370	-17.878	
T6seedFT-086	-8.887			-10.320	

Gene expression data as $2^{-\Delta\Delta Ct}$

Appendix 2, Table 4. Gene expression data as $2^{-\Delta\Delta Ct}$

ID	MAM1_1	MAM1_2	MAM1_3	MAM3_1	MAM3_2	MAM3_3	MAM3_4	CYP79A2	CYP83A1	CYP83A1	CYP83B1	AOP2_1	AOP2_2	AOP2_3	AOP2_2-1	GS-OH_1	GS-OH_2
T1seedlingFT-004	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
T2young leavesFT-004	51.984	47.835	32.447	23.807	30.204	17.959	29.310	4.180	2.124	10.315	0.829	37.100	21.907	18.043	32.672	0.480	0.224
T6young leavesFT-004	8.000	4.521	4.000	3.597	2.508	5.426	5.414	0.022	0.724	4.009	0.091	8.495	3.784	0.899	7.379	0.350	24.991
T4mature leavesFT-004	0.436	0.693	0.427	0.336	0.727	0.200	0.260	0.044	0.065	0.498	0.198	0.524	0.334	0.875	0.314	0.940	35.671
T6mature leavesFT-004	7.763	2.313	2.713	2.319	0.816	3.547	4.037	0.135	0.130	2.224	0.929	3.287	3.212	0.863	2.670	0.324	60.129
T6stemFT-004	0.176	0.035	0.119	0.134	0.004	0.072	0.140	0.006	0.019	0.153	0.002	0.058	0.023	0.006	0.046	0.073	11.832
T6flowerFT-004	0.374	0.214	0.380	0.209	0.120	0.293	0.363	2.412	0.024	0.432	0.035	0.476	0.066	0.026	0.426	0.045	15.032
T6seedFT-004	0.356	0.080	2.842	0.749	0.018	1.000	0.430	0.020	0.023	0.050	0.017	0.040	0.457	0.067	0.027	0.000	7.311
T2inner turnipFT-004	2.783	1.195	1.228	4.103	0.330	5.229	5.802	0.007	0.149	0.845	0.959	0.074	0.036	0.005	0.071	0.128	0.024
T3inner turnipFT-004	1.923	0.386	0.546	0.586	0.017	0.845	0.910	0.007	0.009	0.098	0.062	0.041	0.031	0.001	0.032	0.028	0.161
T6inner turnipFT-004	0.022	0.003	0.224	0.043	0.004	0.002	0.003	0.017	0.001	0.003	0.031	0.055	0.033	0.002	0.003	0.002	45.255
T1seedlingFT-086	4.735	1.033	2.615	1.136	0.008	1.610	1.945	0.141	0.129	1.357	0.191	1.678	0.760	0.751	1.237	0.028	5.229
T2young leavesFT-086	54.821	15.853	29.175	12.611	0.029	22.627	25.049	3.095	0.664	5.040	0.216	38.944	30.065	11.605	29.857	0.001	0.072
T6young leavesFT-086	58.756	13.865	16.223	12.467	0.027	16.186	19.071	0.201	0.966	9.404	0.176	35.671	28.773	4.542	33.128	0.000	12.381
T4mature leavesFT-086	0.357	0.824	0.724	0.360	0.532	0.172	0.129	1.251	0.047	0.584	0.233	0.677	0.457	1.154	0.228	0.000	0.028
T6mature leavesFT-086	26.477	8.734	12.611	14.354	0.221	17.228	18.593	0.052	0.387	7.551	0.722	10.679	18.043	1.995	9.021	0.007	65.194
T6stemFT-086	2.901	0.468	0.667	0.869	0.003	1.042	1.401	0.178	0.051	0.574	0.004	1.165	0.445	0.186	1.107	0.000	0.742
T6flowerFT-086	5.005	0.940	1.628	1.379	0.018	1.682	2.367	0.709	0.095	0.940	0.014	2.417	0.877	0.158	2.158	0.001	0.543
T6seedFT-086	0.026	0.016	0.518	0.127	0.073	0.066	0.065	0.059	0.003	0.001	0.012	0.016	0.094	0.034	0.015	0.001	1.995
T2inner turnipFT-086	11.314	0.929	1.769	3.160	0.001	5.389	5.883	0.015	0.219	2.462	0.312	0.231	0.147	0.025	0.185	0.000	0.147
T3inner turnipFT-086	25.932	2.417	5.966	4.037	0.006	7.890	8.056	0.017	0.077	1.353	0.070	1.643	2.261	0.513	1.154	0.001	0.183
T6inner turnipFT-086	0.249	0.014	0.237	0.037	0.001	0.033	0.046	0.045	0.002	0.019	0.044	0.018	0.069	0.004	0.012	0.000	11.132

