Untargeted metabolic fingerprinting of a loss-of-function and an overexpression mutant reveals involvement in flavonol and pyranoanthocyanidin glycosylation of *BGLU1*, *BGLU3* and *BGLU4* in *Arabidopsis thaliana*

Dissertation

submitted by
Jana-Freja Frommann

for the Degree of
Doctor of Natural Sciences (*Dr. rer. nat.*)
at the Faculty of Biology
Bielefeld University, Germany

July 2020

General remarks

The work of this thesis was performed at the chair of Genetics and Genomics of Plants (CeBiTec & Faculty of Biology, Bielefeld University) from May 2017 until July 2020 under the supervision of Prof. Dr. Bernd Weisshaar and Dr. Ralf Stracke. The UHPLC-DAD-ESI$^+$-QTOF-MS/MS measurements were performed in cooperation with the chair of Chemical Ecology under the supervision of Dr. Rabea Schweiger. The MinION long-read sequencing was performed in collaboration with Dr. Boas Pucker. The cloning of complementation constructs was performed with Lennart Malte Sielmann during his master thesis at the chair of Genetics and Genomics of Plants. Since, the present work is only based on studies in *Arabidopsis thaliana*, the prefix *At* will be omitted for genes and proteins from *Arabidopsis thaliana*.

Statement of authorship

I hereby declare that I am the sole author of this dissertation

Untargeted metabolic fingerprinting of a loss-of-function and an overexpression mutant reveals involvement in flavonol and pyranoanthocyanidin glycosylation of *BGLU1*, *BGLU3* and *BGLU4* in *Arabidopsis thaliana*

and that it is the result of my own investigations, unless otherwise acknowledged in the text. All references and all sources of information have been specifically acknowledged. This dissertation has not been submitted, either in part or whole, for a degree at this or any other university. I have not made any earlier attempts for an unsuccessful promotion.

Bielefeld, 2th July 2020

Jana-Freja Frommann

Abstract

Glycosylation is one of the major mechanisms in plants that tremendously expand the diversity of flavonoids. Besides the well-known uridine diphosphate (UDP)-glycosyltransferases (GTs), a second type of acyl-glucose-dependent GTs acting on flavonoids has been discovered. They were termed GH1-type GTs, based on their amino acid sequence homology to glycoside hydrolases (GHs) typically acting as β-glucosidases (BGLUs). Phylogenetic analyses in Arabidopsis thaliana revealed a cluster of eleven BGLUs (AtBGLU1 to AtBGLU11) including known GH1-type GTs. In this cluster AtBGLU6, known to be a GH1-type GT acting on flavonol substrates, together with AtBGLU1 to AtBGLU5 form a separate branch. In the remaining cluster AtBGLU10 is known to act on anthocyanin substrates. Based on this amino acid sequence similarity analyses, it can be assumed that AtBGLU1 through AtBGLU5 may have similar function as AtBGLU6.

Therefore, this work focused on the functional characterization of the genes *AtBGLU1*, *AtBGLU3* and *AtBGLU4*. In a reverse genetic approach, homozygous loss-of-function T-DNA insertion mutants were identified and characterized by MinION long-read sequencing concerning their T-DNA insertion position. By additionally consulting 35S promotor-driven overexpression mutants, differential gene expression of the three *BGLUs* was verified by RT- and qRT-PCR. The mutants and the corresponding wild type (*A. thaliana* Col-0) were investigated by untargeted metabolic fingerprinting, using UHPLC-DAD-ESI^{+}-QTOF-MS/MS. Samples for metabolite analysis were taken from tissues with high gene expression levels in the wild type. Candidate differential metabolic features with concentration differences between genotypes, correlated to the differential gene expression of the respective BGLU gene, were chosen, being putative GT substrates or products. For *AtBGLU1*, *AtBGLU3* and *AtBGLU4* candidate product compounds, whose MS/MS spectra suggest that they may be glycosylated flavonols were found, namely a triglycosylated kaempferol for *AtBGLU1* and a triglycosylated quercetin for *AtBGLU3* and *AtBGLU4*. In addition, for *AtBGLU3* the putative condensed (epi)gallocatechin–vitisin A candidate product with the respective (epi)gallocatechin–carboxypyranomalvidin substrate as well as the corresponding vitisin B compounds and further condensed flavanol–anthocyanidin substrate compounds were identified in seeds. These condensed compounds were never reported for Arabidopsis before and no (epi)gallocatechin–vitisin A is reported in literature to current stage. For explicit characterization of the compounds, NMR measurements are in progress. To investigate proposed functional redundancy of *AtBGLU1* and *AtBGLU4* in the respective loss-of-function mutants in future work, double mutants of homologous *BGLUs* were generated.

The current results indicate that *AtBGLU1*, *AtBGLU3* and *AtBGLU4* encode GH1-type flavonoid GTs, while *At*BGLU3 seems to be a multifunctional GH1-type GT, acting on several flavonoid substrates, some of them never reported in Arabidopsis before, providing new insights in flavonoid abundance and diversity in Arabidopsis.

Contents

List of Figures

List of Tables

I. List of Abbreviations

2-ME	2-mercaptoethanol
A11	anthocyanidin 11
AAGT	acyl-glucose-dependent anthocyanin glucosyltransferase
ABA	abscisic acid
ADH	alcohol dehydrogenase
A^+–F	anthocyanidin–flavanol
ANS	anthocyanidin synthase
AOMT	anthocyanidin-O-methyltransferase
ApE	A plasmid Editor
API	atmospheric pressure ionization
At	*Arabidopsis thaliana*
AT	acyltransferase
auto-MS/MS	automatic tandem mass spectrometry
bb	backbone
BFF	benzofuran-forming
BGLU	β-glucosidase
BLAST	Basic Local Alignment Search Tool
BLASTn	Basic Local Alignment Search Tool nucleotide
bp	base pair(s)
BY2	Bright Yellow 2
CAT	catalase
CeBiTec	Center of Biotechnology
cDNA	complementary deoxyribonucleic acid
CDS	coding sequence
CHI	chalcone isomerase
CHS	chalcone synthase
CID	collision induced dissociation
Col-0	Columbia-0
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats associated
C_t	cycle threshold
CTAB	cetyltrimethylammonium bromide
DFR	dihydroflavonol reductase
DAD	diode array detector
DC	direct current
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double stranded DNA

DTT	dithiothreitol
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EIC	Extracted Ion Chromatograms
EF1α	elongation factor 1-alpha
ER	endoplasmic reticulum
ESI	electro-spary ionization
Exo-CIP	Exonuclease-Calf Intestinal Phosphatase
F3'H	flavonoid 3'-hydroxylase
F3'5'H	flavonoid-3'5'-hydroxylase
F–A$^+$	flavanol–anthocyanidin
FC	fold change
FG	feruloyl-glucose
FH3	flavanone 3-hydroxylase
FLS	flavonol synthase
F–pyA$^+$	flavanol–pyranoanthocyanidin
GA	gibberellic acid
GH	glycosyl hydrolase
GH1	glycosyl hydrolase family 1
GFP	green fluorescent protein
GM	geometric mean
GPT	glucose-6-phosphate/phosphate translocator
GRP5	glycine-rich protein 5
GSNO	S-nitrosoglutathione
GSNOR	S-nitrosoglutathione reductase
GT	glycosyltransferase
H$_2$O	water
HCl	hydrochloride
HPLC	high performance liquid chromatographie
HRF	heterocyclic ring fission
INRA	National Institute for Agricultural Research
KCl	kalium chloride
KO	knock-out
KOH	kaliumhydroxid
LAR	leucoanthocyanidin reductase
LB	left border
LB medium	lysogeny broth medium
LC	liquid chromatographie
LDH	lactate dehydrogenase
LiCl	lithium chloride

LOX	lipoxygenase enzyme
MgCl$_2$	magnesium chloride
MgSO$_4$	magnesium sulfate
MoNA	MassBank of North America
MRM	Multiple Reaction Monitoring
MS	mass spectrometry
MS medium	Murashige and Skoog medium
MS/MS	tandem mass spectrometry
N$_2$	nitrogen
NaCl	sodium chloride
NAD	nicotinamidadenindinukleotid
Na(HCOO)	natriumformiat
NASC	Nottingham Arabidopsis Stock Center
NEB	New England BioLabs
NMR	nuclear magnetic resonance spectroscopy
NO	nitric oxide
N-terminal	amino-terminal
ONT	Oxford Nanopore Technologies
OX	overexpression
pHBG	*p*-hydroxybenzoyl-Glc
PAL	phenylalanine ammonia lyase
PCR	polymerase chain reaction
PDC	pyruvate decarboxylase
PEG	polyethylenglycol
Pex 4	Peroxin 4
PPDK	pyruvate phosphate dikinase
PSM	plant secondary metabolite
PSPG	plant secondary product glycosyltransferase
PVP	polyvinylpyrrolidone
pyA$^+$–F	pyranoanthocyanidin–flavanol
QM	quinone methide
qRT-PCR	quantitative real time PCR
QTOF	quadrupole time-of-flight
RB	right border
RDA	retro-Diels-Alder
rf	radio-frequency
RNA	ribonucleic acid
RNAi	RNA interference
RNA-Seq	ribonucleic acid sequencing
RNS	reactive nitrogen species

ROS	reactive oxygen species
RT-PCR	reverse transcriptase PCR
SCF	Sequencing Core Facility
SD	standard deviation
SDS	sodium dodecyl sulfate
SE	standard error
SG	sinapoyl-glucose
SIGnAL	The Salk Institute Genomic Analysis Laboratory
SO_2	sulfur dioxide
SOD	superoxide dismutase
SNP	Single Nucleotide Polymorphism
S.O.C	Super Optimal Broth
TAE	Tris-Acetate EDTA
TAIR	The Arabidopsis Information Resource
T-DNA	transfer DNA
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0
TG	transglycosidase
TraVa	Transcriptome Variation Analysis
Tris	Tris(hydroxymethyl)aminomethane
TOF	time-of-flight
UDP	uridine diphosphate
UGT	UDP-glycosyltransferase
UHPLC	ultra-high-performance liquid chromatography
UV	ultraviolet
UV-vis	ultraviolet-visible
Ws-4	Wassilewskija-4
WT	wild type
XGluc	5-bromo-4-chloro-3-indoxyl-β-D-glucopyranoside

1. Introduction

1.1. Plant secondary metabolites

The plant metabolism includes the primary and the secondary metabolism. While the primary metabolism is responsible for the generation of compounds that are needed for plant growth, the metabolites derived from the secondary metabolism (plant secondary metabolites (PSMs)) are defined as compounds, that are not necessary for the cell growth and which have a minor importance for the primary metabolism (Kessler and Kalske, 2018). This definition can led to the assumption, that PSMs are of minor importance for the plant's life. Indeed, in the beginning, they were regarded as detoxification and waste products from the primary metabolism (Reznik, 1960), which is rather wrong, concidering, that PSMs seem to be an important adaptation to the plant's sessile life style. Plants have to adapt very strongly to all changes of their environment. They are known to protect against abiotic stresses like UV-B radiation (Olsson *et al.*, 1998; Solovchenko and Schmitz-Eiberger, 2003), drought and heat (Ramakrishna and Ravishankar, 2011; Berini *et al.*, 2018) and they had been identified to play important roles in the protection against stressors like herbivores and pathogens, either by direct chemical defense, providing toxic, antidigestive, and antinutritive compounds for the antagonist (Duffey and Stout, 1996) or by indirect defense, for example, signaling bad food quality, limited accessibility or detraction (de Moraes *et al.*, 2001; Kessler and Baldwin, 2001). In the same time, they provide beneficial traits, necessary for plant-to-plant communication and to attract pollinators (Raguso, 2008; Schiestl, 2010) as well as pivotal symbionts like rhizobacteria (Bais *et al.*, 2006) and mycorrhizal fungi (Akiyama *et al.*, 2005).

Therefore, plants are dependent on their PSMs in order to protect their life in case of harmful attacks or circumstances and also to improve their life quality and maintain their reproduction. Thus, even if they are not responsible for the energy supply from the primary metabolism, they have a huge impact to increase the fitness of plants (Kessler and Baldwin, 2002).

In 2002, Fiehn stated an estimated amount of 200,000 PSMs being isolated and identified. This appears, in comparison to approximately 391,000 described plant species, a comparatively low number, but need to keep in mind, that PSMs characterization started much later than the characterization of plants, more precisely around 200 years ago with the isolation of morphine. In addition, it has to be taken into account, that only a small proportion of the plants have been profiled, regarding their PSMs, leading to the assumption that an immense number of new compounds is still unknown. Most of the known compounds and compound classes are broadly expressed among members of different plant phyla, accounting thus for a huge amount of different PSMs in each single plant. (Hartmann, 2007; Kessler and Kalske, 2018, and references therein). This diversity is achieved through different mechanisms, including the following. Chemically simple precursors from the primary metabolism can be combined to a large number of new compounds by diverse combinations (Dudareva *et al.*, 2004). Gene families of biosynthetic PSMs genes, which encode for several enzymes with similar function, facilitate the production of many different compounds from a few simple precursors (Tholl *et al.*, 2005). Some genes encode for enzymes, which can produce multiple products out of one precursor compound, just by very small changes in the active site (Köllner *et al.*, 2004).

Modifying enzymes, like carboxyl-, methyl-, acetyl- and glycosyltransferases, that in addition often can use multiple substrates, produce a huge variety of compounds (Negre *et al.*, 2003). In addition, spatially and temporally differential and organ-specific expression of biosynthetic genes, enables to exhibit a much broader diversity of different compounds in the same plant (Pichersky *et al.*, 2006; Dudareva *et al.*, 2013).

Besides the beneficial contribution of PSMs to the fitness of plants, they also seem to assign positive effects on human health, becoming apparent, as throughout history the use of plants and PSMs, as plant-derived pharmaceuticals or as dietary supplements, was applied (Raskin *et al.*, 2002).

According to their biosynthetic pathway in the plant, PSMs are divided into three main groups, namely terpenes, phenolic compounds and alkaloids (Taiz *et al.*, 2014; Mera *et al.*, 2019). The terpenes are the largest group of PSMs. They are also called isoprenoids, due to the basic isoprene structure, and are classified according to the number of their isoprene units. They are built from pyruvate, deriving from the primary carbon metabolism (Vranová *et al.*, 2012; Taiz *et al.*, 2014). Alkaloids are a large diverse group of toxic compounds, including a nitrogen in their basic structure, mostly included in a heterocyclic carbon ring. They are normally synthesized from common amino acids like lysine, tyrosine or tryptophan. Despite their toxicity, many alkaloids are pharmacologically useful, if applied in in low doses (Aniszewski, 2007; Taiz *et al.*, 2014). Phenolic compounds contain the phenylpropanoid derived flavonoids, which are of special interest for this work. Therefore, the phenolic compounds will be introduced in more detail.

1.2. Phenolic compounds

A large group of PSMs span compounds with a phenol group, which is an aromatic ring, containing a functional hydroxyl group (figure 1). This chemically heterogeneous group is called phenolic compounds, or shortly phenolics. They range from small molecules like phenol or phenolic acids over hydroxycinnamic acids, coumarins, flavonoids and lignans to huge highly polymerized tannins (Dai and Mumper, 2010). The role of these compounds is as diverse as their chemical structures, including herbivore and pathogen defense, mechanical support in plant growth, attraction to pollinators and fruit dispersers, absorption of the harmful ultraviolet radiation and also growth reduction of competing plants (Olsson *et al.*, 1998; Lattanzio *et al.*, 2008; Ahuja *et al.*, 2012; John and Sarada, 2012; Valenta *et al.*, 2013; Marsh *et al.*, 2020).

The two basic biosynthetic pathways for phenolics involve the shikimic acid pathway and/or the malonic acid pathway, the first one being most important. In the shikimic acid pathway simple carbohydrate precursors, deriving from glycolysis and the pentose phosphate pathway of the primary metabolism, are converted into three different aromatic amino acids, which are phenylalanine, tyrosine, and tryptophan, while phenylalanine is the precursor for the most abundant phenolic compounds. Therefore, an ammonia molecule is removed from phenylalanine to form cinnamic acid. The step is catalyzed by the intensively studied phenylalanine ammonia lyase (PAL), which branches in a certain extend the primary and secondary metabolism. By further addition of hydroxyl groups and other substituents *trans*-cinnamic acid, *p*-coumaric

acid and their derivatives (like caffeic acids, ferulic acid and sinapic acid) are built, forming a group of simple phenolics called phenylpropanoids. All of them contain a benzene ring and a three-carbon side chain. Further important simple phenolics deriving from phenylpropanoids are the coumarins (also termed as phenylpropanoids) and the benzoic acid derivatives. The structure of the basic phenylpropanoids and benzoic acid derivatives is given in figure 1. From these basic carbon skeletons, plants derive a tremendous amount of more complex phenolic compounds, including lignins and flavonoids, the latter belonging to the largest classes of plant phenolics. (Falcone Ferreyra *et al.*, 2012; Taiz *et al.*, 2014, and references therein).
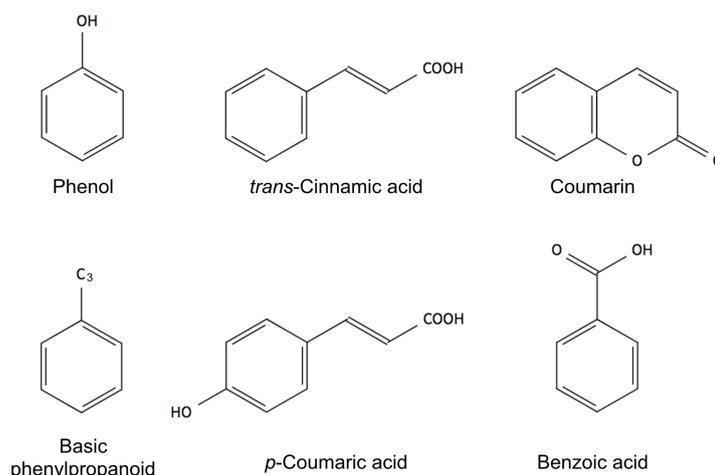
Figure 1 – ***Basic structures of phenolic compounds.*** *The structures of the basic phenolic compounds are given, including the characteristic phenol group. From these basic compounds a tremendous amount of more complex phenolic compounds is derived in plants. Structures are based on Venugopala* et al. *(2013), Taiz* et al. *(2014) and PubChem.*

1.3. Flavonoids

The basic skeleton of flavonoids (figure 3) consists of 15 carbons, which are arranged in two aromatic rings (A and B), linked through a three-carbon bridge (C). The degree of oxidation of the three-carbon bridge, arranged as a heterocyclic ring, is used for the classification of the flavonoids. They are divided into several subgroups like anthocyanidins, flavones, flavanones, flavonols, isoflavonoids and flavanols as the main ones (figure 3). This basic structure of all flavonoids, abbreviated as C_6-C_3-C_6 structure, is derived on the one hand by the mentioned shikimic acid pathway, providing *p*-coumaroyl-CoA from phenylalanine, for the B-ring formation. This B-ring contains the phenylpropanoid three-carbon bridge (C-2, C-3, C-4 in figure 3). On the other hand, the malonic acid pathway provides three malonyl-CoA for the A-ring formation. The reaction of *p*-coumaroyl-CoA and malonyl-CoA is catalyzed by the chalcone synthase (CHS), producing the naringenin chalcone, which is further transformed into the flavanone naringenin by the chalcone isomerase (CHI). This molecule is the precursor for all other flavonoids. The main steps of the biosynthetic pathway of flavonols, flavanols and anthocyanidins are described in figure 2, giving also the chemical structures of the compounds and the catalyzing enzymes for each step. Proceeding from naringenin, the compound is hydroxylated at the 3-position by the flavanone 3-hydroxylase (FH3), to yield dihydrokaempferol. Further hydroxylation by the flavonoid 3'-hydroxylase (F3'H)

leads to dihydroquercetin. Both dihydroflavonols are reduced to leucoanthocyanidins by the dihydroflavonol reductase (DFR). The anthocyanidin synthase (ANS) catalyzes the synthesis of the colored anthocyanidins, while the different types of anthocyanidins are catalyzed by F3'H, the flavonoid-3'5'-hydroxylase (F3'5'H) and the anthocyanidin-*O*-methyltransferase (AOMT). Flavonols are synthesized from dihydroflavonols by the flavonol synthase (FLS), while flavanols derive from leucoanthocyanidins by the activity of the leucoanthocyanidin reductase (LAR). Polymerization of flavanols yield the proanthocyanidins. (Tanaka *et al.*, 2008; Falcone Ferreyra *et al.*, 2012; Taiz *et al.*, 2014; Le Roy *et al.*, 2016). The biosynthesis of the basic flavonoids takes place in the cytosol, where the enzymes are suggested to be anchored in the endoplasmic reticulum (ER) membrane (Grotewold, 2006).
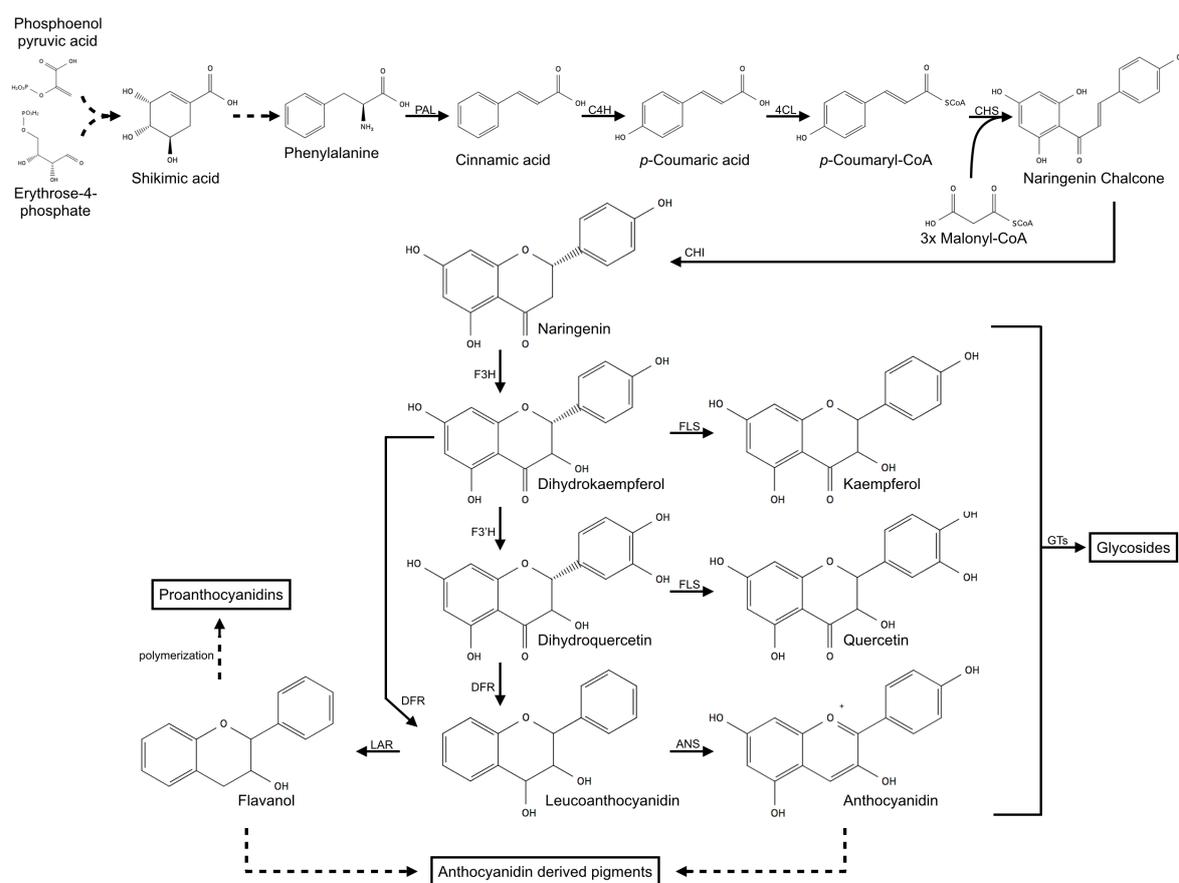
Figure 2 – *Flavonoid biosynthesis pathway. The main steps of the flavonoid biosynthesis pathway in* Arabidopsis thaliana *are depicted, giving the respective intermediates with their chemical structure and marked with arrows the respective catalyzing enzymes. Multiple steps or unknown reactions are indicated by dashed lines. Abbreviations are PAL: phenylalanine ammonia lyase, C4H: cinnamate 4-hydroxylase, 4CL: 4-coumarate:coenzyme A ligase, CHS: chalcone synthase, F3H: flavanone 3-hydroxylase, F3'H: flavonoid 3'-hydroxylase, FLS: flavonol synthase, DFR: dihydroflavonol reductase, ANS: anthocyanidin synthase, LAR: leucoanthocyanidin reductase, GTs: glycosyltransferases. The figure is based on Falcone Ferreyra* et al. *(2012), Roepke (2015) and Nabavi* et al. *(2020).*

Concordant to their diversity and spread across the plant kingdom, flavonoids fulfill diverse functions. This include the functions, that where mentioned above for PSMs, like protection against UV-B radiation, attraction towards pollinators and seed dispersers and defense against pathogens, as well as modulation of chemical plant communication with insects and microbes, for example with rhizobacteria. They are further known as modulators of flower and fruit

color (Lijavetzky *et al.*, 2006; Offen *et al.*, 2006; Griesser *et al.*, 2008) and as protectors against other abiotic stressors, like oxidative and drought stress (Nakabayashi *et al.*, 2014). Apart from this, they are also known for their various health benefits for humans (Tanaka *et al.*, 2008).

To fulfill the various functions the structural diversity of flavonoids is increased by a plenty of different substitutions to the A, B and C rings of the flavonoid skeleton. These are mostly several hydroxyl groups (predominantly at C-3, C-5 and C-7), as well as methyl, glycosyl and acyl groups. The substitutions are catalyzed by isomerases, reductases, hydroxylases, methylases, acyltransferases and glycosyltransferases, with glycosylation as the most abundant modification, resulting in several thousand structures. (Ferrer *et al.*, 2008; Tanaka *et al.*, 2008; Taiz *et al.*, 2014; Le Roy *et al.*, 2016).
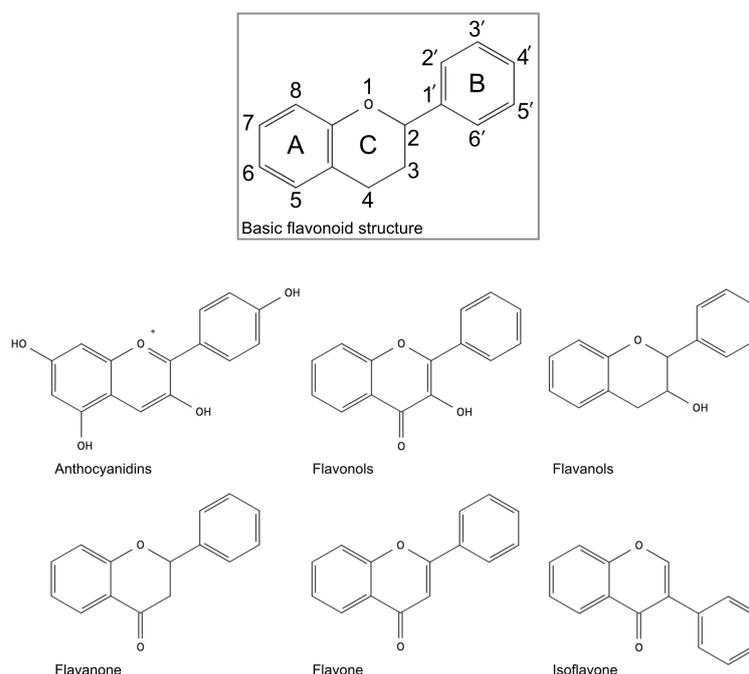
Figure 3 – ***Basic flavonoid structures.*** *The basic C_6-C_3-C_6 flavonoid core structure of the two aromatic rings (A and B) and the three-carbon bridge (C) as heterocyclic ring is depicted in the rectangle. Below basic structures of the main flavonoid subgroups are shown. Structures are based on Panche* et al. *(2016).*

Based on the relevance for this work, flavonols, flavanols and anthocyanidins will be introduced in more detail in the next sections. In addition, the glycosylation of flavonoids will be investigeted.

1.4. Flavonols

Flavonols are mainly divided by their degree of hydroxylation and glycosylation at their aromatic rings (Lillo *et al.*, 2008). The most abundant flavonols are quercetins and kaempferols (figure 2), exhibiting a variety of different biological activities, while the precise function is often not completely understood. One proposed function is the protection of tissue from UV-B radiation (280 - 320 nm) and in this way, protection from DNA mutations and oxidative stress. The UV-B radiation seems to be absorbed in the leaves and stems of plants, while the visible light, which is needed for photosynthesis, can pass. (Olsson *et al.*, 1998; Solovchenko and

Schmitz-Eiberger, 2003; Verdan *et al.*, 2011; Falcone Ferreyra *et al.*, 2012; Taiz *et al.*, 2014). They are also proposed to attract insects and to guide them towards pollen and nectar, by absorbing light at wavelength, invisible for the human eye but visible for insects (Gronquist *et al.*, 2001). As copigments for anthocyanidins, flavonols are also involved in the modulation of flower and fruit color (Asen *et al.*, 1975). Another function, together with flavones, is the proposed induction of rhizobial *nod* genes as well as regulation of auxin transport during the nodulation of rhizobacteria, thus being involved in the mediation and establishment of the symbiosis (Zhang *et al.*, 2009; Falcone Ferreyra *et al.*, 2012). In general, they are obviously involved in polar auxin transport and in this respect, they alter plant development according to their composition in the plant (Ringli *et al.*, 2008; Yin *et al.*, 2014). They also seem to be involved in pollen germination in some plants like *Zea mays* and *Petunia hybrida*, thus, plant fertility (Mo *et al.*, 1992; Mahajan *et al.*, 2011) and in protection against herbivores (Onkokesung *et al.*, 2014).

In addition, flavonols are known to be beneficial for human health, as antioxidants, antiinflammatory, anticancer, neuroprotective and cardioprotective agents and also by disturbing viral replication (Li *et al.*, 2008; Calderón-Montaño *et al.*, 2011; Chen and Chen, 2013).

Flavonols are most abundant in fruits and vegetables. Since, as natural products they are mainly present as a variety of different glycosylated compounds, it can be assumed, that their diverse function is related to diverse glycosylation conditions. For example, for quercetin 300 different glycosides had been identified in 2001, while the precise function of the glycosylation is not known (Jones and Vogt, 2001; Hackman *et al.*, 2008; Li *et al.*, 2008).

1.5. Flavanols

The group of flavanols combines the monomeric flavan-3-ols (catechins, afzelechins and gallocatechins) and flavan-4-ols as well as the oligomeric or polymeric flavan-3-ols, built of two or more monomeric units, called proanthocyanidins or condensed tannins (Hollands *et al.*, 2017; Ottaviani *et al.*, 2018). In contrast to most other flavonoid compounds, flavanols are most abundant as aglycone, but often esterified with gallic acid (Hackman *et al.*, 2008). The predominant flavan-3-ols are differentiated by the degree of their hydroxyl groups (figure 4), while the basic monomer unit is catechin, which exist as four isomers, (+)-catechin, (−)-catechin, (+)-epicatechin and (−)-epicatechin (figure 4). In the following parts for simplification only the term (epi)catechin (or (epi)gallocatechin) will be used, speaking for all four isomers. Flavan-3-ol monomers are bioactive compounds, known to exist in high contents in cocoa, red wine, green tea, red grapes, berries and apples (González-Sarrías *et al.*, 2017).They are most popular for their free radical scavenging activity (Zhu *et al.*, 2002; Wiswedel *et al.*, 2004), predominantly responsible for their positive effect on endothelial function and cardiovascular diseases (one of the main causes of death in developed countries) for humans, including antiinflammatory properties and the ability to lower blood pressure (Leikert *et al.*, 2002; Sies *et al.*, 2005; Rodrìguez-Ramiro *et al.*, 2013; Grassi *et al.*, 2015; Pucci *et al.*, 2017). In addition, they seem to be involved in modulation of intracellular signaling, cytokine regulation and protection against neurodegeneration (Hackman *et al.*, 2008; Nishizawa *et al.*, 2011; Spagnuolo *et al.*, 2018).
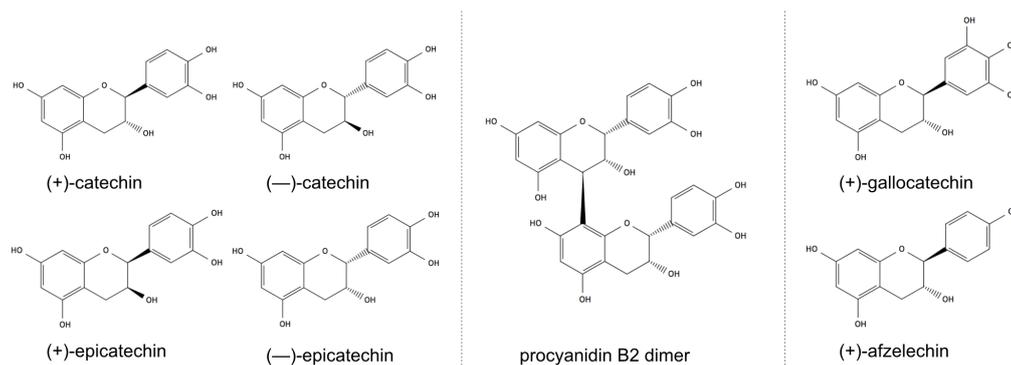
Figure 4 – ***Basic flavanol structures.*** *The basic flavan-3-ol structures are depicted, showing the different catechin isomers (+)-catechin, (–)-catechin, (+)-epicatechin and (–)-epicatechin at the left side, a basic procyanidin B2 dimer of (–)-epicatechin units in the middle and the basic gallocatechin and afzelechins structure from the respective (+)-isomer at the right side. Structures are based on PubChem and Hackman* et al. *(2008).*

1.5.1. Proanthocyanidins or condensed tannins

Besides substitution, a further mechanism, to expand the flavonoid compound diversity is the condensation or polymerization of monomeric flavonoid units. Polymerization of flavanol units yield the so-called proanthocyanidins or condensed tannins, as the most common group of condensed flavonoid compounds. These compounds are brown or nonvisible-colored pigments, often present in specific tissues like seed, peel or bark of many plant species. They were discovered in grapes as the second group of colored compounds, after the monomeric anthocyanidins Somers (1966); Zhao *et al.* (2010a). The term oligomers is used for dimers, trimers or up to heptamers and the term polymers or tannins for larger chains (Li and Deinzer, 2007; Fraser *et al.*, 2012). After treatment with strong acids, these compounds can be hydrolyzed into anthocyanidins, being responsible for the term proanthocyanidins. Proanthocyanidins tend to form complexes with other compounds, like proteins, carbohydrates and alkaloids, which is believed to be the mechanism to interact with biological systems. Since, they bind proteins, biological processes like digestion can be hampered. Therefore, proanthocyanidins are regarded as toxic and thus, being involved in the defense of microorganisms like fungi and bacteria. However, moderate amounts of some tannins can have health benefits for humans, like highly beneficial antioxidant characteristics due to free radical scavenging and antibacterial, antiviral, anticarcinogenic, antiinflammatory, and vasodilatory activities, the latter promoting also cardiovascular health. (Sato *et al.*, 1999; Lotito *et al.*, 2000; Li and Deinzer, 2007; Masuda *et al.*, 2007; Hellström and Mattila, 2008; Taiz *et al.*, 2014; Hollands *et al.*, 2017). Proanthocyanidins are widely distributed in higher plants and are part of the human diet in fruits, berries, nuts and seeds (Li and Deinzer, 2007; Hellström and Mattila, 2008; Taiz *et al.*, 2014; Hollands *et al.*, 2017).

As mentioned, the oligomers are built by the flavan-3-ol monomeric units, while the extension units are the ones, which are connected to another monomeric unit by their C-ring (the lower unit in figure 4), whereas the terminal unit is the respective unit, not linked by the C-ring (upper unit in figure 4; Flamini, 2012). The monomeric units are linked by different interflavonoid linkages like single carbon–carbon (C–C) bonds (C-4 → C-8 or sometimes C-4 → C-6), called B-type linkages or with an additional C-2 → C-7 linkage (or sometimes C-2 → C-5), resulting

in the double A-type linkage (Li and Deinzer, 2007; Hellström and Mattila, 2008; Appeldoorn, 2009; Robbins *et al.*, 2009). The classification of different proanthocyanidins derive on the one hand from these linkages and on the other hand from the respective monomeric units. The most common proanthocyanidins are the procyanidins, exclusively built by the monomers catechin and epicatechin (figure 4). Another common group, the prodelphinidins, are exclusively built by gallocatechin and epigallocatechin. Condensed pigments of different flavanol monomers exist, too. (Li and Deinzer, 2007; Hellström and Mattila, 2008; Appeldoorn, 2009; Fraser *et al.*, 2012; Hollands *et al.*, 2017).

For characterization and identification of the complex proanthocyanidins, LC-MS/MS based methods can replace elaborate sample preparation, since Li and Deinzer (2007) developed a protocol of characteristic fission rules for proanthocyanidins. These rules complete the simple quinone methide (QM) fission, determining the fragments, derived after fission of the monomeric units.
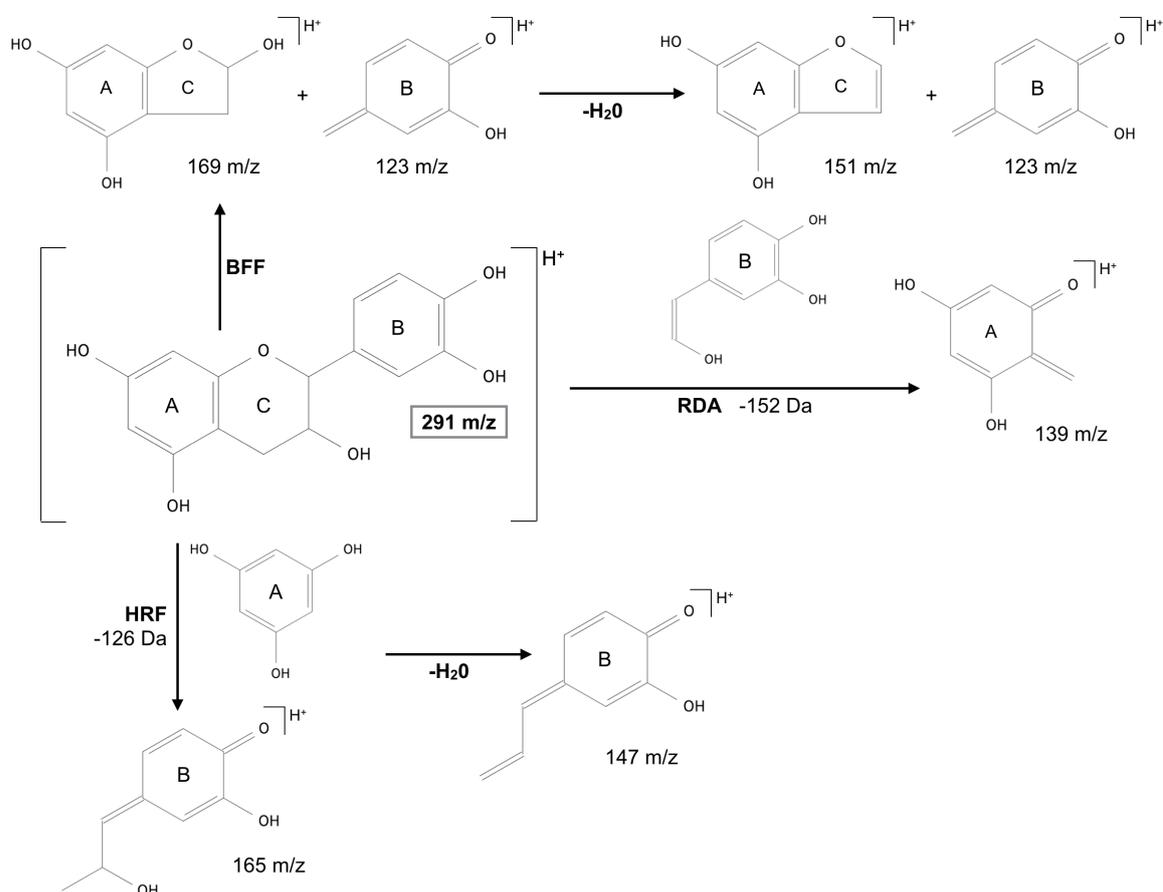
Figure 5 – *Fission rules for proanthocyanidins. For catechin (291 m/z) the proanthocyanidin fission rules are demonstrated, which are BFF fission of ring B with a 122 Da loss or a 140 Da loss, in case of an additional water loss, RDA fission of ring B with a 152 Da loss, HRF fission of ring A with a 126 Da loss and a 144 Da loss, if an additional water is lost. The figure is based on Li and Deinzer (2007) and Flamini (2012).*

They can also be applied to other condensed compounds, including flavanol units. The protocol includes five fission rules, based on retro-Diels-Alder (RDA) fission (loss of ring B), heterocyclic ring fission (HRF, loss of ring A) and benzofuran-forming fission (BFF, loss of ring B). Rule number one says, after fragmentation of the single ring C by HRF fission, no

further RDA fission is possible and the other way around. Rule number two impplies, that the A ring, which is fragmented by BFF fission, cannot fragment by HRF fission and the other way around. Rule number three postulates the neutral loss of HRF fission from ring C, which is always 126 Da. Rule number four defines the neutral losses through RDA fission, which are 136 Da for (epi)afzelechin, 152 Da for (epi)catechin and 168 Da for (epi)gallocatechin. Rule number five defines the neutral losses through BFF fission, which are 106 Da for (epi)afzelechin, 122 Da for (epi)catechin and 138 Da for (epi)gallocatechin. The respective neutral losses for H_2O/BFF fission are 124 Da, 140 Da, and 156 Da. (Li and Deinzer, 2007; Flamini, 2012). The fission rules are shown in figure 5 for catechin as example. With this fission rules, proanthocyanidins can be characterized mainly according to neutral loss information of tandem mass spectrometry data, which will be introduced below.

1.6. Anthocyanidins

The term anthocyanidin is used for the aglycone, while anthocyanins are the respective glycosides. For simplification the term anthocyanidins will be used for aglycones and glycosides in the following work. Only if specificaly the glycosides are addressed, the term anthocyanins will be used. The anthocyanins are most abundant in nature, while the most predominant position for sugar attachments is C-3, followed by C-5. In addition, C-7 glycosylation and in a few cases C-3', C-4' or C-5' glycosylation is known (Fukuchi-Mizutani *et al.*, 2003; Tanaka *et al.*, 2008; Matsuba *et al.*, 2010; Cheng *et al.*, 2014; Taiz *et al.*, 2014). The glycosyl group reduces the reactivity, while making the molecule more stable. In addition, anthocyanidins are often acylated with aromatic phenylacyl acids (*p*-coumaric, caffeic, ferulic, sinapic, gallic or *p*-hydroxybenzoic acids) or aliphatic acids (malonic, acetic, malic, succinic, tartaric or oxalic acids) at the sugar moieties, which can again be glycosylated, to further stabilize the molecule (Zhang *et al.*, 2014; Saigo *et al.*, 2020).

Anthocyanidins built one of the most widespread families of natural pigments in the plant kingdom, absorbing in ultraviolet and visible lights (Saigo *et al.*, 2020). In 2015, Wallace and Giusti stated 27 known anthocyanidin aglycones, accounting for more than 700 different structures, due to the substitution pattern. The most common anthocyanidin aglycones are cyanidin, petunidin, peonidin, pelargonidin, delphinidin and malvidin (Delgado-Vargas *et al.*, 2000; Wallace and Giusti, 2015). Basic chemical structures of these molecules are depicted in figure 6. Anthocyanidins are pigments, responsible for most colors in flowers, leaves, fruits, seeds and other tissues, ranging from orange to red, pink, purple and blue, while the color is on the one hand strongly influenced by the hydroxyl (-OH) and methoxyl (-OCH$_3$) groups at ring B, the first ones accounting for more reddish and the last ones for more bluish colors. On the other hand, the color is influenced by the pH of the vacuole, where the water soluble pigments are stored. The color mainly serves to attract pollinators and seed dispersers. Thus, anthocyanidins are widely distributed in plants, also in such plants, that are consumed in the human diet, mainly crops, beans, vegetables, fruits and berries. (Tanaka *et al.*, 2008; Lee *et al.*, 2009; Cheng *et al.*, 2014; Taiz *et al.*, 2014).
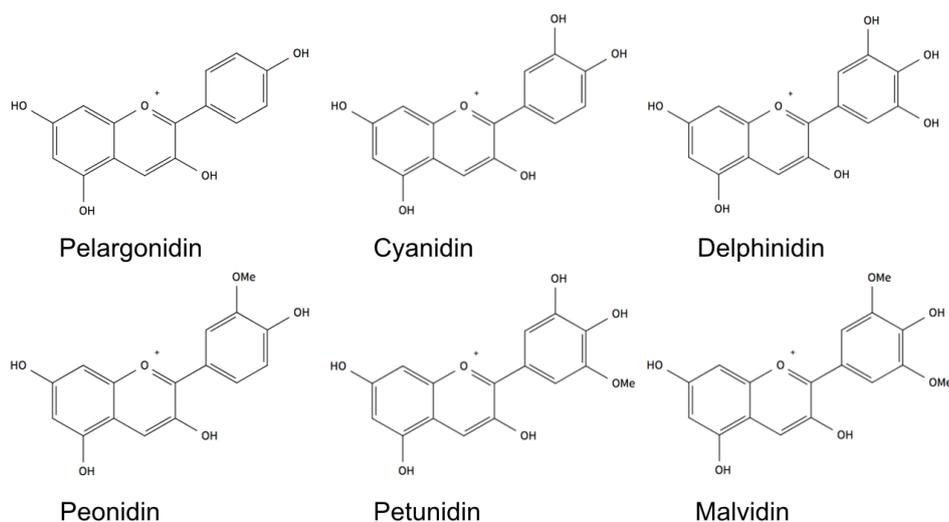
Figure 6 – ***Basic anthocyanidin structures.*** *The basic structures for the main anthocyanidins are depicted, which are pelargonidin, cyanidin, delphinidin, peonidin, petunidin and malvidin. Structures are based on PubChem.*

Further functions of anthocyanidins are the protection of plants against biotic and environmental stress factors, like pathogenic attack and UV radiation (Tanaka *et al.*, 2008). The compounds are especially synthesized under stress conditions, including infections, wounding and UV irradiation. Due to the high antioxidant capacity, they protect the plant against oxidative damage, caused by the stressors. Moreover, they exhibit antibacterial characteristics (Gould *et al.*, 2002) and various health benefits for humans, including antioxidant (Halliwell, 1996; Wang *et al.*, 1997; Pojer *et al.*, 2013), antiinflammatory (Wang and Mazza, 2002; de Pascual-Teresa, 2014) and anticancer (Bomser *et al.*, 1996; Hou, 2003) activities, accounting also for a decreasing risks for coronary heart disease by arterial protection (Colantuoni *et al.*, 1991; Wang *et al.*, 2012), platelet aggregation inhibition (Morazzoni and Magistretti, 1990) and endothelial protection (Youdim *et al.*, 2002), as well as memory and neurological improvement (Andres-Lacueva *et al.*, 2005; Krikorian *et al.*, 2010; Kent *et al.*, 2017), skin protection and therefore anti-aging effects (Soto *et al.*, 2015). The antioxidant activity is believed to protect membranes, proteins, and DNA from free radicals, enhanced by higher intake of anthocyanidins, for example through fruits and vegetables (Halliwell, 1996). Therefore, there exist a broad interest in anthocyanidin rich foods, also to produce supplements for preventative and therapeutic issues (Lee *et al.*, 2009). In addition, the use as natural food colorants is of huge interest (Tanaka *et al.*, 2008; Farr *et al.*, 2018).

In aqueous solutions four anthocyanidin structural types can exist (figure 7), which are pH dependent. These are respectively the flavylium cation (accounting for a red color), the quinonoidal anhydrobase (accounting for a blue color), the colorless carbinol pseudobase (or hemiketal) and the pale yellow chalcone (Liao *et al.*, 1992; Bakker *et al.*, 1997; Heredia *et al.*, 1998; Tanaka *et al.*, 2008). At mild or neutral pH anthocyanidins are predominant in their colorless hemiketal structure, due to reactions with water, while at acidic pH, they are predominant in their cationic flavylium structure, accounting for a positive charge in the molecule (Brouillard and Lang, 1990; Henry, 1996; Heredia *et al.*, 1998; Zhang *et al.*, 2014).
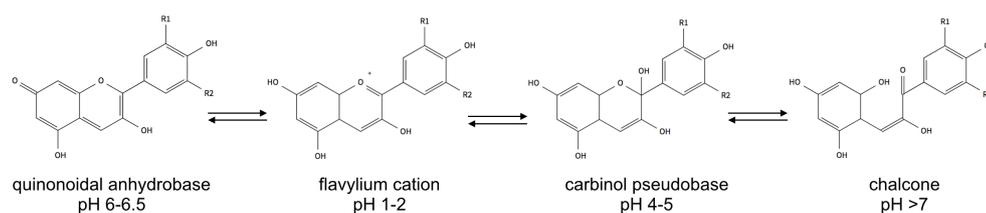
Figure 7 – *Anthocyanidin equilibria in aqueous solution. The four basic anthocyanidin structures are present in equilibrium in aqueous solution under very weakly acidic conditions. The structures account for the quinonoidal anhydrobase, the flavylium cation, the colorless carbinol pseudobase and the pale yellow chalcone. R1/R2: OH or OCH₃, independent of each other. Structures are based on Zhang* et al. *(2014).*

Anthocyanins are very unstable and reactive compounds, which easily degrade for example due to light, oxygen, ascorbic acid and enzymatic actions and which are able to react with other compounds, stabilizing their structure, by copigmentation, self-association, polymerization, and metal complexation (Farr *et al.*, 2018). For example, they react easily with other phenolic compounds, like flavanols, building a very complex group of polymeric pigments. Furthermore, anthocyanidins also react easily with small organic molecules in aqueous surroundings, building another group of monomeric and also polymeric pigments, called pyranoanthocyanidins (Rentzsch *et al.*, 2007b; de Freitas and Mateus, 2011). The condensed compounds deriving from reaction with flavanols, as well as the pyranoanthocyanidins, are of special interest in this work and will be described in more detail in the following sections.

1.6.1. Condensed flavanol and anthocyanidin pigments

Apart from proanthocyanidins, condensed pigments are known for anthocyanidins and flavanols. They were first discovered in wine and thought to build irreversibly during the processing and storage of food and drinks, visible in a color change (Somers, 1971; Bakker *et al.*, 1993; Bakker and Timberlake, 1997; Fulcrand *et al.*, 1998; Remy *et al.*, 2000; Mateus *et al.*, 2002a,b). It was assumed, that the unstable anthocyanidins transform during the wine making process into more stable molecules by specific reactions, yielding oligomeric compounds of anthocyanidins and flavanols (Somers, 1971).

Despite this, these compounds were also found as minor compounds in a few plant samples, for example from red onion, strawberries, black currant seeds, beans, black soybeans, purple corn, fig, pomegranate juice and grape marc, thus making occurrence in nature likely (Lu *et al.*, 2000; Asenstorfer *et al.*, 2001; Lu *et al.*, 2002; Fossen and Andersen, 2003; Andersen *et al.*, 2004; Fossen *et al.*, 2004; McDougall *et al.*, 2005; González-Paramás *et al.*, 2006; Macz-Pop *et al.*, 2006a; Dueñas *et al.*, 2008; González-Manzano *et al.*, 2008; Lee *et al.*, 2009; Sentandreu *et al.*, 2010, 2012; Takahama *et al.*, 2013b). In wine, the condensation is proposed to happen directly, or acetaldehyde mediated (Bakker and Timberlake, 1997; Fulcrand *et al.*, 1998; Asenstorfer *et al.*, 2001). In addition, copigmentation with flavanols or proanthocyanidins is discussed, to happen prior to condensation. Copigments have only little or no color themselves but enhance the color of other pigments. Colorless polyphenols like benzoic and cinnamic acid derivatives, flavanols, tannins, flavones and flavonols can act as such copigments.

The flavanols can interact through their planar electronically unsaturated phenolic moiety with the planar electron-rich flavylium nucleus of the anthocyanidin, resulting in noncovalent adducts, which then could develop covalent links between pigment and copigment in the later process. In the stacking process the copigment removes some water molecules from the solvent shell of the pigment and protects in this way the flavylium pigment from the nucleophilic attack of water, thus, preventing the formation of the colorless hemiketal or chalcone form of the anthocyanidin. This process is called covalent hydration. The color stabilization seems to be the main reason for copigmentation events, because no associations of copigments and colorless anthocyanidins are known and the hydration reactions are very dominant without copigments, resulting in more colorless anthocyanidin species. Also acetaldehyde could be involved in the noncovalent dimerization process, a compound considered in the beginning to be essential for the condensation process with flavanols. (Liao *et al.*, 1992; Brouillard and Dangles, 1994; Escribano-Bailón *et al.*, 1996).

Acetaldehyde mediated condensation results in condensed anthocyanidin–flavanol (A^+–F) molecule, either C–C linked (Somers, 1971) or ethyl-linked (Timberlake and Bridle, 1976; Bakker *et al.*, 1993, figure 8), both formed by a vinyl bond, linking an anthocyanidin and a flavanol or proanthocyanidin dimer (Asenstorfer *et al.*, 2001; Mateus *et al.*, 2002a,b). The condensation reaction takes place between the anthocyanidin flavylium cation and a flavanol moiety, possessing a vinyl residue. This vinyl residue could either derive from the cleavage of ethyl-linked flavanol oligomers, where the condensation had been acetaldehyde mediated, or from direct reactions with acetaldehyde and the C-8 position of the flavanol.

Interflavan-liked malvidin—catechin   Ethyl-linked malvidin—catechin
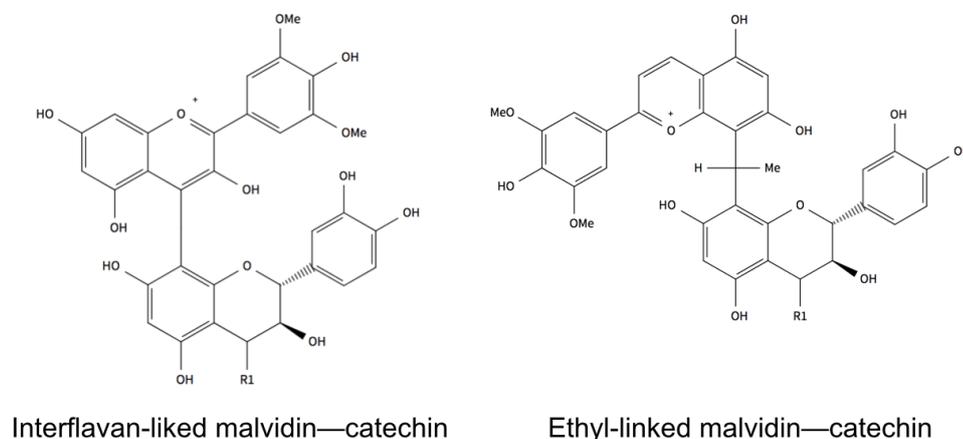
Figure 8 – *Basic structure of malvidin–catechin. The basic structures of a condensed malvidin–catechin molecule is depicted. At the left, the molecule as proposed by (Somers, 1971) with an interflavan-linkage (C–C), at the right, the molecule as proposed by (Timberlake and Bridle, 1976) with an ethyl-linkage. R1: H or flavanol. Structures are based on Remy* et al. *(2000).*

Aldehydes can easily be protonated and then react with the nucleophilic phloroglucinol ring (ring A) of flavanols in aqueous acidic solutions, yielding unstable ethanol adducts, that rearrange, until the addition to the flavylium cation. Protonation in acid surroundings and following dehydration before reaction with the anthocyanidin, builds the ethyl-linkage, resulting in a reddish-blue color of the pigments (Santos-Buelga *et al.*, 1995; Francia-Aricha *et al.*, 1997). If dehydration happens prior to protonation, a vinyl flavanol is build, which

acts as a nucleophile in the reaction with anthocyanidin, forming a C–C bond in less acidic conditions. For both linkages the last step involves cycloaddition (oxidative addition), where the vinyl-flavanol residue binds to the flavylium moiety. (Escribano-Bailón *et al.*, 1996; Francia-Aricha *et al.*, 1997; Es-Safi *et al.*, 1999; Asenstorfer *et al.*, 2001; Mateus *et al.*, 2002a,b; Sánchez-Ilárduya *et al.*, 2012).

In model solution, Dallas *et al.* (1996) could show for the first time, that condensation of anthocyanidins and flavanols also happens without acetaldehyde, but at a slower rate (Eglinton *et al.*, 2004). The direct condensation is proposed to involve nucleophilic addition, yielding either flavanol–anthocyanidin (F–A$^+$) or A$^+$–F molecules, depending on the linkage position of the anthocyanidin. For F–A$^+$ molecules the linkage is established through the C-6 or C-8 of the anthocyanidin and the C-4 of the flavanol. The anthocyanidin in its hydrated hemiketal form (AOH) acts as the nucleophile, while the flavanol unit acts as a positively charged electrophile carbonium ion (F$^+$) through its C-4 position. While the anthocyanidin hemiketal derives after nucleophilic attack of water at the positive charge of C-2 of the flavylium, the carbonium ion derives in acidic conditions by cleavage of a proanthocyanidin interflavanic bond. The condensation yields a colorless flavanol–anthocyanidin (F–AOH), which then dehydrates to the anthocyanidin flavylium form, yielding the F–A$^+$ molecule.

The A$^+$–F molecule is derived by a nucleophilic attack from the C-8 or C-6 of the flavanol at the C-4 position of the charged electrophilic anthocyanidin flavylium ion (A$^+$). The C-4 position of the anthocyanidin can cause a partial positive charge, making the reaction with nucleophiles possible and likely, whereas the flavanol hydroxyl groups at C-5 and C-7 give nucleophilic characteristics to the C-6 and C-8 of the molecule. The reaction can result in a colorless condensed flavene (A–F), that can be oxidized to the flavylium cation, yielding the A$^+$–F molecule, which can be involved in further condensation reactions, for example with other flavanols. The A–F reaction can also result in a colorless bicyclic condensation product, with an A-type linkage, being processed into a yellow xanthylium salt after dehydration. (Liao *et al.*, 1992; Santos-Buelga *et al.*, 1995; Escribano-Bailón *et al.*, 1996; Bakker and Timberlake, 1997; Remy *et al.*, 2000; Fossen *et al.*, 2004; Salas *et al.*, 2004; Dueñas *et al.*, 2006; Sánchez-Ilárduya *et al.*, 2012; Costa *et al.*, 2014). Due to the substitution at C-4 of the flavylium ring, the A$^+$–F compounds are expected to be less reactive than the F–A$^+$ molecules (Somers, 1971).

1.6.2. Pyranoanthocyanidins

Pyranoanthocyanidins are anthocyanidin-derived pigments, which were first discovered in red wine and investigated in model solution and were again thought to build during the processing and storage time under acidic conditions and by the presence of yeast fermentation products (Bakker *et al.*, 1997; Bakker and Timberlake, 1997; Francia-Aricha *et al.*, 1997; Fulcrand *et al.*, 1998; Asenstorfer *et al.*, 2001; Vivar-Quintana *et al.*, 2002; de Freitas and Mateus, 2011). Later, they were also discovered in slight amounts in plants and plant derived samples, like black currant seeds (Lu *et al.*, 2000), roses (Fukui *et al.*, 2002), onions (Fossen and Andersen, 2003), strawberries (Andersen *et al.*, 2004; González-Paramás *et al.*, 2006), figs (Dueñas *et al.*, 2008), beans (González-Paramás *et al.*, 2006; Yoshida *et al.*, 2019), grapes (Zhao *et al.*, 2010b;

Marquez *et al.*, 2012), raisins (Marquez *et al.*, 2013a) and carrot- (Schwarz *et al.*, 2004), orange- (Hillebrand *et al.*, 2004) and cherry- (Rentzsch *et al.*, 2007a) juice, speaking for a much wider spread in nature, than formally assumed. In wine they contribute to wine color, together with other anthocyanidin-derived compounds, while the pyranoanthocyanidins mostly cause hypsochromic shifts, with a maximum wavelength between 495 and 520 nm, resulting in orange hues. Some pyranoanthocyanidins cause bathochromical shifts, causing more bluish hues. (Rentzsch *et al.*, 2007b; de Freitas and Mateus, 2011; Marquez *et al.*, 2013b).

The pyranoanthocyanidin compounds mostly derive from a cycloaddition reaction of small molecules at the flavylium nucleus of the anthocyanidin, linking C-4 and the C-5 hydroxyl group and resulting in an additional pyrano ring D, with the positive charge delocalized over the pyranoanthocyanidin system. These small molecules can be for example pyruvic acid, acetaldehyde, acetoacetic acid, vinylphenols, hydroxycinnamic acids and vinylflavanols, resulting in diverse structures. (Bakker *et al.*, 1997; Bakker and Timberlake, 1997; Fulcrand *et al.*, 1998; Lu *et al.*, 2000; Asenstorfer *et al.*, 2001; Andersen *et al.*, 2004; González-Paramás *et al.*, 2006; Dueñas *et al.*, 2008; He *et al.*, 2010a; de Freitas and Mateus, 2011; Marquez *et al.*, 2013b). According to the respective condensating compounds they can be classified into hydroxyphenol-pyranoanthocyanidins, vitisins (including the carboxypyranoanthocyanidins), flavanol and pyranoanthocyanidin oligomers as well as so-called second generation pyranoanthocyanidins derived from carboxypyranoanthocyanidins like oxovitisins, portisins and pyranoanthocyanidin dimers (Rentzsch *et al.*, 2007b; Nave *et al.*, 2010; de Freitas and Mateus, 2011; Marquez *et al.*, 2013b).

Compared to anthocyanidins, pyranoanthocyanidins increase color stability and lifespan of the molecule, due to less reactivity by the substitutions at C-4, causing a conformation less susceptible to the nucleophilic attack of water (Bakker and Timberlake, 1997; Francia-Aricha *et al.*, 1997; Fulcrand *et al.*, 1998; Asenstorfer *et al.*, 2001). Apart from hydration, the molecule is also resistant towards SO_2 bleaching, pH changes, oxidative degradation and temperature (Bakker and Timberlake, 1997; Francia-Aricha *et al.*, 1997; Fulcrand *et al.*, 1998; Asenstorfer *et al.*, 2001; Vivar-Quintana *et al.*, 2002; Fossen and Andersen, 2003; Andersen *et al.*, 2004; Rentzsch *et al.*, 2007b; de Freitas and Mateus, 2011; Sánchez-Ilárduya *et al.*, 2012; Marquez *et al.*, 2013b).

In the following sections the main pyranoanthocyanidins as well as some special compounds of this class will be described. The basic structures of the most common pyranoanthocyanidins are shown in figure 9.

Hydroxyphenol-pyranoanthocyanidins

The hydroxyphenol-pyranoanthocyanidins include the first compound, where the pyranoanthocyanidin structure was assigned. In this precise case, the cycloaddition happens between an anthocyanidin and a vinylphenol, the latter deriving from decarboxylation of the cinnamic acids, like coumaric acid. The reaction involves the vinyl group of vinylphenol and the C-5 hydroxyl group and C-4 of the flavylium cation, followed by oxidation (Cameira dos Santos *et al.*, 1996; Fulcrand *et al.*, 1996; Vivar-Quintana *et al.*, 2002). Similar structures with different substitutions in the phenol moiety, involving similar mechanisms in the reaction

process, were found, too. Reactions with cinnamic acids (caffeic, coumaric, ferulic or sinapic acid) are abundant and it is proposed, that these cinnamic acid reactions can also happen directly, without decarboxylation (Marquez *et al.*, 2013b). This results for example in pinotin A (Schwarz *et al.*, 2003a), a pyranomalvidin-3-glucoside-catechol, from direct reaction with caffeic acid (Schwarz *et al.*, 2003c; de Freitas and Mateus, 2011). With regard to pinotin A, the hydroxyphenol-pyranoanthocyanidins can in addition be called pinotin-like compounds. (de Freitas and Mateus, 2011; Marquez *et al.*, 2013b).

Vitisins

The vitisins exhibit the simplest structure of pyranoanthocyanidins. They include the most abundant pyranoanthocyanidins in wine, which are the carboxypyranoanthocyanidins. In wine, they are built in a cycloaddition reaction between C-4 and the C-5 hydroxyl group of anthocyanidins and the enol of the yeast fermentation product pyruvic acid, followed by further dehydration and oxidation. If the anthocyanidin is a malvidin-3-glucoside, the compound is called vitisin A. (Bakker and Timberlake, 1997; Chiang *et al.*, 1992; Fulcrand *et al.*, 1998; de Freitas and Mateus, 2011; Marquez *et al.*, 2012). Vitisin A was the first discovered vitisin, which was characterized by Bakker *et al.* (1997) from red wine. This compound is 68 Da bigger than malvidin3-glucoside, which can be explained by an additional C_3O_2 included in the $C_3H_2O_2$ pyran unit linking C-4 and the C-5 hydroxyl group. After Bakker *et al.* (1997), this pyran ring exhibits a keto-function and a hydroxyethylene group (Bakker *et al.*, 1997; Fulcrand *et al.*, 1998). Due to the lack of this compound in fresh grapes, chemical formation from malvidin with a small molecule like an acid or an aldehyde was suggested by Bakker *et al.* (1997). The cycloaddition of pyruvic acid, suggested to derive from the glycolysis during yeast fermentation, was shown by Fulcrand *et al.* (1998), identifying a slightly different structure compared to Bakker *et al.* (1997), where the pyran ring exhibit a carboxyl group instead of the keto and the hydroxyethylene group, confirmed also by other authors (Mateus *et al.*, 2001; Schwarz *et al.*, 2003b).

There exist also pyranoanthocyanidins deriving from the cycloaddition of other carbonyl compounds from yeast fermentation like acetaldehyde and acetoacetic acid, the first one resulting in the vitisin B like compounds, which lack the carboxyl group at position C-10 of the pyrano ring D and which are the smallest known pyranoanthocyanidins. The compound is called vitisin B, if it is derived from malvidin-3-glucoside (Bakker and Timberlake, 1997). The second compound, acetoacetic acid, is very likely to produce methylpyranoanthocyanidins after cycloaddition (He *et al.*, 2006). Even, if this compound is not called a vitisin, it is very similar those, exhibiting a methyl group instead of the carboxyl group in vitisin A like compounds (Marquez *et al.*, 2013b). Vitisin A and its derivatives seem to appear in comparable high concentrations in red wine, thus they are probably the most well studied pyranoanthocyanidins (Asenstorfer *et al.*, 2001).

Oxovitisins

Oxovitisins are vitisins, which did react with water by a nucleophilic attack at the positively charged C-10 position at the D ring, thus transforming the molecule into a neutral pyranone-

anthocyanidin after further decarboxylation, oxidation and dehydration. These compounds are believed to build very slowly, thus, occurring not very abundant (He *et al.*, 2010b). Since, these compounds are not built from anthocyanidins but from carboxypyranoanthocyanidins, they are counted to the second generation pyranoanthocyanidins, as described later.

Flavanol and pyranoanthocyanidin oligomers

Pyranoanthocyanidins condensed with a flavanol were found for the first time in wine, whereas for a long time only pyranoantocyanidin–flavanol (pyA$^+$–F) molecules were found (Asenstorfer *et al.*, 2001; Mateus *et al.*, 2002a,b; He *et al.*, 2010a). In this molecule the anthocyanidin is linked with vinylflavanols (monomers or dimers) by acetaldehyde mediated linkage, resulting in vinylflavanol–pyranoanthocyanidins (Vivar-Quintana *et al.*, 2002). The vinylflavanol is provided either from dehydration of flavanol–ethanol adducts (deriving from the reaction of flavanols and acetaldehyde) or by cleavage of methylmethine-linked flavanol adducts. The reaction is proposed similar to the one for vinylphenols, but in this case the vinylflavanol is added by cycloaddition (Francia-Aricha *et al.*, 1997; Asenstorfer *et al.*, 2001; Mateus *et al.*, 2002b; Marquez *et al.*, 2013b).

Apart from this, one case of flavanol–pyranoanthocyanidins (F–pyA$^+$) is known so far, discovered and described by Nave *et al.* (2010), namely flavanol–vitisin A and flavanol–vitisin B. While the pyA$^+$–F molecules consists mostly of carboxypyranocyanins (Dueñas *et al.*, 2008), Nave *et al.* (2010) described in addition the first condensed pigments including a vitisin compound. For the reaction, instead of a condensation of vitisin and the flavanol, a reaction of a flavanol–anthocyanidin with a pyruvic acid molecule is postulated as more likely. Thus, in a first step the flavanol C-4 carbocation is probably derived by cleavage of a condensed tannin, followed by a C-6 and C-8 nucleophilic attack from the anthocyanidin. Afterwards cycloaddition with small molecules, like for example pyruvic acid in case of the flavanol–vitisin A, could happen, to yield the second pyran ring (Nave *et al.*, 2010).

Portisins

Carboxypyranoanthocyanidins seem to be the major compounds, responsible for so-called second generation pyranoanthocyanidins. These are compounds where anthocyanidins are not directly involved in the formation (Oliveira *et al.*, 2010). One group af the second generation pyranoanthocyanidins are the oxovitisins, as described above. Another group are the portisins, built acetaldehyde mediated from vinylpyranoanthocyanidins, including carboxypyranoanthocyanidin–vinylflavanols (Mateus *et al.*, 2003, 2005) and carboxypyrano-anthocyanidin–vinylphenols (Mateus *et al.*, 2006; Oliveira *et al.*, 2007). These compounds look similar to the vinylflavanol–pyranoanthocyanidins, but derive directly from the carboxypyra-noanthocyanidin, where a carboxypyranoanthocyanidin is linked through a vinyl bridge with a flavanol or phenol, by reaction with vinylphenolic compounds. Again, the flavanol moiety, derives from the decomposition of an ethyl-linked flavanol oligomer or from dehydration of a flavanol-ethanol adduct formed after the reaction of flavanols and acetaldehyde. For the phenol compounds the reaction mechanism is expected to include nucleophilic attack of a hydroxycinnamic acid at the C-10, a formic acid loss and a decarboxylation (Oliveira *et al.*,

2007).

Pyranoanthocyanidin dimers

Another type of second generation pyranoanthocyanidins are the pyranoanthocyanidin dimers. They consist of two pyranoanthocyanidin moieties, which are linked through a methine bridge, probably deriving from a reaction of methylpyranoanthocyanidin and carboxypyranoanthocyanidin, with a mechanism not completely understood. Deprotonation and nucleophilic attack or a formation of a charge-transfer complex are discussed. Those dimers were also synthezised in model solution, linked by a methyl-methine bridge, deriving from a reaction of carboxypyranomalvidin-3-glucoside with ethylpyranomalvidin-3-glucoside (Oliveira *et al.*, 2011). The dimers are responsible for a rare turquoise blue color in acidic solution (Oliveira *et al.*, 2010).
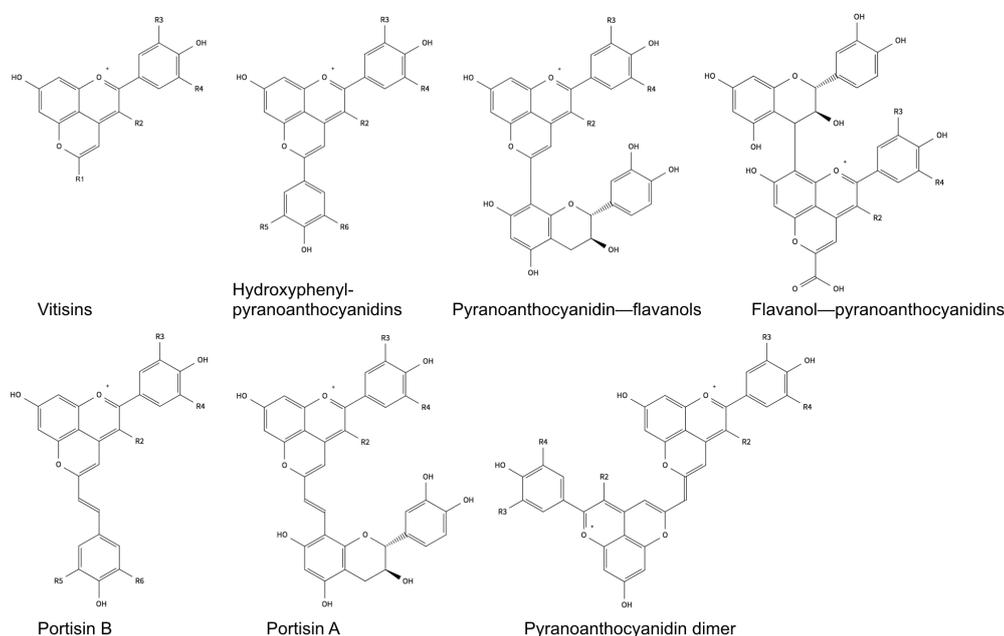
Figure 9 – ***Basic pyranoanthocyanidin structures.*** *The basic structures for the main pyranoanthocyanidins are depicted, which are vitisins, hydroxyphenyl–pyranoanthocyanidins, pyranoanthocyanidin–flavanols, flavanol– pyranoanthocyanidins, portisin A and B and pyranoanthocyanidin dimers. R1: COOH (vitisin A), H (vitisin B) or CH_3 (methylpyranoanthocyanidin); R2: OH or O-Glucose; R3/R4: H, OH or OCH_3 (independent of each other); R5/R6: H, OH or OCH_3 (independent of each other). Structures are based on (Mateus* et al.*, 2002b; Nave* et al.*, 2010; Morata* et al.*, 2019).*

Other pyranoanthocyanidins

Besides the mentioned ones, it seems as if many different types of pyranoanthocyanidins can exist. Two different pyranoanthocyanidins, which are reported only once each in literature are castavinol and rosacyanin B. Rosacyanin B, discovered by (Fukui *et al.*, 2002) in rose petals, is a monomeric pyranoanthocyanidin with a different and more complex structure, derived by a condensation from cyanidin and gallic acid, exhibiting no sugar moiety and with differing characteristics, for example less pH stability. Castavinol was discovered by Castagnino and Vercauteren (1996) in wine and is a diacetyl group added to different anthocyanidins. Further unique molecules are the catechinopyranocyanidins A and B from azuki beans, where the

flavanol and the anthocyanidin are connected by a pyrano ring at C-6 and C-7 of the catechin (Yoshida *et al.*, 2019). Or the pyruvic acid–catechin–malvidin-3-glucoside, where the pyruvic acid substitutes at C-6 and C-8 of the flavanol (Nave *et al.*, 2010).

1.7. Glycosylation of flavonoids

Glycosylation of flavonoids is the major mechanism, by which the structures of flavonoids are diversified (Gachon *et al.*, 2005; Falcone Ferreyra *et al.*, 2012). Glycosylation influence compounds solubility, stability and toxic potential and therefore, change their biological activity, their metabolism and in this way also their bioavailability. The water solubility of the molecule is enhanced by the glycosyl group, compared to the more hydrophobic aglycone, thus, the solubility in aqueous surroundings is enhanced, which is thought to be important for the molecules transport and the storage in the final destination, either vacuole or cell wall. By protecting the reactive nucleophilic groups of the molecule, the glycosyl group stabilizes the molecule. This is especially important for the reactive anthocyanidins, where the glycosylation at the 3-OH position stabilizes the aromatic ring. Because aglycones in general are reactive nucleophilic molecules, they have high damage potential, which makes them often toxic. The attachment of carbohydrates reduce this reactivity, thus, toxicity. Therefore, glycosylation can lead to inactivation or detoxification of these toxic compounds, which is for example used, to store harmful compounds for herbivores and release them through deglycosylation, if necessary. Furthermore, glycosylation regulates the compounds compartmentalization, thus, their intracellular location. Probably, the glycosyl group acts as a marker for accumulation and storage of the compounds in the vacuole or cell wall. Because the glycosyl group reduces the bioactivity compared to the aglycone, which is more involved in biological processes, glycosylation could be an important factor for storage of compounds in inactive cellular compartments, like the vacuole. (Jones and Vogt, 2001; Lim and Bowles, 2004; Gachon *et al.*, 2005; Zhao *et al.*, 2011; Lim and Bowles, 2004; Ishihara *et al.*, 2016; Le Roy *et al.*, 2016).

Deglycosylation for defense purposes is strongly suggested (Morant *et al.*, 2008). Despite, there exist also evidence, that the glycosides themselves are important in the defense process, while it is not always clear if this is due to direct involvement or due to enabling the accumulation of the phenylpropanoids, which are required for plant defense and are in this storage form available at any given time (Armah *et al.*, 1999; Chong *et al.*, 2002; Schulze *et al.*, 2005; König *et al.*, 2014).

The complex involvment of glycosylation becomes clear for the stress response of plants. While glycosides seems to be involved in for example UV-response as well as response to drought and salt stress, probably acting as ROS scavengers (Fini *et al.*, 2012; Sun *et al.*, 2013; Babst *et al.*, 2014; Hectors *et al.*, 2014; Majer *et al.*, 2014; Ahrazem *et al.*, 2015; Peng *et al.*, 2017), it was observed at the same time, that less glycosylation can increase stress tolerance, which makes sense, considering glycosylation to reduce the molecules bioactivity (Kim *et al.*, 2010). Therefore, the role of glycosylation and declycosylation seems to be very complex and dependent on the precise case, making the investigation of the specific function of glycosylation more difficult.

Apart from the already mentioned purposes, the glycosylation status influences the color of

flowers, leaves, seeds and fruits, thus, glycosylation has an indirect influence on pollinator attraction and seed dispersion (Gachon *et al.*, 2005; Griesser *et al.*, 2008; Tanaka *et al.*, 2008). In addition, glycosylation of flavonoids influences fruit flavor and aroma, maybe also in order to defend pathogens (Frydman *et al.*, 2013; Tikunov *et al.*, 2013).

The glycosylation status is mainly regulated by two types of enzymes, the glycosyltransferases (GTs) and the glycoside hydrolases (GHs). The former catalyzes the transfer of sugar moieties from an activated donor molecule to specific acceptor molecules, which can be sugars, lipids, proteins or molecules like phenylpropanoids, thus, flavonoids. (Yonekura-Sakakibara and Hanada, 2011; Le Roy *et al.*, 2016). Glycosylated flavonoids derive by glycosyl linkage through the hydroxyl groups of the aglycone, while glycosylation at C-3 is most common, followed by C-5 or C-7. Flavonoids are mostly *O*-glycosylated, especially flavones, flavonols and anthocyanidins, while for flavones also C-glycosides are common. Flavanols are the only flavonoids, which appear only scarce as glycosides. (Hollman and Arts, 2000; Cheng *et al.*, 2014; Le Roy *et al.*, 2016). According to Hollman and Arts (2000) in plants more than 80 different flavonoid glycosides exist as mono-, di-, tri- or even tetra-glycosylated compounds. The most common sugar moieties are glucose, galactose, rhamnose, xylose, arabinose, fructose or also glucuronic acid, whereas in plants glucose is the most predominant one. Those sugars in D-conformation produce β-glycosides, sugars in L-conformation produce α-glycosides (Hollman and Arts, 2000; Tanaka *et al.*, 2008; Noguchi *et al.*, 2009; Cheng *et al.*, 2014; Saigo *et al.*, 2020). The respective GTs are termed according to the transferred sugar, for example as glucosyltransferases (glucose) or rhamnosyltransferases (rhamnose; Noguchi *et al.*, 2009). GHs regulate the glycosylation status by hydrolyzing and/or rearranging glycosidic bonds. The turnover of flavonoid glycosides due to the reversible glycosylation of these molecules, implemented by both enzymes, is proposed to maintain flavonoid and in general also phenylpropanoid homeostasis as well as to have an impact on plant growth and development, caused by the influence on synthesis, localization and biological activity (Yonekura-Sakakibara and Hanada, 2011; Roepke and Bozzo, 2015; Le Roy *et al.*, 2016).

1.7.1. UDP-glycosyltransferases

GTs and GHs can be subdivided into different families, according to sequence similarities, which are documented in the CarbohydrateActive enZymes (CAZy) database (Lombard *et al.*, 2014) with 111 GT and 167 GH families (June 2020). Concerning GTs, those involved in the glycosylation of secondary metabolites are accounted to the GT1 family. Most of them utilize uridine diphosphate (UDP)-glucose as the activated sugar nucleotide. Therefore, they are termed UDP-glycosyltransferases, abbreviated UGTs (Yonekura-Sakakibara and Hanada, 2011). With respect to flavonoids, they are sometimes called UDP-sugar flavonoid GTs (UFGTs), but in this work, we will remain with the term UGT (Jaakola, 2013; Meng *et al.*, 2019). While the most frequent UDP-sugar donor in plants is UDP-glucose, galactose, xylose, rhamnose, arabinose, or glucuronic acid appear, too (Ross *et al.*, 2001; Osmani *et al.*, 2008; Le Roy *et al.*, 2016). Typically, UGTs perform *O*-glycosylation, but there are also cases of C-glycosylation from UDP-glucose as sugar donor, where the sugar is directly attached to an aromatic carbon atom (Hollman and Arts, 2000; Brazier-Hicks *et al.*, 2009; Wang, 2009).

UGTs are responsible for the above-mentioned influences of glycosylation on secondary metabolites, like stabilization, enhancement of water solubility, deactivation or detoxification of toxic compounds, pathogen defense and biotic stress protection, regulation of glycoside homeostasis, storage and transport (Yonekura-Sakakibara and Hanada, 2011).

Due to the lack of a signal sequence, UGTs of plants are mainly localized in the cytoplasm, where they glycosylate flavonoids and other phenylpropanoids as well as terpenoids, steroids and plant hormones (Bowles *et al.*, 2006). The enzymes exhibit a conserved 44 amino acid sequence, which is according to the PROSITE database (Sigrist *et al.*, 2013, accession number PS00375): [FW]-[2X]-[QL]-[2X]-[LIVMYA]-[LIMV]-[4-6X]-[LVGAC]- [LVFYAHM]-[LIVMF]-[STAGCM]-[HNQ]-[STAGC]-G-[2X]-[STAG]-[3X]-[STAGL]-[LIVMFA]-[4-5X]-[PQR]-[LIVMTA]-[3X]-[PA]-[2-3X]-[DES]-[QEHNR]. The letters in brackets denote the alternative amino acids at this position and X denotes any amino acid. This consensus sequence seems to be responsible, for binding of the UDP moiety of the sugar nucleotide. In plants this consensus motif is termed the plant secondary product glycosyltransferase (PSPG) box, located in the C-terminal region of the proteins (Mackenzie *et al.*, 1997; Offen *et al.*, 2006; Yonekura-Sakakibara and Hanada, 2011). In their tertiary structure, UGTs exhibit two distinct folds, which are termed GT-A and GT-B fold. In plants, only the GT-B fold is present, consisting of two $\beta/\alpha/\beta$ Rossmann-like domains. This means, that each domain contains a central α helix, flanked by β sheets on both sides. The domains are located N- and C-terminal in the protein and are involved in binding of the sugar acceptor and sugar donor. (Wang, 2009; Yonekura-Sakakibara and Hanada, 2011). Therefore, the domains form a cleft as substrate binding site. In this cleft, the mostly invariant nucleotide sugar donor binds at the C-terminal domain with the conserved PSPG-box. The acceptor binds mainly at the N-terminal domain, thus, this domain is less conserved, accounting for a broad range of acceptor molecules. Varying substrate specificity is achieved by several highly variable loops in the N-terminal domain, making diversity in shape and size of the substrate pocket possible (Offen *et al.*, 2006; Osmani *et al.*, 2008; Wang, 2009; Yonekura-Sakakibara and Hanada, 2011). The active site cleft of most UGTs exhibit a conserved histidine, acting as a catalytic residue for enzyme activity. It mediates the deprotonation of the acceptor, which then acts as a nucleophile and attacks the C-1 carbon of the UDP-sugar. Therefore, the UDP moiety is displaced and a β-glycosidic linkage is built. A nearby located, conserved aspartate residue forms a hydrogen bond with the histidine. It probably stabilizes the histidine and balances the charge after deprotonation of the acceptor (Offen *et al.*, 2006; Wang, 2009).

1.7.2. GH1-type glycosyltransferases

Apart from the UDP-dependent glycosylation, Matsuba *et al.* (2010) discovered a new glycosylation reaction in the petals of carnation (*Dianthus caryophyllus*) and delphinium (*Delphinium grandiflorum*), involving the aromatic acyl-glucose 1-*O*-β-D-vanillyl-glucose as novel sugar donor. Due to the glycosylation of anthocyanidin acceptors, these enzymes were termed acyl-glucose-dependent anthocyanin glucosyltransferases (AAGTs). Later, AAGT glycosylation was also found in Canterbury bells (*Campanula medium;* Miyahara *et al.*, 2014) as well as in a monocot, which was African lily (*Agapanthus africanus;* Miyahara *et al.*,

2012). In all cases, a second sugar is transferred on an anthocyanin 3-*O*-glucoside, either at position C-5 (AA5GT) or C-7 (AA7GT), by *O*-glycosidic linkage, resulting in anthocyanin 3,5-diglucosides or anthocyanin 3,7-diglucosides.

Miyahara *et al.* (2013) discovered the first AAGT in *Arabidopsis thalina* (Arabidopsis). In this case, the AAGT catalyzes the glucose transfer to the 4-coumarate moiety of the major anthocyanidin compound cyanidin 3-*O*-[2"-*O*-(2"'-*O*-(sinapoyl) xylosyl) 6"-*O*-(*p*-coumaroyl) glucoside] 5-*O*-[6""-*O*-(malonyl) glucoside] from Arabidopsis, shortly called A11 (Tohge *et al.*, 2005; Miyahara *et al.*, 2013). Thus, for the first time, the new type of glycosyltransferase was shown to glycosylate not directly the anthocyanidin skeleton, but adding a sugar to an acyl moiety on the highly modified anthocyanin.

Ishihara *et al.* (2016) found a second GT of this new type in Arabidopsis, BGLU6. In contrast to earlier studies, this enzyme catalyzes the glucose transfer to a flavonol and not to an anthocyanin. More precisely, it transfers a second glucose residue to an existing glucose moiety at the C-3 position by a $(1 \rightarrow 6)$ glycosidic linkage. This results in a 3-*O*-glucosyl-glucoside 7-*O*-rhamnoside, shortly a 3-*O*-gentiobioside 7-*O*-rhamnoside.

All of these new discovered GTs have in common, that their amino acid sequences exhibit high similarity towards glycoside hydrolase family 1 (GH1) proteins, which typically act as β-glucosidases (BGLUs). According to Luang *et al.* (2013), the term glycosyltransferase is only reserved to enzymes of GT families from CAZy. To be most accurate, the new enzymes should be termed transglucosidases (TGs). The activity of TGs is defined as the transfer of a sugar from a donor other than a nucleotide phosphate or phospholipid phosphate to an acceptor, to form a new glycosidic linkage, thus including all new discovered GTs. Nevertheless, for simplification and in concordance with the literature, the general term GH1-type GT will be used for all enzymes, performing glycosyltransferase activities with homology to GH1 proteins. The GH1-type GTs could have been derived from β-glycosidases early in the angiosperm lineage (Xu *et al.*, 2004; Matsuba *et al.*, 2010; Miyahara *et al.*, 2012). In Arabidopsis there exist 47 annotated *BGLU* genes, termed *BGLU1* through *BGLU47*. Phylogenetic analyses of the amino acid sequences propose a common evolutionary origin of these *BGLUs* (Xu *et al.*, 2004). Those huge *BGLU* gene families are reported also for other plant organisms, like 40 *BGLUs* in the monocot rice (*Oryza sativa*), with probably 34 of them functional (Opassiri *et al.*, 2006), 26 *BGLUs* in maize (*Zea mays;* Gómez-Anduro *et al.*, 2011) and 64 *BGLUs* in rapeseed (*Brassica rapa;* Dong *et al.*, 2019). The purpose of such big gene families could be the facilitation of differential substrate specificity of β-glycosidases, due to many different β-glycosides with different aglycones in plants (Xu *et al.*, 2004).

The respective *BGLU* gene families in each organism are proposed to have emerged by gene duplication or duplication of chromosomal segments. Most likely, they developed after the divergence of monocots and dicots, supported by the fact, that in rice no myrosinases are present, compared to Arabidopsis (Opassiri *et al.*, 2006; Miyahara *et al.*, 2011). The high number of rape *BGLU* genes probably derived due to the genome triplication event in the investigated lineage (Dong *et al.*, 2019). The relatively low number for maize and rice could be due to the lack of glucosinolate metabolism in corn plants, which accounts for pathogen defense in Arabidopsis (Gómez-Anduro *et al.*, 2011).

Phylogenetic analyses on amino acid level from Xu *et al.* (2004) revealed 10 separated clusters for the 47 BGLUs in Arabidopsis. The clusters with the respective BGLUs are shown in table 1. While the function of many of these BGLUs remain still unknown, some of them were experimentally characterized concerning their function. This revealed on the one hand, that the BGLUs in Arabidopsis exhibit diverse functions and substrate specificities. On the other hand it revealed high probability, that the BGLUs in each separated cluster exhibit the same functions by also using the same substrates.

Table 1 – ***Phylogenetic clusters of BGLUs from Arabidopsis.*** *The 10 clusters after phylogenetic analyses of BGLUs from Arabidopsis on amino acid level, with the respective included BGLUs are listed. Data are based on Xu* et al. *(2004).*

Cluster	BGLUs
1	BGLU1 - BGLU11
2	BGLU12 - BGLU17
3	BGLU18 - BGLU25
4	BGLU26, BGLU27
5	BGLU28 - BGLU32
6	BGLU33
7	BGLU34 - BGLU39
8	BGLU40 - BGLU42
9	BGLU43 - BGLU44
10	BGLU45 - BGLU47

According to this, cluster 1 is the most differentiating cluster and should contain GH1-type GTs, including the already characterized BGLU6 and BGLU10. In cluster 2 BGLU15 and BGLU16 were characterized to be β-glucosidases, while the first one is known to act on flavonols (Roepke and Bozzo, 2015; Roepke *et al.*, 2017). Cluster 3 contains a characterized ABA glucoside hydrolase (BGLU18; Lee *et al.*, 2006; Han *et al.*, 2020) and characterized scopolin hydrolases (BGLU21 - BGLU23; Ahn *et al.*, 2010; Yamada *et al.*, 2011). In addition, all BGLUs from cluster 3 contain an ER retention signal at their C-terminal amino acid sequence and the characterized BGLUs were proven to be localized there (Nakazaki *et al.*, 2019). BGLU20 was recently investigated to be necessary for normal pollen development, while the precise mechanism is not known (Dong *et al.*, 2019). Cluster 4 contains an atypical β-thioglucoside glucohydrolase (BGLU26; Bednarek *et al.*, 2009). Thioglucoside glucohydrolase are also called myrosinases. This atypical myrosinase exhibit two conserved glutamic acid residues (Glu), compared to the key glutamine residue (Gln) in typical myrosinases. Such Glu residues are found in BGLU18 - BGLU33, thus it could be, that all of them exhibit similar atypical myrosinase activity, but more specialized, than typical myrosinases (Sugiyama

and Hirai, 2019). In cluster 5 BGLU28 and BGLU30 seem to be myrosinases, acting on glucosinolate substrates (Morikawa-Ichinose *et al.*, 2020; Zhang *et al.*, 2020). BGLU33 is the only BGLU in cluster 6, with a similar function as BGLU18, thus, hydrolyzing ABA glucosides, but being located in the vacuole (Xu *et al.*, 2012). In cluster 7 BGLU34, BGLU35, BGLU37 and BGLU38 were characterized as typical myrosinases, acting on glucosinolate substrates (Barth and Jander, 2006; Islam *et al.*, 2009; Wittstock and Burow, 2010; Zhou *et al.*, 2012), while the other BGLUs in this cluster are proposed to exhibit the same function (Sugiyama and Hirai, 2019). In cluster 8 BGLU42 is not completely characterized, but very likely to hydrolyze phenolic compounds under iron deficiency (Zamioudis *et al.*, 2014). For cluster 9 BGLU44 seems to be a β-mannosidase (Xu *et al.*, 2004), while there exist also predictions for a myrosinase (Tsai *et al.*, 2017). In cluster 10 BGLU45 and BGLU46 were characterized as monolignol glucoside hydrolases (Escamilla-Treviño *et al.*, 2006; Chapelle *et al.*, 2012). The functions of the respective BGLUs in the clusters are supported from phylogenetic analysis, including BGLUs from other species with known function, as performed by Xu *et al.* (2004). In these analyses cluster 1 includes no BGLUs with other known functions. Phylogenetic analysis from Miyahara *et al.* (2011) on amino acid level, including again all 47 BGLUs from Arabidopsis and in addition the AAGTs from carnation and delphinium, revealed again a cluster of BGLU1 to BGLU11, containing the AAGTs. In addition, Miyahara *et al.* (2014) showed clustering of all known AAGTs, including BGLU10. This strongly supports the assumption, that BGLU1 to BGLU11 from Arabidopsis exhibit GH1-type GT activity. A closer look in the cluster, performed by Ishihara *et al.* (2016), revealed a separating cluster around BGLU6, including BGLU1 to BGLU5. It is assumed, that these enzymes could exhibit similar activity as BGLU6, thus probably glycosylating flavonols or other substrates compared to anthocyanidin substrates. To investigate this hypothesis in more detail, in the following work the genes *BGLU1* to *BGLU5* should be characterized concerning their proposed function.

1.8. Functional characterization of unknown genes

To assign a function to a gene, the biochemical activity of the encoded protein must be known, which includes knowledge about the substrates, the enzyme can access, as well as knowledge about the conditions, tissues, cell types and cellular compartments, where the substrate and enzyme are in contact for the biochemical activity. In addition, the role of the substrates and products in the plants are important to know, and which factors regulate the activity of the enzyme and its substrates and products. This complex situation needs the investigation by different combined biochemical, genetic and analytical chemical approaches.

To investigate the function towards (de-)glycosidation, the classical approach is the purification of the encoded enzyme and to investigate the ability to hydrolase or glycosylate certain substrates. However, this approach is already limited due to the fact, that most laboratories' access to glucosides is limited. Therefore, not all possible substrate specificities can be examined, probably missing one important factor. For example, is 4NPGlc often used in enzyme assays for GHs, but for example sorghum (*Sorghum bicolor*) dhurrinase-1 is specific for its substrate dhurrin and is not hydrolyzing 4NPGlc, as well as the scopolin β-glucosidases BGLU21, BGLU22, and BGLU23 (Hösel *et al.*, 1987; Ahn *et al.*, 2010). In addition, this

approach is only straightforward, if the possible substrates are already known, but the access to natural substrates in the plant is difficult to predict. Furthermore, already the purification of the protein can be challenging, if multiple isoenzymes are present. With this respect, successful purification of enzymes should be validated by liquid chromatography-tandem mass spectrometry (LC-MS/MS) of tryptic protein digests (Hua *et al.*, 2015). All these difficulties indicate, that the functional characterization of a gene is not always possible by enzyme purification and following activity assays. To circumvent these problems, data can be gathered from a combination of further experimental approaches, as reviewed by Ketudat Cairns *et al.* (2015) and introduced in the following sections.

The analysis of gene expression, to determine for example the expressing tissue and also expression influencing stress conditions, can provide some first hints towards the function. Gene expression data can be achieved by public microarray and expressed sequence tag (EST) data, as well as public available RNA-Seq data. To perform whole transcriptome sequencing (RNA-Seq) or microdissection experiments is even more accurate, but also expensive. Further approaches to investigate specific expression are for example the fusion of the putative promotor elements with a β-glucuronidase gene, expressed *in planta* and investigated by providing the β-glucuronidase substrate 5-bromo-4-chloro-3-indoxyl-β-D-glucopyranoside (XGluc). The staining gives information about the gene expressing tissue and also in some cases about the expressing cell type. In addition, specific tissues can be investigated by semi-quantitative reverse transcriptase (RT-)PCR or quantitative real-time (qRT-)PCR, where the transcript is reverse-transcribed into cDNA and specific primers amplify the transcript on cDNA levels. (Jefferson *et al.*, 1987; Marone *et al.*, 2001; Wagner, 2013; Berardini *et al.*, 2015; Klepikova *et al.*, 2016; Garrido-Gil *et al.*, 2017).

Apart from the specific expression of the investigated gene, analysis of transcriptome co-expression networks enhance the functional information about a gene. In addition, with sequence similarities on DNA and amino acid level, the hypothetic function can be encircled (Xu *et al.*, 2004; Wu *et al.*, 2018).

To further investigate or proof a certain enzymatic activity cloning of the corresponding cDNA and recombinant expression of the protein, mostly in *Escherichia coli* and *Pichia pastoris*, is often performed (Xu *et al.*, 2004; Opassiri *et al.*, 2006). Even if this approach helps to circumvent the difficult purification with often low protein abundances from plants, it can also lead to difficulties for functional characterization. For example, the *in vitro* assay does not modulate the environmental factors in the plant and buffer components can have activating or inhibiting effects (Luang *et al.*, 2013). In addition, often recombinant protein expression does not work, for example if the organisms are not able to express certain proteins (Ishihara *et al.*, 2016). Furthermore, the knowledge about specific substrates could still lack.

An important further information for functional characterization is the cellular compartment in which the enzyme is located. For β-glucosidases, this is possible with the use of XGluc (Mazzuca *et al.*, 2006), which is cleaved by the enzymes and therefore, shows their location by blue staining. Apart from this, for other enzymes the use of antibodies or fluorescent protein fusions (Swain *et al.*, 1992; Matsuba *et al.*, 2010) give information about the compartment localization in the cell. The simultaneous use of the native promotor and a fluorescent fusion

protein has the advantage, to potentially identify cellular and subcellular localization. By generation of fluorescent protein fusions care must be taken, to not block any signal sequences. To maintain N-terminal signal sequences, the fluorescent protein can be tagged to the C-terminus, but in this case, care must be taken not to block C-terminal localization signals, which are often more difficult to recognize. The subcellular localization can help to understand the function of the protein, while at the same time providing information about substrate accessibility. If subcellular localization is not possible, there exist also some programs that can predict signal peptides for localization.(Matsuba *et al.*, 2010; Miyahara *et al.*, 2012; Nishizaki *et al.*, 2013; Chen *et al.*, 2019).

A very powerful tool for functional characterization is a genetic approach. Loss-of-function mutants can be derived by transposon or T-DNA insertions, by chemically or physically induced mutations, by CRISPR/Cas9 and other genome editing methods as well as by a knockdown of gene expression with RNAi (Miki *et al.*, 2005; Jiang *et al.*, 2012; Xing *et al.*, 2014). These mutants can be investigated for a phenotype. Nevertheless, not each mutant shows a clear phenotype. In addition, concerning for example GH1 enzymes, a functional overlap could prevent a differential phenotype or the phenotype could change only at certain environmental conditions (Palmer and Rodriguez de Cianzio, 1985; Ishihara *et al.*, 2016; Wu *et al.*, 2018). Liquid chromatography mass spectrometry (LC-MS) based metabolite profiling, can be useful, to investigate the chemical phenotype on metabolite level. Metabolite levels can sometimes be differential, even if the physical phenotype is not obvious (Chapelle *et al.*, 2012; Wu *et al.*, 2018).