Appendix 2, Table 5. Gene expression data as 2-ΔΔCt (continued)

ID	GS-OH_3	MYB28_1	MYB28_2	MYB29	MYB34_1	MYB34_2	MYB34_3	MYB34_4	MYB51_1	MYB51_2	MYB51_3	GTR1_1	GTR1_2	GTR1_3	GTR2_1	GTR2_2	GTR2_3
T1seedlingFT-004	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
T2young leavesFT-004	0.034	10.056	3.555	0.126	0.582	0.187	0.251	0.330	2.378	0.973	0.519	0.967	0.742	3.779	0.629	1.084	0.738
T6young leavesFT-004	0.003	1.888	6.364	21.756	0.002	0.001	0.574	0.003	0.180	263.806	2.271	0.544	0.841	1.096	1.468	2.425	1.296
T4mature leavesFT-004	0.060	0.120	0.046	58.485	0.050	0.059	0.294	0.015	13.392	6793.786	0.099	0.418	0.132	0.259	0.769	1.584	0.096
T6mature leavesFT-004	0.003	3.197	3.124	0.195	0.003	0.006	0.348	0.007	1.516	31.197	177.704	0.847	0.750	4.054	2.109	2.116	0.272
T6stemFT-004	0.002	1.718	0.602	0.914	0.000	0.000	0.011	0.001	0.042	1.152	9.691	0.267	0.058	3.224	0.194	0.886	0.662
T6flowerFT-004	0.049	0.822	0.389	1.411	0.008	0.001	0.106	0.010	0.030	4.879	2.266	0.536	0.534	19.096	0.281	0.687	0.135
T6seedFT-004	0.918	0.048	0.032	41.451	0.015	0.000	0.003	0.048	0.016	16.336	0.007	0.744	8.560	11414.569	0.130	14.271	
T2inner turnipFT-004	0.658	0.003	0.176	0.244	1.112	1.888	0.004	0.378	0.064	6.916	0.015	2.896	0.833	7.227	3.541	46.014	8.835
T3inner turnipFT-004	0.056	0.001	0.002	0.961	0.172	0.162	0.007	0.013	0.004	0.354	0.863	0.109	0.061	3.412	0.184	0.377	0.276
T6inner turnipFT-004	0.001	0.009	0.011	4.857	0.001	0.000	0.003	0.003	0.030	22.316	0.244	0.008	0.003	1.379	0.082	0.095	0.113
T1seedlingFT-086	0.015	0.514	2.080	2.732	0.006	0.008	0.463	0.011	0.887	72.505	0.646	0.790	5.976	241.939	0.737	0.925	0.462
T2young leavesFT-086	0.008	1.919	1.146	4.913	0.094	0.009	0.222	0.295	1.417	11.849	0.621	1.140	0.638	3.335	0.306	1.728	0.323
T6young leavesFT-086	0.004	0.927	5.897	1.862	0.001	0.002	0.404	0.016	0.191	131.903	0.580	2.963	0.418	6.056	0.628	13.232	0.670
T4mature leavesFT-086	0.046	0.117	0.011	304.437	0.000	0.085	0.380	0.017	14.655	139.102	0.095	0.199	0.146	2.015	0.744	3.268	0.004
T6mature leavesFT-086	0.008	2.378	8.918	1.178	0.001	0.002	0.421	0.014	0.479	5.566	65.496	0.286	0.539	4.920	1.401	2.558	0.207
T6stemFT-086	0.018	1.945	2.901	0.986	0.001	0.000	0.098	0.005	0.017	5.502	13.177	0.068	0.095	2.305	0.149	0.411	0.163
T6flowerFT-086	0.022	0.912	1.488	0.901	0.001	0.000	0.134	0.007	0.020	2.485	0.322	0.123	0.125	1.937	0.307	0.717	0.506
T6seedFT-086	13.454	0.010	0.003	43.612	0.016	0.001	0.107	0.032	0.009	52.954	0.051	2.532	0.202	4.086	0.374	10.908	1.323
T2inner turnipFT-086	0.634	0.003	1.360	62.973	0.027	0.045	0.015	0.025	0.074	0.827	0.060	1.398	0.383	4.954	2.326	47.754	10.638
T3inner turnipFT-086	0.008	0.005	0.013	2.042	0.018	0.013	0.052	0.084	0.019	0.664	0.645	0.084	0.027	14.115	0.210	1.589	0.817
T6inner turnipFT-086	0.001	0.034	0.367	0.201	0.001	0.000	0.107	0.006	0.028	448.822	0.287	0.048	0.009	5.639	0.080	0.612	0.048