Even more reliability can be achieved, with an additional overexpression or complementation mutant, by searching for differential putative substrate or product levels in the metabolite profil. Due to possible ectopic overexpression or complementation, if the promotor is located outside the normal cells or cellular location, attention must be paid for unusual effects during evaluation of the results (Wang *et al.*, 2011; Kiran *et al.*, 2012). The use of the second mutant can also simplify the search of candidate compounds, if a functional overlap is present. A further possibility to circumvent functional overlap can be the generation of a multiple loss of function mutant for several homologous genes. Nevertheless, for big gene families it could be challenging, to find the redundant genes for the loss-of-function (Iuchi *et al.*, 2007; Kuromori *et al.*, 2009; Yang *et al.*, 2018).

While a genetic approach of two different mutant lines in combination with LC-MS metabolite profiling provides a powerful tool, to identify the potential substrate and/or product compounds of the encoded protein, the identification of the compound can remain challenging, especially for unknown compounds, which are not present in public databases and where no analytical standard is available (Robbins *et al.*, 2009). One of the most accurate identification methods is the structure elucidation with nuclear magnetic resonance (NMR) spectroscopy. The basic principle of this method is based on the nuclear spin from the nuclei of atoms with an uneven mass number and an uneven charge (like ^{13}C). If these nuclei are placed in a strong magnetic field, as done with NMR, the energy levels between the various spin states are splitted. Transitions between adjacent energy levels are connected to the absorption or emission of a photon in the radiofrequency (rf) range, which is measured in NMR. Based on the structural

environment, the photon emission or absorption changes, according to the respective shielding of the surrounding electrons. The NMR frequency is normally reported as a chemical shift (δ) relative to an external standard compound, since the accurate absolute frequencies are difficult to determine. With this method, the spatial composition of atoms in a compound can be determined very precisely. Nevertheless, this approach needs intensive sample purification and huge amounts of the compound, which can be challenging for minor compounds. (MacKenzie and Smith, 2002).

LC-MS based methods, often coupled to NMR, are frequently applied for the identification of new flavonoid compounds (Bakker and Timberlake, 1997; Fulcrand *et al.*, 1998; Stobiecki *et al.*, 2006; Marquez *et al.*, 2013a; Wu *et al.*, 2018). Due to the relevance for this work LC-MS based metabolite profiling will be introduced in the next section in more detail.

1.9. LC-MS based metabolite profiling

Liquid chromatography mass spectrometry (LC-MS) is a technique, that combines the separation and analysis of different compounds in a complex mixture with liquid chromatography (LC) coupled to mass spectrometry (MS). While LC separates the compounds in a sample, MS is used to analyze the samples due to the mass to charge ratio of charged analytes.

Liquid chromatography

Liquid chromatography (LC) is used, to separate non-volatile and thermally labile compounds in a mixture, based on their interactions with a stationary and a mobile phase. The two main chromatographic mechanisms are either normal phase chromatography or reversed phase chromatography. In the first case, the separation is based on the polarity of the analytes, using a polar stationary phase and a non-polar mobile phase. Depending on its polarity, the analyte interacts with the polar stationary phase, while increased polarity cause stronger interaction, thus the elution time is enhanced. In reverse phase chromatography a non-polar stationary phase and an aqueous, moderately polar mobile phase is used. The separation is determined by hydrophobic interactions of the analyte and the stationary phase, with increasing retention time by increasing hydrophobicity of the analyte. The retention time is determined as the timepoint, where the eluent leaves the end of the column. Reversed phase chromatography is used in 95% of LC applications, commonly utilizing columns with a stationary phase based on silica, which has been chemically modified with hydrophobic octadecyl, called C18 column. (Blasco and Picó, 2007; Malviya *et al.*, 2010; Žuvela *et al.*, 2019). The most frequently used liquid chromatography technique is High-Performance Liquid Chromatography, also called High Pressure Liquid Chromatography (HPLC). HPLC consists of the column, that holds the stationary phase, a pump, that moves the mobile phase through the column and a detector, that detects the retention time of the analyte. The analyte mixture is introduced in small volumes to the stream of mobile phase and the elution is often modified by a gradient elution, varying the mobile phase composition by adding gradually more organic solvent into the mobile phase (methanol or acetonitrile), making it less polar. The gradient separates the analyte mixtures due to the affinity of the analyte for the current mobile phase. (Malviya *et al.*, 2010).

A parameters, which determines the performance of HPLC is the internal column diameter, decisive for the loaded sample amount on the column. In addition, low diameters represent higher sensitivity and less solvent consumption. The particle size of the stationary phase, which consists of small silica beads, determines the surface area, while a small size provide more surface area, thus, better separation. At the same time, smaller particles require higher pressure to yield a consistent and reproducible flow rate. Therefore, the pump is an important component of HPLC. Ultra-HPLC (uHPLC) has improved to work at much higher pressures, being able to work with a smaller particle size (less than 2 µm), in this way enhancing the chromatographic resolution (Malviya *et al.*, 2010; Lambert *et al.*, 2015).

The detector of HPLC can be an UV detector, detecting compounds according to their absorbed amount of light, if passing a beam. The output is recorded as a series of peaks, while each peak represents a compound, which passes through the detector and absorbs UV light. The area under the peak is proportional to the amount of substance. While preparative HPLC is used for isolation and purification of a compound, analytical HPLC is used to obtain information about the compound. Depending on the stationary phase, the mobile phase and chemical properties, an analyte has a characteristic retention time. If the peak is clearly separated from the mixture and exhibits a sufficient detection level, the retention time can be used for identification of known compounds. (Malviya *et al.*, 2010). Even more accurate information yield the use of LC combined with MS (Blasco and Picó, 2007).

Mass spectrometry

To analyze compounds in a mixture according to their mass, LC and MS are combined in one measurement. While liquid chromatography (including HPLC) separates the analytes, they are introduced after the elution from the column into the mass spectrometer (MS). MS creates charged ions out of the analytes and detect those. The data can be analyzed due to the mass to charge ratio (m/z) of the analyte, providing information on molecular weight, structure, identity and quantity of the analyte. In addition, the detection is not dependent on UV absorbance. The main challenge in the development of LC-MS analysis, was the transition from LC to MS. While LC produces high pressure and a high gas load for the separation of the analytes, MS requires a vacuum and a limited gas load. In addition, MS requires an elevated temperature, compared to LC with almost ambient temperatures. Also, the mass range is unlimited for LC and limited for a mass analyzer. This challenges were overcome by the use of an interface for the ionization process. (Saibaba *et al.*, 2016). An overview of the basic principle of LC-MS is given in figure 10.

The most common interfaces are atmospheric pressure ionization (API) or electro-spray ionization (ESI). Both are soft ionization techniques, assuring no fragmentation of the analyte upon ionization. In case of API, the liquid, containing the neutral analyte, is vaporized in a heated gas stream, bringing the analyte into the gas phase. In the next step the analyte in the gas phase is ionized. In ESI, the analyte is ionized in the liquid phase inside the electrically charged droplets. These ionized analytes are then transferred into the gas phase. (Wang *et al.*, 2016b). Due to the relevance for this work, a further insight will be given into ESI-MS.

The basic components of ESI-MS are the ion source, the mass analyzer and the detector

(figure 11). The ions are produced in the ionization chamber with the ion source and then transferred into the mass analyzer. This transfer is maintained by several ion optics, which basically focus the ion stream and therefore, maintain a stable trajectory of the ions. In the mass analyzer the ions are sorted and separated, according to their m/z value. The detector system detects and measures the amount of the separated ions, displaying the results in the so-called mass spectrum. The ion optics, the mass analyzer and the detector are kept at high vacuum, to reduce the reactivity and enhance the lifespan of the ions. The ion source itself is kept at atmospheric pressure. From source to the detector a continues pressure gradient is applied, functioning as a pump, to guide the ions from the source through the analyzer to the detector. (Banerjee and Mazumdar, 2012).
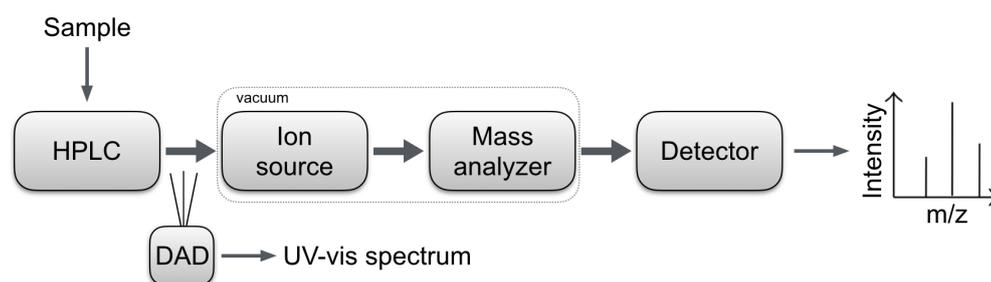
Figure 10 – ***Basic principle of HPLC-DAD-ESI-MS.*** *The basic principle for HPLC-DAD-ESI-MS measurements is depicted. The main setup consists of high performance liquid chromatographie (HPLC) coupled to a diode array detector (DAD), the electro-spray ionization (ESI) source, the mass analyzer and the detector. The detector generates the mass spectrum, while DAD generates the UV-vis spectrum. The figure is based on Banerjee and Mazumdar (2012).*

For the generation of ions at the ion source, the analyte solution is injected by a mechanical syringe pump through a stainless-steel capillary (needle), with a low flow rate. At the tip of the needle a high voltage is applied, called capillary voltage. Whether this voltage is kept in the positive or negative potential determines the ESI mode. In the positive ion mode, the ionization occurs by protonation, in the negative ion mode by deprotonation of the analyte. Due to this strong electric field (called endplate offset), the sample solution is dispersed into an aerosol of highly charged electrospray droplets. A heated drying gas is applied in the same direction, around the capillary, enhancing the nebulization and at the same time guiding the spray from the capillary tip towards the mass spectrometer. In addition, the drying gas evaporates the solvent, reducing the size of the droplets. In the end, the analytes are released from the droplets and pass through a sampling cone into a heated capillary. The heated capillary causes the complete desolvation of the ions, which are led into the mass analyzer. (Kruve *et al.*, 2010; Banerjee and Mazumdar, 2012).

In the mass analyzer, the ions are separated according to their m/z and guided to the detector. The two basic analyzers are the quadrupole analyzer and the time-of-flight (TOF) analyzer. The quadrupole analyzer is composed of four parallel electrical rods, arranged in the corners of a square. A direct current (DC) potential is applied at two of these rods, while the other two are linked by an alternating radio-frequency (rf) potential. The respective applied electric field guides the ions through the detector. For example, a positively charged ion will move towards the negatively charged rod. If the polarity changes, the ion will switch the movement

and undergoes in this way complex oscillation, yielding a trajectory for certain ions towards the detector. In this way, analytes of certain m/z values can be selected. By continuously varying the applied voltage, a scan for various ions is possible. If the rf-only mode is applied, a quadrupole will function as an ion focusing device, meaning, that ions will be guided to other components of the apparatus, whithout selection. (El-Aneed *et al.*, 2009).
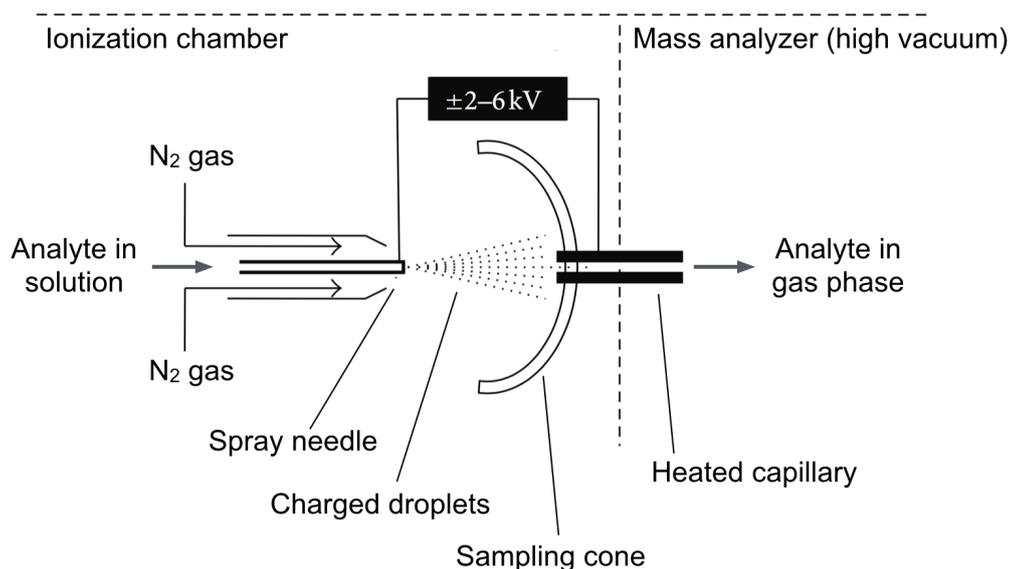
Figure 11 – ***Schematic overview of the ion source.*** *The ionization chamber includes the ion source. The analyte solution is injected into the spray needle and charged by the capillary voltage at the end of the needle. The resulting nebulization yields charged droplets, which are evaporated by the drying gas and guided into the heated capillary towards the mass analyzer. The figure is based on Banerjee and Mazumdar (2012).*

The TOF analyzer separates ions due to their mass, based on the free flight of the ionized molecules in a tube of 1-2 m in length, before reaching the detector. In this analyzer, all ions will reach the detector, thus, no selection of the m/z appears. The resolution is improved on the one hand by a reflectron, which is an electrostatic ion mirror. This is an ion optic device, changing the path of the ions within the TOF. Ions with higher kinetic energies penetrate deeper into the ion mirror, repelling ions gradually, thus improving the resolution, by avoiding misinterpretation at the detector due to different kinetic energies of the ions. In addition, the tube length improves the resolution, because the ions have a longer flight. By the use of the reflectron, the time-of-flight is also enhanced, thus the resolution again improved. The TOF is an important factor for the application of tandem mass spectrometry (MS/MS), if combined with a quadrupole analyzer, which is an important improvement in terms of analyte identification, thus, for LC-MS based metabolite profiling. (El-Aneed *et al.*, 2009).

The application of MS/MS, implies the connection of multiple mass analyzers in a series, including a collision cell. The achieved fragmentation of the analyte in the collision cell, is like a specific fingerprint, allowing valuable information with respect to the molecular structure of the analyte. In such a series, the first analyzer performs ion isolation of a specific m/z (which is the precursor ion). The collision cell is in the following applied for fragmentation of the precursor ion. The final analyzer then separates the produced fragment ions based on their m/z values, which are detected at the detector. A frequently used series is called quadrupole

time-of-flight (QTOF). (El-Aneed *et al.*, 2009).

For QTOF-MS/MS the ions are driven from the heated capillary into the first quadrupole (Q1). This is the ion filter station, where ions of a certain m/z are selected due to the respectively applied voltage. From Q1 the ions are passed on into the second quadrupole (Q2) in rf-only mode, which performs as the collision cell. Inhere, the ions are bombarded by neutral N_2 gas molecules, leading to the fragmentation of the molecules, named collision induced dissociation (CID). Leaving the second quadrupole the product ions of CID and if present unfragmented precursor ions enter the TOF analyzer, where the mass separation of the fragments takes place. Alternatively, Q1 and Q2 can both operate in the rf-only mode, making simple MS analysis with the same instrument possible. For this MS mode a low ion energy as well as a low collision energy (less than 10 eV) must be applied, to prevent collision but at the same time move the ions (Chernushevich *et al.*, 2001; El-Aneed *et al.*, 2009; Allen and McWhinney, 2019). An overview of the elements for QTOF-MS/MS are given in figure 12.
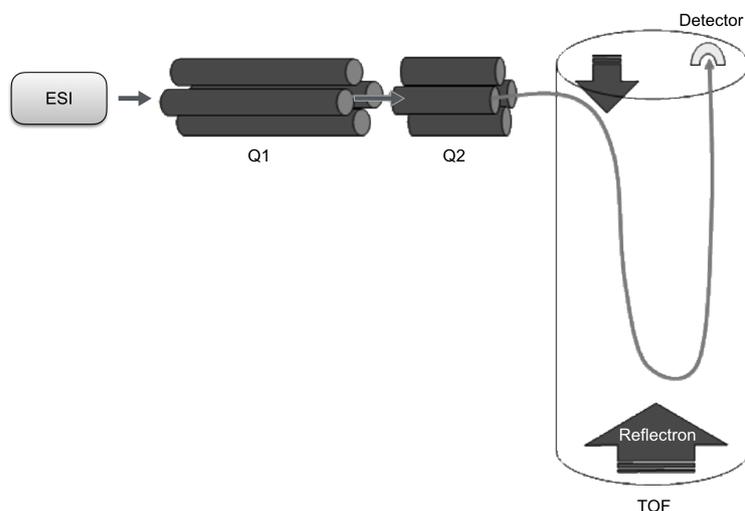
Figure 12 – *Schematic overview of a quadrupole time-of-flight mass analyzer. The ions from the ESI interface are guided towards the first quadrupole (Q1) and further directed into the second quadrupole (Q2), which can act as collision cell. Mass separation according to the m/z is achieved in the time-of-flight (TOF) mass analyzer, including the reflectron and the detector for the gegreation of the mass spectrum. The figure is based on El-Aneed* et al. *(2009).*

The detector is responsible for the output of the measurement. The simplest ion detectors perform a neutralization of the ion, while measuring the resulting current. Since, this is only useful, if the ion flux is large, a more applied detection is the use of an electron multiplier, also applicable for low ion fluxes. The energetic ions strike a so-called conversion dynode (a metal or semiconductor plate) at high voltages, which produces secondary electrons. Those secondary electrons are accelerated and focused onto the second dynode, then to a third dynode and so on, increasing the number of emitted electrons. The dynodes are progressively set to a voltage closer to earth. At the end of the multiplier the output current is converted to a voltage signal, which can be translated to an intensity value, visualized as a mass spectrum of peaks. Until today, it is not completely understood, how the peak height and the analyte are corresponding. But it is assumed, that the peak height in a mass spectrum is directly proportional to the number of corresponding ions, arriving at the detector. The tallest peak

in the spectrum is called the base peak. Based on the isotope differences, the charge of the molecule can be abbreviated, called mathematical charge deconvolution, to estimate the molecular mass of the analyte. Most ESI-MS instruments are equipped with elaborate software for charge deconvolution. Since, m/z is dimensionless, often the biological mass Dalton (Da) is used for the final molecular mass unit. (Banerjee and Mazumdar, 2012).

Since, UV absorption characteristics can provide further insights into the nature and structure of some analytes, LC-MS based methods can be supplemented with a diode array detector (DAD). The DAD monitors the UV-vis absorbance of analytes after LC separation, providing data on selected ultraviolet (UV) and visible (vis) wavelengths (Saibaba *et al.*, 2016).

1.10. Motivation and aim of this study

Flavonoids fulfill diverse functions in plants ranging from protection against biotic and abiotic stress, over attraction of pollinators and seed dispersers to modulation of chemical plant communication. The functional diversity of flavonoids is to some extend achieved by structural diversity. Glycosylation of flavonoids, as the most abundant substitution mechanism for structure diversification, is therefore involved in the maintenance of the various functions of flavonoids, including also beneficial characteristics for human health. Even, if glycosylation of flavonoids has been extensively studied, many questions remain on the precise function of glycosides in the organism, as well as on their biosynthesis. Regarding the biochemical complexity of plants, it is assumed, that many compounds, including flavonoid glycosides, are still not discovered yet. This implies also undiscovered biochemical regulations as well as the often unknown underlying genetic regulations.

The discover of atypical GH1-type glycosyltransferases in several plant organisms, including Arabidopsis, support this assumption. To understand more about the function of these enzymes as well as their regulation and unknown glycosylation reaction, more detailed characterization of these enzymes are necessary. Despite, knowledge about further GH1-type GTs could enhance the understanding.

Phylogenetic analysis revealed apart from the already known GH1-type GTs, BGLU6 and BGLU10, further candidate GH1-type GTs in Arabidopsis. Those include BGLU1 to BGLU5, which cluster with BGLU6 and are suggested, to putatively glycosylate flavonoid substrates other then anthocyanidins, like flavonols in case of BGLU6.

This study focuses on the functional investigation of BGLU1 to BGLU5 in Arabidopsis. The aim is to investigate, if the enzymes function as flavonoid GH1-type GTs and to determine the respective flavonoid substrate, if this function could be confirmed. The functional characterization will be based on a genetic approach as proposed by Miyahara *et al.* (2011), using a T-DNA insertion loss-of-function mutant and an overexpression mutant for each respective *BGLU* gene, to investigate the differential phenotype compared to the wild type. Since, the mutant phenotype, especially for decorating enzymes, is often not distinguishable, metabolic fingerprinting with untargeted UHPLC-DAD-ESI[+]-QTOF-MS/MS will be applied, to investigate the chemical phenotype on differential metabolite levels. Based on the differential metabolite levels, potential substrate and product compounds of the encoded enzymes should be received. Identification of the compounds will be mainly based on the MS/MS fragmentation

pattern. The results should give insights into the function for *BGLU1* to *BGLU5* with respect to the enzymatic reaction of the encoded enzymes as well as the respective utilized substrates of the enzymes.

2. Material and Methods

All solutions and media were prepared with $H_2O_{Milli-Q}$ from Millipore Milli-Q® (Biocel). All chemicals were used with a high degree of purity. All primers were ordered from Invitrogen™ (Paisley, UK) or Metabiom (Martinsried, Germany). All primer sequences are given in the supplements in table S1. Figures of all applied vectors and plasmids are given in the supplement in figures S8 - S11.

2.1. Plant material

Seeds of the Brassicaceae *Arabidopsis thaliana* (Arabidopsis) Col-0 wild type were used from the own laboratory reproduction, originally obtained from the European Arabidopsis Stock Center (NASC, http://www.arabidopsis.info). All mutant plants applied in this study, derived from the laboratory seed collection. Apart from *bglu5-1*, with a Ws-4 background, all other T-DNA insertion lines exhibit Arabidopsis Col-0 background. They were selected with the SIGnAL T-DNA insertion database (The Salk Institute Genomic Analysis Laboratory, http://signal.salk.edu/cgi-bin/tdnaexpress). T-DNA insertion lines *bglu1-1* (GABI_341B12, RSS1100, N432664, ERS4255859) and *bglu3-2* (GABI_853H01, RSS1097, N481877, ERS4255860), for *At1g45191* (*BGLU1*) and *At4g22100* (*BGLU3*) respectively, were obtained from the own laboratory GABI-Kat collection, *bglu4-2* (SALK_029729, RSS1023, N25045, ERS4255861) for *At1g60090* (*BGLU4*) from NASC and *bglu5-1* (FLAG_526E07, RSS1018) for *At1g60260* (*BGLU5*) from the National Institute for Agricultural Research (INRA) Versailles. Homozygous lines were chosen, based on polymerase chain reaction (PCR)-based genotyping (Saiki *et al.*, 1985), also verified by MinION long-read sequencing (Oxford Nanopore Technologies (ONT)). Overexpression mutants, expressing the gene of interest driven by the double enhancer cauliflower mosaic virus 35S (2x35S) promoter (*2x35S:BGLU1* [RSS1131 #6; LB, BASTA^R, 2xpro35S, *att*B1-*AtBGLU1_CDS_without_stop*-*att*B2-cassette, RB], *2x35S:BGLU3* [RSS1132 #44; LB, BASTA^R, 2xpro35S, *att*B1-*AtBGLU3_CDS_without_stop*-*att*B2-cassette, RB], *2x35S:BGLU4* [RSS1133 #91; LB, BASTA^R, 2xpro35S, *att*B1-*AtBGLU4_CDS_without_stop*-*att*B2-cassette, RB] and *2x35S:BGLU5* [RSS1135 #125; LB, BASTA^R, 2xpro35S, *att*B1-*AtBGLU5_CDS_without_stop*-*att*B2-cassette, RB]), were taken from the laboratory seed collection, derived from the master thesis Frommann (2016). Sequences of the overexpression T-DNA of *2x35S:BGLU1*, *2x35S:BGLU3* and *2x35S:BGLU4* are given in the supplements (see appendix A.1 - A.3).

2.2. Growth conditions

If not mentioned differentially all plants were grown in the greenhouse on soil at long day conditions with 14 h of light at 23 °C and 69.6 % humidity. Plants for HPLC and RT/qRT-PCR seed samples for *BGLU3* and *BGLU4* were grown two month at short day conditions

with 8 h of light at 22 °C and 55 % humidity, before being transferred into long day conditions until most siliques were ripe but still closed. For final seed ripening plants were transferred to 14 h of light at 23.8 °C and 55 % humidity, with less watering timepoints, before harvesting. To prevent batch effects (Nygaard *et al.*, 2016) of different genotypes and at the same time crossing between plants of different genotypes, the plants for seed samples were grown in random order at short day conditions and sorted by genotype in the same tray with ongoing regularly changing positions at long day conditions. For HPLC and RT/qRT-PCR samples of *BGLU1*, seeds were surface-sterilized with ethanol and sown on ½ Murashige and Skoog (MS) agar plates (0.5 x MS medium [Sigma], 1.5 % (w/v) sucrose, 0.9 % (w/v) agar, pH 5.7 adjusted with KOH). They were kept for two days at 4 °C in the dark before being transferred to a phytochamber with 16 h of light at 21 °C and 18.5 °C at night for germination. Eight-day old seedlings were transferred to soil and then grown randomly ordered in a phytochamber with 13 h of light at 23 °C and 80 % humidity and 18 °C and 65 % humidity in the night for five to six weeks. *BGLU5* samples were grown as described for *BGLU1* but only used for RT-PCR.

2.3. Seed sterilization

Seed surface sterilization was performed with an ethanol and triton X-100 based method. All centrifugation steps were performed at full speed and room temperature in a table centrifuge. Seeds (up to 300 µl) were transferred to a sterile DNA-purification column without filters, placed on a 2 ml reaction tube to collect the flow through. On top of the seeds around 1 ml 70 % (v/v) ethanol with 0.05 % (v/v) triton X-100 was placed and the seeds were incubated for 10 min at room temperature on the thermoblock at approximately 1200 rpm. Columns were centrifuged, the flow through discarded and 1 ml 100 % ethanol was placed on top of the seeds, followed by again 5 min incubation at room temperature in the thermoblock (1200 rpm). Columns were again centrifuged and the flow through discarded. This last washing step was repeated and the column one more time centrifuged for 1 min to dry the column. Samples were air-dried with open lids for at least 20 min under the clean bench. Seeds were either sowed immediatly or stored with closed lids for up to two weeks at 4 °C.

2.4. Genomic DNA extraction from plant material

2.4.1. Rapid genomic DNA extraction

To investigate the insertion allele or wild type allele of putative heterozygous or homozygous T-DNA insertion single or double mutants, rapid genomic DNA extraction was performed after Edwards *et al.* (1991).

A leave of the size of a fingertip was harvested in a 2 ml screw-cap reaction tube (StarLab). 50 µl H_2O and a spatula tip of Zirkonia Beads (Roth; NO38.1; $\varnothing$ 1 mm) were added and the plant material homogenized in the Precellys® Evolution homogenizer (Bertin Instruments; 4 x 45 s at 9,000 rpm with 15 s pause inbetween). The samples were shortly centrifuged and 400 µl extraction buffer (250 mM NaCl, 200 mM Tris-HCl pH 7.5, 25 mM EDTA pH 8.0, 0.5 % (w/v) SDS) were added. After vortexing, the samples were incubated for 20 min at

65 °C while shaking. The samples were centrifuged at full speed for 5 min and 300 µl of the supernatant were provided with the same volume of isopropyl (alcohol). After incubation at room temperature for 2 min, followed by centrifugation at full speed, the supernatant was discarded and the sediment washed in 400 µl 70 % (v/v) ethanol. The dried sediment was dissolved in 50 to 100 µl of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) over night at 4 °C or by mechanical solvation and incubation for 10 min at 65 °C. The DNA was stored at 4 °C.

To clean up the rapidly extracted DNA, samples were precipitated with ethanol and a sodium acetate based protocol for PCR analysis. Samples were centrifuged for 5 min at full speed and the supernatant transferred into a new reaction tube, to remove suspended solids. $1/10$ volume 3 M sodium acetate pH 5.2 and a 2.5 fold volume of 100 % ethanol were added, the samples vortexed and then centrifuged for 15 min at maximum speed. The supernatant was removed and the sediment washed with 70 % (v/v) ethanol. The dried sediment was again dissolved in 50 to 100 µl of TE buffer as described above.

2.4.2. Genomic DNA extraction with CTAB

For MinION long-read sequencing (ONT) of *bglu1-1*, *bglu3-2* and *bglu4-2*, high molecular weight genomic DNA was extracted from young Arabidopsis rosette leaves using a CTAB-based method. The tissue was grounded by mortar and pestle in liquid nitrogen and six spatula spoons plant material were transferred into 400 µl CTAB buffer I (2 % CTAB, 100 mM Tris-HCl pH 7.9, 20 mM EDTA pH 8.0, 1.4 M sodium chloride, 0.5 % (w/v) PVP) including 4 µl of 1 M DTT. Samples were inverted and shaken for 15 min at 65 °C. After cooling down at room temperature, the same volume dichloromethane was added, the samples inverted and centrifuged. The upper phase was mixed with two volumes (800 µl) CTAB buffer II (1 % CTAB, 50 mM Tris-HCl pH 7.9, 10 mM EDTA pH 8.0, 0.25 % (w/v) PVP) inverted and incubated for 30 min at room temperature for precipitation of DNA. After centrifugation for 20 min the supernatant was completely removed, and the sediment completely resolved in 200 µl of 1 M sodium chloride by incubation at 37 °C. DNA precipitation was done by the same volume of isopropyl (alcohol). After incubation for 30 min and centrifugation for 20 min, the supernatant was completely removed, and the sediment washed with 70 % (v/v) ethanol. The sediment was air-dried, to remove any residual alcohol and then solved in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) including 10 % (v/v) RNase A (10 mg/ml, Sigma) to digest any residual RNA in the sample. Digestion took place for 5 min at 65 °C and then at 4 °C over night. A second DNA precipitation was done by adding one volume of 1 M sodium chloride to the sample, followed by the same volume of isopropyl (alcohol). Samples were inverted and incubated for 30 min at room temperature, centrifuged for 30 min and the supernatant completely removed. The sediment was washed in 70 % (v/v) ethanol, air-dried and resolved in water (LiChrosolv®, LC-MS Grade, Merck). This method was successful for genomic DNA isolation of *bglu3-2*, where the DNA was directly applied to sequencing (see 2.5.1). For *bglu1-1* the method was improved by adding 2 % 2-mercaptoethanol (2-ME) instead of DTT into the CTAB I buffer and expanding the precipitation steps to 45 min. For *bglu4-2* the method was improved after Siadjeu *et al.* (2020).

2.5. Loss-of-function T-DNA insertion line characterization

To investigate the function of *BGLU1*, *BGLU3*, *BGLU4* and *BGLU5* by a genetic approach including a T-DNA insertion and an overexpression line, suitable loss-of-function T-DNA insertion lines needed to be chosen. Therefore, homozygous lines were selected, based on PCR-based genotyping results, done in preliminary works. For PCR, the wild type allele primers RS1302 and RS1303 for *bglu1-1*, RS1319 and RS1320 for *bglu3-2*, RS1259 and RS1260 for *bglu4-2* and RS1251 and RS1230 for *bglu5-1*, as well as insertion allele primers RS1848 and RS1845 for *bglu1-1*, RS1846 and RS1847 for *bglu3-2*, RS1259 and RS629 for *bglu4-2* and RS1242 and RS1237 for *bglu5-1* were used. Homozygosity was in addition verified by MinION long-read sequencing (ONT), due to the sequence gap after alignment of the long-read sequences to the reference sequence TAIR10 (see 2.5.2).

Homozygous T-DNA insertion mutants *bglu1-1*, *bglu3-2* and *bglu4-2* were characterized concerning the exact T-DNA insertion position and the T-DNA border sequences by MinION long-read sequencing (ONT, see 2.5.1). In addition, the sequencing was used to proof the presence of only one T-DNA insertion in each line.

2.5.1. MinION long-read sequencing

For MinION long-read sequencing (ONT) of *bglu1-1*, *bglu3-2* and *bglu4-2* high molecular weight genomic DNA was extracted from young Arabidopsis rosette leaves using the CTAB-based method as described in 2.4.2. To improve the sequencing results, for *bglu1-1* and *bglu4-2* short DNA fragments were depleted with the Short Read Eliminator kit (SRE kit; Circulomics) prior to sequencing, according to the manufacturer's instructions. Preceding sequencing, quality control of the DNA was performed via NanoDrop™ 2000 measurement, electrophoresis in a 1% or 0.8% agarose gel (2.18) and Qubit fluorometric quantitation, the latter one by using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific). The library preparation followed the SQK-RAD004 (*bglu3-2*) or SQK-LSK109 (*bglu1-1*, *bglu4-2*) protocol (ONT). Sequencing was performed on R9.4.1 and R10 flow cells. Live base calling was performed on the GridION by Guppy (ONT).

2.5.2. Sequence analysis after long-read sequencing

Long-reads of each *bglu* line were subjected to Canu v1.8 (Koren *et al.*, 2017) for *de novo* genome assembly with previously described parameters (Pucker *et al.*, 2019). Contigs were polished as described in (Siadjeu *et al.*, 2020). BLASTn v2.8.1 (Altschul *et al.*, 1990) was applied to identify T-DNA insertions in the genome sequences of the three *bglu* lines, based on the bait sequences of pAC161 (GenBank: AJ537514.1) and the respective T-DNA sequence between LB and RB (see supplements appendix A.4) for *bglu1-1* and *bglu3-2* and pROK2 (http://signal.salk.edu/pBIN-pROK2.txt-new) and the respective T-DNA sequence between LB and RB (see supplements appendix A.5) for *bglu4-2*. Global alignments of the sequence of T-DNA loci and the corresponding sequence extracted from TAIR10 were generated via MAFFT v7.299b (Katoh and Standley, 2013). Based on the long-read sequencing results, primers were designed to amplify the T-DNA borders by PCR, using one primer in the genomic

DNA sequence and primers in the adjacent T-DNA sequence. Primers were RS1302 and F067 for the first border and RS1845 and RS1846 for the second border in *bglu1-1*, 3144 and RS1320 for the first and RS1847 and T35S for the second border in *bglu3-2* and RS1259 and 3144 for the first and 3144 and JFF001 for the second border in *bglu4-2*. PCR was performed with the Q5® High Fidelity DNA Polymerase (NEB). 25 µl reaction volume contained 1x Q5® Reaction Buffer (NEB), 200 µM dNTPs, 0.5 µM each primer, 0.02 $^{U}/_{µl}$ Q5® High-Fidelity DNA Polymerase (NEB) and 1 µl (first primer combination of *bglu1-1*) or 2 µl (for all other reactions) genomic DNA from the ONT sequencing sample as template. The cycling protocol 30 s, 98 °C—10 s, 98 °C; 30 s, 60 °C (second border *bglu1-1*, *bglu3-2*) or 61 °C (first border *bglu4-2*) or 64 °C (second border *bglu4-2*) or 70 °C (first border *bglu1-1*); 15 s (first border *bglu3-2*) or 55 s (first border *bglu1-1*) or 1 min (second border *bglu1-1*) or 2 min and 40 s (second border *bglu3-2*) or 2 min (*bglu4-2*), 72 °C—2 min, 72 °C with 35 cycles was performed. The results were analyzed by electrophoresis in a 1 % agarose gel (see 2.18). Due to multiple bands for the second border of *bglu3-2* and for both borders of *blgu4-2*, the respective bands of the borders were precipitated from the gel with the NucleoSpin™ Gel and PCR Clean-up Kit (Macherey-Nagel), according to the manufacturer's instructions, eluting the samples in 15 µl (*bglu3-2*), 20 µl (first border *bglu4-2*) or 25 µl (second border *bglu4-2*).

All PCR products were cleaned up with the Exo-CIP™ Rapid PCR Cleanup Kit (NEB), according to the manufacturer's instructions. 3 µl of each cleaned up sample were then submitted to Sanger sequencing at the Sequencing Core Facility (SCF) at the Center of Biotechnology (CeBiTec) to determine the exact orientation, position and nucleic acid sequence of both T-DNA borders. Primers for sequencing were the same primers as for the described PCRs. Differing from this, additional sequencing primers F068 and B097 for the second border of *bglu1-1* and B230 and RS1848 for the second border of *bglu3-2* were needed. Sequencing results were analyzed with Sequencher v5.4 (http://www.genecodes.com) and A plasmid Editor (ApE, https://jorgensen.biology.utah.edu/wayned/ape/).

2.6. Database derived expression prediction

HPLC, RT- and qRT-PCR samples derived from tissue with the respective investigated *BGLU* gene naturally expressed in the wild type. Expression predicition was performed with data from The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org, Nov 2015; Huala *et al.*, 2001) and the Arabidopsis eFP Browser (http://www.bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi, Nov 2015) as well as RNA-Seq expression data from the TraVa website (Klepikova *et al.*, 2016, http://travadb.org/browse/, 2018). For final experiments, samples from tissue predicted with the TraVa database, were chosen. For *BGLU1* and *BGLU5* at least 10 mm long rosette leaves, for *BGLU3* dry seeds and for *BGLU4* 24 h water-soaked seeds (see 2.19.1) were used. According to the RNA-Seq data, the chosen tissues should exhibit the highest expression of the respective gene in the wild type. For *BGLU2* almost no expression was predicted, therefore it was skipped in all further analyses.

2.7. RNA isolation

For semi-quantitative reverse transcriptase (RT-) and quantitative real time (qRT-)PCR, RNA isolation was performed with the Plant Total RNA Kit (protocol A; Spectrum™) according to the manufacturer's instructions, including the DNase I digestion step. Differing from the protocol, the initial incubation was performed at 65 °C for 20 min at 1,000 rpm, to improve the extraction. In addition, in case of seed samples, longer centrifugation steps were needed, if the columns were blocked.

RNA Isolation was done, using 100 mg mortared six-week-old rosette leaves for *BGLU1* and *BGLU5* and 70 – 90 mg mortared seeds for *BGLU3* and *BGLU4*, the latter 24 h soaked on a wet filter paper (see 2.19.1). In addition, for *BGLU3* soaked seed samples were prepared as described for *BGLU4*, too.

For RT-PCRs the RNA samples from seeds were precipitated with LiCl and sodium acetate pH 5.2 after Suzuki *et al.* (2004). Therefore, samples were mixed with 1/3 volume of 8 M LiCl and stored over night at -80 °C. The samples were transferred to a new reaction tube, centrifuged for 30 min at 4 °C and 12,000 x g and the supernatant was discarded. The sediment was dissolved in 50 µl RNase-free H_2O and 1/10 volume of 3 M sodium acetate pH 5.2 was added, along with 1 volume isopropyl (alcohol). After 10 min at room temperature the samples were centrifuged for 15 min at 4 °C and full speed. The supernatant was discarded and the sediment washed in 300 µl 70 % (v/v) ethanol. The ethanol was completely removed, samples air-dried and the sediment dissolved in 25 µl RNase-free H_2O, before storing at -80 °C.

2.8. cDNA synthesis

cDNA of all RNA samples (see 2.7) was generated, using the ProtoScript® II First Strand cDNA Synthesis Kit (New England BioLabs (NEB)) according to the manufacturer's instructions of the standard protocol, using 1 µg of RNA if possible, otherwise the maximum volume, and the $d(T)_{23}$ VN primer mix. Due to better reaction conditions the synthesis was performed in the mastercycler epigradient (Eppendorf). Successful cDNA synthesis was controlled by amplifying the constitutively expressed *ACTIN 2* (*At3g18780*) gene, with PCR, using a *Taq* DNA polymerase (Chien *et al.*, 1976) with ThermoPol™ Buffer (NEB) and the primers RS469 and RS470 in the same mastercycler. 25 µl volume contained 1x ThermoPol™ Buffer (NEB), 200 µM dNTPs, 0.2 µM each primer, 0.625 units/25 µl PCR *Taq* DNA polymerase (NEB) and 1 µl (*BGLU1*) or 2 µl (*BGLU3* and *BGLU4*) cDNA as template. The cycling protocol 30 s, 95°C—30 s, 95°C; 1 min, 58°C; 1 min, 68°C—5 min, 68°C with 25 cycles was performed. The results were analyzed by electrophoresis in a 1 % agarose gel (see 2.18).

2.9. Semi-quantitative reverse transcriptase PCR

To investigate the differential gene expression of *BGLU1*, *BGLU3*, *BGLU4* and *BGLU5* in all chosen lines for HPLC measurements, semi-quantitative reverse transcriptase (RT-)PCR was performed with samples from the HPLC sample batch, prepared together with those as described in 2.19.1. Sample growth, RNA isolation, cDNA synthesis and *ACTIN 2* control were performed as mentioned above (2.2, 2.7, 2.8). All primer pairs were designed intron and

T-DNA insertion spanning in the respective loss-of-function T-DNA insertion line, to test the presence of an intact transcript. For primer design see 2.11.

The RT-PCR was performed using a *Taq* DNA polymerase with ThermoPol® Buffer (NEB), as described in 2.8. As template 1 µl (*BGLU1*) or 2 µl (*BGLU3* and *BGLU4*) cDNA were applied. The gene specific primers were RS1302 and RS1390 (*BGLU1*), RS1319 and JFF010 (*BGLU3*), RS1260 and RS1261 (*BGLU4*) and RS1229 and RS1230 (*BGLU5*). The cycling protocol differed from the one in 2.8 by the annealing temperatures of 58 °C (*BGLU1* and *BGLU4*) or 60 °C (*BGLU3* and *BGLU5*), the extension times of 1 min (*BGLU4*) or 2 min (*BGLU1*, *BGLU3* and *BGLU5*) and by 35 cycles.

2.10. Quantitative real time PCR

For more sensitive investigation of the differential gene expression of *BGLU1*, *BGLU3* and *BGLU4* in all chosen lines for HPLC measurements, a quantitative real time (qRT-)PCR was performed with a new sample set, which was treated like the HPLC samples (see 2.19.1). Samples for *BGLU3* and *BGLU4* were prepared from the same seed batch of the HPLC samples, but at a later timepoint, thus involve 12 (*BGLU3*) or 15 (*BGLU4*) month old seeds. All qRT-PCR samples were prepared in biological triplicates. Sample growth, RNA isolation, cDNA synthesis and *ACTIN 2* control were performed as mentioned above (2.2, 2.7, 2.8). Due to strong homology of the *BGLU* genes, specific primer design for qRT-PCR needed the combination of several design approaches (see 2.11). At the end primers JFF032 and JFF033 for *BGLU1*, JFF059 and RS1320 for *BGLU3* and JFF072 and JFF073 for *BGLU4* were chosen. All primer pairs are intron and T-DNA insertion spanning in the respective loss-of-function T-DNA insertion line, to test the presence of an intact transcript. Each primer pair was checked by RT-PCR for specificity, using the same PCR approach as described in 2.9 choosing the annealing temperatures of 52 °C (*BGLU1* and *BGLU3*) or 47 °C (*BGLU4*), an extension time of 30 s and 1 µl cDNA as template for all samples. Results were analyzed by electrophoresis in a 2 % agarose gel (see 2.18).

The qRT-PCR was performed with a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad) in technical triplicates, using the Luna® Universal qPCR Master Mix (NEB) according to the manufacturer's instructions, but taking only half of the recommended amount. As template 1 µl cDNA of a ¼ dilution with water (LiChrosolv®, LC-MS Grade, Merck) were applied. Three housekeeping genes were chosen as reference genes, to normalize the *BGLU* expression levels, which were *ACTIN 2* (primers RS469 and RS470), *PEROXIN 4* (*Pex4*; *At5g25760*; primers RS935 and RS936) and *EF1α* (*At5g60390*; primers JFF028 and JFF029). The cycling program was 2 min, 95 °C—15 s, 95 °C; 30 s, 60 °C—5 s, 60 °C and 1 min 95 °C with 39 cycles.

Expression levels were determined by relative quantification of *bglu* or *2x35S:BGLU* expression compared to the wild type, using the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001). Calculations were performed with the mean of the technical replicates, using

$$\Delta\Delta C_t = \Delta C_t \; (treated \; sample) - \Delta C_t \; (untreated \; sample) \tag{1}$$

and

$$\Delta C_t = C_t \; (gene \; of \; interest) - C_t \; (reference \; gene) \tag{2}$$

The treated samples in this case are the respective *bglu* or *2x35S:BGLU* samples. The untreated samples are the respective wild type samples. If C_t values of the *bglu* samples were zero, for calculations the maximum cycle number added by one (= 40) was taken. All C_t values are listed in table S2 in the supplements. For the reference genes, the geometric mean (GM) of two reference genes (*Pex4* and *EF1α*), with the formula

$$GM = \sqrt[n]{x_1 * x_2 * ... * x_n} \tag{3}$$

was taken, where n stands for the number of samples and x for the mean of the respective reference gene (Vandesompele *et al.*, 2002). Furthermore, in the analysis the standard deviation (SD) was exchanged by the standard error (SE; Cumming *et al.*, 2007), with the formula

$$SE = \frac{\sigma}{\sqrt{n}} \tag{4}$$

with σ as standard deviation and n as the number of samples. In addition, assuming to have uncorrelated variables, the additive error after Taylor (1997) was used to calculate the asymmetrically distributed errors of the ΔC_t values relative to the average value (Livak and Schmittgen, 2001), with

$$error(a + b) = \sqrt{error(a)^2 + error(b)^2}. \tag{5}$$

2.11. Primer design

Primer design was performed with ApE, using the integrated primer tool, if not mentioned differently. Specificy of primers was then controlled with TAIR BLAST (https://www.arabidopsis.org/Blast/).

Specific qRT-PCR primers were designed by different approaches, each time intron and T-DNA insertion spanning for the respective *bglu* line, in order to investigate the expression of the intact transcript. The primers were designed for an amplicon length of 50 to 150 bp, if possible and 18 to 24 nucleotides in length. Primer dimerization and self-dimerization was avoided and checked with the Multiple Primer Analyzer (Thermo Fisher Scientific, https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html). Primers JFF032 and JFF033 for *BGLU1* were designed with ProbeFinder v2.53 (https://qpcr.probefinder.com/input.jsp). Primers JFF059 and RS1320 for *BGLU3* were designed manually from the cDNA sequence and checked as described above, the amplicon size was 350 bp, due to homology problems for smaller amplicons. Primers JFF072 and JFF073 for *BGLU4* were designed with the python script find_primers.py (https://github.com/hschilbert/Primer_design). The input fasta file was a primer list designed with the primer tool of ApE on the cDNA sequence, 300 bp upstream or downstream of the T-DNA insertion, for the forward and reverse primer

respectively. Best primers from both output lists were double checked with TAIR BLAST for specificity. Again, the amplicon size was 300 bp, due to homology problems for smaller amplicons.

2.12. Generation of loss-of-function double T-DNA insertion mutants

Due to probable gene redundancy in the *BGLU* gene cluster, affecting in the present work especially *bglu1-1* and *bglu4-2*, apart from the loss-of-function single T-DNA insertion mutants, loss-of-function double T-DNA insertion mutants were generated by crossing of the respective homozygous loss-of-function single T-DNA insertion mutants. Precisely, heterozygous lines for *bglu1-1_5-1*, *bglu1-1_3-2* and *bglu3-2_4-2* were generated, as well as homozygous lines for *bglu1-1_3-2* and *bglu3-2_4-2*.

Loss-of-function single T-DNA insertion mutants were grown in the greenhouse as described in 2.2 until first flowers opened, but closed buds still remained. Crossing was performed in two separate batches, using both loss-of-function mutants as the mother and father plant, respectively. From the actual mother plant, cautiously the sepals, petals and stamina with pollen sac were removed with a fine pincette from closed buds, without hurting the pistil. Stamina with pollen sac were taken from the actual father plant with a fine pincette and pollen were transferred to the free pistil of the mother plant. The mother plant was kept in the greenhouse for two days without light, to prevent withering of the pistil. After one day pollination of the mother plant was repeated. After two days the light was turned on and the plants were grown, until silique ripening. Siliques from crossing were separately wraped in small paper bags and harvested after opening of the siliques. The seeds were sown on soil in the greenhouse and grown as described in 2.2.

PCR based genotyping was performed, if rosette leaves were at least 1 to 2 cm in size. DNA extraction was performed as described in 2.4.1. PCRs were performed on the respective wild type *BGLU* allele and the respective insertion *bglu* allele from each parental line. Primers for the wild type allele were RS1302 and RS1303 (*BGLU1*), RS1319 and RS1320 (*BGLU3*), RS1259 and RS1260 (*BGLU4*) and RS1251 and RS1230 (*BGLU5*). Primers for the insertion allele were RS1845 and RS1848 (*bglu1-1*), RS1846 and RS1847 (*bglu3-2*), RS1259 and 3144 (*bglu4-2*) and RS1242 and RS1237 (*bglu5-1*). PCRs were performed with a homemade *Taq* DNA polymerase. 25 µl reaction volume contained 1x PCR buffer (10x: 500 mM KCl, 100 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$), 200 µM dNTPs, 0.2 µM each primer, 0.5 µl *Taq* DNA polymerase (homemade) and 2 to 3 µl genomic DNA as template. The cycling protocol 3 min, 94 °C—30 s, 94 °C; 30 s, 58 °C (*BGLU1*, *bglu1-1*, *BGLU3*, *bglu3-2*) or 55 °C (*BGLU4*, *bglu4-2*, *BGLU5*, *bglu5-1*); 1 min (*BGLU3*) or 1 min and 30 s (*BGLU1*, *bglu1-1*, *bglu3-2*, *bglu5-1*) or 2 min (*bglu4-2*) or 3 min and 30 s (*BGLU4*), 72°C—5 min, 72°C with 35 cycles was performed. The results were analyzed by electrophoresis in a 1 % agarose gel (see 2.18), expecting for heterozygous lines an amplicon for each allele.

Seeds from the heterozygous lines *bglu1-1_3-2* and *bglu3-2_4-2* were sown on soil and plants were grown in the greenhouse as described above. Due to expected self-pollination of the heterozygous plants in the next generation, genotyping was performed as described above, this time selecting for homozygous lines, showing exclusively amplicons for the respective

insertion alleles.

2.13. Cloning of genomic DNA

Full length genomic DNA cloning of *BGLU1*, *BGLU3* and *BGLU4* was performed by using the GATEWAY® technology, for following respective *BGLU* complementation lines.

Genomic DNA was extracted with a CTAB based method as described in 2.4.2 from Arabidopsis Col-0 rosette leaves. Differing from the described protocol, only half of the amount and 10 % 2-ME instead of 1 M DTT was used. The second DNA precipitation was done as described in 2.4.1.

For PCR amplification of the full length genomic DNA sequences of *BGLU1*, *BGLU3* and *BGLU4*, the primers were designed around 1700 bp upstream of the translation start and around 350 bp downstream of the translation termination, if possible. The sequence reference was based on the TAIR10 annotation and it was assured, not to locate primers in the region of the up- or downstream following genes. This was achieved with the phyton script seqex.py (https://github.com/bpucker/MolecularMethodsInGenomeResearch), yielding upstream of the translation start 1773 bp (*BGLU1*), 123 bp (*BGLU3*) or 938 bp (*BGLU4*) and downstream of the translation termination 321 bp (*BGLU1*), 381 bp (*BGLU3*) and 398 bp (*BGLU4*). The primers were JFF035 and JFF036 (*BGLU1*), JFF037 and JFF038 (*BGLU3*) and JFF039 and JFF041 (*BGLU4*), exhibiting all *attB*-recombination site overhangs. PCRs were performed with a Q5® High Fidelity DNA Polymerase (NEB) as described in 2.5.2 with a 50 µl reaction volume and 2 µl genomic DNA as template. The annealing temperatures were 68 °C (*BGLU1*), 66 °C (*BGLU3*) or 61 °C (*BGLU4*) and the extension time was 2 min (*BGLU3*) or 3 min (*BGLU1* and *BGLU4*). The PCR results were analyzed by electrophoresis in a 1 % agarose gel (see 2.18) and the respective bands purified from agarose gel, using the NucleoSpin® Gel and PCR Clean-Up kit (Macherey-Nagel) according to the manufacturer's instructions. All amplicons were recombined in the GATEWAY® vector pDONR™/Zeo with BP clonase (Invitrogen™), according to the manufacturer's instructions, using electrocompetent *E.coli* TOP10 cells for transformation as described in 2.15, selecting with zeocin (50 µg/ml). Plasmids were then isolated as described in 2.16 and checked by enzymatic restriction (see 2.17) and Sanger sequencing at the SCF. Primers for sequencing were unil, revl, JFF004, JFF056, RS1303, RS1389, RS1390, RS1393, RS1474, RS1484, RS1845 and SeLB for *BGLU1*, unil, revl, JFF031, RS1392, RS1503 and RS1847 for *BGLU3* and unil, revl, JFF001, JFF006, F306, RS1260 and RS1261 for *BGLU4*.

2.14. Cloning of genomic complementation constructs

For the generation of complementation lines of *bglu1-1*, *bglu3-2* and *bglu4-2* the genomic sequences from 2.13 were introduced into the vector pMDC123 (Curtis and Grossniklaus, 2003) using LR clonase and the GATEWAY® technology (Invitrogen™), according to the manufacturer's instructions, with electrocompetent *E. coli* TOP10 cells for transformation as described in 2.15, selecting with ampicillin (100 µg/ml). Plasmids were then isolated as described in 2.16 and checked by enzymatic restriction (see 2.17). For *BGLU3* the recombined sequence was checked by Sanger sequencing at SCF, using the primers unil, revl, JFF005,

JFF024, JFF025, JFF030, JFF031, KT04, P35S, RS1320 and T35S.

In addition, genomic complementation constructs for *bglu1-1*, *bglu3-2* and *bglu4-2* with a green fluorescent protein (GFP)-tag were generated. The coding sequence of GFP was introduced prior to the translation termination in the genomic *BGLU1*, *BGLU3* and *BGLU4* sequences in pDONR™/Zeo (see 2.13) with Gibson assembly® (Gibson *et al.*, 2009). Therefore, primers were designed with the NEBuilder® Assembly Tool v2.2.6 (https://nebuilder.neb.com/#!/, Nov 2019) using the respective entry clones as backbone sequence and the GFP sequence from the vector pMDC83 (Curtis and Grossniklaus, 2003) as insert sequence. The resulting primers for the backbone (bb) were JFF042 and JFF043 (*BGLU1*), JFF046 and JFF047 (*BGLU3*) and JFF050 and JFF051 (*BGLU4*). The resulting primers for the GFP insert were JFF044 and JFF045 (*BGLU1*), JFF048 and JFF049 (*BGLU3*) and JFF052 and JFF053 (*BGLU4*). The PCRs were performed with the Q5® High Fidelity DNA Polymerase (NEB), as described in 2.13, using 2 µl plasmid DNA in a 1/500 dilution with water as template. The annealing temperatures were 59 °C (*BGLU1* bb), 59.3 °C (*BGLU3* bb), 56.4 °C (*BGLU4* bb) or 60.6 °C (GFP) the extension time was 1 min (GFP), 3 min (*BGLU3* bb), 3.5 min (*BGLU4* bb) or 4 min (*BGLU1* bb). The PCR products were purified from a 1 % agarose gel after electrophoresis (2.18), using the NucleoSpin® Gel and PCR Clean-Up kit (Macherey-Nagel) according to the manufacturer's instructions. The DNA concentration was determined with NanoDrop™ 2000, before doing a DpnI digestion (5 µl PCR product, 1 µl CutSmart buffer (NEB), 0.5 µl DpnI (NEB), 3.5 µl water (LiChrosolv®, LC-MS Grade, Merck)) by incubation for 25 min at 37 °C and inactivation for 20 min at 80 °C, to remove residual plasmid DNA. These products were directly used for the Gibson assemly®. 5 µl DNA, containing three times more insert than backbone on equimolar levels, were mixed with 15 µl master mix (320 µl IRB [2 M Tris-HCl, 2 M MgCl$_2$, 1 M DTT, 100 mM NAD, 10 mM dNTPs, 50 % (v/v) PEG-8000], 0.64 µl T5 exonuclease [10 ^{U}/$_{µl}$, NEB], 20 µl Phusion High-Fidelity DNA Polymerase [2 ^{U}/$_{µl}$, NEB], 160 µl *Taq* DNA ligase [40 ^{U}/$_{µl}$, NEB], 700 µl H$_2$O$_{MilliQ}$). After incubated for 1 h at 50 °C, the Gibson assembly products were transformed into *E.coli* TOP10 cells as described in 2.15 and selected with zeocin (50 ^{µg}/$_{ml}$). Plasmids were isolated as described in 2.16 and checked by a restriction digest as described in 2.17. For *BGLU3* the restriction digest was done after following LR reaction. Plasmids from Gibson assembly were sequenced with Sanger sequencing at the SCF. Primers for sequencing were unil, revl, JFF002, JFF003, JFF004, JFF022, JFF036, JFF056, RS1303, RS1389, RS1390, RS1393, RS1474, RS1484 and SeLB for *BGLU1* and unil, revl, JFF001, JFF006, JFF026, F306, F307, KT04, RS1239, RS1259, RS1260, RS1261 and RS1267 for *BGLU4*. Primers for *BGLU3* were the same as described above.

The final genomic complementation constructs (with and without GFP-tag) were transfromed into the respective T-DNA insertion lines *bglu1-1*, *bglu3-2* or *bglu4-2* by Lennart Sielmann in his master thesis, using the *Agrobacterium tumefaciens* mediated floral dip method (Clough and Bent, 1998) and generating putative complementation lines, which have to be validated in future work.

2.15. Transformation of electrocompetent *E. coli* TOP10 cells

The products of the Gateway® BP- and LR-reaction were applied to transformation of electrocompetent *E. coli* TOP10 cells (F– *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G, Invitrogen™). For transformation *E. coli* TOP10 cells were thawed on ice and mixed with 2 µl reaction product. The cells were then transferred to a precooled electroporation cuvette (1 mm electrode distance) and electroporation took place in the "EcI" modus of the MicroPulser™ (BIO-RAD). After electroporation cells were transferred into 1 ml of S.O.C medium (20 mM glucose, 10 mM NaCl, 10 mM MgCl$_2$, 10 mM MgSO$_4$, 2.5 mM KCl, 20 g/l tryptone, 5 g/l yeast extract) and incubated for 1 h at 37°C and 900 rpm in the thermomixer for regeneration. 100 µl of the cells were plated on a LB agar plate (10 g/l tryptone (Bacto™), 5 g/l yeast extract (Bacto™), 5 g/l NaCl, 15 g/l agar) with the appropriate antibiotic to select positively transformed cells. The remaining cells were sedimented, resuspended in the reflux and plated as well. Dried and sealed plates were incubated upside down at 37 °C over night.

2.16. Plasmid isolation

All following centrifugation steps were performed at full speed in a table centrifuge. For plasmid isolation 3 ml of LB medium (10 g/l tryptone (Bacto™), 5 g/l yeast extract (Bacto™), 5 g/l NaCl) with the appropriate selecting antibiotic, were inoculated with a single colony and incubated at 37 °C and 220 rpm over night.

2 ml of the over night culture were centrifuged for 2 min, the supernatant completely removed and 100 µl of solution I (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0) were added to resuspend the sediment by vortexing. Afterwards, 200 µl solution II (0.2 M NaOH, 1 % (w/v) SDS) were added for lysis and carefully mixed, by just inverting the tubes. After 5 min at room temperature, 150 µl solution III (3 M sodium acetate pH 4.8) for neutralization were added and carefully mixed, by just inverting the tubes.

Centrifugation for 5 min followed, before transferring the supernatant into a new reaction tube. Samples were mixed carefully with 1 ml of icecold (-20 °C) 100 % ethanol by inverting the tubes. Centrifugation for 10 min followed and the supernatant was discarded. The sediment was washed in 1 ml 70 % (v/v) ethanol and air-dried samples were resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) with 0.1 mg/ml RNaseA (Sigma). Low-copy plasmids (pDONR™/Zeo backbone) were resuspended in 30 to 50 µl, other plasmids in 50 to 100 µl TE buffer. The samples were incubated for 10 min at 37 °C and 700 rpm in the thermomixer and afterwards at least 1 h at 4 °C without shaking for RNaseA digestion.

2.17. Enzymatic restriction of plasmids

Enzymatic restriction for plasmids, was performed prior to sequencing, to select for putatively successful cloning. All restrictions were performed with enzymes from (NEB). In a volume of 15 µl, 2 µl to 6 µl of isolated plasmids were mixed with the appropriate 1x NEB buffer and at least 1 unit of enzyme to digest 1 µg DNA in 1 hour. Therefore, 0.3 µl enzyme for 20 U/µl, 0.6 µl enzyme for 10 U/µl and 1.2 µl enzyme for 5 U/µl were chosen. The restriction was

performed for at least 1 h at the recommended temperature of the restriction enzymes. The chosen restriction enzymes for the respective plasmids are shown in table 2. All restriction digests were analyzed in an agarose gel according to section 2.18. The restriction for *BGLU3* with GFP in pDONR™/Zeo did not yield positive results, in a second run, this restriction was skipped and only restriction after the LR reaction was performed, with following sequencing.

Table 2 – *Restriction enzymes. All enzymes (NEB) for plasmid restriction are listed.*

Gene	Plasmid	Enzymes
BGLU1	pDONR™/Zeo	PvuII
BGLU1-GFP	pDONR™/Zeo	PvuII
BGLU3	pDONR™/Zeo	BsrGI
BGLU3-GFP	pDONR™/Zeo	ClaI, HindIII, XmaI
BGLU4	pDONR™/Zeo	EcoRI, HindIII
BGLU4-GFP	pDONR™/Zeo	ClaI, EcoRV
BGLU1	pMDC123	EcoRV
BGLU1-GFP	pMDC123	EcoRV
BGLU3	pMDC123	NdeI
BGLU3-GFP	pMDCS123	PvuII
BGLU4	pMDC123	PvuII
BGLU4-GFP	pMDC123	NcoI-HF, XbaI

2.18. Gel electrophoresis

Gel electrophoresis was, according to the fragment size, performed with 0.8 - 2% agarose gels, using 0.1 µg/ml ethidium bromide for visualization. The gels were run with 1x TAE-buffer (50x: 242 g/l Tris, 57.1 ml/l acetic acid, 10 ml/l of 0.5 M EDTA pH 8.0) and 10 V per centimeter electrode distance. Probes were loaded with 5x gel loading buffer (20 mM EDTA pH 8.0, 50% (w/v) glycerol, 0.08% (w/v) xylene cyanol) to a 1x concentration. DNA ladders (NEB) were used as size standard, according to the expected fragment size (1 kb, 2-log or 100 bp). The visualization with UV-light was performed with a trasnilluminator (Intas®).

2.19. Metabolite analysis

To investigate the function of *BGLU1*, *BGLU3* and *BGLU4*, phenotypical changes of differential metabolite levels were investigated by a genetic approach, using a loss-of-function T-DNA insertion line, an overexpression line and the wild type for each respective gene. The differential metabolite levels were analyzed by untargeted ultra-high performance liquid chromatography–diode array detection–positive electro-spray ionization–quadrupole time-of-flight–tandem

mass spectrometry (UHPLC-DAD-ESI$^+$-QTOF-MS/MS) metabolic fingerprinting, searching for compounds, that exhibit differential peaks, expected for putative substrate or product compounds of the encoded enzymes.

2.19.1. UHPLC-DAD-ESI$^+$-QTOF-MS/MS sample preparation

Plants were grown as described in 2.2. For each investigated *BGLU* gene the HPLC sample set contained the respective T-DNA insertion line *bglu1-1*, *bglu3-2* or *bglu4-2*, the respective overexpression line *2x35S:BGLU1*, *2x35S:BGLU3* or *2x35S:BGLU4* and the Arabidopsis Col-0 wild type. All samples were prepared in four replicates. Harvesting and further preparation was performed in a random order, to avoid batch effects, as described for the sample growth. Each *BGLU1* sample contained four to six at least 10 mm long rosette leaves (around 200 mg) of six-week-old Arabidopsis plants, immediately frozen in liquid nitrogen after harvesting with a sterile pincette into a 2 ml Eppendorf$^®$ safe-lock microcentrifuge tube and stored at -80 °C until use. The choice of the 2 ml Eppendorf$^®$ safe-lock microcentrifuge tube was a critical parameter for later grounding. Harvesting was done 6 h after artificial sunrise in the growth chamber, to adjust the metabolism to day conditions. Plants from overexpression samples were genotyped afterwards, to confirm the transgene. The DNA extraction for genotyping was performed as described in 2.4.1. The PCR based genotyping was done with a *Taq* DNA polymerase (NEB) as described in 2.9, using 2 µl genomic DNA as template, an annealing temperature of 55 °C and an extension time of 2 min. The transgenic located primer was RS1434 and the gene specific primer RS1390.

All *BGLU3* and *BGLU4* samples were taken from a batch with combined seeds of several plants to have enough seeds for all four replicates. For each *BGLU3* sample 150 mg dry seeds were transferred after harvesting into a 2 ml Eppendorf$^®$ safe-lock microcentrifuge tube and immediately frozen in liquid nitrogen before storing at -80 °C. For each *BGLU4* sample 120 mg of dry seeds after harvesting were sown on a wet filter paper (7 x 7 mm, 1 ml H_2O_{MilliQ}) in a sterile Petri dish. The Petri dish was closed with parafilm and the seeds were soaked for 24 h in the dark at 4 °C, before being washed from the filter paper with H_2O_{MilliQ} and transferred into a 2 ml Eppendorf$^®$ safe-lock microcentrifuge tube. Samples were centrifuged for 1 min at full speed, water was discarded and the soaked samples immediately frozen in liquid nitrogen and stored at -80 °C.

For *BGLU5* HPLC samples were prepared as described for *BGLU1* but not further analyzed, thus skipped in the following sections.

All HPLC samples were freeze-dried for 48 h and then grounded in a ball mill (Retsch MM301) with 7 mm steel balls for 3 min at 30 Hz. 10 mg $\pm$ 1 µg of dry powder was transferred into a new microcentrifuge tube and stored in the exicator. For extraction of polar and semi-polar metabolites, dry samples were treated three times with 333 µl of 90 % (v/v) methanol (90 ml methanol [LC-MS grade, Fisher Scientific UK Limited, Loughborough, UK or Th. Geyer GmbH & Co. KG, Renningen, Germany], 10 ml H_2O_{MilliQ}) including luteolin-7-*O*-glucoside (0.6 mg per 100 ml, Extrasynthese) as internal standard. After 5 min of vortexing, samples were centrifuged for 10 min at 13,200 rpm. Each supernatant was combined and pooled supernatants were then filtered with Phenex^{TM} syringe filters (Phenomenex$^®$, 0.2 µm) into

HPLC autosampler glass vials and stored at -80 °C until measurements. For each ten samples one blank with only the extraction solution was prepared in the same way as described above. All extraction steps were performed at 4 °C or on ice.

2.19.2. Untargeted UHPLC-DAD-ESI$^+$-QTOF-MS/MS metabolic fingerprinting

Metabolic extracts were analyzed with ultra-high performance liquid chromatography (UHPLC: Dionex UltiMate 3000, Thermo Fisher Scientific, San José, CA, USA) coupled to diode array detection (DAD), positive electrospray ionization (ESI$^+$) and quadrupole time of flight mass spectrometry (QTOF: compact, Bruker Daltonics, Bremen, Germany). 1 to 6 µl extract were injected to a Kinetex XB-C18 column (1.7 µm, 150 mm x 2.1 mm), including a guard column (Phenomenex$^®$, Torrance, CA, USA), which protects the column from remaining impurities and suspended solids in the extract. The precise injection volume for each sample is given in table 3. Extracts were separated at 45 °C with a flow rate of 0.5 ml min^{-1}. The mobile phase was a gradient from A (0.1 % formic acid [p.a., eluent additive for LC-MS, ~98 %, Honeywell Research Chemicals, Fluka] in H$_2$O$_{MilliQ}$) to B (0.1 % formic acid in acetonitrile [LC-MS grade, Fisher Scientific, Loughborough, UK or HiperSolv CHROMANORM, VWR, Fontany-sous-Bois, France]), with an initial concentration of 2 % B. For the standard gradient, B was linearly increasing to 30 % B in 20 min, before increasing to 75 % B within 9 min. In case of measurements for enhanced peak separation two different flat gradients, increasing B more slowly, were applied, with B linearly increasing to 30 % B in 40 min, before increasing to 75 % B within 18 min, for the first enhanced separation (called 58 min gradient). In the second alternative separation B was linearly increasing to 30 % B in 80 min, before increasing to 75 % B within 36 min (called 116 min gradient). For the standard method prior to each sample, at 0.3 min, a sodium formate (Na(HCOO))-based calibration solution was introduced to the ESI sprayer for mass axis recalibration. This was also the case for standard measurements in the flat gradient (58 min and 116 min). For most measurements of seed samples (*BGLU3* and *BGLU4*) and for the ESI$^-$ mode, this recalibration was performed after sample separation. For the standard gradient this took place between 29 min and 33.5 min, with 32.5 min to 33.5 min marked as recalibration segment. For the 58 min gradient, this took place between 58.0 min and 63.5 min, with 62.5 min to 63.5 min marked as the recalibration segment. In addition, for these measurements, between 6.2 min and 7.0 min the sample was piped into the waste and the recalibration solution into the ESI sprayer. This method was called *cutLC* and all measurements with this method are shown in table 3. The reason to use this method was a very huge peak in this area in seed samples, causing a detection overload with normal measurement parameters. Therefore, a separate measurement with standard parameters was performed, using *BGLU3* seed samples in a ^{1}/$_9$ dilution (table 3). For all measurements the final step after separation was equilibration with 100 % B, before returning to initial conditions.

Mass spectra were gained in general in the positive ESI mode as line spectra, from 50 to 1300 m/z at 1 to 8 Hz. The spectra rate for each measurement is shown in table 3. For the ESI source a nebulizer pressure of 3 bar with end plate offset of 500 V and capillary voltage of 4500 V was used. The drying gas was N$_2$ at a flow rate of 12 L min^{-1}, heated to 275 °C. For

the MS mode the quadrupole ion energy was set to 4 eV, the low mass cut-off at 90 m/z and the collision energy at 7 eV, with a transfer time of 75 µs and a pre-pulse storage of 6 µs. The low mass cut-off removes all masses below 90 m/z. The transfer time and pre-pulse storage determine the transmission of the ions through the quadrupoles into the TOF. For MS/MS spectra the auto-MS/MS mode was used with N_2 as the collision gas. The mass range was 50 to 1300 m/z, with collision energies of 20 eV for 100 m/z, increasing to 30 eV for 500 m/z, to 40 eV for 1000 m/z and then to 50 eV for 1300 m/z. Three precursor ions per cycle were allowed to enter into the collision cell, without a charge state restriction. Active precursor exclusion was performed after two spectra and precursors were reconsidered for fragmentation after 0.2 min or if the current intensity was at least three times higher than the previous one. This assures, that following isomers with the same m/z are fragmented too and that fragmentation is repeated, if the peak is not finished. The smart exclusion threshold was set at 3, guaranteeing, that a peak is only considered for fragmentation, if the peak is three times higher, than the baseline. The absolute threshold for a peak was set at 2000 counts. The isolation width was set to 2 Da for 100 m/z, increasing to 4 Da for 500 m/z, then to 6 Da for 1000 m/z and then to 8 Da for 1300 m/z. The isolation width is the mass window, the MS detector uses to isolate the precursor ions.

Chosen samples were measured in addition in the ESI$^-$ mode. This changed the capillary voltage in the ESI source to 3500 V, while all other parameters remained. The mass spectra in the negative ion mode were gained at 1 Hz and 6 µl of samples were injected. For candidate features with very small peaks, the measurement was repeated with 1 Hz for line spectra, to enhance the peak intensities.

Table 3 – **_LC and MS parameters._** *All varying parameters for LC and MS measurement are listed in the following table.*

Measurement	Method	Gradient	Injection	Spectra rate
BGLU1 fingerprinting	standard	standard	6 µl	8 Hz
BGLU1 peak enhancement	standard	standard	6 µl	1 Hz
BGLU1 ESI$^-$	cutLC	standard	6 µl	1 Hz
BGLU3 fingerprinting	cutLC	standard	3 µl	5 Hz
BGLU3 flat gradient 1	cutLC or standard	58 min	3 or 6 µl	5 Hz or 2 Hz
BGLU3 flat gradient 2	standard	116 min	6 µl	2 Hz
BGLU3 fingerprinting $1/9$	standard	standard	1 µl	8 Hz
BGLU3 ESI$^-$	cutLC	standard	6 µl	1 Hz
BGLU4 fingerprinting	cutLC	standard	3 µl	5 Hz
BGLU4 peak enhancement	cutLC	standard	3 µl	1 Hz
BGLU4 ESI$^-$	cutLC	standard	6 µl	1 Hz

2.19.3. Data processing of the metabolic fingerprint

Mass axis calibration, peak picking and spectral background subtraction was performed with Compass DataAnalysis v4.4 (Bruker Daltonics), using the Find Molecular Feature algorithm, for all metabolic fingerprint and peak enhancement measurements. For standard data processing the signal-to-noise threshold was set to 3, the correlation coefficient threshold to 0.75, the minimum compound length as 22 spectra and the smoothing width to 6. Data processing parameters for all measurements, including parameter variations, are shown in table 4. For *BGLU3* and *BGLU4 cutLC* measurements the split-multiple-compound method for peak picking was chosen, to separate a close peak from the internal standard, exhibiting the same m/z. For bucket generation allowed ion types were $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$, $[M+K]^+$, $[M-H_2O+H]^+$, $[M-CO_2+H]^+$, $[M+Na_4-H]^+$, $[2M+H]^+$, $[3M+H]^+$, $[2M+Na]^+$, $[2M+NH_4]^+$ and $[2M+K]^+$. Peaks from the same compound were aligned in one bucket with Compass ProfileAnalysis v2.3 (Bruker Daltonics), choosing the m/z with the highest intensity as bucketing basis. In the following procedure, each bucket is called a single feature. In addition, advanced bucketing was used, allowing deviations of 0.1 or 0.2 min of the retention time and 6 mDa of the m/z, for peaks in one bucket. For cutLC measurements, all peaks between 6.2 min and 7.0 min were skipped in the bucket generation. Therefore, for the *BGLU3* measurement with the 1/9 dilution, only peaks between 6.0 min and 13.0 min (to include the internal standard, eluting around 12.0 min) were considered.

Table 4 – ***Peak picking and bucket generation.*** *All varying parameters for peak picking and bucket generation are listed in the following table.*

Measurement	S/N	correlation coefficient	minimum compound length	smoothing width	RT shift [min]
BGLU1 fingerprinting	3	0.75	22	6	0.1
BGLU1 peak enhancement	1	0.75	5	0	0.1
BGLU3 fingerprinting	3	0.75	15	4	0.2
BGLU3 fingerprinting 1/9	3	0.75	22	6	0.2
BGLU4 fingerprinting	3	0.75	15	4	0.2
BGLU4 peak	1	0.75	5	0	0.1

Table 4 – continued

Measurement	S/N	correlation coefficient	minimum compound length [spectra]	smoothing width	RT shift [min]
enhancement					

2.19.4. Identification of candidate features by fold changes

To identify candidate features with differential metabolite levels, expected for substrate or product compounds for the encoded enzymes of the respective *BGLU* gene, the fold changes of all features were calculated, by comparing the peak intensities of the *2x35S:BGLU* or *bglu* samples to the wild type. All samples were adjusted to the highest intensity of the internal standard with ProfileAnalysis v2.3 (Bruker Daltonics) and then adjusted to the blank by selecting all samples, where the mean of all replicates is 50 times higher compared to the mean of the blank.

$$Sample_{mean} > 50 * Blank_{mean} \qquad (6)$$

To discard measurements without effect, only features with at least one sample group showing values above zero in 75 % of the replicates were selected.

$$Count = \frac{Replicates \ > \ 0}{n} \qquad (7)$$

$$Feature_{selection} = Count > 74.0 \qquad (8)$$

These samples were adjusted to the exact dry weight.

$$Feature_{adjusted} = \frac{Feature_{selection}}{dry \ weight} \qquad (9)$$

Next, features with at least one sample group showing in 75 % of the replicates values above zero between *bglu* and wild type samples and/or between *2x35S:BGLU* and wild type samples, were chosen, as described above. The fold changes (FC) were then calculated by dividing the mean of the *bglu* (KO) or *2x35S:BGLU* (OX) samples through the mean of the wild type (WT) samples, respectively.

$$FC = \frac{OX_{mean} \ or \ KO_{mean}}{WT_{mean}} \qquad (10)$$

If the fold change is zero (due to no peaks, picked by the algorithm), the features are qualitatively lower compared to the wild type (called qual_low). Features are qualitatively higher compared to the wild type, if no peaks for the wild type sample were picked (called qual_high). The data were manually checked for features, which are in comparison to the wild type, higher in *2x35S:BGLU* samples (FC $\geq$ 1.5) and lower or not present in the *bglu* samples (FC $\leq$ 0.5), accounting for potential product features of the respective encoded BGLU protein. In addition, chromatogram data and the differential expression results from

RT- and qRT-PCR were considered for the selection. To check also for potential substrate features of the respective encoded BGLU protein, data were also checked for features lower in *2x35S:BGLU* samples and higher in *bglu* samples, compared to the wild type. For substrate features no fold change treshold was set. In case of qualitative fold changes but small peaks visible in the chromatogram of the main candidate features, the peak areas were manually integrated in DataAnalysis and the area adjusted to the peak area of the internal standard and to the dry weight. Fold changes were then calculated from these data.

2.19.5. Metabolite identification approach

Candidate product and substrate features were analyzed in Compass DataAnalysis v4.4 (Bruker Daltonics). Extracted Ion Chromatograms (EICs) were taken for each feature in each sample, allowing a peak mass drift of 0.005 m/z. Only candidate features with an EIC peak pattern in concordance with the fold change and differential gene expression result were kept and analyzed in more detail. If possible, the chemical formula was derived with Smartformula Manually (allowing: three sulfur and six phosphors, all allowed ion types from bucket generation, mass tolerance of 0.002 Da, mSigma lower than ten and error (ppm) lower than five). The MS/MS spectrum was analyzed in DataAnalysis regarding the fragment ion peaks and the respective neutral losses between the fragments (fragmentation pattern). In cases, where peaks were to small and therefore, no automatic MS/MS fragmentation was achieved or if more MS/MS information based on different collision energies for fragmentation were needed, peaks were targeted fragmented using the MRM mode (see 2.19.6).

Information about dominant fragment ion peaks and common neutral losses were derived from literature research. For structural formula prediction, the features were also compared with public databases using MetFrag (Ruttkies *et al.*, 2016), choosing spectral similarity (MoNA) and exact spectral similarity (MoNA) for an alignment with MassBank (https://massbank.eu/MassBank/). Results were checked for high similarity and if all dominant fragment ion peaks of the MS/MS spectrum were explained.

For all candidate substrate and product features the UV-vis spectrum was analyzed after background subtraction in DataAnalysis, for additional compound information based on the UV-vis absorption.

2.19.6. MRM fragmentation

Multiple Reaction Monitoring (MRM) fragmentation was performed with an alternating MS measurement. The MS measurement was performed as described in 2.19.2. For MRM all differing parameters from 2.19.2, including the method, gradient, injection volume, spectra rate for line spectra, collision energy and isolation width are given in table 5. For *BGLU1* targeted fragmentation was performed with the candidate product feature 741.2221 m/z at 10.82 min. This peak was also targeted fragmented in the negative ion mode, assuming a mass subtraction of 2.015651 Da from $[M+H]^+$ to $[M-H]^-$, thus 739.2065 m/z.

For *BGLU3* targeted fragmentation was performed with the candidate product feature 773.2116 m/z at 10.51 min. In addition, for this feature targeted fragmentation was performed at different collision energies and for some of these measurement a low mass cut-off of

350 m/z in the MS mode as well as a more flat gradient for advanced peak separation. This fragmentation was performed for the feature 611.1592 m/z at 10.52 min, too. The peak of 773.2116 m/z was also targeted fragmented in the negative ion mode, as described above, thus fragmenting 771.1959 m/z.

For *BGLU3* several other features were targeted fragmented, namely 411.7178 m/z at 11.87 min, 441.7099 m/z at 9.45 min, 441.7102 m/z at 9.96 min, 448.7181 m/z at 11.05 min, 460.1686 m/z at 11.76 min, 440.7207 m/z at 12.45 min, 432.7049 m/z at 14.62 min, 467.1764 m/z at 12.89 min, 494.1968 m/z at 19.11 min, 536.7408 m/z at 17.07 min, 757.2166 m/z at 11.42 min, 757.2152 m/z at 11.63 min, 943.3804 m/z at 13.74 min, 660.3742 m/z at 13.21 min, 402.1451 m/z at 13.11 min, 795.3419 m/z at 17.03 min, 287.1382 m/z at 7.25 min, 633.2899 m/z at 19.68 min and 575.2851 m/z at 18.19 min. For the candidate feature product 866.4164 m/z at 11.75 min in addition to fragmentation with the auto-MS/MS mode, targeted fragmentation was performed at different collision energies.

For *BGLU4* targeted fragmentation was performed for the candidate product feature 757.2149 m/z at 11.71 min and the candidate substrate features 611.1583 m/z at 11.34 min. The first peak was also targeted fragmented in the negative ion mode, assuming a mass of 755.1992 m/z m/z for [M-H]$^-$. Apart from these candidate features, 529.2079 m/z at 27.45 min, 507.1473 m/z at 14.37 min and 465.1003 m/z at 8.75 min were targeted fragmented. All data from MRM measurements were recalibrated with the recalibration segment manually in DataAnalysis.

Table 5 – ***MRM fragmentation.*** *All varying parameters for targeted MRM fragmentation are listed.*

Feature [m/z]	Method	Gradient	Injection [µl]	Spectra rate [Hz]	Collision energy [eV]	Isolation width [Da]
741.2221	standard	standard	6	1	35	5
739.2065 ESI$^-$	cutLC	standard	6	1	35	5
773.2116	cutLC	standard	5	2	35	5
773.2116 ESI$^-$	cutLC	standard	6	2	35	5
773.2116 58 min	standard	58 min	6	2	10, 20, 30	5
773.2116	cutLC	standard	6	2	40, 45	5
771.1959 ESI$^-$	cutLC	standard	6	1	35	5
611.1592 58 min	standard	58 min	6	2	10, 20, 30	4.5

Table 5 – continued

Feature [m/z]	Method	Gradient	Injection [μl]	Spectra rate [Hz]	Collision energy [eV]	Isolation width [Da]
411.7178	cutLC	standard	5	2	30	4
441.7102	cutLC	standard	5	2	30	4
448.7181	cutLC	standard	5	2	30	4
460.1686	cutLC	standard	5	2	30	4
440.7207	cutLC	standard	5	2	30	4
432.7049	cutLC	standard	5	2	30	4
467.1764	cutLC	standard	5	2	30	4
494.1968	cutLC	standard	5	2	30	4
536.7408	cutLC	standard	6	2	30	4
757.2166	cutLC	standard	6	1	35	5
757.2152	cutLC	standard	6	1	35	5
943.3804	cutLC	standard	6	2	40	6
660.3742	cutLC	standard	6	2	35	4.5
402.1451	cutLC	standard	6	2	30	3.5
795.3419	cutLC	standard	6	2	37	5
287.1382	cutLC	standard	6	2	25	3
633.2899	cutLC	standard	6	2	35	4.5
575.2851	cutLC	standard	6	2	30	4.5
866.4164	cutLC	standard	6	2	15, 25	6
866.4164	standard	standard	6	3	40, 50, 60	6
757.2149	cutLC	standard	6	1	35	5
755.1992 ESI^-	cutLC	standard	6	1	35	5
611.1583	cutLC	standard	6	1	35	4.5
529.2079	cutLC	standard	6	1	30	4
507.1473	cutLC	standard	6	1	30	4
465.1003	cutLC	standard	6	1	30	4

3. Results

3.1. Identification and characterization of T-DNA insertions in chosen *bglu* lines

For investigations on phenotypic differences regarding specific metabolite levels in different mutant lines of the respective *BGLU* genes, compared to the wild type, homozygous loss-of-function T-DNA insertion lines of *BGLU1*, *BGLU3* and *BGLU4* were successfully characterized. T-DNA insertion mutant alleles *bglu1-1* (GABI_341B12) for *BGLU1*, *bglu3-2* (GABI_853H01) for *BGLU3* and *bglu4-2* (SALK_029729) for *BGLU4* were identified in the Arabidopsis Col-0 background. The position, the orientation and the border sequences of the T-DNA insertion were identified by MinION long-read sequencing and Sanger sequencing. The results from the MinION long-read sequencing are given in table 6, by giving the coverage and average read length from the sequencing, as well as the number of contigs, the average, minimum and maximum contig size and the N50 value from the assembly. The N50 value is the minimum contig length to cover 50 % of the genome with contigs larger or equal to the N50 value (Miller *et al.*, 2010). The results show a clear improvement of the long-read sequencing, if the SRE kit was applied, to remove short reads. The best result was achieved for *bglu1-1*, while the small changes in the DNA isolation method did not change or improve the results and probably are not crucial for the isolation quality. The sequencing results are submitted to the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena), with ERS4255859 (*blgu1-1*), ERS4255860 (*bglu3-2*) and ERS4255861 (*bglu4-2*). All results were sufficient for determiation of the T-DNA insertions.

Table 6 – ***MinION long-read sequencing.*** *The results of the MinION long-read sequencing are listed, by giving the coverage and average read length from the sequencing, as well as the number of contigs, the average, minimum and maximum contig size and the N50 value from the assembly.*

Line	Coverage	Avg. read length [nt]	Contigs	Contig size [nt]	N50 [nt]
bglu1-1	93x	13329	132	avg: 957464 min: 10093 max: 16099356	14257951
bglu3-2	15x	3401	638	avg: 178834 min: 10046 max: 2792360	526830
bglu4-2	68x	9894	409	avg: 305874 min: 1101 max: 14749264	11246391

According to the SIGnAL T-DNA insertion database T-DNA insertions for *bglu1-1* and *bglu4-2* were expected to be located in the third intron and for *bglu3-2* in the sixth exon.

MinION long-read sequencing confirmed single T-DNA insertions in each investigated line at the expected positions. The only slight difference appeared for *bglu3-2*, were only one border is located in the predicted sixth exon and the other border instead in the sixth intron. Precise sequences of both T-DNA borders in the genomic region of the respective *bglu* line were analyzed by PCR amplification and following Sanger sequencing. Figure 13 gives an overview of all three *bglu* lines with the respective T-DNA insertion position and the T-DNA border sequences. In the following description of the border sequences, the genomic *bglu* sequence is underlined and the T-DNA border sequence indicated in bold. In almost all cases between both sequences additional nucleotides are present, as a result from the insertion event.
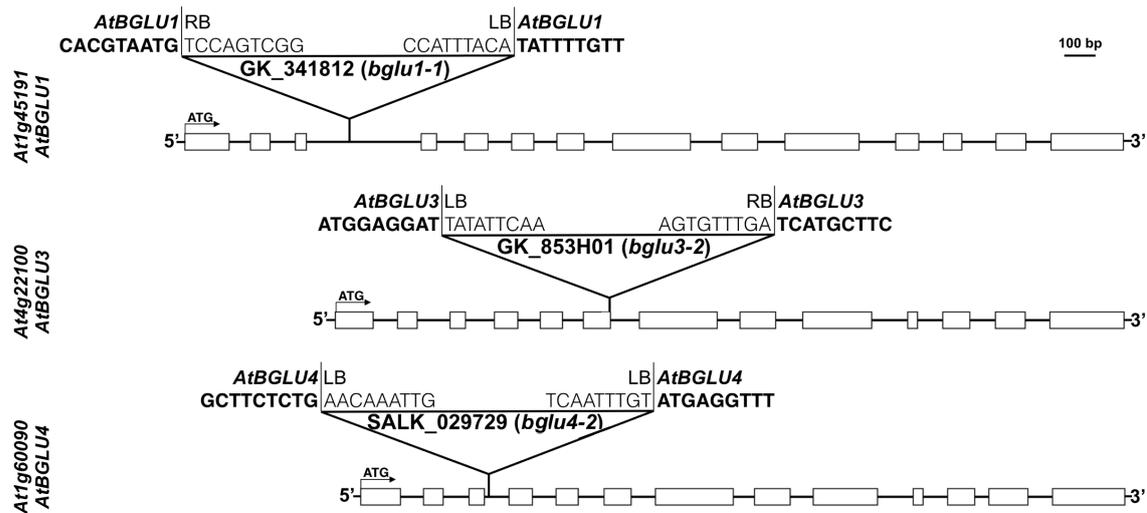
Figure 13 – *Overview of* **A. thaliana bglu** *T-DNA insertion lines.* *T-DNA insertion alleles of* BGLU1, BGLU3 *and* BGLU4 *are depicted. Boxes indicate exons of the wild type gene structure, lines in between indicate the respective introns. ATG marks the start codon. Triangles indicate the T-DNA insertion for* bglu1-1*,* bglu3-2 *and* bglu4-3*, showing the position in the respective* BGLU *gene, as well as the T-DNA insertion borders with the respective sequences of the T-DNA insertion and the genomic sequence (bold). Sequences in between, not matching genomic or T-DNA insertion sequences are not shown.*

Bglu1-1 exhibits on chromosome 1 one right T-DNA border (RB) at position 17116579 and one left T-DNA border (LB) at position 17116598. The RB exhibiting the border sequence C ACGTAATGAAGACAATATGGGAGACTAGTCTCCCTCATACTCATTGTATGTCCA TATGGGAGACTAGTC**TCCAGTCGG**, the LB the border sequence **CCATTTACA**T GAGTACCATATTTTGTT.

Bglu3-2 shows on chromosome 4 a LB at position 11709058 and a RB at position 11708997, with the LB border sequence ATGGAGGATAAATCATC**TATATTCAA** and the RB border sequence **AGTGTTTG**ATATAATTCTTCTCATGCTTC.

Bglu4-2 exhibits on chromosome 1 two LBs, the first one at position 22156000, the second one at position 22156017. The first LB sequence is GCTTCTCTG**AACAAATTG** and the second one **TCAATTTGT**CCGCAATGTGTTGTTGTCTATGAGGTTT.

Apart from the described T-DNA insertion characterization, MinION long-read sequencing could in addition confirm the homozygosity of the T-DNA insertion lines, after alignment of the reads to the genomic sequence of the respective *BGLU* gene (data not shown).

For further characterization these *bglu* lines were analyzed concerning the expression level of

the intact transcript of the respective *BGLU* gene, where the T-DNA is located.

3.2. Differential gene expression of *BGLU1*, *BGLU3*, *BGLU4* and *BGLU5* between mutant lines and Arabidopsis wild type

To compare phenotypic differences between the Arabidopsis wild type and *bglu* and *2x35S:BGLU* mutant lines on differential metabolite levels, *bglu* and *2x35S:BGLU* mutant lines of *BGLU1*, *BGLU3*, *BGLU4* and *BGLU5* were characterized concerning the respective gene expression levels of the intact transcript. Gene expression levels were determined by comparison of the respective amount of transcript in the *bglu* or *2x35S:BGLU* lines with the wild type. Gene expression was analyzed in tissues, with the highest expected expression in the wild type.

According to RNA-Seq data from the TraVa database, *BGLU1* shows highest expression in 10 mm rosette leaves with 813 read counts by median normalization (Lin *et al.*, 2016), apart from lower expression levels in other tissues. The gene expression level increases in growing rosette leaves but is comparably low in leaves at the flowering timepoint, thus samples for gene expression analysis were taken prior to flowering, from leaves which were at least 10 mm long. The expression of *BGLU3* is restricted to seeds, showing highest expression in dry seeds with 4006 read counts by median normalization and a lower expression in seeds of yellowing siliques (1800 read counts by median normalization). Concerning *BGLU4*, the expression is also restricted to seeds, but this time to soaked and germinating seeds, with highest expression after being soaked for 24 h (4062 read counts by median normalization), decreasing in the next days of soaking and germination. Since, the highest expression levels for *BGLU5* were found in rosette leaves (up to 3157 read counts by median normalization), the samples were taken as described for *BGLU1*. For *BGLU2* almost no expression was predicted with Trava (less then 15 read counts by median normalization). Further investigations were excludes, due to the presumption of a pseudogene.

The respective gene expression of intact transcripts was examined in two different approaches. The first one was a RT-PCR approach, where samples were harvested at the same time point and from the same sample batch as the HPLC samples. Since, *BGLU5* was not further examined regarding the metabolite fingerprint, only RT-PCRs were done and the results are shown in the supplements. The PCR results in the agarose gel did not show any crucial expression of intact transcript in all four *bglu* lines (figure 14A, 14B, 14C and S6), only for *bglu1-1* a neglectable slight band was seen in the agarose gel. The expression level of the overexpression lines was clearly higher compared to the wild type for *2x35S:BGLU1* (figure 14A), *2x35S:BGLU4* (figure 14C) and *2x35S:BGLU5* (figure S6), based on the strong bands in the agarose gel. Only for *2x35S:BGLU3* the expression level was not clearly distinguishable from the wild type in the agarose gel, but showing already for the wild type strong PCR bands (figure 14B). For *BGLU3* expression of the intact transcript was in addition examined in soaked seed samples, because it was in question, if transcript levels as well as protein translation are present in dry seeds. Therefore, the whole sample set, was prepared and examined for dry and soaked seed samples. For soaked seeds, similar results as for dry seeds were achieved, concerning the expression of the intact *BGLU3* transcript, examined by RT-PCR. Despite the

predicition from the TraVa database, gene expression was also present in the 24 h soaked wild type seeds, but less then in dry seeds. The results are shown in figure S5 in the supplements. Due to stronger expression of *BGLU3* in dry seeds in the wild type, further focus was put towards dry seed samples in the ongoing experiments.
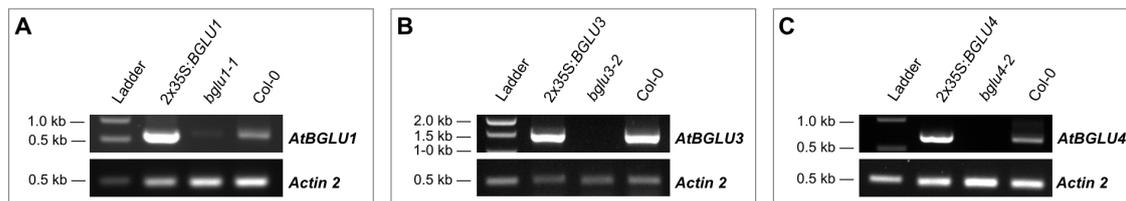
Figure 14 – ***RT-PCR results of* BGLU1, BGLU3 *and* BGLU4.** *RT-PCR results for the intact transcript of* BGLU1 *are shown in A, RT-PCR results for the intact transcript of* BGLU3 *are shown in B and RT-PCR results for the intact transcript of* BGLU4 *are shown in C. For each* BGLU *gene, the* 2x35S:BGLU *line, the* bglu *line and the Col-0 wild type are presented, including the* Actin 2 *control. The bands of the* Actin 2 *control of* BGLU3 *are cut from the original figure for visualization. Therefore, the original gel is shown in figure S5 in the supplements.*

To achieve more sensitive gene expression data for the intact transcript of each examined *BGLU* gene, qRT-PCR measurements were performed in a new sample batch but treated as the HPLC samples. In general, the results confirmed the RT-PCR results. The relative expression level of intact transcripts from the *bglu* lines or the *2x35S:BGLU* lines compared to the wild type was determined by the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001). With this method, the wild type expression level is set at 1 and down-regulation is indicated by fold changes between 0 and 1 and up-regulation by all fold changes above 1. The results are shown in figure 15. For visualization, the y-axis of the plot was broken after the fold change of 70, compressing the SE of the overexpression samples of *BGLU1* and *BGLU4* between the relative expression of 70 and 145. *ACTIN 2* was discarded from the analysis, due to C_t values in the negative control, differing less than ten cycles from the samples C_t values. Even, if the melting curves indicate another PCR product (data not shown), according to the manufacturer's instructions, fluorescence of other PCR products, appearing ten cycles or less before or after the product of interest, could interfere with the fluorescence of the PCR product of interest (Thermo Fisher, 2016). Slight contaminations in the negative control of reference genes can appear very abundantly, due to frequent applications of these genes in the laboratory. Such slight contaminations, differing more than ten cycles from the product of interest, were counted as minor influence and were ignored.

For *BGLU1* the fold change of the relative gene expression in *2x35S:BGLU1* is 67.2721 and 0.0423 in *bglu1-1*. For *BGLU3* the fold change of the relative gene expression in *2x35S:BGLU3* is 25.1242 and 0.0012 in *bglu3-2*. The fold change of the relative *BGLU4* gene expression in *2x35S:BGLU4* is 35.7993 and 0.0005 in *bglu4-2*.