Appendix 2, Table 6. Gene expression data as 2- $\Delta\Delta C_t$ (continued)

ID	GTR2_4	GTR3_1	GTR3_2	GTR3_3
T1seedlingFT-004	1.000	1.000	1.000	1.000
T2young leavesFT-004	0.293	5.010	1.375	13.693
T6young leavesFT-004	0.951	5.959	1.669	3.265
T4mature leavesFT-004				
T6mature leavesFT-004	0.207	7.129	0.976	4.773
T6stemFT-004	0.365	104.483	0.427	2003.283
T6flowerFT-004	0.000	71.267	1.262	93.269
T6seedFT-004	7.082		0.606	
T2inner turnipFT-004	7.797	27.370		92.547
T3inner turnipFT-004	0.246	1.153	0.031	4.294
T6inner turnipFT-004	0.097	8.478	0.043	2.806
T1seedlingFT-086	0.316		0.522	2.346
T2young leavesFT-086	0.148		0.557	7.567
T6young leavesFT-086			1.958	
T4mature leavesFT-086	0.125	1.587	0.826	0.197
T6mature leavesFT-086	0.179	4.118	0.339	0.468
T6stemFT-086	0.108	55.587	0.098	19.044
T6flowerFT-086		45.134	0.000	
T6seedFT-086			0.018	
T2inner turnipFT-086	7.722	25.711	0.004	82.749
T3inner turnipFT-086	0.467	8.507	0.077	25.187
T6inner turnipFT-086			0.015	23.892