The strong down-regulation of the intact *BGLU* gene in the *bglu* mutants, leads to the assumption of either very low or no translation of the encoded protein. Thus, very low levels of the enzyme's product metabolite (if even measurable) compared to the wild type, are expected. For the *2x35S:BGLU* lines a strong up-regulation of the respective *BGLU* gene was measured. This leads to the assumption of enhanced translation of the encoded protein, compared to the wild type, and therefore enhanced enzyme activity and higher levels of the

enzyme's product metabolite, compared to the wild type. With respect to the RT-PCR results the expression of *BGLU3* in *2x35S:BGLU3* is compareably high in relation to the wild type. Evaluation of these results should keep in mind, that the seeds for the qRT-PCR results are 12 month older, than the ones of the RT-PCR. The relative expression for *BGLU3* could have changed, if *BGLU3* expression is decreasing in the wild type in older seeds. In general, the results clearly show the expected tendency of higher gene expression in *2x35S:BGLU* lines, compared to the wild type and lower or no gene expression in the *bglu* lines, compared to the wild type.
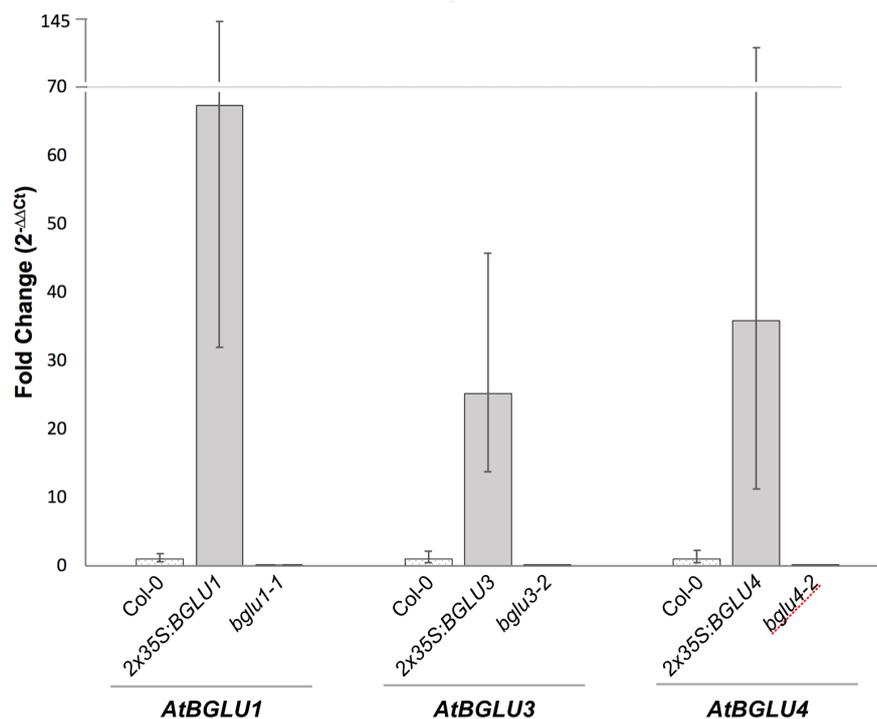
Figure 15 – *qRT-PCR results of differential* **BGLU** *gene expression. The bars visualize the differential* BGLU *gene expression of the intact transcripts of* BGLU1, BGLU3 *and* BGLU4 *in the respective* 2x35S:BGLU *and* bglu *lines, compared to the wild type. The error bars indicate the asymmetrically distributed, cumulative standard error (SE). The y-axis was broken after the fold change of 70, compressing the SE of the overexpression samples of* BGLU1 *and* BGLU4 *between the relative expression of 70 and 145.*

3.3. Candidate product and substrate features

Untargeted metabolic fingerprinting of loss-of-function T-DNA insertion and overexpression mutants as well as the wild type was applied, to search for features, showing differantial peaks in the respective *2x35S:BGLU* and *bglu* line, compared to the wild type, accounting for potential product or substrate features of the encoded protein. Therefore, fold changes for *BGLU1*, *BGLU3* and *BGLU4* from peak intensities of *2x35S:BGLU* or *bglu* lines type were calculated from the processed features of untargeted metabolic fingerprinting, compared to the wild. For *BGLU1* from 3316 unprocessed features, 775 features remained after processing for fold change calculations, with 666 features of *2x35S:BGLU1* and 677 features of *bglu1-1*. For *BGLU3* untargeted metabolic fingerprinting revealed 6209 features, which were reduced to 1306 features after processing, with 1193 features of *2x35S:BGLU3* and 1127 features of

bglu3-2 for fold change calculations. 4374 unprocessed features for *BGLU4* were reduced to 1286 processed features for fold change calculations, with 1181 features for *2x35S:BGLU4* and 1148 features for *bglu4-2.*
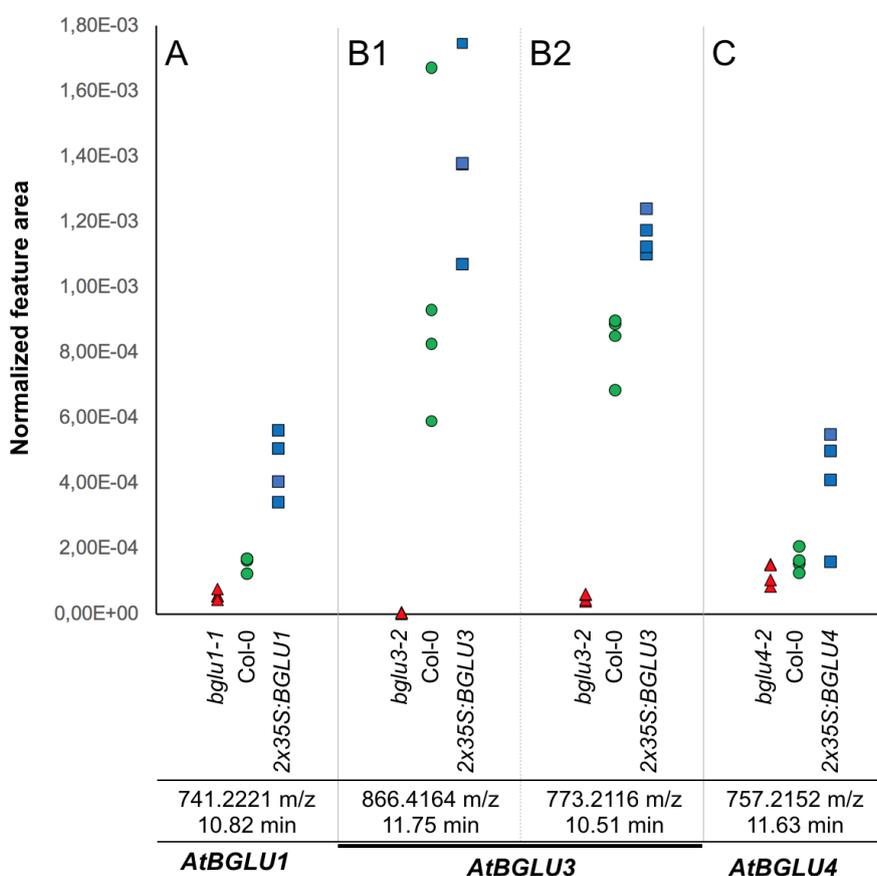
Figure 16 – *Main candidate product features. Peaks of the main candidate product features for* BGLU1, BGLU3 *and* BGLU4*, visualized by their normalized, manually integrated feature area. Peaks of* bglu *samples are indicated by red triangles, peaks of* 2x35S:BGLU *samples by blue rectangles and peaks of the wild type samples by green circles. In A, the main candidate product feature for* BGLU1 *is depicted, in B1 and B2 the main candidate product features for* BGLU3 *and in C the main candidate product feature for* BGLU4.

Fold changes were manually investigated for features with higher *2x35S:BGLU* peaks and simultaniously lower or absent *bglu* peaks, expected to derive from up- or downregulation of the respective gene, for candidate product compounds or, the other way around, for candidate substrate compounds. A treshold for peaks, possibly deriving from genetic upregulation of the encoded protein, was set to a peak fold change of 1.5 or higher. Peaks expected to derive from genetic downregulation, had a fold change treshold of 0.5 or lower. All candidate features for possible product or substrate compounds, were in addition checked with chromatogram data and compared to the RT- and qRT-PCR results. Features with most concordant fold changes and expression patterns were preferred as candidate product or substrate features. In addition, bigger peaks were preferred over smaller peaks, if multiple features came into question. In cases of qualitative fold changes (no peaks picked for *2x35S:BGLU*, *bglu* or wild type with the algorithm) but still very small peaks, visible in the chromatogram for the main candidate features, manual integration of the peak area was performed and fold changes were calcualted from these data. Manually integrated and normalized peak areas were chosen

for visualization of the main candidate features, while small peaks were visualized with the enhanced line spectra (figure 16 and 17).
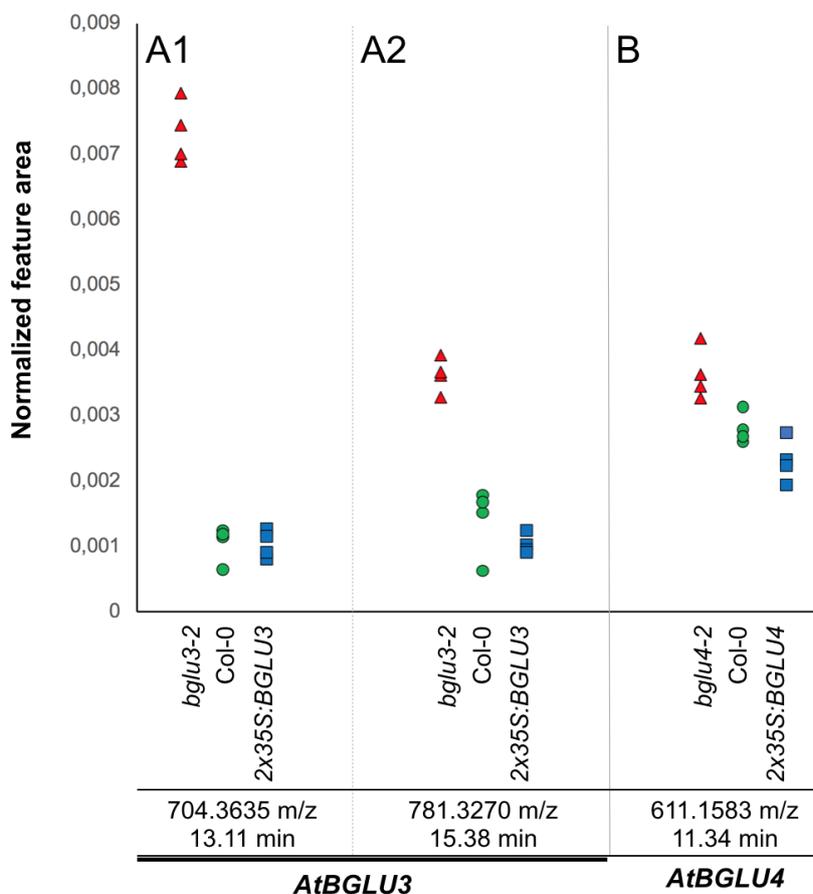
Figure 17 – *Main candidate substrate features. Peaks of the main candidate substrate features for* BGLU3 *and* BGLU4*, visualized by their normalized, manually integrated feature area. Peaks of* bglu *samples are indicated by red triangles, peaks of* 2x35S:BGLU *samples by blue rectangles and peaks of the wild type samples by green circles. In A1 and A2, the main candidate substrate features for* BGLU3 *are depicted, in B the main candidate product feature for* BGLU4*.*

In total for *BGLU1* and *BGLU4* one main candidate product feature was found, respectively (figure 16A and 16C) and for *BGLU4* in addition one main candidate substrate feature (figure 17B). For *BGLU1* no candidate substrate feature was found. For *BGLU3* two main candidate product features (figure 16B1 and 16B2) and two main candidate substrate features (figure 17A1 and 17A2) were found.

For *BGLU1* no candidate product feature lacking *bglu1-1* peaks, could be found. The main product feature of interest is 741.2221 m/z at 10.82 min (figure 16A), exhibiting distinct higher *2x35S:BGLU1* peaks compared to the wild type, with a fold change of 5.9967. Peaks for *bglu1-1* were not picked with the algorithm but visible in the chromatogram. Manual integration of the enhanced peak area resulted in fold changes of 2.9483 for *2x35S:BGLU1* and 0.3768 for *bglu1-1*. To investigate functional redundancy by homolog *BGLUs* in future experiments, a heterozygous double mutant of *bglu1-1* and *bglu5-1* and a homozygous double mutant for *bglu1-1* and *bglu3-2* was successfully generated (data not shown). To confirm the differential metabolite levels of 741 m/z by a third mutant line, plasmids for complementation

of genomic *BGLU1* in *bglu1-1* were successfully cloned with and without a GFP-tag for the generation of complementation mutants. Figures of the plasmids are given in the supplements (figure S8 - S11).

Table 7 – **Substitutions on 866 m/z.** *Possible substitutions on 866 m/z, resulting in the masses of other double charged candidate product compounds are listed. For calculations, it was assumed, that the feature of the double charged compound results from the $[M+H]^+$ ion. Apart from this, they could also result from other adduct types.*

Feature	Neutral mass [Da]	Single charged mass [m/z]	Double charged mass [m/z]	Neutral mass [Da]	Functional group
866.4164 m/z 11.75 min	865.4	866.4	433.7	0	starting compound
432.7049 m/z 14.62 min	863.4	864.4	432.7	2	-2 H (deprotonation)
411.7178 m/z 11.87 min	821.4	822.4	411.7	44	$-CO_2$ (decarboxylation)
440.7207 m/z 12.45 min	879.4	880.4	440.7	+14	$+CH_2$ (methylation)
440.7209 m/z 12.90 min	879.4	880.4	440.7	+14	$+CH_2$
441.7099 m/z 9.45 min	881.4	882.4	441.7	+16	+O (hydroxylation)
441.7102 m/z 9.96 min	881.4	882.4	441.7	+16	+O
448.7181 m/z 11.05 min	895.4	896.4	448.7	+30	+O, $+CH_2$
536.7408 m/z 17.07 min	1071.4	1072.4	536.7	+ 206 Da	not known
460.1686 m/z 11.76 min	918.2	919.2	460.1	+ 52.8	+ unknown adduct (UA)
467.1764 m/z 12.89 min	932.2	933.2	467.1	+52.8, +14	+ UA, $+CH_2$
494.1968 m/z 19.11 min	986.2	987.2	494.1	+120.8	+ UA, $+CH_2$, + O, + $2H_2O$, +2H

For *BGLU3* most candidate product features showed only slightly higher fold changes than 1 for *2x35S:BGLU3* peaks. Therefore, not only features with FC $\geq$ 1.5 were chosen for this mutant. For *bglu3-2* the candidate product features show no or only very low peaks. All candidate product features were compared with the soaked *BGLU3* seed samples for accordance (data not shown). This revealed several candidate product features, leaving after further investigations two main candidate features, which are 866.4164 m/z at 11.75 min (figure 16B1) and 773.2116 m/z at 10.51 min (figure 16B2). For the first feature the fold change was 1.1574 for *2x35S:BGLU3* and qualitative (qual_low) for *bglu3-2*. After manual integration the fold changes were 1.3889 for *2x35S:BGLU3* and 0.0023 for *bglu3-2*. This feature was predominant as double charged compound (feature 433.7128 m/z at 11.76 min, fold changes table 8). For the 773 m/z feature the fold changes were 1.1711 for *2x35S:BGLU3* and 0.0634 for *bglu3-2* and 1.4001 for *2x35S:BGLU3* and 0.0608 for *bglu3-2* after manual integration.

The remaining features with candidate product peak pattern include mainly double-charged compounds (441.7099 m/z at 9.45 min, 441.7102 m/z at 9.96 min, 448.7181 m/z at 11.05 min, 460.1686 m/z at 11.76 min, 411.7178 m/z at 11.87 min, 440.7207 m/z at 12.45 min, 440.7209 m/z at 12.90 min, 432.7049 m/z at 14.62 min, 467.1764 m/z at 12.89 min, 494.1968 m/z at 19.11 min and 536.7408 m/z at 17.07 min), probably deriving by methylation, carboxylation or hydroxylation from the main candidate feature 866 m/z (explained by common neutral masses), while some of them elute as a couple at the same retention time, exhibiting a neutral mass difference of 52.8 Da, probably speaking for an unknown adduct type (table 7). Almost all of these double charged features show only low peaks and occur in no or just very little amounts as single-charged compounds. Thus, they were not counted as the main compound of interest. The fold changes are given in table 8.

Apart from the double charged peaks, there exist two minor single charged candidate product features, 757.2166 m/z at 11.42 min with the fold changes of 1.2009 for *2x35S:BGLU3* and 0.4874 for *bglu3-2* as well as 757.2152 m/z at 11.63 min with the fold change 4.4796 for *2x35S:BGLU3* and qual_low *bglu3-2*, or 1.3279 (*2x35S:BGLU3*) and 0.12104 (*bglu3-2*) after manual integration. In relation to a main candidate substrate feature (see below) the feature 943.3804 m/z at 13.74 min seems to be a potential candidate product feature, despite the unexpected fold cange of 0.6182 for *2x35S:BGLU3* (0.7346 after manual integration). *Bglu3-2* shows no peaks for this feature.

The first main candidate substrate feature for *BGLU3* is 704.3635 m/z at 13.11 min (figure 17A1), also predominant as a double charged compound (feature 330.6915 m/z at 13.11 min, fold changes table 8). The fold change is 6.4811 for *bglu3-2* and 0.9842 for *2x35S:BGLU3*, as well as 7.0455 for *bglu3-2* and 0.9868 for *2x35S:BGLU3* after manual integration. The second main candidate substrate feature is 781.3270 m/z at 15.38 min (figure 17A2) with the fold changes 2.8970 (*bglu3-2*) and 0.7732 (*2x35S:BGLU3*), as well as 2.6131 (*bglu3-2*) and 0.7346 (*2x35S:BGLU3*) after manual integration.

Apart from these two main candidate features, additional minor candidate substrate features were found, which are 660.3742 m/z at 13.21 min, 402.1451 m/z at 13.11 min, 619.2742 m/z at

17.37 min, 575.2851 m/z at 18.19 min, 287.1382 m/z at 7.25 min, 633.2899 m/z at 19.68 min and 795.3419 m/z at 17.03 min as the most interesting ones. The feature 402 m/z elutes at the same time as the feature 704 m/z and could be an adduct of this one.

To confirm the differential metabolite levels of the BGLU3 candidate substrate and product compounds by a third mutant line, plasmids for complementation of genomic *BGLU3* in *bglu3-2* were successfully cloned with and without a GFP-tag for the generation of complementation mutants. Figures of the plasmids are given in the supplements (figure S8 - S11).

As for *BGLU1*, the fold changes of *BGLU4* did not reveal a candidate product feature, lacking the *bglu4-2* peaks. The main candidate feature 757.2149 m/z at 11.71 min (figure 16C) is most concordant to the RT- and qRT-PCR results, exhibiting a qualitative fold change (qual_high) for *2x35S:BGLU4* and also no picked peaks for *bglu4-2*, due to low peak intensities. After manual integration of the enhanced peak area the fold changes were 2.5331 (*2x35S:BGLU4*) and 0.7701 (*bglu4-2*). Two further minor candidate product features were 529.2079 m/z at 27.45 min and 507.1473 m/z at 14.37min, which are less concordant with the RT- and qRT-PCR results. The fold changes are listed in table 8.

Based on the results of the main candidate product feature, two possible substrate features were found in the data and analyzed in more detail, which are 611.1583 m/z at 11.34 min as the main one (figure 17B), with fold changes of 1.3567 (*bglu4-2*) and 0.7867 (*2x35S:BGLU4*) or 1.3058 (*bglu4-2*) and 0.8263 (*2x35S:BGLU4*) after manual integration and 611.1584 m/z at 8.75 min with fold changes of 1.4339 (*bglu4-2*) and 0.7782 (*2x35S:BGLU4*).

To investigate functional redundancy by homolog *BGLUs* in future experiments, a homozygous double mutant for *bglu4-2* and *bglu3-2* was successfully generated (data not shown). To confirm the differential metabolite levels of the candidate substrate and product compound of BGLU4 by a third mutant line, plasmids for complementation of genomic *BGLU4* in *bglu4-2* were successfully cloned with and without a GFP-tag for the generation of complementation mutants. Figures of the plasmids are given in the supplements (figure S8 - S11).

All candidate features and the respective fold changes are listed in table 8. The peak intensities adjusted to the internal standard from the metabolic fingerprint are listed in table S3 - S5 in the supplements.

Table 8 – ***Fold changes of candidate features.*** *Fold changes of the normalized peak intensities of all candidate features from the metabolic fingerprint are listed, as well as the fold changes from the normalized area, if available. For the features 741.2221 m/z 10.82 min and 757.2149 m/z 11.71 min, the area of the enhanced peaks were chosen for fold change calculations.* qual *represents a qualitative fold change.*

Feature	*2x35S:BGLU* fingerprint	*bglu* fingerprint	*2x35S:BGLU* area	*bglu* area
741.2221 m/z 10.82 min	5.9967	-	2.9483	0.3768
866.4164 m/z 11.75 min	1.1574	qual_low	1.3889	0.0.0023
773.2116 m/z 10.51 min	1.1711	0.0634	1.4001	0.0608
433.7128 m/z 11.76 min	1.1959	qual_low	1.2007	0.0016

continued next page

Table 8 – continued

Feature	2x35S:BGLU fingerprint	bglu fingerprint	2x35S:BGLU area	bglu area
441.7099 m/z 9.45 min	1.1679	qual_low	-	-
441.7102 m/z 9.96 min	1.1902	qual_low	-	-
448.7181 m/z 11.05 min	1.1322	qual_low	-	-
460.1686 m/z 11.76 min	1.2287	qual_low	-	-
411.7178 m/z 11.87 min	1.2141	qual_low	-	-
440.7207 m/z 12.45 min	0.9470	qual_low	-	-
440.7209 m/z 12.90 min	1.3719	0.0702	-	-
432.7049 m/z 14.62 min	3.1654	qual_low	-	-
467.1764 m/z 12.89 min	1.6617	qual_low	-	-
494.1968 m/z 19.11 min	1.5560	qual_low	-	-
536.7408 m/z 17.07 min	4.0852	2.4988	-	-
757.2166 m/z 11.42 min	1.2009	0.4874	-	-
757.2152 m/z 11.63 min	4.4796	-	1.3279	0.12104
943.3804 m/z 13.74 min	0.6182	qual_low	0.7346	0
704.3635 m/z 13.11 min	0.9842	6.4811	0.9868	7.0455
330.6915 m/z 13.11 min	0.9960	5.0214	0.9738	5.0267
781.3270 m/z 15.38 min	0.7732	2.8970	0.7346	2.6131
660.3742 m/z 13.21 min	0.8468	4.0721	-	-
402.1451 m/z 13.11 min	0.6262	4.0739	-	-
619.2742 m/z 17.37 min	0.8064	1.5352	-	-
575.2851 m/z 18.19 min	0.8926	1.6637	-	-
287.1382 m/z 7.25 min	-	6.7767	-	-
633.2899 m/z 19.68 min	0.9749	1.9358	-	-
795.3419 m/z 17.03 min	-	3.18403	-	-
757.2149 m/z 11.71 min	qual_high	-	2.5331	0.7701
529.2079 m/z 27.45 min	2.3515	-	-	-
507.1473 m/z 14.37min	1.7563	0.3143	-	-
611.1584 m/z 8.75 min	0.7782	1.4339	-	-
611.1583 m/z 11.34 min	0.7867	1.3567	0.8263	1.3058

3.4. Identification of the main BGLU1 candidate product feature

The candidate *BGLU1* product feature 741.2221 m/z at 10.82 min exhibit only small peaks, which were not fragmented in the auto-MS/MS mode. Thus, targeted fragmentation in the MRM mode was performed (figure 18).
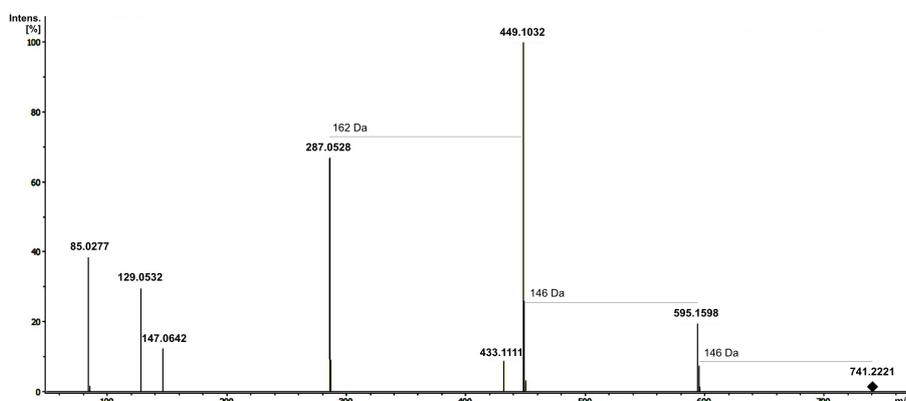
Figure 18 – ***MS/MS spectrum of feature 741.2221 m/z at 10.82 min.*** *The MS/MS spectrum of the main BGLU1 candidate product feature 741.2221 m/z at 10.82 min. The characteristic fragment ion peaks are 595 m/z with a neutral loss of 146 Da, 449 m/z with a neutral loss of 146 Da and 287 m/z with a neutral loss of 162 Da. The precursor ion is indicated by a rhombus.*

The resulting MS/MS spectrum shows apart from the precursor ion of 741 m/z, the three dominant fragment ions 287 m/z, 449 m/z and 595 m/z. The neutral losses between the fragments are 146 Da (741 m/z – 595 m/z), 146 Da (595 m/z – 449 m/z) and 162 Da (449 m/z – 287 m/z), speaking assumably for a hexose, and two deoxyhexoses attached to the aglycone. In the positive ion mode 287 m/z is a characteristic fragment ion peak for either a kaempferol (flavonol) or cyanidin (anthocyanidin) backbone. Attempts to distinguish both aglycones were performed with measurements in the negative ion mode (ESI$^-$) (Sun *et al.*, 2012). The MS spectrum for ESI$^-$ did only reveal a dominant [M–2H]$^-$ peak (figure 19A). The MS/MS spectrum for ESI$^-$ did only show one fragment ion of 577 m/z, apart from the precursor ion, with a neutral loss of 162 Da (figure 19B).

The results in the positive ion mode are in concordance with the results of Wu *et al.* (2018), where, like our case, no identification of the compound with public databases through MetFrag was possible. The best two hits in PubChem (Kim *et al.*, 2019) are a 3-[6-*O*-(3-*O*-α-L-rhamnopyranosyl-α-L-rhamnopyranosyl)-β-D-glucopyranosyloxy]-4',5,7-trihydroxyflavone and a kaempferol 3-*O*-[α-L-rhamnopyranosyl(1→2)-β-D-galactopyranosyl]-7-*O*-α-L-rhamnopyranoside. For both compounds a different fragmentation pattern is expected with a dominant 433 m/z fragment ion. Nevertheless, these hits indicate, that the candidate product compound could be a glycosylated kaempferol derivative. The UV-vis spectrum exhibits two peaks, at 226 nm and 343 nm, underlining the hint on a glycosylated flavonol (Lin and Harnly, 2007). Quispe *et al.* (2013) found a compound with the same MS/MS fragmentation pattern and suggest a kaempferol 3-(2"-rhamnosyl rutinoside). All these hints need to be further investigated and validated.
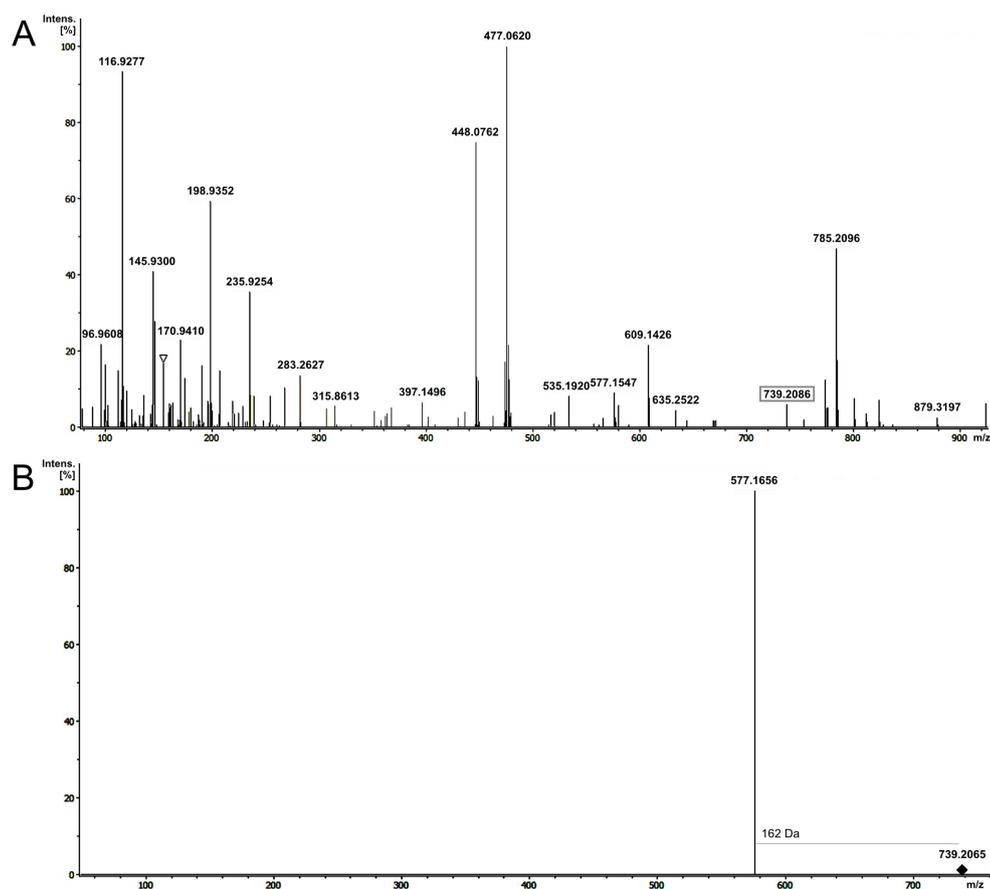
Figure 19 – *Candidate* **BGLU1** *feature with ESI⁻*. *The MS spectrum of 739.2065 m/z at 10.82 min (A) shows only the [M–2H]⁻ peak and no [M–2H + H₂O]⁻ peak. The MS/MS spectrum (B) shows apart from the precursor ion only one fragment ion peak of 577 m/z, with a neutral loss of 162 Da. The precursor ion is indicated by a rhombus.*

3.5. Identification of BGLU3 candidate product and substrate features

3.5.1. Candidate product feature 866.4164 m/z at 11.75 min and candidate substrate feature 704.3635 m/z at 13.11 min

The MS/MS spectrum of the main BGLU3 candidate product feature 866.4164 m/z at 11.75 min (figure 20A) shows apart from the precursor ion 866 m/z the main fragment ion 822 m/z. The neutral loss of 44 Da (866 m/z – 822 m/z) could account for the loss of a CO_2 group, thus a compound with a very fragile carboxyl group is likely.
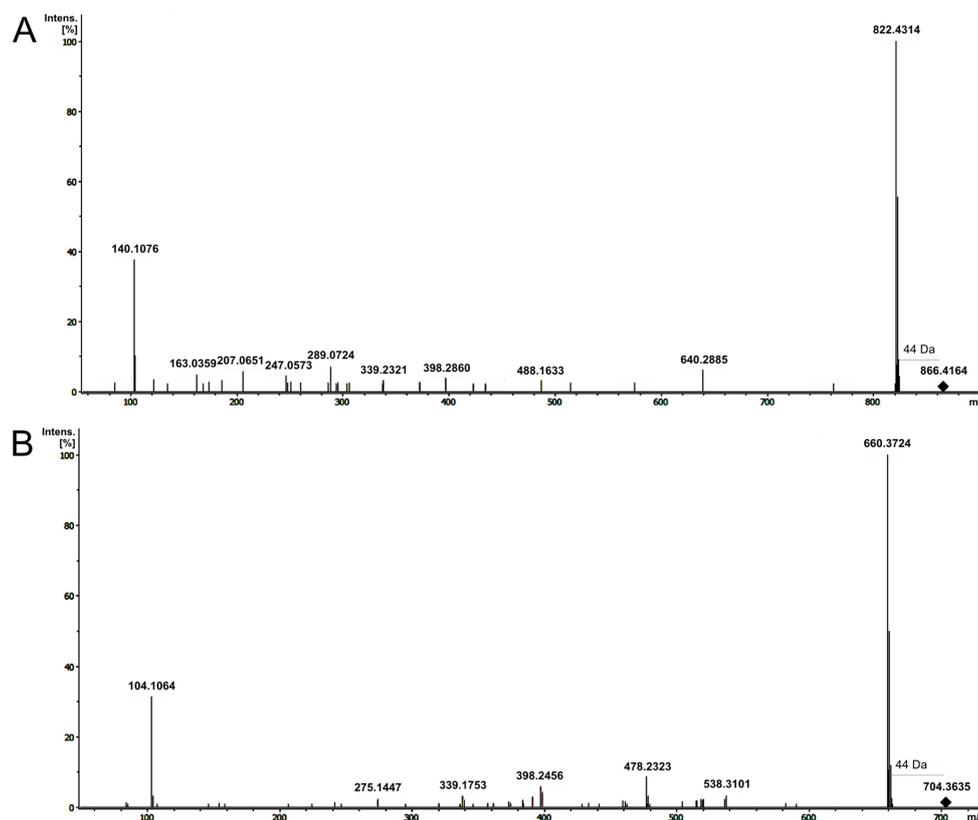
Figure 20 – *Candidate* **BGLU3** *features 866.4164 m/z at 11.75 min and 704.3635 m/z at 13.11 min. The MS/MS spectrum of the main BGLU3 candidate product feature 866.4164 m/z at 11.75 min (A) shows the characteristic main fragment ion 822 m/z with a neutral loss of 44 Da. The main BGLU3 candidate substrate feature 704.3635 m/z at 13.11 min is 162 Da smaller than the candidate product feature 866 m/z and shows a similar MS/MS fragmentation (B), with the characteristic main fragment ion 660 m/z with a neutral loss of 44 Da. The respective precursor ions are indicated by a rhombus.*

The second most dominant fragment ion is 104 m/z, which gives at first no additional information. In addition, there are several smaller fragment ions, namely 163 m/z, 207 m/z, 247 m/z, 289 m/z, 339 m/z, 398 m/z, 488 m/z and 640 m/z as the biggest ones. The 163 m/z peak could derive from a charged hexose group. A 289 m/z peak is known from fragmentation of oligomeric proanthocyanidins, if the upper unit is an (epi)catechin. In addition, the mass 866 m/z is just 1 Da smaller than a procyanidin trimer. The other frament ions, as well as the respective neutral losses between those peaks provide at a first glance no further compound information and attempts to identify the compound with MetFrag, did not lead to any convincing results.
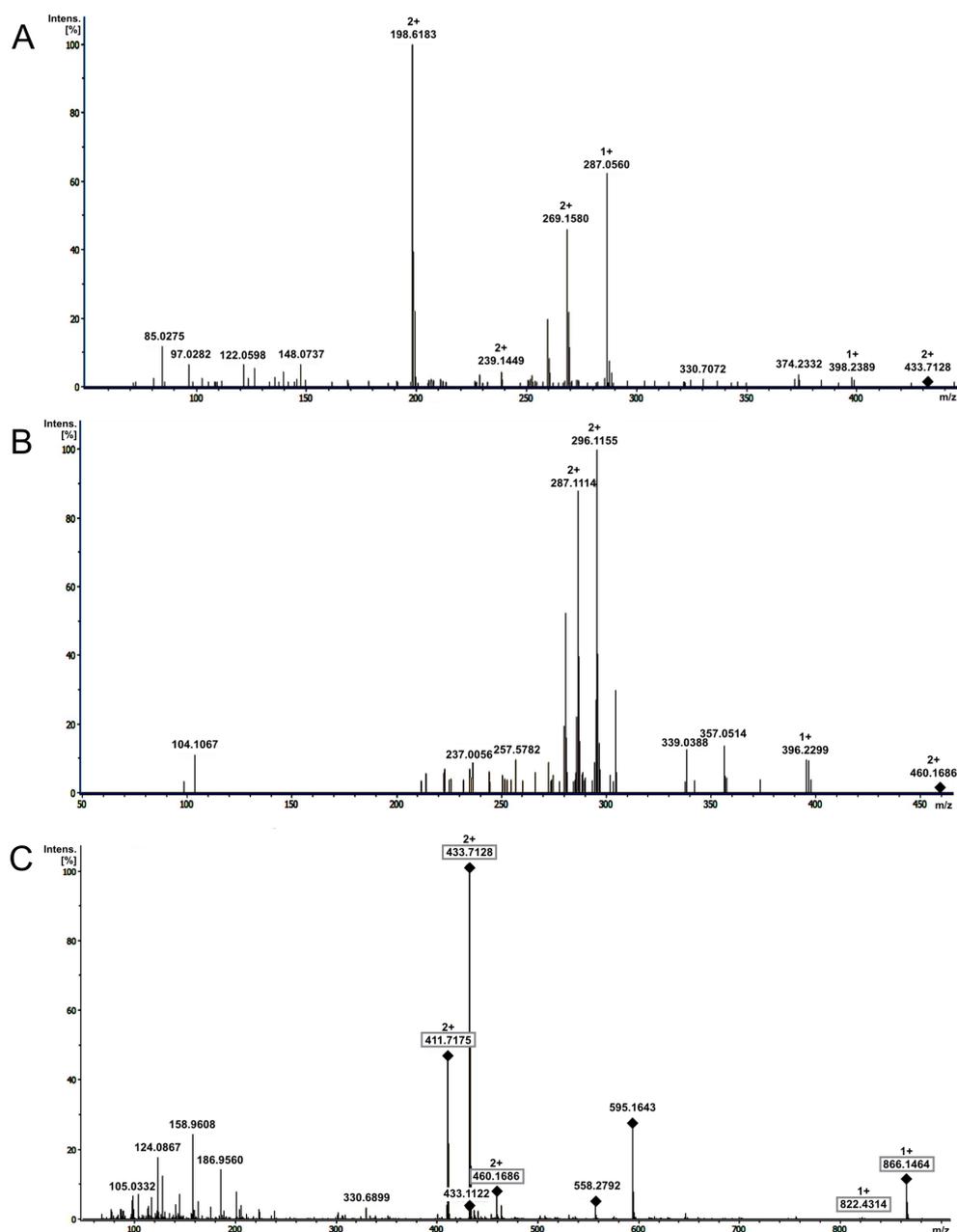
Figure 21 – *Candidate **BGLU3** features eluting with 866.4164 m/z. The MS/MS spectrum of the double charged BGLU3 candidate product feature 433.7128 m/z at 11.76 min is shown in A, exhibiting also the single charged fragment ion 398 m/z. The MS/MS spectrum of the double charged BGLU3 candidate product feature 460.1686 m/z at 11.76 min, eluting at the same time, is shown in B. The MS spectrum (C) shows all peaks, belonging to the candidate product feature 866.4164 m/z at 11.75 min (depicted with rectangels). All peaks, with known charge are indicated by 1+ (single-charged) or 2+ (double-charged). Precursor ions are indicated by a rhombus.*

Targeted fragmentation at different collision energies (60, 50, 40, 25 and 15 eV) should enhance information of the MS/MS spectrum. Characteristic fragment ions (table 9) for lower collision energies are, apart from the precursor ion 866 m/z, once more 289 m/z and 822 m/z and in addition 147 m/z, 575 m/z and 721 m/z, as the main ones. Again, the 575 m/z peak, differs only slightly from an oligomeric proanthocyanidin, in this a procyanidin A-type (2 Da less) or B-type (4 Da less) dimer. In addition, the predominance as a double-charged compound, is a characteristic trait for higher molecular mass proanthocyanidins, mainly found in seeds.

Table 9 – ***Targeted fragmentation of 866 m/z.*** *Targeted fragmentation of 866.4164 m/z was performed at different collision energies in the MRM mode. All main fragment ions are listed for each chosen collision energy. The 866 m/z precursor ion is indicated in bold.*

Collision energy [eV]	Main fragment ions [m/z]
15	**866**, 822, 721, 575, 289
25	**866**, 822, 721, *703*, 575, 289, 147
40	**866**, 822, 640, 575, 398, 339, 104
50	**866**, 822, 700, 640, 520, 478, 398, 374, 339, 251, 104
60	**866**, 822, 700, 640, 520, 478, 398, 374, 339, 251, 211, 104

Higher collision energies (table 9) yielded a smaller 822 m/z fragment ion and higher 339 m/z and 398 m/z fragment ions, as well as again a high 104 m/z and a small 640 m/z fragment ion, suggesting 339 m/z and 398 m/z to be characteristic fragment ions of the compound. Additional fragments were 251 m/z, 374 m/z, 478 m/z, 520 m/z and 700 m/z. The respective neutral losses between the fragment ion peaks and the peaks itselves are mostly not reported for oligomeric proanthocyanidins and no common neutral losses for oligomeric proanthocyanidins, as defined by Li and Deinzer (2007), are present. Nevertheless, the compound shares some characteristics with oligomeric proanthocyanidins, suggesting the presence of another oligomeric compound, like condensed compounds of flavanols and anthocyanidins (F–A$^+$ or A$^+$–F). Applying the fragmentation rules of Sánchez-Ilárduya *et al.* (2012) for condensed flavanols and anthocyanidins (RDA fission, HRF fission and a partial fragmentation of the upper unit), the 640 m/z fragment ion can result from RDA fission of an upper (epi)gallocatechin unit (-168 Da), if also a CO_2 loss (-44 Da) and a loss of a methyl group (-14 Da) happens. The 398 m/z fragment ion peak could derive from the characteristic partial fragmentation of an upper (epi)gallocatechin unit (-262 Da), together with a CO_2 (-44 Da) and a hexose (-162 Da) loss. No characteristic HRF fission (-126 Da) for an upper flavanol unit was found. A protonated (epi)gallocatechin unit would exhibit the mass of 307 m/z or 306 m/z if condensed with another flavonoid unit through a C–C bond. The neutral loss of 306 Da with a neutral loss of hexose (-162 Da) yields 398 m/z, too. Thus this fragment ion could also be characteristic for a naturally charged anthocyanidin unit, lacking a hexose unit and exhibiting as a free aglycone one more hydrogen, thus a 399 m/z. No anthocyanidin of the mass 399 m/z is known so far, and in addition anthocyanidins are not known to contain fragile carboxyl groups.
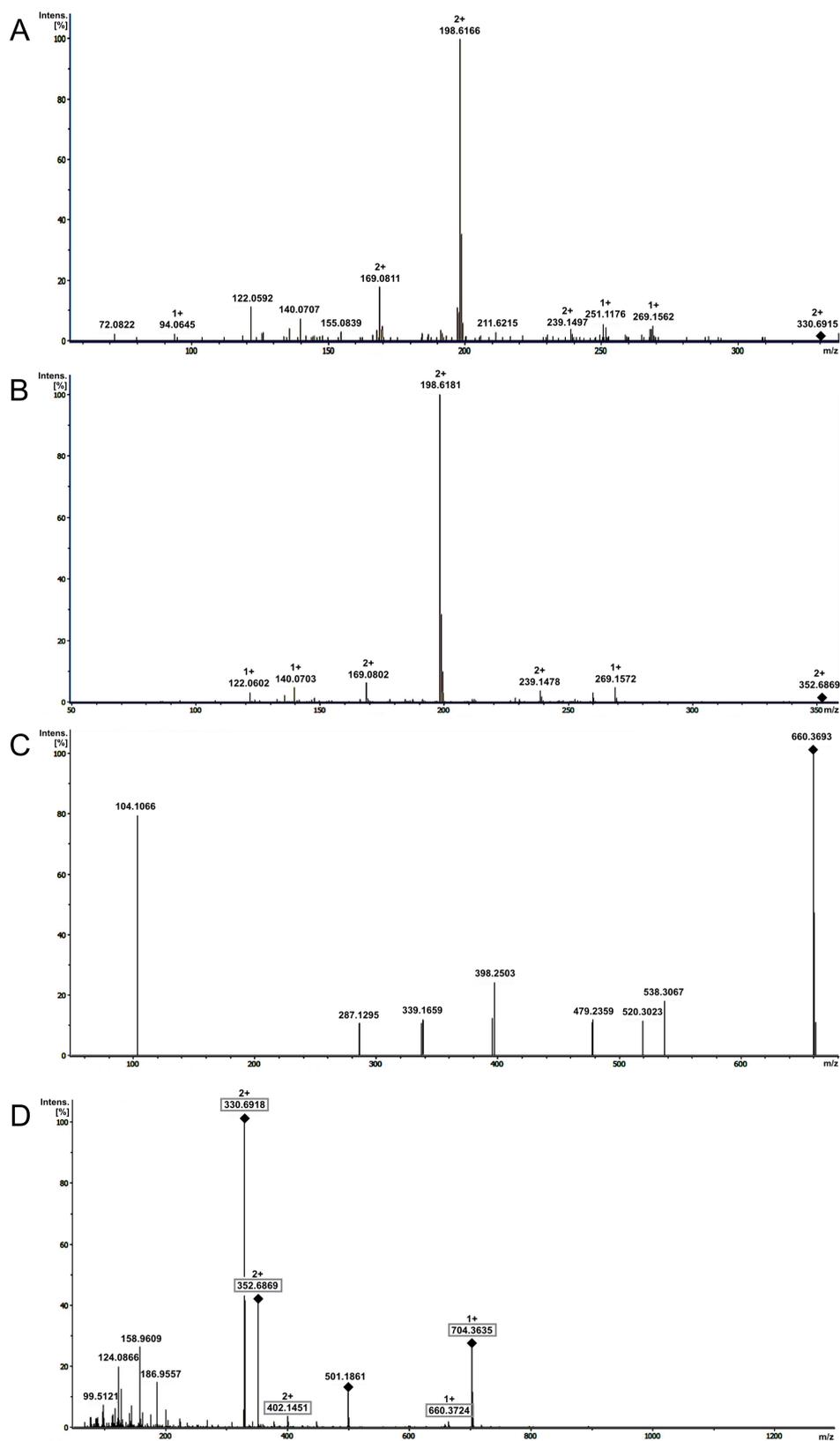
Figure 22 – **Candidate BGLU3 *features eluting with 704.3635 m/z at 13.11 min.*** *The MS/MS spectrum of the double-charged* BGLU3 *candidate substrate feature 330.6915 m/z at 13.11 min is shown in A. The MS/MS of 352.6869 m/z ([M+H]$^{2+}$) from the same bucket is shown in B. The MS/MS in C derives from the single-charged [M-CO$_2$+H]$^+$ adduct, also dominant as the double-charged compound 330 m/z ([M-CO$_2$ +H]$^{2+}$). The MS spectrum (D) shows all peaks, belonging to the candidate substrate feature 704.3635 m/z at 13.11 min (depicted with rectangels). All peaks, with known charge are indicated by 1+ (single-charged) or 2+ (double-charged). Precursor ions are indicated by a rhombus.*

Instead anthocyanidin-derived pyranoanthocyanidins are known for these characteristics. A protonated vitisin A ion exhibits a mass of 561 m/z, fragmenting into 399 m/z after the neutral loss of a hexose (vitisidin A), as well as a fragile carboxyl group. Therefore, the characterisitic fragment ions of the main BGLU3 candidate product compound a proposed to derive from an (epi)gallocatechin–carboxypyranomalvidin-3-*O*-glucoside (figure 23). Therefore, the remaining fragment ions of the auto-MS/MS fragmentation are explained as followed. 289 m/z can result from a vitisin A loss and an additional oxygen loss (-16 Da) at the (epi)gallocatechin unit. 247 m/z can derive from the charged partial 262 Da loss (described above), with an additional oxygen loss (262 Da + H - 16 Da). 339 m/z can derive from an (epi)gallocatechin loss, a CO_2 loss, a hexose loss and a radical methyl loss (866 m/z - 306 Da - 44 Da - 162 Da - 15 Da). 448 m/z can derive from the loss of (epi)gallocatechin and a partial ring B loss of vitisin A. 207 m/z can derive from a vitisin A loss and a partial ring B loss of (epi)gallocatechin. 104 m/z is characteristic for auto-MS/MS fragmentation and MRM fragmentation with higher collision energies. The fragmentation is not completely clear. The main known fission points for MS/MS fragmentation are given in figure 24.
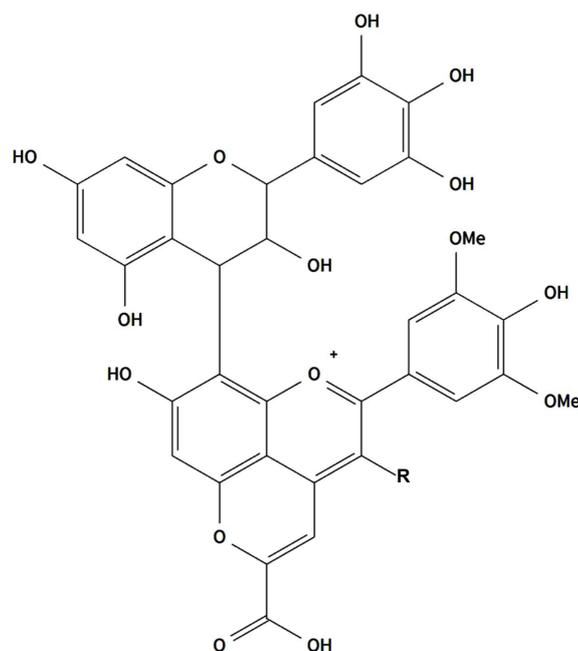
Figure 23 – *Predicted structure for 866 m/z and 704 m/z. 866 m/z is predicted to be an (epi)gallocatechin–carboxypyranomalvidin-3-O-glucoside (R: O-glucose) as the main product compound of BGLU3. In relation to this, 704 m/z is predicted to be the respective substrate compound, thus an (epi)gallocatechin–carboxypyranomalvidin (R: OH).*

All these results are in concordance with the candidate BGLU3 substrate feature 704.3635 m/z at 13.11 min, probably an (epi)gallocatechin–carboxypyranomalvidin (figure 23). The compound is 162 Da smaller, than the 866 m/z candidate product compound, predominant as the double-charged compound and accumulates in *blgu3-2*, where BGLU3 is supposed to be nonfunctional. The MS/MS spectrum (figure 20B) resembles the one of 866 m/z with one main fragment ion of 660 m/z from a 44 Da CO_2 loss (704 m/z – 660 m/z). In addition, the main fragment ions 104 m/z, 339 m/z, 398 m/z and 478 m/z are the same as described for 866 m/z. 478 m/z can account for the RDA fission of the upper (epi)gallocatechin unit

and 398 m/z for the vitisidin A unit or for the partial loss of (epi)gallocatechin (660 m/z - 262 Da). 339 m/z and 104 m/z are proposed as for 866 m/z. Further fragment ion peaks are 275 m/z and 538 m/z. 275 m/z can derive, from a vitisidin A loss with the second carbon from the C–C bond and an additional water loss (866 m/z - 573 Da - 18 Da). 538 m/z can derive from the 138 Da loss of ring B of the (epi)gallocatechin unit and two additional methyl losses from vitisidin A (704 m/z - 138 Da - 2x14 Da).