Appendix 3: Wilcoxon and t-tests

T-test

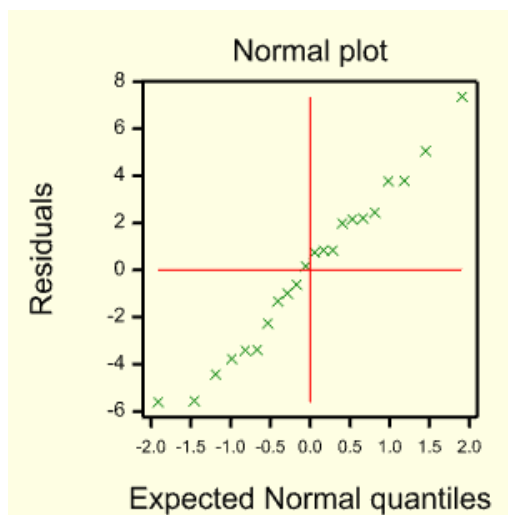
Normality verifications

Shapiro-Wilk test for Normality

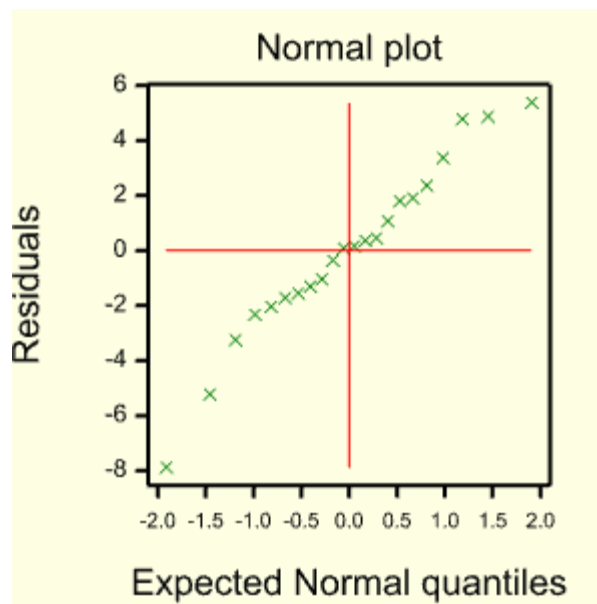
Data variate: MAM3_Bra013009
Test statistic W: 0.9657
Probability: 0.611

Shapiro-Wilk test for Normality

Data variate: GS_OH__Bra022920
Test statistic W: 0.9269
Probability: 0.106



Appendix 3, Figure 1. Q-Q plot for GS_OH_BRA022920 (GS-OH_1) gene expression data (- Δ Ct)



Appendix 3, Figure 2. Q-Q plot for MAM3_BRA013009 (MAM3_2) gene expression data ($-\Delta Ct$)

Equal variances verification

Variances and degrees of freedom

Accession	Var_	d_f
FT-004	15.28	10
FT-086	6.59	10

Bartlett's test for homogeneity of variances

Chi-square 1.64 on 1 degrees of freedom: probability 0.201

Variances and degrees of freedom

Accession	Var_	d_f
FT-004	16.92	10
FT-086	8.76	10

Bartlett's test for homogeneity of variances

Chi-square 1.01 on 1 degrees of freedom: probability 0.314

T-tests

Two-sample t-test

Variate: MAM3_Bra013009
Group factor: Accession

Test for equality of sample variances

Test statistic F = 1.93 on 10 and 10 d.f.

Probability (under null hypothesis of equal variances) = 0.31

Summary

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
FT-004	11	-9.45	16.92	4.113	1.240
FT-086	11	-12.97	8.76	2.960	0.892

Difference of means: 3.519
Standard error of difference: 1.528

95% one-sided confidence interval for difference in means: (0.8842, ...)

Test of null hypothesis that mean of MAM3_Bra013009 with Accession = FT-004 is not greater than mean with Accession = FT-086

Test statistic t = 2.30 on 20 d.f.

Probability = 0.016

Two-sample t-test

Variate: GS_OH__Bra022920
Group factor: Accession

Test for equality of sample variances

Test statistic F = 2.32 on 10 and 10 d.f.

Probability (under null hypothesis of equal variances) = 0.20

Summary

Standard

Standard error

Sample	Size	Mean	Variance	deviation	of mean
FT-004	11	-4.106	15.28	3.909	1.1785
FT-086	11	-11.225	6.59	2.566	0.7738

Difference of means: 7.119
Standard error of difference: 1.410

95% one-sided confidence interval for difference in means: (4.688, ...)

Test of null hypothesis that mean of GS_OH__Bra022920 with
Accession = FT-004 is not greater than mean with Accession = FT-086

Test statistic $t = 5.05$ on 20 d.f.

Probability < 0.001

Wilcoxon tests on GLS

Normality verifications

Shapiro-Wilk test for Normality

Data variate: Pro
Test statistic W: 0.8943
Probability: < 0.001

Shapiro-Wilk test for Normality

Data variate: Napol
Test statistic W: 0.7457
Probability: < 0.001

Shapiro-Wilk test for Normality

Data variate: Napin
Test statistic W: 0.8715
Probability: < 0.001

Shapiro-Wilk test for Normality

Data variate: %4hBrass
Test statistic W: 0.6661
Probability: < 0.001

Shapiro-Wilk test for Normality

Data variate: Canapin
Test statistic W: 0.9168
Probability: < 0.001

Shapiro-Wilk test for Normality

Data variate: Erucin
Test statistic W: 0.4197
Probability: <0.001

Shapiro-Wilk test for Normality

Data variate: Brass
Test statistic W: 0.7584
Probability: <0.001

Shapiro-Wilk test for Normality

Data variate: Naturtiin
Test statistic W: 0.8474
Probability: <0.001

Shapiro-Wilk test for Normality

Data variate: Berteroin
Test statistic W: 0.4797
Probability: <0.001

Shapiro-Wilk test for Normality

Data variate: %4mBrass
Test statistic W: 0.8453
Probability: <0.001

Shapiro-Wilk test for Normality

Data variate: NeoBrass
Test statistic W: 0.7840
Probability: <0.001

Wilcoxon rank-sum tests on all GLS

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Pro
Group factor: Accession

Value of U: 1814.0 (first sample has higher rank sum).

Normal approximation: 4.35 ($p < 0.001$)
Adjusted for ties: 4.35 ($p < 0.001$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Napol
Group factor: Accession

Value of U: 266.0 (first sample has higher rank sum).

Normal approximation: 9.84 ($p < 0.001$)
Adjusted for ties: 9.85 ($p < 0.001$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Napin
Group factor: Accession

Value of U: 541.0 (second sample has higher rank sum).

Normal approximation: 8.86 ($p < 0.001$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: %4hBrass
Group factor: Accession

Value of U: 1228.0 (first sample has higher rank sum).

Normal approximation: 6.43 ($p < 0.001$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Canapin
Group factor: Accession

Value of U: 2024.0 (first sample has higher rank sum).

Normal approximation: 3.61 ($p < 0.001$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Erucin
Group factor: Accession

Value of U: 2681.0 (second sample has higher rank sum).

Normal approximation: 1.28 ($p=0.201$)
Adjusted for ties: 1.34 ($p=0.179$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Brass
Group factor: Accession

Value of U: 2837.0 (second sample has higher rank sum).

Normal approximation: 0.73 ($p=0.467$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Naturtiin
Group factor: Accession

Value of U: 3016.0 (second sample has higher rank sum).

Normal approximation: 0.09 ($p=0.927$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Berteroin
Group factor: Accession

Value of U: 2060.0 (first sample has higher rank sum).

Normal approximation: 3.48 ($p<0.001$)
Adjusted for ties: 3.59 ($p<0.001$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: %4mBrass
Group factor: Accession

Value of U: 2649.0 (first sample has higher rank sum).

Normal approximation: 1.39 ($p=0.164$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: NeoBrass
Group factor: Accession

Value of U: 2177.0 (second sample has higher rank sum).

Normal approximation: 3.07 ($p=0.002$)
(under null hypothesis that group FT-004 is equal to group FT-086).

Sample sizes: 78, 78.

Wilcoxon rank-sum tests on GLS with significant differences between accessions

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Pro
Group factor: Accession

Value of U: 1814.0 (first sample has higher rank sum).

Normal approximation: 4.35 ($p<0.001$)
Adjusted for ties: 4.35 ($p<0.001$)
(under null hypothesis that group FT-004 is not greater than group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: %4hBrass
Group factor: Accession

Value of U: 1228.0 (first sample has higher rank sum).

Normal approximation: 6.43 ($p<0.001$)
(under null hypothesis that group FT-004 is not greater than group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Berteroin
Group factor: Accession

Value of U: 2060.0 (first sample has higher rank sum).

Normal approximation: 3.48 ($p < 0.001$)

Adjusted for ties: 3.59 ($p < 0.001$)

(under null hypothesis that group FT-004 is not greater than group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Canapin

Group factor: Accession

Value of U: 2024.0 (first sample has higher rank sum).