Additional information were gathered by the UV-vis spectrum, while the spectrum shows only very low peaks and is probably not completely clean, due to the small peak intensities and overlapping peaks. Two peaks were visible, for 866 m/z around 280 nm and 344 nm and for 704 m/z around 220 nm and 330 nm, the first one in concordance with flavanol absorption and the second one with pyranoanthocanidin absorption.
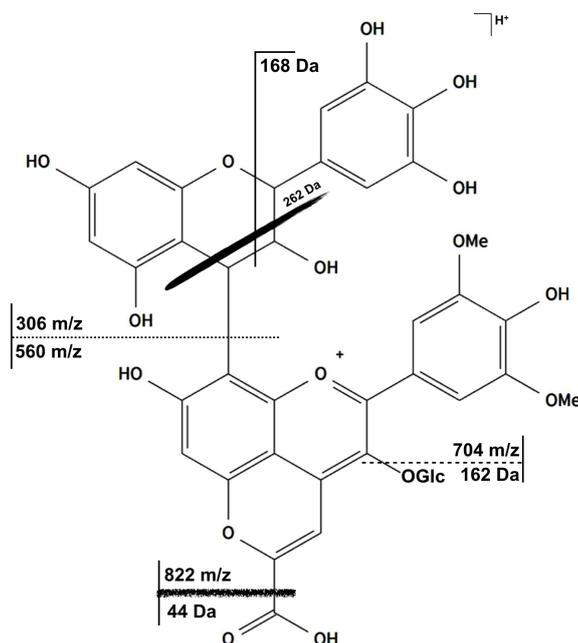
Figure 24 – ***Main MS/MS fragments of 866 m/z.*** *The main fission points for MS/MS fragmentation of 866 m/z are depicted. OGlc:* O-*glucose.*

The MS spectrum of 866 m/z (figure 21C) contains several peaks at the same retention time, with three of them belonging to features, depicted as candidate product features for *BGLU3*, namely the double charged 433 m/z peak and a double charged 411 m/z peak sorted in to the same bucket, as well as the double charged 460 m/z peak. To investigate, if these peaks belong all to the compound of 866 m/z or if they derive from other compounds, a measurement with an improved separation by a flat gradient of 58 min was performed, revealing, that all three peaks are not separated and seem to belong to the same compound of 866 m/z, while other compounds were separated (figure 25) and also chage their order. 433 m/z is the double-charged compound of 866 m/z ($[M+H]^{2+}$), while 411 m/z as double-charged compound ($[M-CO_2+H]^{+2}$) seems to account for the dominant 822 m/z fragment ion of 866 m/z, which is probably fragmenting in the quadrupole Q1 prior to targeted fragmentation and therefore treated as precursor ion, due to the fragile CO_2 group. This is supported by the small 822 m/z peak in the MS spectrum of 822 m/z. The compound of 460 m/z is supposed to be 52.8 Da

bigger than 866 m/z, thus could by an unknown adduct type of 866 m/z ($[M+52.8\ Da]^{2+}$) as mentioned above. The MS/MS spectra of 433 m/z and 460 m/z (figure 21) show both some single-charged fragment ion peaks as seen in the MS/MS spectrum of 866 m/z, speaking for the same compounds, too.
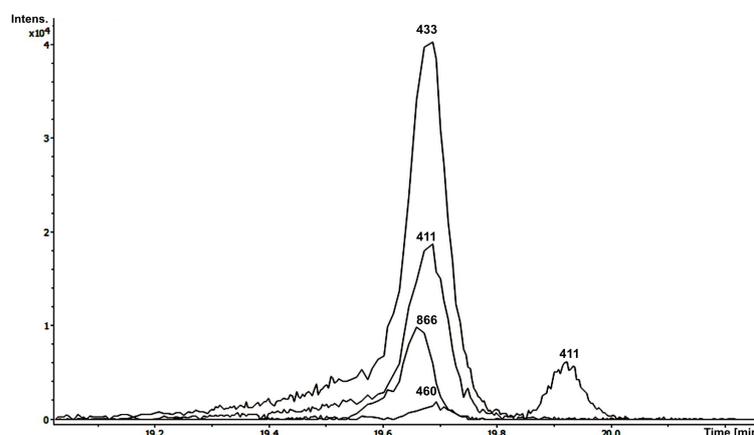
Figure 25 – *Flat gradient for features eluting with 866 m/z. The features 433.7128 m/z at 11.76 min (433), 866.4164 m/z at 11.75 min (866) and 460.1686 m/z at 11.76 min (460) as well as the first 411 m/z peak (411) elute at the same time point and are not seperated by the flat gradient of 58 min. Therefore, it is assumed, that these peaks belong to the same compound. In contrast the second 441 peak belongs to the feature 411.7178 m/z at 11.87 min.*

The MS spectrum of 704 m/z (figure 22D) revealed, like for 866 m/z the double charged $[M+H]^{2+}$ ion and a $[M-CO_2+H]^{+2}$ adduct. In this case the $[M-CO_2+H]^{+2}$ adduct is predominant (330 m/z) and the $[M+H]^{+2}$ ion (352 m/z) sorted into the bucket of the 330 m/z feature. The MS/MS spectra (figure 22A and 22B) clearly indicate the same compound. As single-charged compound the $[M+H]^+$ ion (704 m/z) is predominant, although, a peak for the $[M-CO_2+H]^+$ adduct exist, as already seen for 866 m/z. The respective MS/MS for this adduct (figure 22C) shows the respective characteristic peaks as already described for 704 m/z. This results underline the fragility of the predicted CO_2 group, probably even more fragile at the aglycone.

3.5.2. Candidate product feature 773.2116 m/z at 10.51 min

For the identification process the candidate BGLU3 product features 773.2116 m/z at 10.51 min was targeted fragmented in the MRM mode. The resulting MS/MS spectrum (figure 26) shows apart from the precursor ion 773 m/z the two main fragment ions 303 m/z and 449 m/z. The neutral losses between the fragment ion peaks are 146 Da (449 m/z - 303 m/z) and two times 162 Da, making 324 Da (773 m/z - 449 m/z), indicating a deoxyhexose and two hexoses or a caffeoylglucoside attached to the aglycone.
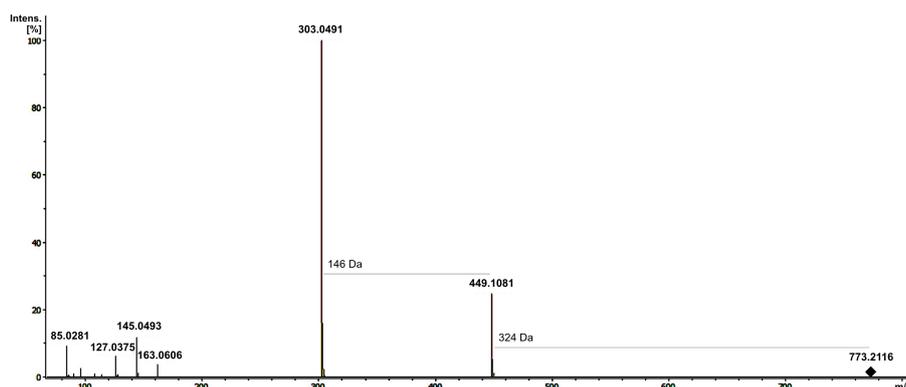
Figure 26 – *MS/MS spectrum of feature 773.2116 m/z at 10.51 min.* *The MS/MS spectrum of the* BGLU3 *candidate product feature 773.2116 m/z at 10.51 min is depicted. The characteristic fragment ions are 449 m/z with a neutral loss of 324 Da (2 x 162 Da) and 303 m/z with a neutral loss of 146 Da. The precursor ion is indicated by a rhombus.*

Targeted fragmentation with different collision energies did not change the fragmentation pattern to show a 611 m/z fragment ion peak (data not shown). In the positive ion mode 303 m/z is a characteristic fragment ion peak for either a quercetin (flavonol) or delphinidin (anthocyanidin) backbone. Measurements in the negative ion mode (ESI⁻) were performed to distinguish both backbones. The MS spectrum for ESI⁻ shows only a dominant [M–2H]⁻ and no [M-2H + H$_2$O]⁻ peak (figure 27A). The MS/MS spectrum (figure 27B) show apart from the precursor ion of 771 m/z, only two fragment ions of 625 m/z and 447 m/z, the first one exhibiting a neutral loss of 146 Da and the last one exhibiting a neutral loss of 178 Da, which accounts for 162 Da and an oxygen (16 Da). Therefore, an unusual and uncomplete fragmentation seems to have happened, also seen by the huge precursor ion peak and the missing aglycone peak. Even if the anthocyanidin aglycone could not be completely excluded, it is not supported from the results, too.

For identification attempts of the compound using MetFrag, the best hit in PubChem suggests according to the SpectraBase™ (https://spectrabase.com) a quercetion-3-*O*-α-rhamnopyranosyl-(1→6)-β-glucopyranosyl-(1→6)-β-galactopyranoside, thus a rhamnose (neutral loss of 146 Da), a glucose and a galactose (both neutral losses of 162 Da) attached to a quercetin backbone.
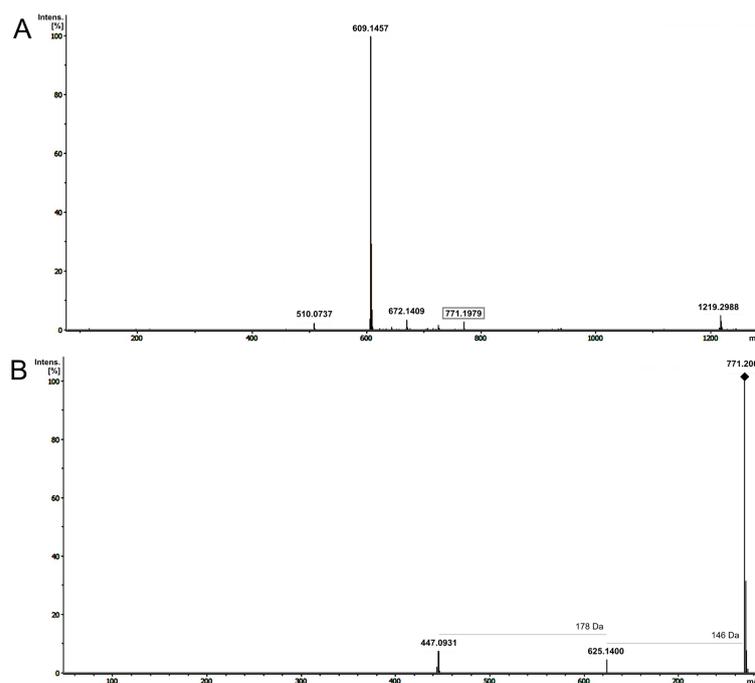
Figure 27 – *Candidate* **BGLU3** *feature with ESI⁻*. *The MS spectrum of 771.1959 m/z at 10.51 min (A) shows only the [M–2H]⁻ peak and no [M–2H + H₂O]⁻ peak. The MS/MS spectrum (B) shows apart from the precursor ion only the two fragment ions 625 m/z with a neutral loss of 146 Da and 447 m/z with a neutral loss of 178 Da (162 Da + 16 Da). The precursor ion is indicated by a rhombus.*

The second-best hit is a quercetin-3-*O*-glucose-(1→3)-rhamnose-(1→6)-glucoside, showing in the Global Natural Products Social Molecular Networking (GNPS) database (Wang *et al.*, 2016a) a 465 m/z instead of a 499 m/z fragment ion peak.
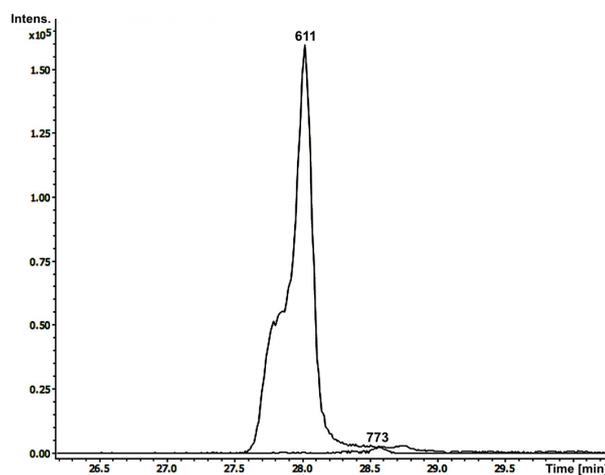
Figure 28 – *Flat gradient for feature 773.2116 m/z at 10.51 min.* *The features 773.2116 m/z at 10.51 min (773) is separated from the dominant 611 m/z peak (611) by the 116 min flat gradient. Therefore, it is assumed, that these peaks derive from different compounds.*

There are several more suggested quercetin hits, giving a strong hint towards a quercetin compound. The UV-vis spectrum shows two peaks around 265 nm and 365 nm, fitting the expectation for flavonols.

In the MS spectrum a 611 m/z compound elutes together with the 773 m/z compound, exhibiting a very huge peak, compared to 773 m/z and a similar MS/MS spectrum (data

not shown). To ensure, that these two peaks are not deriving from the same compound, a measurement for enhanced separation was performed in a flat gradient of 116 min. The results clearly indicate two separate compound peaks (figure 28).

3.5.3. Further double-charged candidate product features

The remaining double-charged candidate product features 441.7099 m/z at 9.45 min, 441.7102 m/z at 9.96 min, 448.7181 m/z at 11.05 min, 411.7178 m/z at 11.87 min, 440.7207 m/z at 12.45 min, 440.7209 m/z at 12.90 min, 432.7049 m/z at 14.62 min, 467.1764 m/z at 12.89 min, 494.1968 m/z at 19.11 min and 536.7408 m/z at 17.07 min exhibit mostly small peaks and very small or no peaks for the predicted $[M+H]^+$ single-charged peaks. Therefore no single-charged peak was picked by the algorithm. Because these peaks exhibit nevertheless the expected peak pattern for product compounds, it is assumed, that these peaks could partly derive from similar compounds as 866 m/z, differing in the carboxylation, hydroxylation or methylation pattern (table 7).
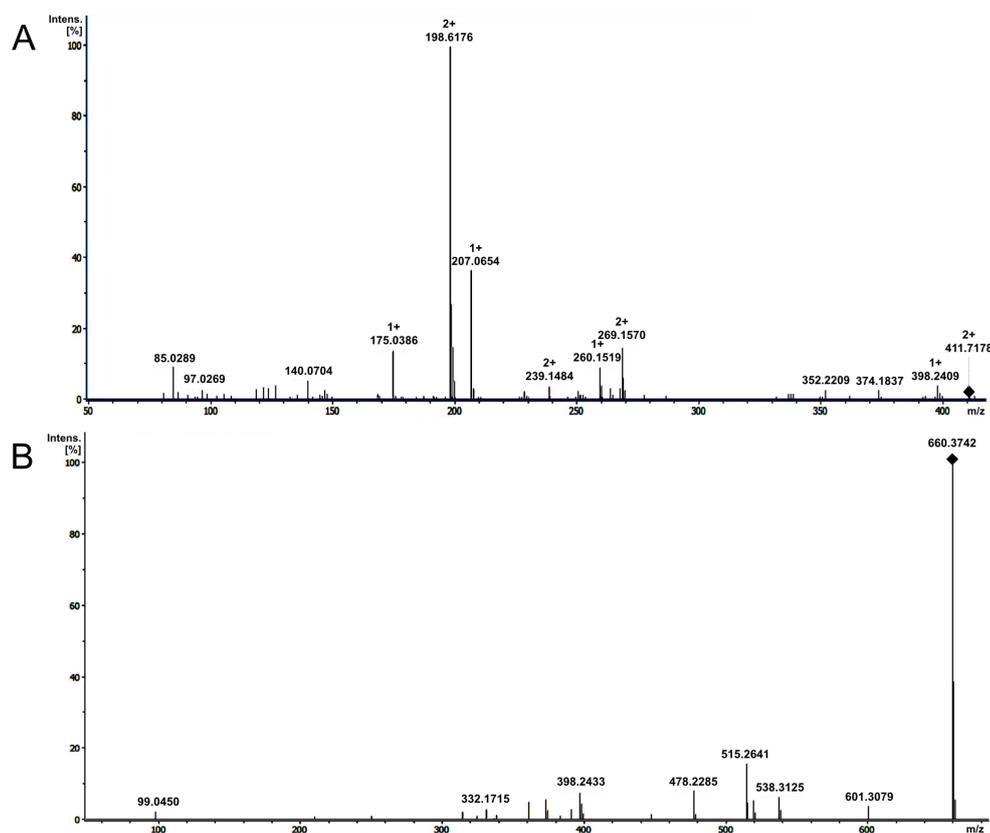
Figure 29 – *Candidate* **BGLU3** *features 411.7178 m/z at 11.87 min and 660.3742 m/z at 13.21 min.* The MS/MS spectrum of the double-charged candidate BGLU3 product feature 411.7178 m/z at 11.87 min is shown in A. The MS/MS spectrum of the single-charged candidate BGLU3 substrate feature 660.3742 m/z at 13.21 min in B. The precursor ions are indicated by a rhombus.

With this respect, the feature 411.7178 m/z at 11.87 min exhibits a decarboxylation, compared to 866 m/z. Vitisin B lacks the CO_2 group of vitisin A. A single-charged 822 m/z peak was not picked by the algorithm, but could be found by extracting the EIC of the mass 822 m/z. The very small peak elutes shortly after the peak 866 m/z at around 11.90 min, thus at the same

retention time as 411 m/z. The peak was not fragmented, thus, only the MS/MS spectrum from the double-charged compound is available (figure 29A). Most characteristic fragment ion peaks are similar to the ones from 433 m/z (figure 21A), thus, indicating a similar compound. In addition, the single-charged feature 660.3742 m/z at 13.21 min is likely to account for the respective substrate peak, not picked as the double-charged compound this time. Again, the MS/MS spectrum (figure 29B) is very similar to the one of 704 m/z (figure 20B), assumably accounting for a similar compound. In contrast to the respective $[M-CO_2+H]^+$ peak 660 m/z for 704 m/z, this feature elutes a little bit later from the column, as seen for 411.7178 m/z at 11.87 min in figure 25. This indicate a slightly differing conformational state between a vitisin B derived compound or a CO_2 fragment loss.

Due to the lack of the single-charged compound peaks and thus, double-charged and single-charged fragment ions in the MS/MS, the identification of the other double-charged candidate product features was not possible. All MS/MS spectra of the additional double-charged candidate product features are shown in figure S1 in the supplements.

3.5.4. Further single-charged candidate product features

The additional two single-charged candidate BGLU3 product compounds 757.2166 m/z at 11.42 min and 757.2152 m/z at 11.63 min were targeted fragmented in the MRM mode. The MS/MS spectrum of 757 m/z at 11.42 min (figure 30B) shows apart from the precursor ion 757 m/z the three main fragment ions 303 m/z, 449 m/z and 611 m/z. The neutral losses between the fragment ion peaks are 146 Da (757 m/z - 611 m/z), 162 Da (611 m/z - 449 m/z) and 146 Da (449 m/z - 303 m/z), indicating a deoxyhexose, a hexose and another deoxyhexose attached to the aglycone, which could be a quercetin or a delphinidin, as described above.

The MS/MS spectrum of 757 m/z at 11.63 min (figure 30A) shows apart from the precursor ion 757 m/z the two main fragment ions 287 m/z and 433 m/z. The neutral losses between the fragment ions are 324 Da, accounting for two times 162 Da (757 m/z - 433 m/z) and 146 Da (433 m/z - 287 m/z), indicating a deoxyhexose and two hexoses or a caffeoylglucoside attached to the aglycone, which could be a kaempferol or a cyanidin, as described above. Measurements in the negative ion mode, to distinguish flavonol and anthocyanidin aglycones were not performed, because the features were not chosen as the main candidate product compounds for BGLU3 activity.
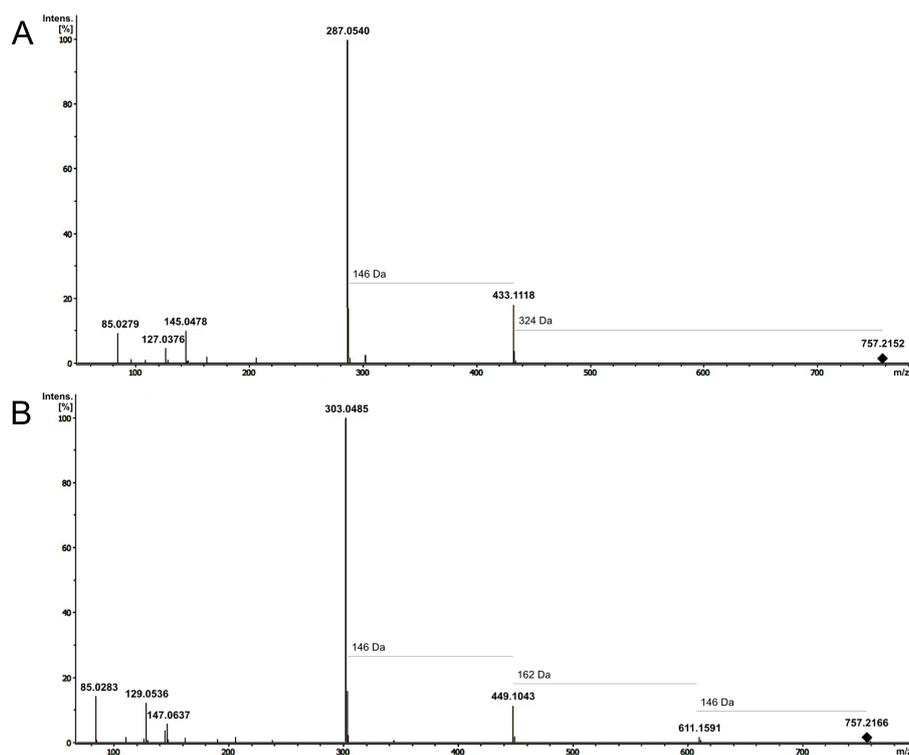
Figure 30 – *Candidate* **BGLU3** *features 757.2152 m/z at 11.63 min and 757.2166 m/z at 11.42 min.* *The MS/MS spectrum of the* BGLU3 *candidate product feature 757.2152 m/z at 11.63 min (A) shows the main fragment ions 433 m/z with a neutral loss of 324 Da (2 x 162 Da) and 287 m/z with a neutral loss of 146 Da. The MS/MS spectrum of the* BGLU3 *candidate product feature 757.2166 m/z at 11.42 min (B) shows the main fragment ions 611 m/z with a neutral loss of 146 Da, 449 m/z with a neutral loss of 162 Da and 303 m/z with a neutral loss of 146 Da. The respective precursor ions are indicated by a rhombus.*

3.5.5. Candidate substrate feature 781.3270 m/z at 15.38 min

The MS/MS spectrum of the second main candidate BGLU3 substrate feature 781.3270 m/z at 15.38 min (figure 31) shows the main fragment ions 619 m/z, 601 m/z, 575 m/z, 478 m/z, 452 m/z and 287 m/z. 619 m/z derives from a neutral loss of 162 Da, probably a hexose loss. An additional H_2O loss (-18 Da) results in the 601 m/z fragment ion peak. Another neutral loss of 44 Da results in the 575 m/z fragment ion peak. A neutral loss of 288 Da from 575 m/z results in the 287 m/z peak. A neutral loss of 288 Da and an additional oxygen loss (-16 Da) result in a 478 Da peak and a neutral loss of 123 Da from 575 m/z leads to the predominant 452 m/z fragment ion. While 288 Da is a characteristic neutral loss of an upper (epi)catechin unit in condensed pigments, a neutral loss od 123 Da can derive from a B ring loss of (epi)catechin. The fragment ion 287 m/z can account for a cyanidin, thus, the MS/MS spectrum hints towards a condensed flavanol–anthocyanidin compound, including an (epi)catechin and a cyanidin. A condensed (epi)catechin–cyanidin compound with a hexose moiety is 44 Da smaller than 781 m/z (287 m/z + 288 Da + 162 Da). Compared to cyanidin, malvidin is 44 Da bigger, due to the additional OCH_2 and CH_2 groups. Probably, the compound of 781 m/z is an (epi)catechin–malvidin-3-*O*-glucoside.
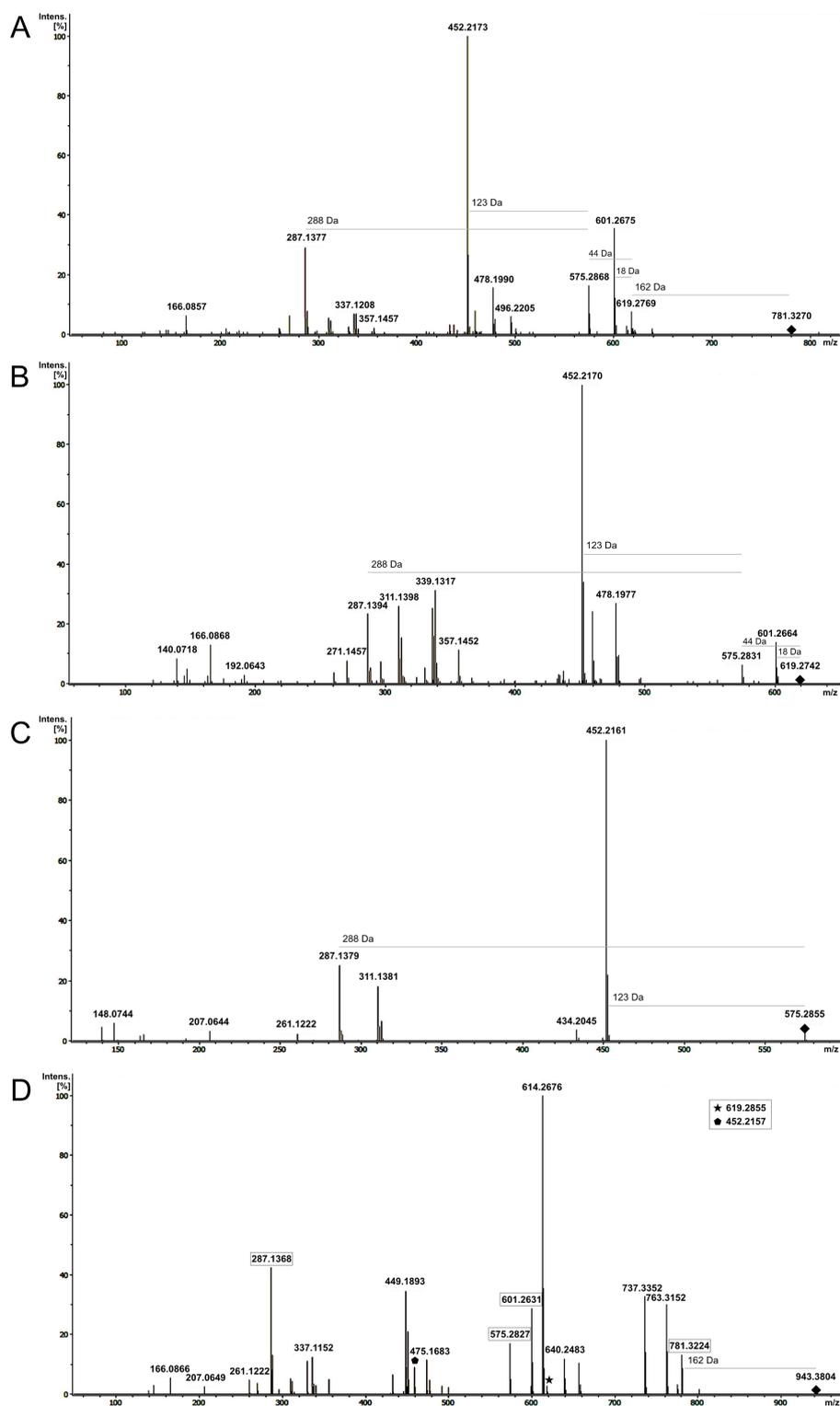
Figure 31 – *Candidate* **BGLU3** *features related to 781.3270 m/z at 15.38 min.* *The MS/MS spectrum of the main candidate* BGLU3 *substrate feature 781.3270 m/z at 15.38 min (A) shows the main fragment ions 619 m/z (neutral loss of 162 Da), 601 m/z (neutral loss of 18 Da), 575 m/z (neutral loss of 44 Da), 452 m/z (neutral loss of 123 Da) and 287 m/z (neutral loss of 288 Da). The candidate substrate feature 619.2742 m/z at 17.37 min is 162 Da smaller and shows a very similar MS/MS fragmentation pattern (B). The MS/MS spectrum of the candidate substrate feature 575.2851 m/z at 18.19 min (C), shows again a related fragment ion pattern. The possible product compound from the 781 m/z substrate compound could be 943.3804 m/z at 13.74 min. Apart from other dominant fragment ions, the MS/MS spectrum (D) shows all fragment ions characteristic in the MS/MS spectra above (indicated by boxes), as well as the neutral loss of 162 Da for the fragment ion 781 m/z. The respective precursor ions are indicated by a rhombus.*

781.3270 m/z at 15.38 min and the additonal candidate substrate compound 619.2742 m/z at 17.37 min seem to be the same compound, differing by one hexose moiety, as seen by the MS/MS spectrum of 619 m/z (figure 31B). In addition to this, the candidate substrate compound 575.2851 m/z at 18.19 min could be involved in the formation of the other two compunds, as seen by the similar MS/MS spectrum (figure 31C). The same case is proposed for the candidate substrate compound 287.1382 m/z at 7.25 min (for the MS/MS spectrum see S2A).

The data, include a very small peak with 943.3804 m/z at 13.74 min, being 162 Da bigger than 781 m/z. The MS/MS spectrum after targeted fragmentation (figure 31) shows all main peaks of 781 m/z, but with other additional main fragment ion peaks. Thus, the compound could be a possible product of 781 m/z glycosylation by BGLU3.

3.5.6. Further candidate substrate features

Concerning the remaining features for BGLU3 with an expected substrate peak pattern, the feature 402.1451 m/z at 13.11 min could be an adduct of 704 m/z due to elution at the same retention time (figure 22D).
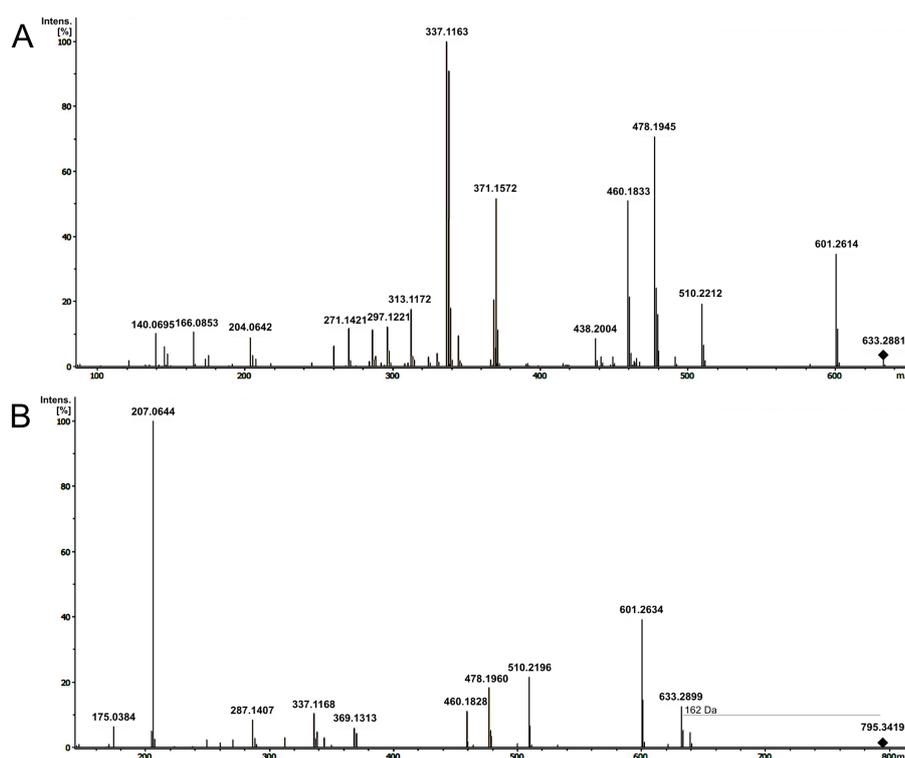
Figure 32 – *Candidate* **BGLU3** *features 633.2899 m/z at 19.68 min and 795.3419 m/z at 17.03 min. The MS/MS spectrum of the two* BGLU3 *candidate substrate features 633.2899 m/z at 19.68 min (A) and 795.3419 m/z at 17.03 min (B) differing by a neutral loss of 162 Da are depicted. They exhibit similar MS/MS fragmentation patterns and could derive from related compounds. In addition, the fragment ion peaks hint on a condensed compound of a flavanol and an anthocyanidin. The respective precursor ions are indicated by a rhombus.*

Double-charged fragment ions in the MS/MS spectrum (figure S2B) indicate a double-charged compound. Although, the calculated single-charged peak does not exhibit a neutral loss of 52.8 Da towards 704 m/z, the dominant fragment ions are similar to those, seen in the MS/MS

spectrum of 460 m/z (figure 21B), which elutes at the same time as 866 m/z. No further information could be derived for this compound.

Like 781.3270 m/z at 15.38 min and 619.2742 m/z at 17.37 min, the features 795.3419 m/z at 17.03 min and 633.2899 m/z at 19.68 min differ by a neutral mass of 162 Da, probably one hexose. According to the MS/MS spectrum (figure 32A and 32B) after targeted fragmentation, both features could derive from related compounds. In addition, the characteristic fragment ion peaks ressemble those of 781 m/z and 619 m/z, speaking for a similar condensed compounds, which are 14 Da smaller.

3.6. Identification of BGLU4 candidate product and substrate features

The main *BGLU4* candidate product feature 757.2149 m/z at 11.71 min was targeted fragmented in the MRM mode, due to low peak intensities. The resulting MS/MS spectrum (figure 33A) shows apart from the precursor ion 757 m/z the main fragment ions 303 m/z and 449 m/z. The neutral losses between the fragments are 308 Da, deriving from 146 Da and 162 Da, while the order is not known (757 m/z - 449 m/z) and 146 Da (449 m/z - 303 m/z), suggesting the attachment of two deoxyhexoses and one hexose or one deoxyhexose and a *p*-coumaroylglucoside to the aglycone.
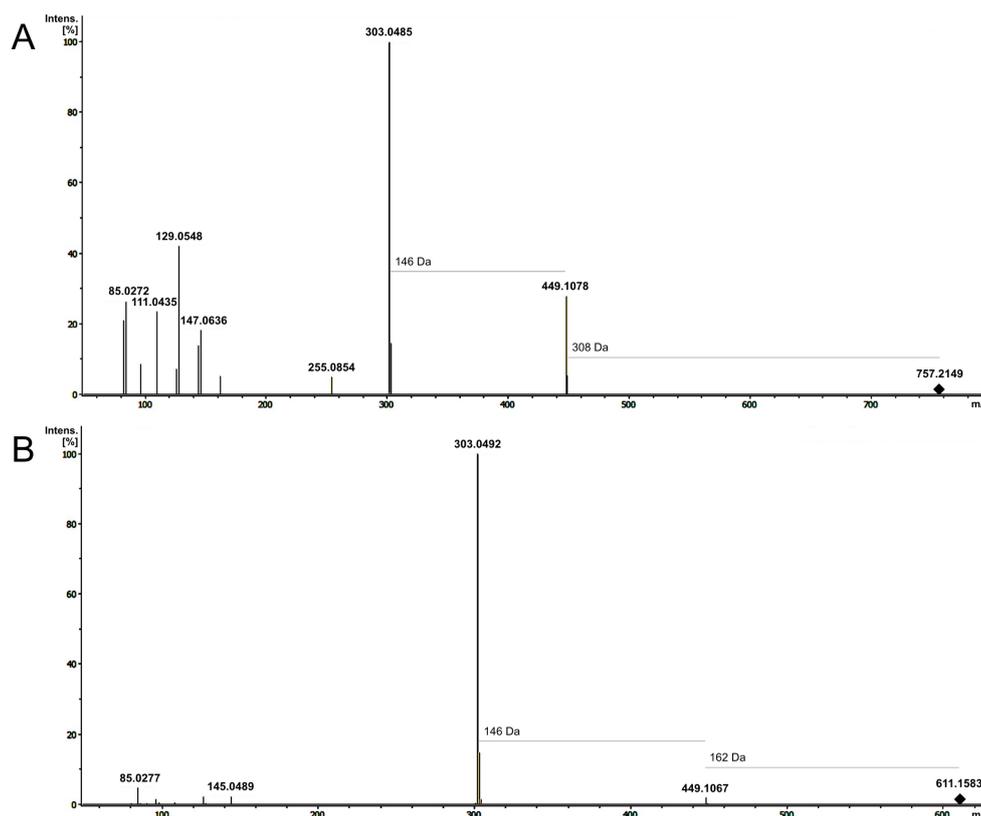
Figure 33 – *Candidate BGLU4 features 757.2149 m/z at 11.71 min and 611.1583 m/z at 11.34 min. The MS/MS spectrum of the BGLU4 candidate product feature 757.2149 m/z at 11.71 min (A) shows the main fragment ions 449 m/z with a neutral loss of 308 Da (162 Da + 146 Da) and 303 m/z with a neutral loss of 146 Da. The MS/MS spectrum of the BGLU4 candidate substrate feature 611.1583 m/z at 11.34 min (B) shows the main fragment ions 449 m/z with a neutral loss of 162 Da and 303 m/z with a neutral loss of 146 Da. The respective precursor ions are indicated by a rhombus.*

Since, the 303 m/z aglycone in the positive ion mode could derive either from a quercetin

(flavonol) or delphinidin (anthocyanidin) backbone, measurements in the negative ion mode (ESI$^-$) were performed to distinguish both backbones. The MS spectrum for (ESI$^-$) shows no dominant [M-2H + H$_2$O]$^-$ apart from the [M-2H]$^-$ peak (figure 34A). The MS/MS spectrum (figure 34B) exhibits, apart from the precursor ion 755 m/z, the three fragment ions 609 m/z, 447 m/z and 301 m/z, with the same neutral losses from the positive ion mode (146 Da [775 m/z - 609 m/z], 162 Da [609 m/z - 447 m/z] and 146 Da [447 m/z - 301 m/z]), this time revealing the precise order between 162 Da and 146 Da. No radical aglycone was seen. Therefore, it is assumed, that the compound is a quercetin with a deoxyhexose, a hexose and another deoxyhexose attached.
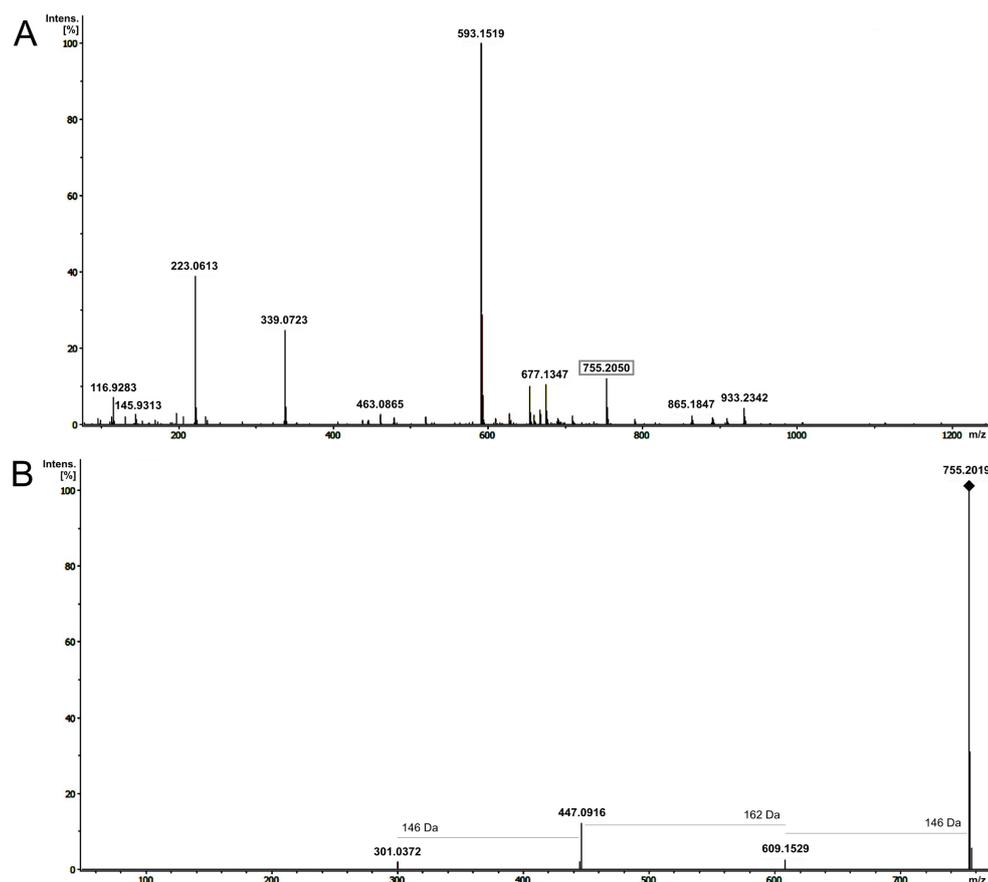
Figure 34 – **Candidate BGLU4 feature with ESI$^-$.** *The MS spectrum of 755.1992 m/z at 11.71 min (A) shows only the [M–2H]$^-$ peak and no [M–2H + H$_2$O]$^-$ peak. The MS/MS spectrum (B) shows apart from the precursor ion the fragment ions 609 m/z with a neutral loss of 146 Da, 447 m/z with a neutral loss of 162 Da and 301 m/z with a neutral loss of 146 Da. The precursor ion is indicated by a rhombus.*

Attempts to identify the compound with MetFrag, derived many hits in PubChem, the best ones all glycosylated quercetins. The best hit was a quercetin 3-*O*-[α-L-rhamnopyranosyl(1→2)-β-*D*-galactopyranosyl]-7-O-α-L-rhamnopyranoside (PubChem CID 57397680), followed by a quercetin 3-rhamnosyl-(1→4)-rhamnosyl-(1→6)-glucoside (PubChem CID 44259158) and a quercetin 3-rhamnosyl-(1→2)-rhamnosyl-(1→6)-glucoside (PubChem CID 44259157). Speaking again for a glycosylated quercetin as the candidate compound, supported by the UV-vis spectrum, showing a 247 nm and a 387 nm peak.

757.2149 m/z at 11.71 min elutes very close to two other compounds with the same mass of 757 m/z, already described as candidate product compounds for BGLU3 (see 3.5.4). In the

soaked seed samples, only the 757 m/z peak at 11.71 min shows the expected product peak pattern for BGLU4, which is not seen in the dry seed samples for BGLU3. The first 757 m/z peak at 11.42 min shows the same MS/MS fragmentation pattern (figure 30B) as the BGLU4 candidate feature. Therefore, it is possible, that both compounds are two isomers, with a different spacial structure, that could be glycosylated by two different enzymes.

In relation to the candidate product feature, two candidate substrate features with the mass of 611 m/z were found due to fold changes. The MS/MS spectrum of the candidate substrate feature 611.1584 m/z at 8.75 min (figure S3) revealed fragment ions of 303 m/z and 465 m/z, not accounting for the substrate of BGLU4, to generate the candidate product compound. The MS/MS spectrum of the candidate substrate feature 611.1583 m/z at 11.34 min (figure 33B) shows fragment ions at 303 m/z and 449 m/z, thus, being a possible substrate for glycosylation by BGLU4.

Further features with a peak pattern for candidate BGLU4 product compounds were 529.2079 m/z at 27.45 min and 507.1473 m/z at 14.37 min. The MS/MS spectra are shown in the supplements in figure S4. No further analysis were done for these features.

4. Discussion

4.1. BGLU1, BGLU3 and BGLU4 as possible GH1-type glycosyltransferases

Most known glycosylation reactions for flavonoids are mediated by the cytosolic UGTs. In the recent years, starting in 2010 with the discovery of Matsuba *et al.*, more and more evidence was rising up, that there exists at least one further glycosylation mechanism, mediated by GTs with amino acid homology to GH1 β-glycosidases, termed GH1 type GTs. Thus, it must be assumed, that not all glycosylation mechanism for flavonoids in plants are clarified, including knowledge about the enzyme, the reaction itself, the utilized substrates, the compartmentalization and also the genetic regulation. Moreover, with regard to the huge amount of flavonoid glycosides and the already known variety of different structures, it must be assumed, that many genes, encoding for GTs or flavonoid modifying enzymes in general, are not characterized yet. The probable presence of many unidentified minor flavonoid compounds support this assumption even more. (Matsuba *et al.*, 2010; Ishihara *et al.*, 2016; Wu *et al.*, 2018). Enhanced knowledge about the new type of GTs, their substrates and their encoding genes, can support the understanding of plant integral regulatory systems for key metabolic traits like glycosylation of flavonoids. In addition, further insights into the function and diversity of flavonoid glycosides in plants are provided (Wu *et al.*, 2018).

This work focused on the functional characterization of three genes in Arabidopsis, *BGLU1*, *BGLU3* and *BGLU4*, assumed to encode for GH1-type GTs, based on amino acid similarities, investigated in earlier studies (Miyahara *et al.*, 2011; Ishihara *et al.*, 2016). The approach was an untargeted metabolite fingerprint of a respective *bglu* T-DNA insertion line and a respective *2x35S:BGLU* overexpression line compared with the Col-0 wild type. Candidate product or substrate compounds with the respective differential peak pattern for the metabolite were chosen for compound identification, to generate information about the enzymatic activity of the encoded proteins. This revealed a proposed GH1-type GT activity of all three encoded

proteins BGLU1, BGLU3 and BGLU4 using flavonoid substrates.

4.1.1. BGLU1 as possible flavonol GH1-type glycosyltransferase

Identification approaches for the BGLU1 candidate product compound of 741 m/z provide strong hints towards a triglycosylated kaempferol, mainly based on the MS/MS fragmentation pattern and results from the best database hits, supported by the proposition of Routaboul *et al.* (2006) to find kaempferols more likely in leaves, compared to quercetins in seeds. A distinct identification of the compound was not possible, due to many database hits on the one hand and due to different expected fragmentation patterns for the best hits, compared to the candidate product compound, accounting for a different glycosylation pattern.

Further information about the BGLU1 candidate product compound were gained by Wu *et al.* (2018). With 309 Arabidopsis accessions, they performed untargeted LC-MS-based metabolomic profiling with metabolic genome-wide association studies, using differing metabolite levels at different environmental conditions, namely the control as well as light and temperature stress as traits. This resulted in 123 environment-specific, highly resolved metabolite quantitative trait loci out of more than 3000 semi-polar metabolites. One of the differential metabolite levels could be traced back to *BGLU1* by SNP m27661 and seven other significantly associated SNPs, located in this gene. The respective metabolite feature 741.2220 m/z in the positive ion mode is in concordance with our feature, including the MS/MS spectrum with the main fragment ions 287 m/z, 449 m/z, 595 m/z and the precursor ion 741 m/z. In addition, the results were confirmed with another T-DNA insertion line (SALK_060948), again concordant with our results, meaning decreased metabolite levels, but no lack of the metabolite in the loss-of-function T-DNA insertion line. It must be emphasised, that in the RT and qRT-PCR slight amounts of intact *BGLU1* transcript was found for *bglu1-1*. These slight amounts were regarded as neglectable and are not expected to result in such high metabolite levels, while the peaks for *bglu1-1* samples indicate obvious compound production. Wu *et al.* (2018) explain these peaks by possible functional redundancy from other homologous *BGLU* genes. Miyahara *et al.* (2013) propose such a functional redundancy for the AAGT BGLU10, too, because the proposed product A11 did also accumulate to some extend in the *bglu10* mutant. For the functional redundancy, they propose *BGLU9* as a candidate AAGT gene, because in prior work they suggested AAGT activity for BGLU9 (Miyahara *et al.*, 2011) and both amino acids show high similarity. For BGLU1, based on phylogenetic analysis, BGLU3 is the closest homolog. Therefore, in this work, a homozygous double mutant of *bglu1-1* and *bglu3-2* was generated by crossing. In future studies it will be promising to investigate, if this double mutant results in the lack of metabolite levels in *bglu3-2*. Based on expression levels on the TraVa database, the very strict expression of *BGLU3* in seeds make a redundancy of these two genes less likely. Taking the expressing tissues into account, the closest homolog of *BGLU1* would be *BGLU5*. Accordingly, a heterozygous double mutant of *bglu1-1* and *bglu5-1* was generated in this work by crossing. In future work, a homozygous line should be generated and investigated for metabolite levels of the candidate product compound.

As for the results of Wu *et al.* (2018), the small peak of the candidate product compound is located close to a main peak of 741 m/z, which probably belongs to the already characterized

3-*O*-[6"-*O*-(rhamnosyl) glucoside] 7-*O*-rhamnoside or 3-*O*-[2"-*O*-(rhamnosyl) glucoside]-7-*O*-rhamnoside (Tohge *et al.*, 2005), exhibiting another glycosylation pattern.

Additionally, Wu *et al.* (2018) revealed co-mapping of the metabolite with the locus *UGT78D1*, encoding an UDP-glycosyltransferase, which transfers an UDP-rhamnose to the 3-OH position of kaempferol and quercetin (Jones *et al.*, 2003). These findings, together with the results from this work, including the knowledge, that the investigated metabolite is the putative product of BGLU1 enzymatic activity, presumably exclude the glycosylation of the 3-OH position by BGLU1 and in addition suggest, that the investigated compound is *O*-rhamnosylated at C-3. Further common glycosylations for flavonols, are either an extension of the sugar unit at C-3 or a 7-*O*-glycosylation (Hectors *et al.*, 2014; Ishihara *et al.*, 2016). Findings of Kerhoas *et al.* (2006) suggest, that there exists a kaempferol diglycoside in Arabidopsis seeds of 595 m/z with the fragmentation pattern of 449 m/z and 287 m/z, exhibiting a rhamnose at C-3 and a hexose at C-7. They suggest this in contrast to a compound with a hexose at C-3 and a rhamnose at C-7, exhibiting the fragment ions 433 m/z and 287 m/z, which fit to the above-mentioned main 741 m/z peak, eluting close to the candidate product compound.

The fragmentation of the C-3 rhamnosylated and C-7 hexosylated compound fit to the fragmentation pattern of the candidate BGLU1 product compound. If one rhamnose is located at C-3 and the hexose located at C-7, accounting for 595 m/z, there exists a strong probability, that *BGLU1* transfers a deoxyhexose, probably another rhamnose. The additional glycosylation of the predicted rhamnose moiety at C-3 would yield a similar glycosylation reaction as for BGLU6, transferring a hexose to an existing hexose moiety at C-3. This would be in concordance with the expectations based on the MS/MS spectrum, where a deoxyhexose is expected to fragment first, thus, being spatially located on a more external position. 3-*O*-dirhamnosides were found for quercetion and kaempferol in other organisms (Ahn *et al.*, 2014; Neugart *et al.*, 2015).

Nevertheless, it cannot be excluded, that the hexose is attached to the rhamnose at C-3, which would suggest, that BGLU1 rhamnosylates C-7, which is known for other GH1-type GTs in other organisms (Matsuba *et al.*, 2010; Miyahara *et al.*, 2012). Such 3-*O* and 7-*O*-rhamnosylated kaempferol compounds are known for seeds and leaves in Arabidopsis (Veit and Pauli, 1999; Kerhoas *et al.*, 2006). Besides, Quispe *et al.* (2013) found a 741 m/z compound with the same fragmentation pattern in *Vasconcellea quercifolia*, suggesting a kaempferol 3-(2"-rhamnosyl rutinoside) as the kaempferol derivative to manghaslin. Manghaslin exhibits a rhamnose at C-3, with another rhamnose and a glucose attached to this C-3 rhamnose. In the study, manghaslin was characterized with ^{13}C NMR and the 741 m/z compound was suggested to be the kaempferol derivative, because the compound exhibit 16 Da less, also in all fragment ions, with the same order of neutral losses. Other glycosylation patterns cannot be excluded. The suggested structures are shown in figure 35. So far no GH1-type GT transferring a sugar moiety other than glucose, like rhamnose, was reported (Sasaki *et al.*, 2014), making BGLU1 probably the first GH1-type flavonoid rhamnosyltransferase. This enzymatic activity needs to be investigated with enzyme activity assays, which will be discussed later.

It is known, that some anthocyanidins are indistinguishable from some flavonols in the positive ion mode. For example, cyanidin glycosides and kaempferol glycosides, both exhibiting a 287

m/z aglycon fragment ion. In this case, the anthocyanidin accounts for a $[M]^+$ ion and the flavonol for a $[M+H]^+$ ion, due to the natural positive charge of anthocyanidins. Even the MS^n often leads to the same fragmentation patterns (Sun *et al.*, 2012). In case of BGLU1, there exists for example also a compound of 595 m/z with the MS/MS fragment ions of 449 m/z and 287 m/z, accounting for a cyanidin-3-*O*-rutinoside (Tian *et al.*, 2005; Dincheva *et al.*, 2013). In addition, Miyahara *et al.* (2011) could show, that *BGLU1* expression was induced in anthocyanin-inducing medium, postulating, that *BGLU1* is involved in the anthocyanin biosynthesis.
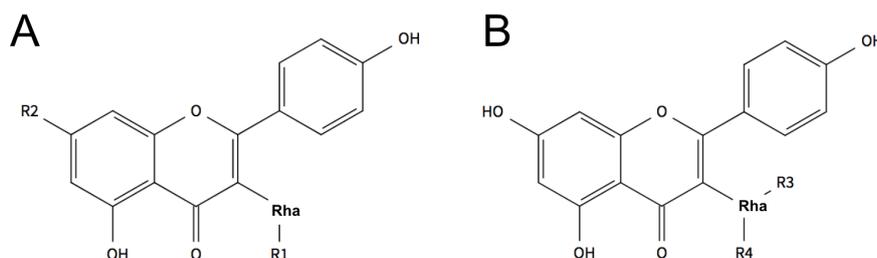
Figure 35 – ***Possible structures of 741.2221 m/z at 10.82 min.*** *The proposed structures for 741.2221 m/z at 10.82 min, with possible sugar attachment sites are depicted. A suggest a C-3 and C-7 glycosylation, with a rhamnose attached to C-3. R1 could be a rhamnose or glucose, while rhamnose is favored. R2 should be a glucose, if R1 is a rhamnose and vice versa. B shows the proposed kaempferol 3-(2"-rhamnosyl rutinoside) structure from Quispe et al. (2013), with R3 and R4 a rhamnose and a hexose, respectively. Rha: rhamnose.*

In order to distinguish both compounds, the UV-vis molecular absorbance is often measured with DAD. Anthocyanidins exhibit their absorption maxima around 480 - 540 nm and flavonols absorb only below 400 nm. The distinction by this approach is not always possible, because DAD is not as sensible as MS and therefore, detection of small quantities or co-eluting anthocyanidin and flavonol compounds can be challenging (Sun *et al.*, 2012). In case of the candidate BGLU1 product compound the UV-vis spectra hinted towards a glycosylated flavonol, but due to the low peak intensity and because the measurement setup did not measure above 400 nm, these data have to be consided with care.

Despite, a measurement in the negative ion mode can be useful for the distinction, because anthocyanidins seem to exhibit some characteristics in the negative ion mode, which are different compared to flavonols. These are characteristic dominant anthocyanidin $[M–2H]^-$ and $[M–2H + H_2O]^-$ peaks in the MS and for some anthocyanin glycosides also the double charged $[M–2H]^{2-}$ and $[M–2H + H_2O]^{2-}$ peaks. In contrast, flavonols are only dominant with their $[M–H]^-$ peak. In addition, the fragmentation pattern is more complex and shows more frequently the radical aglycon in the MS/MS spectrum, in case of anthocyanidins. The complex fragment ions are proposed to derive from the different conformational states of anthocyanidins in liquid solutions. (Sun *et al.*, 2012). Measurements for the candidate BGLU1 product compound in the negative ion mode yielded a very small peak in the MS and only an uncomplete MS/MS fragmentation pattern. The neutral loss of 162 Da is not in concordance with the neutral loss of 146 Da in the positive ion mode but could also account for another fragmentation order in the negative ion mode. The uncomplete fragmentation could derive from the small peak intensity of the BGLU1 candidate product compound or due to non-optimal collision energies in the MRM mode. Even if the presence of a cyanidin

backbone cannot be excluded from these data, the lack of a dominant $[M–2H + H_2O]^-$ peak or double charged peaks do not support a cyanidin backbone, especially if taking into account the co-mapping of *BGLU1* with *UGT78D1*, which is an UGT, known to act on flavonols. Miyahara *et al.* (2013) did not see any anthocyanidin profile differences for *bglu1* loss-of-function mutants in leaves, supporting, that BGLU1 acts not on anthocyanidins.

Because there exist on the one hand many different combination possibilities for the sugar types and sugar attachment positions of triglycosylated flavonoids, and on the other hand flavonol and anthocyanidin aglycones are not always distinguishable, for explicit structure determination and identification of substituent positions, either the measurements with a standard or structural confirmation, for example by NMR measurements, would be necessary (Wu *et al.*, 2018). Flavonoid standards for triglycosides are difficult to obtain and due to many combinatorial possibilities, many standards would be necessary. NMR measurements are challenging for compounds with low peak intensities, suggesting low compound levels. The compound needs to be completely pure, thus, extensive sample purification is necessary. In addition, the measurement requires several magnitudes more of analyte compared to HPLC. Therefore, the method is very time consuming for small peaks (Del Rio *et al.*, 2004).

Nevertheless, the purification of the candidate BGLU1 product compound for NMR measurements is in progress, in collaboration with the work group of Prof. Dr. Norbert Sewald at the chemical department. In the still ongoing work, 2.5 kg Arabidopsis rosette leaves of the *2x35S:BGLU1* overexpression line were extracted with dichloromethane/methanol, ethyl acetate and butanol, respectively, finding the target compound in the butanol fraction with reduced metabolite background. Separation of the butanol fraction with sephadex columns of two different pore sizes, successfully removed further main background metabolites. The current main challenge is the separation of the compound from further compounds of the mass 741 m/z, while remaining enough sample for NMR measurement.

Apart from the mentioned methods, there are reports about the identification of sugar positions and even sugar linkages for flavonoid compounds, using HPLC-MS^n (Cuyckens *et al.*, 2001; Hvattum and Ekeberg, 2003; Cuyckens and Claeys, 2005; Ablajan *et al.*, 2006; Ablajan, 2011; van der Hooft *et al.*, 2011). This analysis is mainly based on the investigation of the ratios between radical aglyca and non-radical aglyca in the positive and negative ion mode, by a complex measurement protocol. Due to the low peak intensities of the candidate compound, this could be challenging, because in the negative ion mode no aglycone peak was seen.

However, if the compound purification is not successful or does not yield enough compound for sufficient resolution with NMR measurements, this could be an additional attempt for the structure elucidation.

For *BGLU1* no candidate substrate compound fitting to the product compound was found by fold changes. A similar result is reported in Ishihara *et al.* (2016), where the substrate levels of BGLU6 seemed to be independent from the activity of BGLU6. The authors suggest an unknown mechanism, that could prevent the accumulation of certain flavonol glycosides.

4.1.2. BGLU3 as a possible multifunctional flavonoid GH1-type glycosyltransferase

Identification approaches for the candidate BGLU3 product feature 866 m/z and candidate BGLU3 substrate feature 704 m/z led in the beginning to the assumption, of BGLU3, glycosylating proanthocyanidins, since the 289 m/z fragment ion of 866 m/z is known from fragmentation of oligomeric proanthocyanidins, if the upper unit is an (epi)catechin (Li and Deinzer, 2007; Flamini, 2012) and several peaks (precursor and fragment ions) exhibit similar masses of proanthocynidins. In addition, the predominance as a double-charged compound, is a characteristic trait for higher molecular mass proanthocyanidins, mainly found in seeds (Routaboul *et al.*, 2006; God̄evac *et al.*, 2010).

Because the neutral losses of the MS/MS fragmentation did not account for proanthocyanidins as defined by Li and Deinzer (2007), a compound, sharing some characteristics with oligomeric proanthocyanidins, was suggested as likely, leading to condensed compounds of flavanols and anthocyanidins. This guided the assumption, that BGLU3 could transfer a glucose moiety to a condensed (epi)gallocatechin–carboxypyranomalvidin molecule by *O*-glycosidic linkage at the C-3 position of the carboxypyranomalvidin, like a β-D-glucopyranose as shown by (Bakker *et al.*, 1997). This results in an (epi)gallocatechin–carboxypyranomalvidin-3-*O*-glucoside product, or with other words an (epi)gallocatechin–vitisin A.

Condensed compounds of flavanols and anthocyanidins, as well as pyranoanthocyanidins and their condensations with flavanols are mostly known from wine, where they were first discovered (Somers, 1971; Fulcrand *et al.*, 1996), but also reported for some plants, like strawberries (Fossen *et al.*, 2004), black currant seeds (McDougall *et al.*, 2005), beans (Macz-Pop *et al.*, 2006a) or purple corn (González-Manzano *et al.*, 2008). The very abundant wine compound vitisin A, which is a carboxypyranomalvidin glucoside, derived by the cycloaddition of malvidin-3-*O*-glucoside and pyruvic acid, was first characterized by Bakker *et al.* (1997) and Fulcrand *et al.* (1998). Both describe only slightly different structures, with an hydroxyethylene group attached to the pyran ring in the first case and an attached carboxyl group in the second case. The proposed structure from Fulcrand *et al.* (1998) was adapted for the structure proposition of the candidate BGLU3 compounds, due to the described characteristic 44 Da CO_2 loss, which was also shown for the present results. The CO_2 loss also supports the assumption of the flavanol as the upper unit. Furthermore, the described characteristic 399 m/z fragment ion peak of the vitisin A aglycone, fits to the characteristic 398 m/z fragment ion peak from this work, exhibiting one hydrogen less, from the predicted C–C bond with an (epi)gallocatechin. The formation of the dominant fragment ion 104 m/z cannot be sufficiently explained. It is proposed, that the fragment could derive from the pyrano ring D with an additional H_2O, because the structural composition of malonic acid includes the same elements, with a mass of 103 Da, yielding a charged 104 m/z molecule (Amorim *et al.*, 2009). This suggestion is just hypothetic and no reports for this fragment from pyranoanthocyanidins were found in literature. (Fulcrand *et al.*, 1998; Marquez *et al.*, 2013a; Lambert *et al.*, 2015).

In the suggested condensed compound, vitisin A accounts for a mass of 560 m/z (one hydrogen is lost by the C–C bond) and the (epi)gallocatechin for a mass of 306 Da (one hydrogen is lost by the C–C bond), as a charged flavanol unit. 560 m/z and 306 m/z

yield 866 m/z, while the natural charged vitisin A unit accounts for a double-charged compound in the positive ion mode, which is the predominant condition of the present candidate compounds. It is assumed, that the (epi)gallocatechin–vitisin A compound derives by a cycloaddition of (epi)gallocatechin–malvidin with pyruvic acid, deriving from the glycolysis, as discussed later. Therefore, the compound formation starts with a direct condensation between (epi)gallocatechin and malvidin, followed by cycloaddition with pyruvic acid. For the direct condensation, the anthocyanidin needs to be in its hydrated hemiketal form. This hydrated form is not likely for pyranoanthocyanidins, whose characteristics are the resistence towards hydration. Therefore, the reaction of flavanols and vitisins is very unlikely and could not be confiremd in model solution (Nave *et al.*, 2010). A proposed cycloaddition mechanism was adapted from Fulcrand *et al.* (1998), as shown in figure 36.
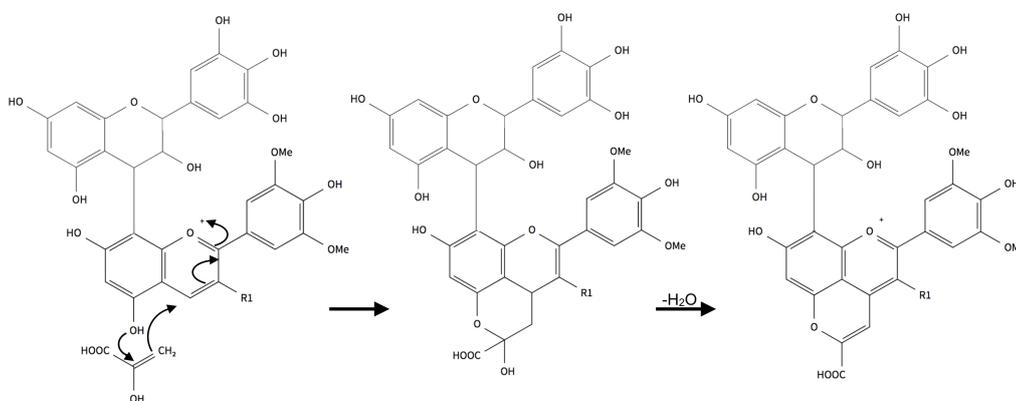
Figure 36 – ***Proposed cycloaddition of pyruvic acid and (epi)gallocatechin–malvidin.*** *The proposed mechanism for the cycloaddition of pyruvic acid and (epi)gallocatechin–malvidin is based on Fulcrand* et al. *(1998). R1= OH for (epi)gallocatechin–malvidin, R1 = OGlc for (epi)gallocatechin–malvidin-3-*O*-glucoside.*

The presence of (epi)gallocatechin as upper terminal unit is suggested, based on the fragmentation rules of Sánchez-Ilárduya *et al.* (2012), to distinguish between F–pyA$^+$ and pyA$^+$–F molecules, using retro-Diels–Alder (RDA) fission, heterocyclic ring fission (HRF) and a partial fragmentation of the upper unit. The retro Diels–Alder (RDA) fragmentation derives through the cleavage of the 1 and 3 bonds of the flavanol C-ring (loss of 168 Da for (epi)gallocatechin). The cleavage of the 1 and 4 bonds of the C-ring of the flavanol, losing a phloroglucinol ring (loss of 126 Da), takes place in the HRF fission. The partial fragmentation of the upper flavanol unit derives by cleavage of the 2 and 4 bonds of the flavanol C-ring (loss of 262 Da for (epi)gallocatechin). The loss of 288 Da indicate a broken B-type interflavanic bond, corresponding to the complete loss of the flavanol unit ((epi)catechin), while also the loss of 290 Da is possible and this fragmentation cannot distinguish between F–pyA$^+$ and pyA$^+$–F molecules (Flamini, 2012; Sánchez-Ilárduya *et al.*, 2012). In general, the fragmentation pattern of the MS/MS spectrum can be difficult to evaluate for these condensed compounds and different authors describe different rules for the evaluation. According to Sánchez-Ilárduya *et al.* (2012) the above decribed rules are sufficient for the characterization, because the phloroglucinol ring loss of 126 Da loss ($C_6H_6O_3$, ring A) is characteristic for the upper flavanol unit, since a lower flavonol unit cannot lose the ring A (González-Paramás *et al.*, 2006; Sánchez-Ilárduya *et al.*,

2012). In addition, the partial loss of the upper unit should not happen for pyA$^+$-F molecules and RDA fission was not observed by Sánchez-Ilárduya *et al.* (2012), thus, this fragmentation is probably less abundant in those molecules. According to Sentandreu *et al.* (2012), the MS/MS between F-pyA$^+$ and pyA$^+$-F molecules is not sufficiently distinguishable, but only the MS^3 fragmentation of the aglycones. This fragmentation yields more characteristic fragments, with five characteristic ions for F-pyA$^+$ and only two characteristic ions for pyA$^+$-F molecules (Sentandreu *et al.*, 2012). Since, no MS^3 fragmentation was performed for our measurements, the evaluation was based on the results of the MS/MS spectrum, guided by the rules of Sánchez-Ilárduya *et al.* (2012).

No neutral loss of 126 Da was found, despite being expected for an upper flavanol unit. Nave *et al.* (2010) could not see this fragment for flavanol–carboxypyranoanthocyanidins from wine, too. The RDA fragmentation with a 168 Da loss was also shown by McDougall *et al.* (2005) in black currant seeds, exhibiting F-pyA$^+$ compounds with an (epi)gallocatechin as the upper unit. Therefore, the results were proposed to hint towards an upper (epi)gallocatechin unit. Since, F-pyA$^+$ isomers elute before pyA$^+$-F, due to higher polarity, this would also be a good point to investigate, but is not possible with the present data, where only one molecule is known to be present (Sánchez-Ilárduya *et al.*, 2012).

From the HPLC results, no proposition can be made, if an epigallocatechin or catechin is present in the compound as upper unit and if this is the (+) or (−) conformation. (+)isomers would elute earlier in the MS, thus can be distinguished, if (+) and (−) isomers are present (Sánchez-Ilárduya *et al.*, 2012), which is not the case for the candidate compounds. Because in nature (+)-catechin, (−)-epicatechin, (+)-gallocatechin and (−)-epigallocatechin are dominant, it is assumed, that the BGLU3 candidate product compound is either an (+)-gallocatechin or an (−)-epigallocatechin (Hollman and Arts, 2000).

Since, compounds formed by C4–C8 linkage are more probable than C4–C6 linkages, due to higher stabilization of the charge in this formation, this was suggested to be the proposed linkage for the BGLU3 candidate compounds, too (Sánchez-Ilárduya *et al.*, 2012).

Additional support for the F-pyA$^+$ compound provided the UV-vis spectra. Two peaks were visible, for 866 m/z around 280 nm and 344 nm and for 704 m/z around 220 nm and 330 nm. The first peak is known for proanthocyanidins, thus should represent the flavanol unit (Es-Safi *et al.*, 1999; Li and Deinzer, 2007; Lin and Harnly, 2007; Tava *et al.*, 2019). In general, for pyranoanthocyanidins, a shift in the maximum absorption wavelength compared to anthocyanidins is expected, due to the additional pyran ring. For vitisin A hypsochromic spectral shifts from malvidin 3-glucoside of 18-19 nm are known, mostly reported as a peak maximum around 514 nm and being responsible for a more orange hue compared to red (Bakker *et al.*, 1997; Bakker and Timberlake, 1997; Francia-Aricha *et al.*, 1997; Mateus *et al.*, 2002a; Dueñas *et al.*, 2008; Nave *et al.*, 2010). The flavanol unit is not expected, to change the UV-vis spectrum for the pyranoanthocyanidin unit significantly, because the latter is the chromophore group (Nave *et al.*, 2010). This was shown by Nave *et al.* (2010) for (epi)catechin–vitisin A compounds, with an absorption maximum of 514 nm, which is like the expected maximum peak for vitisin A on its own. In general, pyranoanthocyanidins are characterized by this maximum absorption peak between 495 and 520 nm (Bakker *et al.*, 1997; Lu *et al.*, 2000; Mateus *et al.*, 2002a;

Andersen *et al.*, 2004; Fossen *et al.*, 2004; Marquez *et al.*, 2013a). This could also characterize dimeric flavanol anthocynidin and certain flavanol pyranoanthocyanidin compounds, while sometimes for those also a second peak around 425–470 nm is reported (Santos-Buelga *et al.*, 1995; Mateus *et al.*, 2002a; Fossen *et al.*, 2004; He *et al.*, 2010a; Marquez *et al.*, 2013a), causing a shift to higher wavelength, thus, a bathochromic shift, like portisins or other compounds from acetaldehyde-mediated condesation, resulting in a bluish color, with absorption peaks between 530 and 540 nm (Mateus *et al.*, 2002a; Marquez *et al.*, 2013a; Dueñas *et al.*, 2008; Nave *et al.*, 2010; Sánchez-Ilárduya *et al.*, 2012). Since, these wavelengths were not recorded in our measurement, the compounds cannot be evaluated according to this maximum wavelength. Nevertheless, a second characteristic of the pyranoanthocyanidins is the appearance of a new peak. This peak is reported in the absorption area around 352-370 nm, which is proposed to derive by the 4-substituted anthocyanidin and are sometimes reported as very low UV absorption bands (Bakker and Timberlake, 1997; Fulcrand *et al.*, 1998; Fossen and Andersen, 2003; Andersen *et al.*, 2004; Farr *et al.*, 2018). Thus, the second peaks of the candidate compounds could derive from the pyranoanthocyanidin unit. The slightly lower absorption maxima could derive due to the low peaks, hampering precise estimation.

So far, no condensed flavanol–anthocyanidin or flavonol–pyranoanthocyanidin, or pyranoan-thocyanidins in general, are reported for Arabidopsis. A detailed discussion about the so far unknown compounds in Arabidopsis seeds will follow in later sections. Due to the proposed restricted expression of *BGLU3* to seeds, the proposed glycosylation reaction could be re-stricted to seeds, too. Therefore, also the suggested substrate and product compound could be present only in seeds.

The candidate BGLU3 product compound 773 m/z is probably a triglycosylated quercetin, based on the results from the MS/MS spectrum, the UV-vis spectrum and the best database hits. Nevertheless, identification of the compound with the used databases did not yield sufficient results. In addition, as described above for *BGLU1*, it cannot be excluded, that the 303 m/z fragment ion could derive from an anthocyanidin, in this case delphinidin (Sun *et al.*, 2012). As for *BGLU1*, measurements in the negative ion mode, yielded an uncomplete MS/MS spectrum, without the aglycone peak. Thus, the presence of a radical aglycone is not known. In addition, the MS/MS spectrum showed neutral losses of 146 Da and 162 Da + 16 Da in contrast of two 162 Da neutral losses, expected from the positive ion mode. Probably the low peak intensities or non-optimal collision energies in the MRM mode resulted in unexpected fragmentation patterns. Nevertheless, no dominant $[M–2H + H_2O]^-$ peak or double charged peaks in the MS were found, making a flavonol more likely.

This is supported, by the results from the UV-vis spectra, which are in concordance with glycosylated flavonols and suit less for an anthocyanin (Lin and Harnly, 2007; Tava *et al.*, 2019). However, these results should keep in mind, that the compound peak is very small, making the analysis of distinct UV-vis spectra less likely. Probably the spectra are also not very clean, due to a very huge overlapping 611 m/z peak, showing characteristics of a glycosylated flavonol, with the same MS/MS spectrum as 773 m/z, but lacking 162 Da. Measurements with a flat gradient covering 116 min of elution, proved the 611 m/z peak to belong to a different compound, thus not being a fragile fragment of 773 m/z.

The MS/MS spectrum of the candidate compound always yielded a neutral loss of 324 Da, independend of changing collision energies. This neutral loss is expected to derive from two hexoses, which would yield an additional 611 m/z fragment ion, if fragmenting separately. Measurements with a timsTOF™ (Bruker Daltonics; Garabedian *et al.*, 2018) provided such a very small 611 m/z fragment ion peak, probably due to the improved separation technique. An identification was still not possible. Since, this measurement was only performed with one replicate, it can only give a hint. All in all, it can be resumed from these results, that the two expected hexose moieties are in a spatial position where they fragment mostly together.

Apart from two hexoses, the neutral loss of 324 Da can also derive from a caffeoylglucoside (Dai *et al.*, 2017), for example a quercetin 3-*O*-α-[6'''-caffeoylglucosyl-β-(1→2)-rhamnoside] (PubChem, CID: 11274280). Sugar attachments acylated with aromatic or aliphatic acids such as *p*-coumaric, caffeic or ferulic substituents, for example delphinidin 3-*O*-caffeoylglucoside (Pinasseau *et al.*, 2017), are very common for anthocyanidins, too (Tian *et al.*, 2005).

If two hexose or one caffeoylglucoside is attached to C-3 and if the candidate product compound is a quercetin, it is likely, that a deoxyhexose, like rhamnose, is located at the C-7. For the activity of BGLU3 the most probable reactions are either the proposed rhamnosylation at C-7, as seen for other GH1-type GTs (Matsuba *et al.*, 2010; Miyahara *et al.*, 2012) or the attachment of a second hexose to the sugar moiety at C-3, as known for BGLU6 (Ishihara *et al.*, 2016). Because it is likely, that BGLU3 transfers a hexose moiety to the proposed (epi)gallocatechin-carboxypyranomalvidin, it is suggested, that in this case BGLU3 transfers a hexose, too, probably a second hexose at the C-3 position, to yield a gentobiose, as known for BGLU6. In addition, the MS/MS spectrum for the BGLU6 product exhibits the same characteristics, with a very little peak for one hexose loss, but a huge peak for the gentobiose loss. Ishihara *et al.* (2016) suggest for the BGLU6 gentobiose product a common flavonol triglucoside (1→6) glycosidic linkages, which could also be the linkage type for this BGLU3 candidate product compound. Since, the suggested compound looks like the product compound of BGLU6 (figure 37), probably BGLU3 mediates the sugar transfer in another tissue, compared to BGLU6.
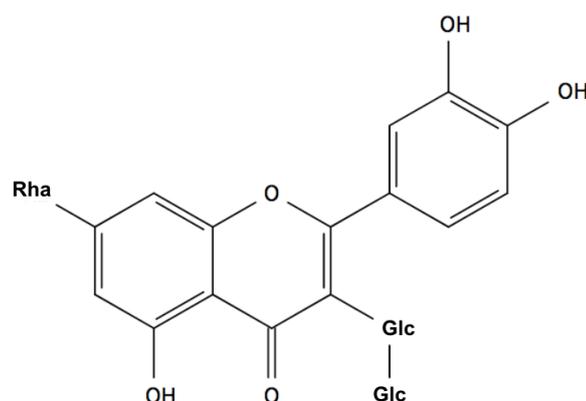
Figure 37 – *Proposed structure of 773.2116 m/z at 10.51 min.* *The proposed structure of 773.2116 m/z at 10.51 min with a gentobiose at C-3 and a rhamnose at C-7. Rha: rhamnose, Glc: glucose*

As already discussed for *BGLU1*, the LC-MS/MS measurement with available standards for

distinct structure elucidation would be useful, but complicated by the few available standards of triglycosylated flavonoids and the huge amount of substitution possibilities. Due to the very small peak, purification of sufficient amounts of the compound out of Arabidopsis seeds for NMR measurements, could be difficult. Nevertheless, growing of plant material for this purpose is in process.

The additional double-charged candidate BGLU3 product compounds are suggested to derive either from further minor compounds differing from the 866 m/z compound by common neutral masses (Dietz *et al.*, 2019), with very small single-charged peaks, being glycosylated by BGLU3. Apart from this, these compounds could derive from substitutions of 866 m/z in further biosynthesis pathways, thus not directly involving BGLU3 activity. In addition, both cases are possible. The first case is supported by the double-charged 411 m/z feature, probably deriving from (epi)gallocatechin–vitisin B, which is 44 Da smaller as (epi)gallocatechin–vitisin A (Bakker and Timberlake, 1997). The respective single-charged substrate feature is 660 m/z. Nave *et al.* (2010) report such a condensed vitisin B compounds in wine together with vitisin A, too. The similar structure of both vitisins make a glycosylation by the same enzyme likely. Due to the fact, that in both MS/MS spectra the single-charged 398 m/z fragment ion is visible, despite the CO_2 group should not be present, the assumption as described for 866 m/z and 704 m/z, that this peak can derive from the 262 Da neutral loss of the partial upper (epi)gallocatechin unit, is supported. The formation of (epi)gallocatechin–vitisin B is supposed to derive by the same mechanism as described above for (epi)gallocatechin–vitisin A, but in this case by cycloaddition of acetaldehyde.

Due to the neutral loss of 52.8 Da between some features, eluting at the same time, an unknown adduct type is possible to appear for some compounds. This is likely for 433 m/z and 460 m/z, described in the results, as well as for 440.7209 m/z at 12.90 min and 467.1764 m/z at 12.89 min. Features with the same m/z but different retention times are believed to derive from isomers or diastereoisomersim, which is known as a typical characteristic for oligomeric proanthocyanidins, thus probably also condensed flavanol–anthocyanidin compounds (Li and Deinzer, 2007).

Apart from the candidate BGLU3 substrate feature 704 m/z, further candidate substrate features exhibit a MS/MS fragmentation pattern indicating condensed pigments, more precisely flavanol–anthocyanidins. The candidate substrate features 781 m/z, 619 m/z and 575 m/z seem to belong to related compounds. 781 m/z, exhibiting the highest intensities, was chosen as the main substrate candidate compound. A minor peak with an unexpected fold change for a BGLU3 product compound was found, which is 162 Da bigger than 781 m/z and seems to be a realted compound due to the MS/MS spectrum. Therefore, it was assumed, that BGLU3 is transferring a hexose moiety to a condensed flavonol–anthocyanidin compound, which already exhibits a hexose moiety, too. The structure of the compound for the feature 781 m/z could not be completely clarified, but there are strong hints for an (epi)catechin–malvidin-glycoside, probably again with a hexose at the C-3 position (figure 38).

The dominant neutral loss of 123 Da from a B ring loss of (epi)catechin, suggested in the results, is described in Flamini (2012). The H_2O loss between 619 m/z and 601 m/z seems to be characteristic for these condensed flavanol–anthocyanidin compounds (Salas *et al.*, 2004;

Dueñas *et al.*, 2006; Lee *et al.*, 2009). The 44 Da neutral loss in the MS/MS spectrum is suggested to derive from an OCH_2 and CH_2 loss from malvidin, but these losses are not found in literature. There is only an (epi)catechin–cyanidin molecule reported by Lee *et al.* (2009), which is 44 Da smaller, thus exhibits 737 m/z. Nevertheless, in literature there are different (epi)catechin–malvidin-glucosides reported, exhibiting a mass of 781 m/z with different fragmentation patterns (Vivar-Quintana *et al.*, 2002; Salas *et al.*, 2004; Dueñas *et al.*, 2006; González-Paramás *et al.*, 2006; Sánchez-Ilárduya *et al.*, 2012), for example one compound, with a dominant 481 m/z fragment ion, deriving from a 138 Da loss, which should account for an atypical linkage (Dueñas *et al.*, 2006) or a compound with the fragment ions 601, 493, 467 and 373 m/z, which were derived from MS^3 (Salas *et al.*, 2004), thus are probably not comparable. Therefore, it is assumed, that the BGLU3 candidate substrate compound could be a condensed (epi)catechin–malvidin-3-*O*-glucoside, deriving from 289 Da ((epi)catechin), 330 Da (malvidin) and 162 Da (hexose). Probably these condensed pigments in literature show slightly differences to each other, like different linkages, resulting in different MS/MS fragmentation patterns.

Because 781 m/z, 619 m/z and 575 m/z and 287 m/z exhibit all substrate behavior, concerning their peak pattern, it is likely, that the lack of further processing of 781 m/z from BGLU3 in the *bglu3-2* line could lead to less processing of the other compounds. In addition, it cannot be excluded, that BGLU3 could be responsible for the probable hexose attachment at 619 m/z, to yield the 781 m/z compound. If 781 m/z is a substrate of BGLU3 and if BGLU3 catalyzes the transfer of a second hexose moiety, the F–A^+ structure is supported, due to probable steric hindrance in the A^+–F molecule (Sentandreu *et al.*, 2012). Such condensed compounds with two hexoses were for example found in pomegranate (Sentandreu *et al.*, 2010). Another option is, that BGLU3 transfers a second hexose at the 5-OH position. Such (epi)catechin–cyanidin diglucosides are reported for example by González-Paramás *et al.* (2006) as an (epi)catechin–cyanidin 3-glucoside-5-glucoside.
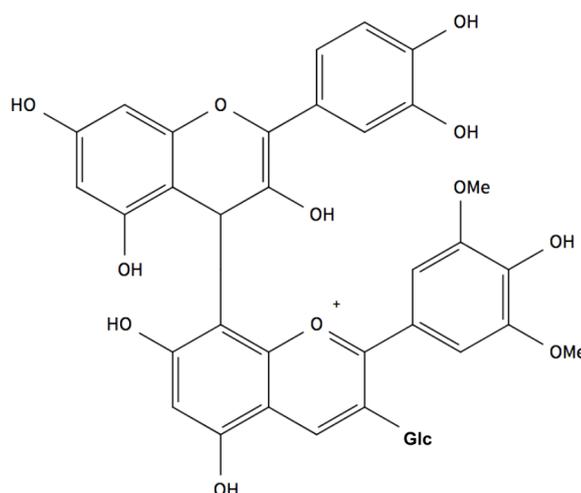
Figure 38 – ***Proposed structure of 781.3270 m/z at 15.38 min.*** *The proposed structure of 781.3270 m/z at 15.38 min for an (epi)catechin–malvidin-3-*O*-glucoside. Rha: rhamnose, Glc: glucose*

The candidate substrate features 633 m/z and 795 m/z could derive as well from related

condensed flavanol–anthocyanidin compounds, separated by 162 Da and 14 Da smaller as the above described compounds. Possibilities are (epi)catechin–petunidin or (epi)catechin–europinidin, while the first one is more likely, because it is one of the major anthocyanidins in plants (Tanaka *et al.*, 2008). A product feature for this compound was not found.

Summarizing the results for BGLU3, there is a strong hint, that BGLU3 could act as a GH1-type GT, glycosylating different flavonoid compounds from at least three different flavonoids subgroups, namely flavanols (including kaempferol and quercetin, based on the results for the single-charged 757 m/z candidate BGLU3 product compounds), condensed flavanol—anthocyanidins and flavanol—pyranoanthocyanidins, the latter two reported for the first time in Arabidopsis.

Most known GH1-type GTs are reported for their strict substrate preferences (Matsuba *et al.*, 2010; Miyahara *et al.*, 2012, 2014), but there are reports for other GH1-type GTs, using several substrates, like BGLU6, glucosylating kaempferol (K3GG7R), quercetin (Q3GG7R) and isorhamnetin (I3GG7R; Ishihara *et al.*, 2016) and Os9BGlu31 from rice, accepting phenylpropanoids, flavonoids and phytohormone glycoconjugates as acceptors (Luang *et al.*, 2013). There are also hints for BGLU7 and BGLU8 to accept flavonols and anthocyanidins (Wu *et al.*, 2018). Ketudat Cairns *et al.* (2015) state this multifuctionality to be possibly a characteristic trait of GH1 enzymes, due to the fact that there seem to exist much more glycosylated compounds, than GH1 enzymes (glycosyltransferases and hydrolases) are known. This could be a mechanism to diversifly respond to different environmental factors. In addition, for some enzymes, different work groups identified different functions, as seen for BGLU10, which could be explained by this hypothesis, too.

For the candidate product compound 866 m/z the expression and metabolite levels between wild type and overexpression were less differential, than for candidate products of BGLU1 and BGLU4. This supports the statement of Ishihara *et al.* (2016), proposing, that for some flavonoids the accumulation of a certain level is prevented, thus probably in the *2x35S:BGLU3* overexpression line. One could the other way around also propose, that for some compounds a certain treshold of compound is always ensured, as seen for the candidate BGLU3 substrate compound 704 m/z in the wild type and the overexpression mutant. The relatively high expression of *BGLU3* in *2x35S:BGLU3* compared to the wild type as shown by the qRT-PCR results, can be explained by the use of seeds, one year older compared to the HPLC and RT-PCR samples. Probably, in the wild type, BGLU3 expression decreases in older seeds.

Studies of the *Brassica rapa BGLUs* by Dong *et al.* (2019), showed an upregulation of *BrBGLU15* and *BrBGLU16* in fertile buds, which are considered as orthologs of *AtBGLU3* and *AtBGLU4* respectively. The authors suggest an involvement of these genes in pollen development, as shown for *BrBGLU10/AtBGLU20*. For *BGLU3* and *BGLU4* this is not further investigated. Nevertheless, those genes could be involved in similar or same functions. Due to very low peak intensities of 773 m/z and 757 m/z and to some extend also 866 m/z, it was considered to investigate, if these peaks appear in rapeseeds, too. Rapeseeds can be ordered in huge amounts and show much more biomass, than Arabidopsis seeds. For NMR measurements, where huge amounts of sample are necessary for the compound purification of minor compounds, the experiment with Arabidopsis seeds will be challenging. Therefore, it

was considered, if these compounds could be isolated from rapeseed. Nevertheless, due to the high isomerism of glycosylated flavonoids, the attempt to identify the compounds in other organisms like rapeseed, need to be regarded with care, because, it is almost not possible, to determine, if the two compounds are the same, unless, NMR is performed. Despite, if experiments in this direction are planned, *BrBGLU15* and *BrBGLU16* are strong candidates in rapeseed for the homologous function of *BGLU3* and *BGLU4*, making the appearance of the BGLU3 candidate compounds in rapeseed possible (Dong *et al.*, 2019).

Expression analysis of Cao *et al.* (2017) did not reveal *BGLU3* expression in seeds, but for those *BGLUs* expressed in this tissue, they postulate a potential role in seed establishment or germination, which could be true for *BGLU3*, too and will be discussed below.

4.1.3. BGLU4 as a possible flavonol GH1-type glycosyltransferase

The results for the candidate product compound 757 m/z of BGLU4 hint towards a triglycosylated quercetin, including data from the MS/MS spectrum, the UV-vis spectrum as well as the results from the best database hits. A distinct identification of the compound structure was not possible, as described for the other putative flavonol compounds of *BGLU1* and *BGLU3*. Due to the fact, that the peak is very small, the proposition for a glycosylated flavonol from the UV-vis spectrum (Lin and Harnly, 2007; Tava *et al.*, 2019) are to be handled with care, too. As described for *BGLU3* the 303 m/z fragment ion peak could also account for a delphinidin backbone. Thus, further investigations as described above, were done in the negative ion mode, yielding a clear MS/MS spectrum without a radical aglycon and no dominant $[M–2H + H_2O]^-$ peak or double charged peaks in the MS, supporting the presence of a flavonol.

The MS/MS spectrum in the positive ion mode, did only reveal a neutral loss of 308 Da, apart from the 146 Da neutral loss, supposingly deriving from one hexose and one deoxyhexose, without information about the fragmentation order. In addition, the neutral loss of 308 Da could also account for a *p*-coumaroylglucoside (Sánchez-Ilárduya *et al.*, 2012). The measurement in the negative ion mode, provided more information, giving a fragmentation pattern with the neutral loss order of 146 Da, 162 Da and again 146 Da, being in favor of a deoxyhexose and a hexose moiety, instead of a *p*-coumaroylglucoside. Even, if no definite information about the sugar attachment sites are provided from the data, the fragmentation order is the same as described by Tohge *et al.* (2005) for the 3-*O*-[2"-*O*-(rhamnosyl) glucoside]-7-*O*-rhamnoside, mentioned already for BGLU1. For the respective diclycoside of 595 m/z, Kerhoas *et al.* (2006) suggest the hexose attachment to the C-3 and the deoxyhexose attachment to the C-7. Transferring this to our candidate product compound, a second deoxyhexose could be attached to the hexose at C-3 (figure 39). This would suggest the transfer of a deoxyhexose, possibly a rhamnose, to the C-7 position or to the hexose moiety at C-3, as described for BGLU1, for BGLU4. These suggestions are highly hypothetical and there exist many other possibilities for the attachment of the sugars, as well as for the sugar types. To really identify the spatial structure, it is necessary to either measure standards, which are difficult to obtain for triglycosylated flavonoids and which is challenging, due to the vast amount of glycosylation possibilities. Alternatively, it would be necessary to isolate the compound and perform NMR

measurements. This could be very challenging, due to the very small peak of 757 m/z, suggesting low compound concentrations and making isolation of the needed amount of pure compound in Arabidopsis seeds difficult. As described for *BGLU1*, there is also an option to perform a specific fragmentation protocol, but due to the low peak intensities and the difficulty to derive a purer and more concentrated compound, it could be also very challenging and time consuming. As described for *BGLU3* the isolation of the compound from other species, where seeds are much better to achieve, like for example rapeseed with the candidate *BGLU4* homolog *BrBGLU16*, could be considered. But as already mentioned, this approach is very sensitive to errors and false positives.
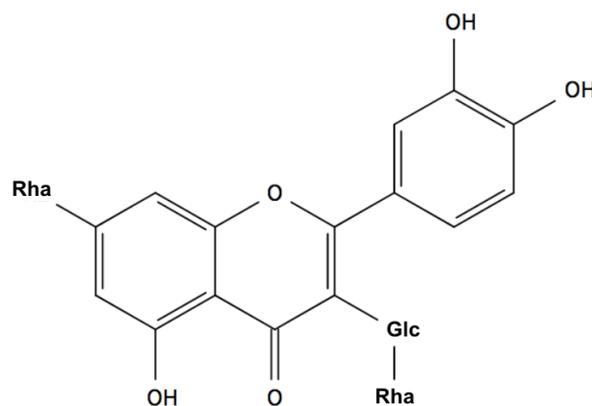
Figure 39 – *Proposed structure of 757.2149 m/z at 11.71 min.* *The proposed structure of 757.2149 m/z at 11.71 min with a glucose at C-3, substituted with a rhamnose and another rhamnose at C-7. Rha: rhamnose, Glc: glucose.*

Findings of Jeong *et al.* (2018) give valuable hints towards the possible function of *BGLU4*. In the work, they investigated the regulation of genes in response to sucrose. Sucrose is a crucial compound for the growth and development of plants and it is known to affect flavonoid and anthocyanidin accumulation (Jeong *et al.*, 2018, and references therein). They identified the transcription factor AtMyb56, to be involved in the response to high sucrose levels, by probably regulating the sucrose-induced *AtGPT2* expression in response to the circadian cycle. GPT2 is a plastidial glucose-6-phosphate/phosphate translocator (GPT) protein, mediating the transport of glucose-6-phosphate from the cytosol to plastids for starch biosynthesis. The activity alters the level of free maltose and thereby the accumulation of anthocyanidins in plants. In addition, flavonols were shown to accumulate after sucrose treatment. The authors could show, that several genes were upregulated in the *atmyb56* mutant after sucrose treatment, including *BGLU4*, with a 10-fold upregulation compared to the wild type. In the mutant the sucrose response is expected to be disturbed, causing osmotic stress. *BGLU4* could be involved in the osmotic stress response. (Jeong *et al.*, 2018). The expression of *BGLU4* in soaked seeds, which probably are naturally exposed to osmotic stress, would support the functional hypothesis for *BGLU4*. Probably, BGLU4 is involved in stress response during germination. As described for BGLU3, Cao *et al.* (2017) postulate a potential role in seed establishment and germination for *BGLU* genes, expressed in seeds. The latter supporting the functional hypothesis for *BGLU4*, being predominantly expressed in soaked seeds. Since, Jeong *et al.* (2018) investigated the upregulation of *BGLU4* expression in roots, probably

the product compound of BGLU4 is more abundant in different Arabidopsis tissues after osmotic stress treatment. Therefore, it would be promising to investigate the behaviour of *BGLU4* expression in response to stress and investigate in parallel differential metabolite levels. Presumably, under certain conditions the metabolite level is significantly increased, making compound purification simpler.

As described for BGLU1, no candidate product compounds were found, lacking the peaks of the *bglu4-2* mutant. The RT- and qRT-PCR results did not account for whole transcript presence. As described above, those findings are in concordance with findings from Wu *et al.* (2018) for BGLU1 and Miyahara *et al.* (2013) for BGLU10, suggesting for BGLU4 functional redundancy. Based on amino acid sequence homology, BGLU5 would be the closest homolog, possibly accounting for the functional redundancy. Attempts, to generate homozygous double mutants of *bglu4-1* and *bglu5-1* were not successful, probably, due to the close location on the same chromosome, making the chance for a cross-over less likely. Based on the expression data from RNA-Seq results at the TraVa database, *BGLU3* is also a candidate for functional redundancy, due to the strict expression in seed. In addition, BGLU3 seems to glycosylate to some extend an isomer of the candidate BGLU4 product compound, which is 757.2166 m/z at 11.42 min, making glycosylation of the BGLU4 candidate product by BGLU3, in case of a BGLU4 activity loss, even more likely. Since, this isomer exhibits a certain metabolite level in the *bglu3-2* sample, too, probably *BGLU4* could account for this reaction in case of the activity loss of BGLU3. To get more insights into this functional redundancy, homozygous double mutants of *bglu3-2* and *bglu4-2* were generated by crossing. In future experiments, these double mutants should be investigated on metabolite levels of the respective candidate product compounds of BGLU3 and BGLU4. In general, the investigation of functional redundancy of the highly homolog BGLUs would profit of several double or multiple mutants for the members of this gene family, which is difficult to obtain by crossing, especially, if the *BGLU* genes are in close proximity, as described for *BGLU4* and *BGLU5*. Mutant generation with CRISPR-Cas9 could be a possibility for faster progress (Miao *et al.*, 2013).

One candidate substrate compound was found for BGLU4. This compound shows only slightly higher peaks in the *bglu4-2* line, but since there are also peaks for the putative product compound, it is likely, that the substrate is still consumed by another enzyme, due to the postulated functional redundancy. In addition, as described for BGLU1 and BGLU6 (Ishihara *et al.*, 2016), the substrate levels seem to be in some cases independent from the enzymes activity.

4.2. Further candidates of GH1-type glycosyltransferases in Arabidopsis

According to phylogenetic analysis on amino acid level for all 47 BGLUs in Arabidopsis and all characterized GH1-type GTs from other organisms, BGLU1 to BGLU11 could exhibit GH1-type GT activity. For BGLU6 and BGLU10 this was already shown (Miyahara *et al.*, 2013; Ishihara *et al.*, 2016). The aim of this work was the functional characterization of BGLU1 to BGLU5, clustering around BGLU6. For BGLU1, BGLU3 and BGLU4 the results indicate, that those enzymes should perform GH1-type glycosylation, too. BGLU2 and BGLU5 were excluded from further analysis, which will be discussed later. BGLU7 to BGLU9 cluster

around BGLU10, while BGLU11 behaves like an outlier in the cluster (figure 40). For BGLU7
and BGLU8, Wu *et al.* (2018) provided further insights into the functional characterization
by untargeted metabolomic fingerprinting and metabolic genome-wide association studies, as
performed for BGLU1. For both genes, *BGLU7* and *BGLU8*, three associated metabolites were
discovered. Two of them exhibit 741 m/z in the positive ion mode, as for the candidate product
compound of BGLU1, but with different retention times. Due to the provided EIC data,
they could account for the small peaks, eluting shortly after the *BGLU1* 741 m/z candidate
product peak, again close to the proposed kaempferol 3-*O*-[2"-*O*-(rhamnosyl) glucoside]
7-*O*-rhamnoside. These metabolites co-mapped, as for *BGLU1*, with the *UGT78D1* locus,
encoding the UDP-rhamnose transferase, which is active at the 3-OH position of kaempferol
and quercetin (Jones *et al.*, 2003). Compared to BGLU1, *BGLU7* and *BGLU8* could be
involved in the glycosylation of the same compound at other positions, like the expected
C-7 hexosylation, or they could glycosylate the same compound at the same position, but in
different compartments, as discussed in more detail in later sections.
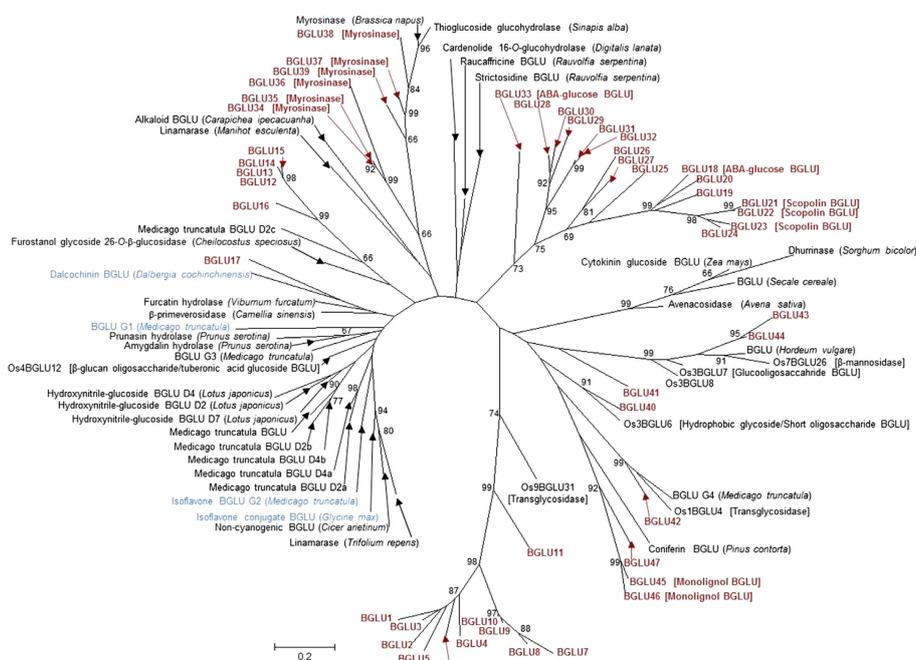
Figure 40 – ***Phylogenetic tree of BGLUs.*** *Phylogenetic tree on amino acid level for BGLUs from different organism based on Roepke and Bozzo (2015).*

In addition, they could glycosylate 741 m/z compounds with differing glycosylation patterns.
Different from the hypothesis, that BGLU7 to BGLU11 glycosylate anthocyanidins (Ishihara
et al., 2016), BGLU7 and BGLU8 seem to glycosylate flavonols, too. Moreover, the third
associated metabolite co-mapped to the locus of *UGT79B2* and *UGT79B3*, encoding again
UDP-rhamnose transferases, but using cyanidin and cyanidin 3-*O*-glucoside as acceptors (Li
et al., 2017). The respective metabolite exhibits 903 m/z in the positive ion mode and is,
according to the mass, expected to exhibit in addition to the probable hexose and rhamnose
another hexose and deoxyhexose, if cyanidin is the aglycone. *BGLU7* and *BGLU8* could be
involved in this glycosylation. Therefore, they should glycosylate anthocyanidins and flavonols.