Normal approximation: 3.61 ($p < 0.001$)

(under null hypothesis that group FT-004 is not greater than group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Napol

Group factor: Accession

Value of U: 266.0 (first sample has higher rank sum).

Normal approximation: 9.84 ($p < 0.001$)

Adjusted for ties: 9.85 ($p < 0.001$)

(under null hypothesis that group FT-004 is not greater than group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

Variate: Napin

Group factor: Accession

Value of U: 541.0 (second sample has higher rank sum).

Normal approximation: 8.86 ($p < 0.001$)

(under null hypothesis that group FT-004 is not less than group FT-086).

Sample sizes: 78, 78.

Mann-Whitney U (Wilcoxon rank-sum) test

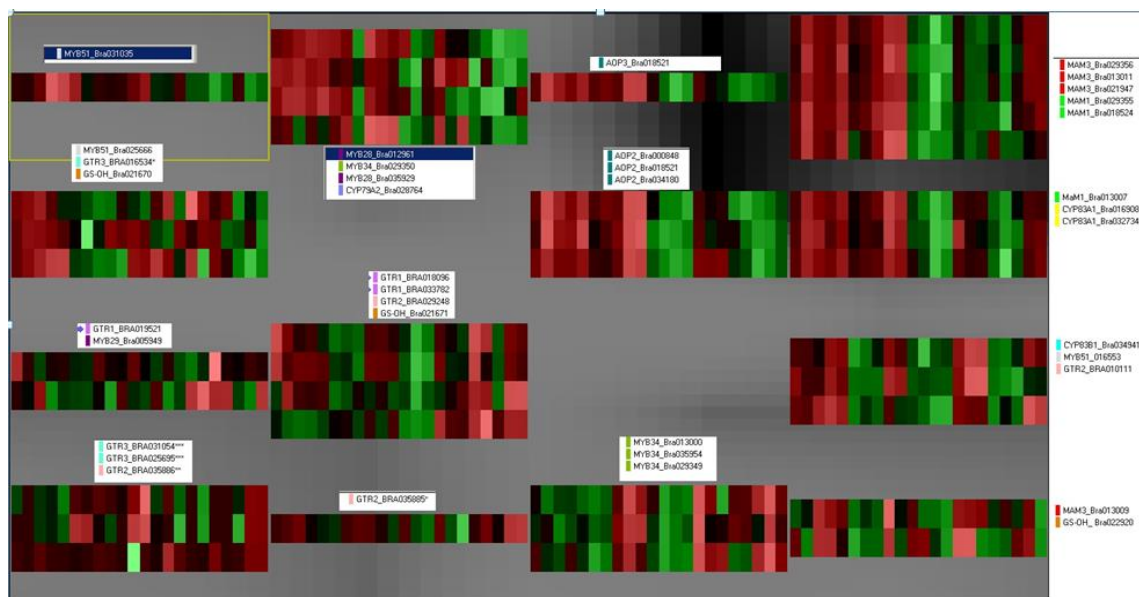
Variate: NeoBrass

Group factor: Accession

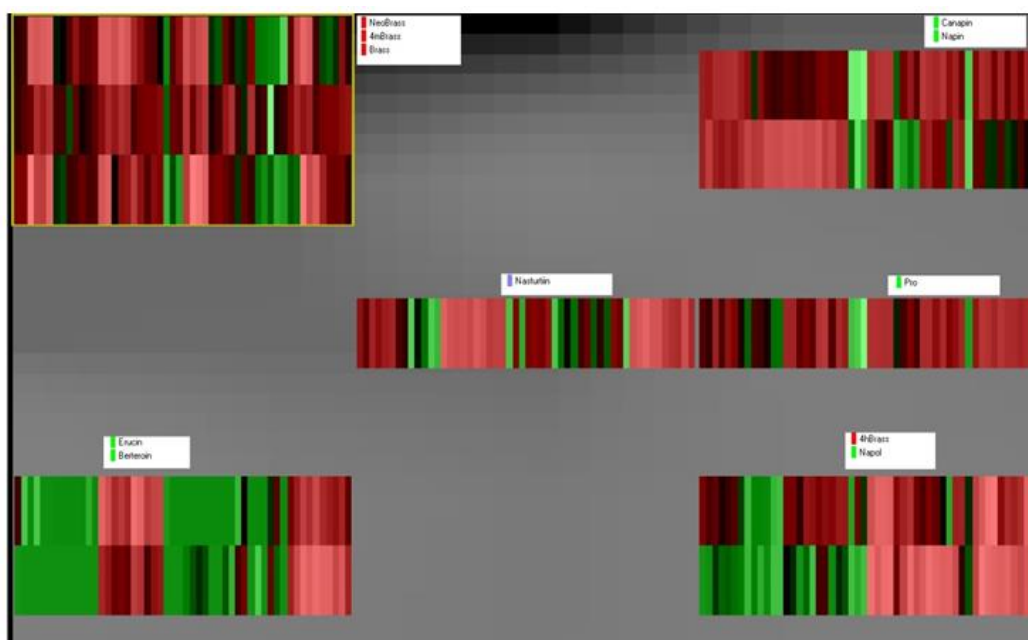
Value of U: 2177.0 (second sample has higher rank sum).

Normal approximation: 3.07 ($p=0.001$)
(under null hypothesis that group FT-004 is not less than group FT-086).
Sample sizes: 78, 78.

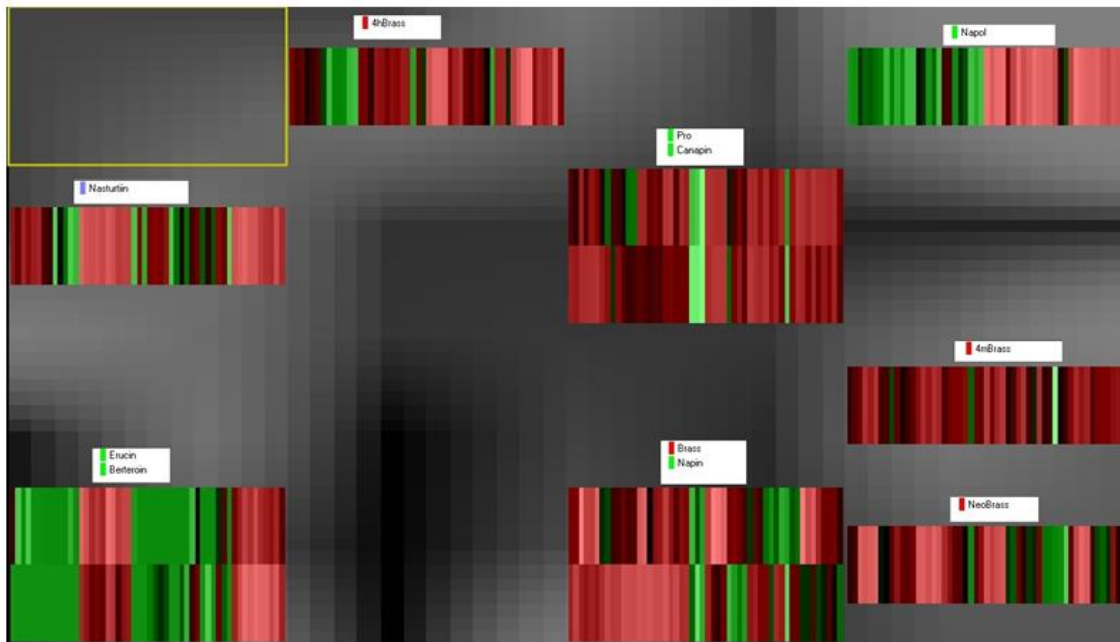
Appendix 4: SOM maps not shown in the report



Appendix 4, Figure 1. 4 x 4 SOM for GLS biosynthetic genes



Appendix 4, Figure 2. 3 x 3 SOM for GLS



Appendix 4, Figure 3. 4 x 4 SOM for GLS