Such an UGT that is accepting flavonols and anthocyanidins, was also reported by (Tohge *et al.*, 2005) and by Meng *et al.* (2019). In addition, a broader substrate acceptance was reported for other GH1-type GTs as discussed above. Despite, members of the cluster BGLU7 to BGLU11 seem to glycosylate flavonols, the glycosylation of anthocyanidins by BGLU1 to BGLU6 was not shown, so far, if accounting pyranoanthocyanidins as different compounds. *BGLU1*, *BGLU6*, *BGLU7* and *BGLU9* are suggested to encode for proteins like GH1-type GTs, involved in the anthocyanidin biosynthesis or their glycosylation, by Miyahara *et al.* (2011). All of them were expressed at higher levels in anthocyanin-inducing medium, leading to this assumption. For *BGLU1* and *BGLU6* this was not shown so far and the results from Ishihara *et al.* (2016) and this work indicate rather the glycosylation of flavonols. BGLU9 is suggested by Miyahara *et al.* (2013) to act as a GH1-type GT like BGLU10, accounting for the postulated functional redundancy in the *bglu10* mutant.

BGLU11 clusters with BGLU1 to BGLU10 but ranges a little bit like an outlier in this group, clustering with the hydroxyisourate hydrolase (HIUH-ase) from *Glycine max*, an ureide-metabolizing enzyme found in soybean root nodules, leaves, stems, and roots (Raychaudhuri and Tipton, 2002; Xu *et al.*, 2004). Homologs were also found in rice (Opassiri *et al.*, 2006) and maize (Gómez-Anduro *et al.*, 2011) BGLUs. Thus, BGLU11 could be like an outranging candidate in the BGLU1 - BGLU11 cluster, acting rather as a hydrolase. Nevertheless, phylogenetic analysis of Roepke and Bozzo (2015) shows homology to the rice BGLU Os9BGlu31 (figure 40), which is characterized as transglycosidase, meaning the enzyme exhibits GH1-type GT activity. This enzyme shows also little hydrolase activity, thus probably, BGLU11 could catalyze both reactions. This was also shown for BGLU10 (Wang *et al.*, 2011; Miyahara *et al.*, 2013), as discussed below.

Summarizing these findings, there exist strong hints, that the cluster of BGLU1 to BGLU11 is built by GH1-type GTs, also if for some members functional characterization needs to be done and in addition, enzymatic studies for BGLU1, BGLU3 and BGLU4 need to confirm their activity.

4.2.1. GH1-type GT candidates *BGLU2* and *BGLU5*

The discovery of the 47 *BGLU* genes in Arabidopsis by Xu *et al.* (2004) was based on the one hand on the predicted amino acid sequences. Apart from this, they examined the genes based on the intron–exon organization. The typical structure was 13 exons and 12 introns, or 12 exons and 11 introns, accounting for most of the *BGLUs* and being very similar to the intron–exon structure of later examined *BGLUs* from rice, maize and rapeseed, predominantly exhibiting 12 exons and 11 introns (Xu *et al.*, 2004; Opassiri *et al.*, 2006; Gómez-Anduro *et al.*, 2011; Dong *et al.*, 2019).

For the candidate GH1-type GTs examined in this work *BGLU3*, *BGLU4* and *BGLU5* exhibit 13 exons and 12 introns, same as *BGLU6*. *BGLU1* exhibits in this work 14 exons and 13 introns, while TAIR shows several intron–exon structures. *BGLU2* exhibits only 8 exons and 7 introns. According to Xu *et al.* (2004), there exist 8 probable pseudogenes in the 47 predicted *BGLU* genes of Arabidopsis, namely *BGLU1*, *BGLU2*, *BGLU5*, *BGLU6*, *BGLU14*, *BGLU36*, *BGLU39* and *BGLU43*, based on the intron–exon structure and expression data, suggesting

the lack of key motifs essential for *β*-glycosidase activity in the predicted polypeptide product, due to reading frame shifts and/or premature translation termination. For *BGLU6* it was shown by Ishihara *et al.* (2016), that the gene exhibits a premature translation termination in some Arabidopsis accessions, including Col-0, transforming *BGLU6* in these accessions into a pseudogene, as defined by (Snyder and Gerstein, 2003). Nevertheless, in other accessions, like Ler, *BGLU6* is a functional gene and could be characterized, to encode a GH1-type GT (Ishihara *et al.*, 2016). Due to the atypical intron–exon structure of *BGLU2*, lacking the first four exons and introns of a typical *β*-glycosydase gene, and because almost no expression could be shown by RNA-Seq, according to the TraVa database, *BGLU2* was proposed to exhibit no functional protein product, thus being a real pseudogene. Therefore, the gene was not further investigated in this work. Because *BGLU1* and *BGLU5* exhibit both clear expression according to RNA-Seq data from the TraVa database and a normal intron–exon structure, they were further investigated. Miyahara *et al.* (2011) could show a complete open reading frame for *BGLU1* cDNA, in concordance with the result of this work and was therefore postulated not to be a pseudogene. At the end, the results from this work indicate the functionality of *BGLU1*. According to TAIR, there exist different transcripts for *BGLU1*, if one of these transcripts accounts for a pseudogene, and was observed from Xu *et al.* (2004), this could be an explanation for the differing results.

In contrast, expression of *BGLU5* did always reveal intron number 8, without further splicing, reducing the 13 exons to 12 exons. This introduces a premature translation termination, which led to the assumption, that *BGLU5* is either a pseudogene in Col-0 (Snyder and Gerstein, 2003), exhibits another function, or is only functional after alternative splicing (**?**), probably occurring in certain conditions. Due to uncertainties about the functionality of this gene, further functional investigation would imply a more complex approach, compared to the other three *BGLUs* in the investigated cluster. Therefore, *BGLU5* was not further analyzed by metabolic fingerprinting. In comparison to BGLU6, which is a pseudogene in some accessions, in expression analysis of Miyahara *et al.* (2011), the promotor of *BGLU6* was active, although the cDNA contains a premature stop codon in the middle of the open reading frame. Therefore, it was postulated, that *BGLU6* evolved recently as a pseudogene by a point mutation. This could also be the case for *BGLU5*, showing still strong expression according to the TraVa database and the RT-PCR results from this work. It would be interesting to investigate the situation for *BGLU5* also in other accessions, probably finding accessions with a functional *BGLU5* gene for further investigations.

4.2.2. GH1-type GTs from other organisms

The first two discovered GH1-type GTs were discovered in carnation and in delphinium by Matsuba *et al.* (2010). Both enzymes did glycosylat anthocyanins in crude protein extract from the respective plants with the purified recombinant protein, showing the 5-*O*-glucosylation reaction of anthocyanin-3-*O*-glucoside in carnation and the 7-*O*-glucosylation reaction of anthocyanin-3-*O*-glucoside in delphinium, mediated by the new discovered enzymes. Special about both enzymes was on the one hand the homology to GH1 *β*-glycosidases on amino acid level and on the other hand, the use of acyl-glucose instead of UDP-sugars as sugar-donors, 1-

O-*β*-D-vanillyl-glucose in this precise case. Therefore, these enzymes were termed acyl-glucose dependent anthoyanin glycosyltransferases (AAGTs). In case of delphinium the AAGT was the first GT, known to glycosylate the C-7 position of anthocyanin-3-*O*-glucoside (Miyahara *et al.*, 2012).

In the following years, more of these novel glycosyltransferases were dicovered. Two of them, from African lily and Canterbury bells, further AAGTs glycosylating at C-7 of anthocyanin-3-*O*-glucoside or in case of Canterbury bells 3-*O*-rutinoside, suggesting that AAGTs have a functional specialization for this C-7 glycosylation reaction. In contrast, for flavonols UGTs are known for the C-7 glycosylation reaction (Miyahara *et al.*, 2012, 2014).

Another type of GH1-type GTs was discovered by Nishizaki *et al.* (2013), again in delphinium. Similar to BGLU10 in Arabidopsis, the glycosylation is not performed directly on the aglycone or on another sugar moiety, but the sugar is transferred to a *p*-hydroxybenzoic acid (pHBA) molecule of delphinidin 3-*O*-rutinoside-7-*O*-(6-*O*-(*p*-hydroxybenzoyl)-glucoside), to form the two 7-polyacylated anthocyanins, violdelphin and cyanodelphin. Thus, the enzymes were called AA7BG-GT1 (DgBGLU7) and AA7BG-GT2 (DgBGLU10) respectively, while the abbreviation AA7BG-GT stands for acyl-glucose dependent anthocyanin 7-*O*-(6-*O*-(*p*-hydroxybenzoyl)-glucoside) glucosyltransferase (Nishizaki *et al.*, 2013). Their function could be confirmed by a natural homozygous double knockout mutant for the AA7BG-GT genes, with a premature translation termination (Ishii *et al.*, 2017).

Another type of GH1-type GT is described as transglucosidase (TG) for rice, termed Os9BGlu31 (Luang *et al.*, 2013). This enzyme also shows high sequence similarity to GH1 enzymes. Like *BGLU4*, *Os9BGlu31* shows highest expression in germinating seeds after 12 and 24 h. In addition, the expression is increased under drought stress, thus probably it is involved in the adaptation to environmental changes, similar to *BGLU4*.

4.3. The function of GH1-type GTs in plants

GH1-type GTs belong to a big gene familiy, present in many different organisms. It is known, that there exist a broad range of *β*-glycosidases, exhibiting different sugar specificities, probably to fulfill diverse functions. Therefore, it can be assumed, that there exist also GTs with a broad range of sugar donor or acceptor specificity, including GH1-type GTs (Miyahara *et al.*, 2011). Since, there exist also UGTs, it is of interest, to understand the different functionality of GH1-type GTs.

In the beginning, all GH1-type GTs were shown to exhibit strict acceptor recognition, as AAGTs from carnation, delphinium, African lily and Canterburry bells, suggesting GH1-type GTs being involved in very specific reactions (Matsuba *et al.*, 2010; Miyahara *et al.*, 2012, 2013; Nishizaki *et al.*, 2013; Miyahara *et al.*, 2014). On the other hand, some GH1-type GTs seem to exhibit a much broader acceptor substrate specificity, like Os9BGlu31, to some extend BGLU6 and also very likely BGLU7 and BGLU8 as well as BGLU3, as shown in this work (Luang *et al.*, 2013; Ishihara *et al.*, 2016; Wu *et al.*, 2018). Since, for Os9BGlu31 the acceptors a highly diverse, including free phenolic acids, phytohormones, and flavonoids, the broad range of substrate tolerance was proposed as a functional mechanism, to equilibrate the free phenolic acid and phenolic acid conjugate levels in plants (Luang *et al.*, 2013).

With these characteristics, GH1-type GTs and UGTs are not really distinguishable, since, also for UGTs exist some enzymes for very specific reactions, like aglycone glycosylation in Arabidopsis from a flavonol 7-*O*-glucosyltransferase (UGT73C6; Jones *et al.*, 2003), a flavonol 7-*O*-rhamnosyltransferase (UGT89C1; Yonekura-Sakakibara *et al.*, 2007), a flavonoid 3-*O*-glucosyltransferase (UGT78D2; Tohge *et al.*, 2005), a flavonol 3-*O*-rhamnosyltransferase (UGT78D1; Jones *et al.*, 2003), a flavonol 3-*O*-arabinosyltransferase (UGT78D3; Yonekura-Sakakibara *et al.*, 2008) and an anthocyanin 5-*O*-glucosyltransferase (UGT75C1; Tohge *et al.*, 2005). Specific glycosylation of other glycoside moieties at C-3, exist for example from a flavonoid 3-*O*-glucoside 2"-*O*-glucosyltransferase (UGT79B6) in pollen of Arabidopsis (Yonekura-Sakakibara *et al.*, 2014), or UGT707B1 from *Crocus sativa* (Trapero *et al.*, 2012), UGT3 from *Catharanthus roseus* (Masada *et al.*, 2009) and NSGT1 in tomatoes Tikunov *et al.* (2013). Such a regioselectivity for UGTs was also demonstrated with different coumarins (Lim *et al.*, 2003) and predicted for flavonoids (Jackson *et al.*, 2011), showing, that these enzymes predominantly transfer the sugar to very specific positions, if multiple possible sugar binding sites are presented from an acceptor. Investigations of Vogt and Jones (2000) towards this regioselectivity showed, that many glycosyltransferases are regioselective or regiospecific rather than highly substrate specific, indicating, that the glycosylation position is more crucial for the specificity, than the precise substrate, accounting for a broad substrate specificity. In combination with GHs, accepting mostly also a broad range of substrates, the turnover of glycosylated metabolites is highly enhanced and diversified. This might be the mechanism, by which the evolution of novel secondary products can be achieved (Vogt and Jones, 2000; Tanaka *et al.*, 2008; Le Roy *et al.*, 2016). If GH1-type GTs are also regioselective, the glycosylation of BGLU3 candidate product compound by BGLU3 at the same positions in the molecule are likely, thus probably glycosylation at C-3 or another sugar moiety at C-3. Apart from BGLU3, no GH1-type GT is known so far, to assumably catalyze the first sugar transfer to the aglycone, as shown for many UGTs. Therefore, a tendency for GH1-type GTs to glycosylate flavonoid glycosides, can be seen. It can be assumed, that genes of GH1-type GTs belong predominantly to *late-acting* flavonoid biosynthetic genes, as shown for *AA7GT* from African lily, being active later in the flavonoid biosynthetic pathway and also more late in the development (Miyahara *et al.*, 2012). This is also supported by Zhang *et al.* (2014), suggesting, that GH1-type GTs are likely to act late in the flavonoid biosynthesis, probably after transportation of already decorated flavonoids in the vacuole. These suggestions are based on the findings of the first AAGTs in carnation and delphinium, suggesting in addition, restriction to specialized decoration. The assumption of GH1-type GTs acting in vacuoles could be verified for many enzymes, with exception of BGLU6, making a functional localization in the vacuole likely, as discussed later.

Therefore, the specific activity of GH1-type GTs could probably be based on certain compartments, rather than substrates. In addition, investigations for BGLUs, including putative GH1-type GTs, revealed *BGLU* expression in many different tissues, induced or increased by different factors. Specificity of *BGLUs* was predominantly achieved by tissue specificity and in addition, differential stress response (Cao *et al.*, 2017). Probably, individual tissues express their specific GTs in response to certain environmental conditions.

It is known, that the accumulation of glycosylated flavonoids is stress induced (Olsen *et al.*, 2009), supposing a function in protection against stress derived damages. Especially for flavonol triglycosides, a functional role in UV protection is proposed, which could include ROS scavenging activity, while triglycosylated flavonols with 7-*O*-rhamnosylation, accumulating after UV stimulation, seem to be involved in the long-term UV acclimation. It is postulated, that this 7-*O*-rhamnosylation is specifically catalyzed by UGT89C1, as a time-dependent response to certain environmental conditions. (Hectors *et al.*, 2014; Ishihara *et al.*, 2016). In Arabidopsis, flavonols are present as di- or triglucosides (rhamnose and glucose) in almost all tissues (Saito *et al.*, 2013), probably involved in the described UV protection, but very likely exhibiting even more functions.

In general, even if most structures of the main flavonoid mono-, di-, or triglycosides are known so far, the function, which is related to the structure is mostly unknown. In addition, many genes involved in the glycosylation pathway, are still unknown. Therefore, enhanced functional information about flavonoid glycosides is needed, to understand the function of the catalyzing enzymes and therefore, the encoding genes in more detail (Pollastri and Tattini, 2011; Ishihara *et al.*, 2016). Stress response is the postulated main function for glycosides, while the precise mechanism is often unknown. Apart from ROS scavenging, for volatile phenylpropanoids for example, it is also known, that triglycosides are more abundant in ripe fruits, preventing the release of the volatiles after tissue damage, due to a more complicated cleavage, showing the involvment of sugar moieties in compound stability Tikunov *et al.* (2013).

The involvement of glycosylation in the stress response could explain, why many unknown minor glycosylated flavonoid compounds in plants are present, as also shown in this work. Each specific stress condition would probably lead to higher expression of certain genes, encoding proteins for certain substitutions, resulting in higher compound concetration of certain secondary metabolites, as also shown by Wu *et al.* (2018) for *BGLU1*. This stress response was also investigated by Cao *et al.* (2017), suggesting, that each *BGLU* could account for a distinct role in plants, according to the respective environmental condition. For the *BGLUs* of Arabidopsis, investigated in this work, in seedlings *BGLU1* was shown to probably participate in salt response. Apart from this, BGLU1 is probably involved in the light and heat stress response, as indicated by results of Wu *et al.* (2018). In addition, Cao *et al.* (2017) could show that *BGLU3*, *BGLU6* and *BGLU11* were upregulated under salt stress. *BGLU7* and *BGLU9* were up-regulated under cold stress. Because the transcriptional regulation of genes is highly influenced by the 5'-flanking regions, mostly due to sequence elements in the few hundred base pairs upstream of the transcription start, Cao *et al.* (2017) investigated stress-responsive *cis* elements in this regions for *BGLUs*. Several elements were found for the *BGLUs* investigated in this work, which are one ABA related G-box for *BGLU1*, ABA and GA related elements as well as one element related to anaerobic stress for *BGLU3*, GA related elements and elements related to anaerobic and heat stress for *BGLU4* as well as ethylene related and anaerobic and heat stress related elements for *BGLU5*. This supports the assumption, that GH1-type GTs could be involved in stress response. Such *cis* elements were also investigated for *BGLUs* in maize, finding several elements, like auxin, salt-induced, heat shock, jasmonate signalling and antioxidant response elements. Nevertheless, the expression data suggest, that ther should

exist additional *cis* elements and unknown factors that determine *BGLU* gene transcription (Gómez-Anduro *et al.*, 2011).

Even if no GH1-type GTs are known in maize, the homology to BGLUs could yield further insights into the purpose of the novel GTs. Maize *BGLUs* too, show characteristic expression response due to tissue specificity and response to different abiotic conditions, unrelated to different genetic background, while highest gene expression variance was seen for tissues. This could led to the assumption, that huge *BGLU* gene families are related to tissue specific expression of different homologous genes. Because environmental variance has also more impact as genetic background, different laboratories can have completely different results for the same gene, just due to environmental factors, which could be an explanation for varying expression patterns, too (Gómez-Anduro *et al.*, 2011). Such tissue specificity could also be shown for flavonol glycosides, as well as a temporally and developmentally dependent distribution of flavonol glycosides (Routaboul *et al.*, 2006; Stracke *et al.*, 2010).

In summary, differential function of GH1-type GTs compared to UGTs could be related to tissue specificity, environmental conditions and also time- and development dependent specificity. As stated by Le Roy *et al.* (2016), the major challenge for future research will be, to obtain more knowledge and thus better understanding about the individual regulation of the multigenic GT/GH BGLU familie during development and also in response to abiotic and biotic stress.

Compartmentation in the vacuole

Because there exist accession specific flavonoid glycosides, like kaempferol 3-*O*-rhamnosyl-glucosylglucoside-7-*O*-rhamnoside in C24 (Stobiecki *et al.*, 2006) or kaempferol-hexosylhexoside-deoxyhexoside in Ms-0 (Tohge and Fernie, 2010) and therefore accession specific GTs, like BGLU6 (Ishihara *et al.*, 2016), as well as tissue specific flavonoids, it is likely that also tissue specific enzymes exist, as discussed before and supported by the presence of a pollen-specific 3-*O*-glucoside:2''-*O*-glucosyltransferase in Arabidopsis (Stracke *et al.*, 2010; Yonekura-Sakakibara *et al.*, 2014). Therefore, also organelle specific flavonoid glycosides and the respective modifying enzymes could exist.

All known GH1-type GTs exhibit a predicted putative signal peptide to the vacuole at the N-terminus (Matsuba *et al.*, 2010; Miyahara *et al.*, 2012; Luang *et al.*, 2013; Miyahara *et al.*, 2013; Nishizaki *et al.*, 2013; Miyahara *et al.*, 2014; Ishihara *et al.*, 2016). Most enzymes, where subcellular localization studies with a GFP-reporter were performed, were detected in the vacuole of onion epidermal cells after microprojectile bombardment, including AA7GT, AA7BG-GT1 and AA7BG-GT2 of delphinium and to some extend AA5GT from carnation (Matsuba *et al.*, 2010; Nishizaki *et al.*, 2013). Os9BGlu31 from rice was localized in the vacuoles in maize mesophyll protoplasts after polyethylene glycol-mediated transformation and in rice calli, whereas the mesophyll protoplasts did only yield poor signals (Luang *et al.*, 2013). A clear outlayer is BGLU6, which was localized in BY2 protoplasts only in the cytoplasm, despite the putative transit peptide for vacuolar localization at the N-terminal amino acid sequence (Ishihara *et al.*, 2016). Localization in the cytoplasm was also shown for BGLU1, BGLU3 and BGLU4 in previous studies (Frommann, 2016; Meistrowitz, 2016). In this work,

several attempts to generate a vacuolar GFP based control for subcellular localization in the vacuole of BY2 protoplasts, were not successful, including GRP5, used as control in onion epidermal cells by Matsuba *et al.* (2010). In addition, the repeated subcellular localization of BGLU4 revealed subcellular localization of the protein in the assumed vacuole. Tamura *et al.* (2003) investigated, that GFP in vacuoles of higher plants can result in problems by exposure to light, due to degradation of GFP in the acidic pH of vacuoles. Therefore, the lack of vacuolar localization with GFP could derive from a technical effect, making the use of other controls and/or other organisms for transient expression necessary. For example a control of vacuolar membarens could be a more sufficient solution, as described in Appelhagen *et al.* (2015). In addition, the different results suggest, that BGLU1, BGLU3 and BGLU4 could probably be located in the vacuole and in the cytoplasm, as described for AA5GT in carnation (Matsuba *et al.*, 2010), perhaps due to synthesis in the cytoplasm and transportation and accumulation in the vacuole. Because transportation of BGLUs through endomembranes into the vacuole is possible, the endomembrane control described in Appelhagen *et al.* (2015) would be a good investigation for future research, too. Matsuba *et al.* (2010) suggest, that GH1-type GTs could be transported across endomembranes and then function in noncytosolic compartments. The transportation could derive by the vesicle trafficing route, where donor and acceptor are also transported directly into the vacuole. The enzymatic reaction is then postulated to take place in the vacuole, where the sugar donor is also located, as discussed in more detail in the next section. Apart from this, glycosylation could also happen in the ER on the way to the vacuole, probably involving an alternative route for anthocyanidins, through the ER-to-vacuole protein-sorting route, reviewed by Grotewold and Davies (2008). In general, the transport mechanisms for enzymes and flavonoids into the vacuole are not well understood and need further investigations. For flavoinoids in Arabidopsis, several transportation mechanisms are suggested. Including a flavonoid binding glutathione S-transferase (GST)-like protein, binding and transporting the falvonoids with an unknown molecular mechanism (Kitamura *et al.*, 2004), direct transport via ER-derived vesicles independ on GST activity, an ATP-dependent transport mechanism, observed in Arabidopsis seedlings (Poustka *et al.*, 2007), or the involvment of a vacuolar flavonoid/H^+-antiporter transporter, shown to be necessary for vacuolar accumulation of proanthocyanidins and for anthocyanidin accumulation *in vitro* (Marinova *et al.*, 2007).

Since, most GH1-type GTs are located in the vacuole and they posses a putative signal peptide for transportation into the vacuole, the glycosylation reaction is mostly postulated to appear in this compartment (Nishizaki *et al.*, 2013). This is also in concordance with the estimated reaction optimum of Os9BGlu31 at pH 4.5, which is the pH of vacuoles in plants (Luang *et al.*, 2013). Such predictions of the theoretical isoelectric point of BGLUs was also performed in maize, often in concordance with the biological pH of the predicted compartment localization (Gómez-Anduro *et al.*, 2011).

Because, BGLUs in general exhibit to huge extend signal peptides at their N-terminus, which indicate localization in the vacuole (Matsuba *et al.*, 2010; Nakano *et al.*, 2014), it is likely, that GH enzymes like β-glycosidases are mainly located in the vacuole, too. This propose, that in the vacuole the glycosylation pattern is modified by the activity of GHs and GH1-type

GTs, on the one hand by declycosylation and on the other hand even by further glycosylation in the vacuoles themselves. In contrast UGTs lack such a signal peptide and these enzymes are expected and also shown to be located mainly in the cytoplasm, where the respective glycosylation reaction happens (Carter *et al.*, 2004; Lim and Bowles, 2004; Yonekura-Sakakibara and Hanada, 2011). This leads to the suggestion, that UGTs are responsible for the first glycosylations of flavonoids and that intensive elongation an specific glycosylations are performed in the vacuole, where the flavonoids are stored, by the activity of GH1-type GTs (Carter *et al.*, 2004; Luang *et al.*, 2013; Sasaki *et al.*, 2014).

Acyl-glucose as sugar donor

Because most of the GH1-type GTs were characterized to use acyl-glucose as sugar donors, instead of UDP-sugars, the vacuolar localization of the enzymes and the respective glycosylation reaction in this compartment is further supported. Bacause acyl-glucoses are known to be accumulated in vacuoles. This was for example shown by Oba *et al.* (1981) or by Sharma and Strack (1985).

For AA5GT and AA7GT from carnation and delphinium, Matsuba *et al.* (2010) could show, that several acyl-glucose molecules were utilized in enzyme assays, including benzoyl-glucose derivatives and 1-*O*-β-hydroxycinnamoyl-glucoses (HCA-Glcs), while the highest specificity was shown for the benzoic acid derivative 1-*O*-β-D-vanillyl-glucose, apart from activity for 1-*O*-β-D-isovanillyl-glucose, 1-*O*-β-D-sinapoyl-glucose, 1-*O*-β-D-caffeoyl-glucose and 1-*O*-β-D-4-coumaroyl-glucose, as well as 1-*O*-β-D-feruloyl-glucose and 1-*O*-β-D-p-hydroxybenzoyl-glucose (pHBG) for the recombinant proteins. Since, 1-*O*-β-D-vanillyl-glucose could be isolated from the respective plants, this was suggested to be the sugar donor *in vivo*.

For African lily and Canterbury bells a broad range of acyl-glucose molecules was accepted by the AAGTs, too. Since, from the petals of African lily and Canterbury bells no acyl-glucose could be isolated, the donor substrate is difficult to estimate. For AAGT from African lily the authors suggest feruloyl-glucose (FG) as the acyl-glucose donor molecule *in vivo* (Miyahara *et al.*, 2012), in contrast to *p*-hydroxybenzoyl-glucose in Canterbury bells (Miyahara *et al.*, 2014). For the AA7BG-GTs in delphinium the donor acyl-glucose substrate *p*-hydroxybenzoyl-Glc (pHBG) was detected also in the petals, thus being suggested to be the substrate *in vivo* Nishizaki *et al.* (2013).

In rice the preferred donor substrate of Os9BGlu31 seems to be 1-*O*-feruloyl-β-D-glucose, while again, the enzyme showed a broad donor spectrum ranging from 4NP-glycosides, 1-*O*-acyl-β-D-glucose esters, like FG, 1-*O*-(4-coumaroyl)-β-D-glucose, 4HBG, 1-*O*-sinapoyl-β-D-glucose, and 1-*O*-vanillyl-β-D-glucose, to the flavonoid glucosides phloridizin and apigenin 7-*O*-glucoside and also GA$_4$-GE (Luang *et al.*, 2013).

BGLU10 of Arabidopsis did also utilize different acyl-glucoses, preferring HCA-Glcs over benzoyl-glucoses, with highest activity towards sinapoyl-glucose (SG; Miyahara *et al.*, 2013). Plants from the Brassicaceae order are supposed to accumulate predominantly SG, therefore this is the major candidate for the natural donor substrate *in vivo* in Arabidopsis (Lorenzen *et al.*, 1996; Baumert *et al.*, 2005; Fraser *et al.*, 2007; Miyahara *et al.*, 2013; Sasaki *et al.*, 2014). BGLU6 is suggested to utilize acyl-glucose derivatives, like glucose esters of sinapic or benzoic

acid, too. But no enzyme assays were possible, because no recombinant protein in *E. coli* or *Pichia pastoris* was expressed and no functional crude extract was derived. Nevertheless, sinapoyl-glucose is supposed the *in vivo* donor substrate, as for BGLU10 (Ishihara *et al.*, 2016). The main reason for this assumption is the co-expression of BGLU6 with UGT84A. UGT84A expression is induced by UV-B radiation and it is involved in converting hydroxycinnamates to 1-*O*-glucose esters. It is the main contributor to the production of 1-*O*-sinapoyl glucose and overexpression lines show increased levels of sinapoyl glucose in seeds and seedlings. Therefore, this could be the donor substrate (Meissner *et al.*, 2008; Yonekura-Sakakibara *et al.*, 2012; Ishihara *et al.*, 2016). These findings led to the assumption, that BGLU1, BGLU3 and BGLU4 could utilize SG as sugar donor, too.

The localization of modifying enzymes in the compartment, where donor substrate accumulation takes place was also shown for acyltransferases (ATs), as reviewed by Sasaki *et al.* (2014). As suggested for UGTs and GH1-type GTs, two types of ATs with distinct compartmentalization and donor substrate utilization exist, too. ATs acting in the cytoplasm use acyl-CoA as donor molecule (D'Auria, 2006), while ATs acting in the vacuole use acyl-glucose as donor molecules, also mainly acyl-1-*O*-β-D-glucose (Hause *et al.*, 2002; Nishizaki *et al.*, 2013). It is likely that ATs and GTs located in the vacuole utilize the same donor molecules, either transferring the glucose moiety for the glycosylation reaction or the acyl moiety for the acylation reaction. This was shown by Nishizaki *et al.* (2013) for *p*-hydroxybenzoylglucose in delphinium for the biosynthesis of violdelphin, where the donor acyl-glucose substrate *p*-hydroxybenzoyl-Glc (pHBG) was used in the biosynthesis of these compounds for both, the addition of glucose and the addition of pHBA, thus they called pHBG as a zwitter donor (Nishizaki *et al.*, 2013; Ishii *et al.*, 2017). This double function as donor molecule is also supposed for sinapoyl-glucose, for example for the biosynthesis of the anthocyanin A11 (Yonekura-Sakakibara *et al.*, 2012; Miyahara *et al.*, 2013; Ishihara *et al.*, 2016).

This shows, that instead of being only a simple storage organ, the vacuole seems to be the location of a variety of enzymatic reactions, probably for the release of compounds, only needed under specific conditions (Gould *et al.*, 2002; Sasaki *et al.*, 2014).

Interestingly, the glycosylation reaction to generate acyl-glucose itself was shown to be catalyzed through an UGT (Nishizaki *et al.*, 2014), therefore, should take place in the cytosol (Yonekura-Sakakibara and Hanada, 2011). In Arabidopsis UGT84A1, UGT84A2 and UGT84A3 were shown to be involved in such an acyl-glucose production, while UGT84A1 is involved in *p*-hydroxybenzoylglucose (pHBG) formation, UGT84A2 in SG production and UGT84A3 in the synthesis of *p*-coumaroylglucose (Lim *et al.*, 2001; Meissner *et al.*, 2008; Yonekura-Sakakibara *et al.*, 2012). In this respect, the production of acly-glucose molecules and the utilization of the molecules for further glycosylation reactions is separated by different compartment, which is another supporting factor, for localization of GH1-type GTs in the vacuole.

4.4. Emergence of GH1-type GTs

In order to investigate the evolution of the GH1 hydrolase gene family, phylogenetic analysis by Cao *et al.* (2017) was done with 304 sequences of plant GH1 hydrolases. In the big phylogenetic

tree, they could show twelve branches. Because GH1 sequences from one organism appeared in distant braches, it is assumed, that BGLUs could have multi-ancestral origins and diverse functions in plant development and stress responses. Already in green algae, GH1 hydrolase sequences were found, suggesting, that GH1 hydrolases diverged before the evolution of land plants. In addition, it is suggested that high divergence between the lineages of dicots and monocots appeared by frequent gene duplication (Cao *et al.*, 2017).

In contrast to this high divergence between dicots and monocots, GH1-type GTs appear in both lineages, as shown for AA7GT in African lily and Os9BGlu31 in rice for monocots and the other known GH1-type GTs in dicots (Miyahara *et al.*, 2012; Luang *et al.*, 2013), suggesting, that GH1-type GTs developed in angiosperms before the divergence of monocots and dicots, thus in an early stage of angiosperm development (Miyahara *et al.*, 2012). This is also supported by the presence of the conserved catalytic residue motif in all enzymes (Matsuba *et al.*, 2010; Miyahara *et al.*, 2012; Luang *et al.*, 2013; Ishihara *et al.*, 2016).

Furthermore, due to the strong homology, GH1-type GTs are suggested to be derived from β-glycosidases in this early stage of angiosperm development (Miyahara *et al.*, 2012). While clustering of GH1-type GTs appears mostly in one clade, suggesting the spread of GH1-type GTs by gene duplication, phylogenetic analysis of Nishizaki *et al.* (2013) revealed, that the AABG-GTs in delphinium are separated from the other known AAGTs, assuming that they did not arise from gene duplication of DgAA7GT. The authors propose, that the change from hydrolase to transferase activity could have occured at higher frequency at different GH1 protein members. This could be the reason, why acyl-glucose dependent GTs could occur also for other secondary metabolite substrates in plants, other then anthocyanidins, as shown for Os9BGlu31 from rice (Luang *et al.*, 2013) and BGLU6 (Ishihara *et al.*, 2016), as well as BGLU1, BGLU3 and BGLU4 in this work, in Arabidopsis .

The determination of the factors, that transform hydrolase activity into transferase activity are of huge interest and intensively studied. To understand the mechanism in more detail, the hydrolase activity of β-glycosidases will be discussed in the next section.

4.4.1. β-glycosidases

β-glucosidases usually catalyze the hydrolysis of β-O-glycosidic bonds between the anomeric carbon of a monosaccharide and the nonreducing end of a carbohydrate (aglycone) or glycoside molecule (Opassiri *et al.*, 2006; Miyahara *et al.*, 2011), while in each catalytic cycle the enzymes remove a carbohydrate from the non-reducing end of their substrates, thus, releasing β-glucose and the associated aglycone (Czjzek *et al.*, 2000; Xu *et al.*, 2004; Marana, 2006a; Ketudat Cairns *et al.*, 2015).

The enzymes are believed to be involved in physiologically important processes like biotic and abiotic stress response, herbivore defense, phytohormone activation, lignification, and cell wall remodeling, by liberation, thus, activation of many physiologically important compounds, like phytohormones and defense compounds. Therefore, in plants many different β-glycosidases, like β-galactosidases, phospho-β-galactosidases, phospho-β-glucosidases, and thioglucosidases are present. All of them with diverse activities, often coupled to strong aglycone specificity (Opassiri *et al.*, 2006; Gómez-Anduro *et al.*, 2011; Dong *et al.*, 2019).

All GH1 enzymes belong to the GH-A clan, exhibiting an $(\alpha/\beta)_8$ barrel secondary structure and all enzymes show the catalytic acid/base and the nucleophile consensus sequences with the highly conserved peptide motifs (TF/L/MNEP and I/VTENG), located at the C-terminal ends of strands 4 and 7, respectively. These peptide motifs exhibit the two conserved catalytically active glutamate residues (E), serving as the catalytic acid/base and catalytic nucleophile for the hydrolysis of β-glycosidic bonds (Henrissat *et al.*, 1996; Sanz-Aparicio *et al.*, 1998; Czjzek *et al.*, 2000; Zouhar *et al.*, 2001; Chuenchor *et al.*, 2008; Roepke and Bozzo, 2015). The side chain of the glutamate residues, located at the bottom of the active site, contain two carboxylic acids, which are necessary for the hydrolysis reaction. The reaction is called a retaining mechanism and can be divided in two steps, a glycosylation and a deglycosylation step. The glycosylation step reveals a covalent intermediate between enzyme and substrate, which is a glycosyl-enzyme intermediate. Therefore, the nucleophilic E in the motif I/VTENG attacks the anomeric carbon (C-1) of the sugar substrate, to form the covalent intermediate. At the same time protonation of the glycosidic oxygen by the acid catalyst E in the TF/L/MNEP motif releases the aglycone. In the deglycosylation step the covalent intermediate is hydrolyzed, thus, the sugar is displaced from the enzyme by water and another nucleophile displacement, assisted by the catalytic acid/base. For this reaction the second catalytic E, which is now an anion and therefore a base catalyst, removes a proton from water. The resulting -OH group performs a nucleophilic attack on the covalent bond between the glycone and the enzyme, releasing the glycone while regenerating the nucleophilic E (Czjzek *et al.*, 2000, 2001; Zouhar *et al.*, 2001; Xu *et al.*, 2004; Marana, 2006b; Luang *et al.*, 2013; Roepke and Bozzo, 2015). Myrosinases, a special member of the GH1 family, lack the catalytic acid/base residue. These enzymes contain the homologous motif TINQL, in which glutamine has replaced glutamic acid. Myrosinases hydrolyze glucosinolate, where the aglycones is an excellent leaving group. In addition, the catalytic base activity is replaced by the co-factor ascorbate (Burmeister *et al.*, 2000). In contrast, the nucleophilic glutamic acid residue seems to be absolute necessary for the hydrolytic activity (Turan, 2008).

4.4.2. Converting hydrolases into transferases

For β-glucosidase hydrolase activity the two glutamic acids were shown to by necessary by crystal structures (Burmeister *et al.*, 1997; Sanz-Aparicio *et al.*, 1998; Czjzek *et al.*, 2000; Verdoucq *et al.*, 2004; Chuenchor *et al.*, 2011). Therefore, attempts were done to convert the hydrolaze activity into a glycosidase activity, by mutations of the conserved catalytic glutamic acids, being successful with mutant enzyme proteins, where the catalytic nucleophile residue was replaced by different non-nucleophilic ones, showing less hydrolase activity, a broader substrate tolerance and an increased transglucosidase activity (Mayer *et al.*, 2000; Perugino *et al.*, 2004; Hancock *et al.*, 2005, 2006; Shaikh and Withers, 2008; Wang and Huang, 2009; Matsuba *et al.*, 2010). Nevertheless, this cannot be the mechanism for GH1-type GT activity, naturally occurring in plants. GH1-type GTs still contain the conserved catalytic motifs, including the nucleophilic glutamic acid, while being able to transfer a sugar moiety from a sugar donor to a flavonoid acceptor molecule (Matsuba *et al.*, 2010). In addition, Luang *et al.* (2013) could show by single substitution mutants, that the absence of the acid/base

is not critical for transglucosidase activity. This leads to the assumption, that probably the glutamate residues are not important to distinguish between glycosylation and hydrolysis. Therefore, the search for GH1-type GTs cannot be laid only on homology research of the primary sequence. As described for many β-glycosidases, there exist no general rule for assigning the substrate specificity for a family GH1 β-glucosidase based on its sequence or even its structure. The structures must be evaluated for their fit to a particular substrate of interest. Crystal structures would be necessary for GH1-type GTs, to clearly investigate the catalytic mechanism. Nevertheless, the primary sequence can give some first hints, if a substrate can fit or not (Miyahara *et al.*, 2011, 2014; Ketudat Cairns *et al.*, 2015)

Based on the known mechanism for hydrolaze activity and investigations of Os9BGlu31 transglucosidase activity, Luang *et al.* (2013) suggest for the GT activity an inversion mechanism, compared to the retaining mechanism of GHs, which is described as a simple displacement mechanism. It is suggested, that in the moment, while the glycosidic bond of the donor substrate is broken, as described for β-glycosidases in the section above, the acceptor attacks from the opposite site of the sugar. In case of a hydrolysis, this acceptor is the above-mentioned water molecule, in case of GTs, the acceptor is another molecule, to which the sugar is transferred, releasing the sugar moiety from the glycosyl-enzyme intermediate. Therefore, GH1-type GT activity would represent the glucosyl transfer activities of GH1 β-glucosidases. Apart from this, glycon transfer probably can also happen by further unknown mechanisms (Miyahara *et al.*, 2011; Luang *et al.*, 2013).

This led to the assumption, that the enzymatic activities of GTs and GHs are very similar. Since there exist cases, where GTs catalyze reversible reactions and also GHs could exhibit transglucosidase activities in addition to their hydrolase activity, it cannot be excluded, that both enzymes can perform both reactions under certain circumstances (Opassiri *et al.*, 2004; Lairson *et al.*, 2007; Luang *et al.*, 2013; Ketudat Cairns *et al.*, 2015).

In addition, two different work groups assigned two different functions for BGLU10, which is in one case a hydrolase activity (Wang *et al.*, 2011) and in the other case a glycosyltransferase activity (Miyahara *et al.*, 2013). In the first case, it was shown, that *BGLU10* was induced by ABA, drought and salt treatment, and that the respective loss-of-function line was more susceptible to drought stress, while two respective overexpression lines were less susceptible. Therefore, they suggested, *BGLU10* could be an ABA-GE β-glucosidase, releasing ABA from its glucosyl ester in the plant. On the other hand, Miyahara *et al.* (2013) could show, that BGLU10 seems to transfer the sugar moeity to A9, to generate A11, using a loss-of-function mutant and enzyme assays with crude protein extracts. This raises the question, if BGLU10 can perform both reactions, dependent on the respective environmental conditions, as shown for maize *BGLUs*. The gene was expressed in various tissues and was induced by various stress factors, thus could be involved in different stress responses (Gómez-Anduro *et al.*, 2011; Wang *et al.*, 2011).

In summary, GH1-type GT and β-glucosidase activity could derive by a closely related mechanism, which differs probably only by slightly different conformational states of the catalytic pocket. If GH1 enzymes could perform both reactions under different environmental conditions or if glycosydase or hydrolase activity is restricted to specific enzymes needs further

investigations.

4.5. Flavanol–anthocyanidin, pyranoanthocyanidin and flavanol–pyranoanthocyanidin formation in Arabidopsis seeds

The results for the potential substrates of BGLU3 glycosylation, suggest the presence of condensed flavanol–anthocyanidin (F–A$^+$) and condensed flavanol–pyranoanthocyanidin (F–pyA$^+$) compounds in Arabidopsis seeds, the last one suggesting also the presence of pyranoanthocyanidins in Arabidopsis seeds. All these compounds were never described for Arabidopsis before. The presence of flavanol–pyranoanthocyanidins, where the pyranoanthocyanidin is the lower extension unit, are described only for wine and an (epi)gallocatechin–carboxypyranomalvidin is described for the first time, according to literature. This rises the question, if the occurence of natural flavanol–anthocyanidins and flavanol–pyranoanthocyanidins, as well as pyranoanthocyanidins, are likely to occur in Arabidopsis seeds, and in addition, what function those compounds could have in plant seeds.

The above listed molecules were first discovered in wine, correcting the formaly assumed degradation of anthocyanidins in wine or model solution as rather a condensation of this anthocyanidins to new compounds, than a degradation (Rivas-Gonzalo *et al.*, 1995).

Carboxypyranoanthocyanidins are reported in a few cases to derive from plant material, like vitisin A or vitisin A like compounds in dried grapes (Marquez *et al.*, 2012) and raisins (Marquez *et al.*, 2013a) and as 5-carboxypyranocyanidin and 5-carboxypyranopetunidin in onions (Fossen and Andersen, 2003) and figs (Dueñas *et al.*, 2008) or as 5-carboxypyranopelargonidin in strawberries (Andersen *et al.*, 2004). The term 5-carboxy, derives from the location of the carboxyl group on the second pyran ring (Rentzsch *et al.*, 2007b).

Although, condensed pigments of flavanols and pyranoanthocyanidins are only known from wine (Asenstorfer *et al.*, 2001; Mateus *et al.*, 2002a,b; He *et al.*, 2010a; Nave *et al.*, 2010), condensed pigments of flavanols and anthocyanidins are known to appear apart from wine (Remy *et al.*, 2000; Salas *et al.*, 2004; Sánchez-Ilárduya *et al.*, 2012) also in plant material like strawberries (Fossen *et al.*, 2004; González-Paramás *et al.*, 2006), black currant seeds (McDougall *et al.*, 2005), beans (Macz-Pop *et al.*, 2006a), black soybean (Lee *et al.*, 2009), purple corn (González-Manzano *et al.*, 2008), fig (Dueñas *et al.*, 2008) and pomegranate (Sentandreu *et al.*, 2010). In wine, Nave *et al.* (2010) discovered the first and, to current stage, only F–pyA$^+$ condensed pigments with the flavanol as the upper unit, in condensation with a pyranoanthocyanidin. The pyranoanthocyanidins consists of vitisin A and vitisin B respectively. Thus, these condensed pigments are in concordance with compounds found in this work in Arabidopsis seeds, but with a catechin as the flavanol unit. According to the findings of Nave *et al.* (2010), the formation of these vitisin A and vitisin B containing condensed pigments is most likely to appear by a condensation of a flavanol–anthocyanidin compound with a pyruvic acid, compared to a condensation reaction between a flavanol and a vitisin A. The reason is proposed to derive from the reduced reactivity of pyranoanthocyanidins. Therefore, for the formation of the candidate BGLU3 substrate compound of 704 m/z in the first step a condensation reaction between an (epi)gallocatechin and a malvidin aglycone is proposed. This was also stated by He *et al.* (2010a), proposing the formation of vinyl-linked

pyranoanthocyanidin–flavanols to derive from ethyl-linked anthocyanidin–flavanol pigments, which are more unstable and react to the stable intermediates.

For a long time, the formation of condensed flavanol–anthocyanidin (and vice versa) compounds as well as pyranoanthocyanidins were thought to appear only during storage and processing of food and drinks under acidic conditions and by the presence of fermentation products, to build more stable compounds (Somers, 1971; Bakker and Timberlake, 1997; Fulcrand *et al.*, 1998). In case of pyranoanthocyanidins in plant samples, there exist also assumptions, that these compounds derive due to sample handling and preparation. Findings in corn, hint on other mechanisms, for the formation of these compounds, because for some condensed compounds the precursor compounds were not detected, or the distribution of the condensed compounds was not in concordance with the distribution of the free compounds. They also show, that in one endosperm the anthocyanins were 75% pelargonidin derivatives, while the condensed pigments mainly contained cyanidin. This should led to the assumption, that maybe both condensation possibilities occur, in the plant itself and during food processing (González-Manzano *et al.*, 2008; Sentandreu *et al.*, 2010). This is also supported by the fact, that in this work, it was shown, that the levels of flavonol–pyranoanthocyanidins were changing in concordance with loss-of-function and overexpression lines, thus the glycosylation mechanism should have taken place in the seeds before sample processing. If the compounds just derive due to processing effects, such differential metabolite levels are rather unlikely.

Investigations of raisins by (Marquez *et al.*, 2013a) lead to the strong assumption, that the formation in plant tissue could derive by a dehydration-stress induced pathway during the drying process, which turns the aerobic grape metabolism into an anaerobic one. In raisins pyranoanthocyanidins as well as anthocyanidin–flavanol compounds were identified, one of these compounds was vitisin A. Because the formation of these compounds needs the presence of molecules like pyruvic acid, acetoacetic acid and acetaldehyde, which in wine and juice should derive from fermentation processes, but were not found in grape must, it is assumed that they are synthesized during grape drying, too, thus accounting for the presence in raisins. The authors state that a change in the grape metabolism during the drying process, converting the aerobic metabolism into an anaerobic one, is likely. The dehydration process is a stress situation for plant cells, due to weight loss, resulting from the loss of water and this changes the plant metabolism. The stress response to water loss involves a membrane permeability change, due to the activity of the lipoxygenase enzyme (LOX) at the beginning of the water loss. LOX is known to degrade cell membranes, increase ion leakage, decreased lipid content (due to oxidation of fatty acids into 6-carbon (C6) alcohols and aldehydes) and increase water loss, which causes the change of an aerobic to an anaerobic metabolism. The anaerobic metabolism activates enzymes, which are able to degrade sugar and/or malic acid into pyruvic acid and related compounds as well as the alcohol dehydrogenase enzyme (ADH) and the pyruvate decarboxylase (PDC). While pyruvic acid is for example a necessary compound for vitisin A formation, for vitisin B and flavanol–anthocyanidin condensation, acetaldehyde is an important compound. Acetaldehyde could derive during an anaerobic metabolism, if pyruvic acid is converted into acetaldehyde by decarboxylation from PDC. ADH converts acetaldehyde reversibly into ethanol, which serves as the anaerobic energy supply. (Marquez

et al., 2013a). Therefore, acetaldehyde is a naturally in plants occurring compound, making acetaldehyde mediated condensation of flavanols and anthocyanidins possible (Rasheed *et al.*, 2018).

In seeds, strict anaerobic metabolism is not likely to happen, because plants are not able to live under strict anoxia and seeds are not germinative, due to the toxic effect of ethanol for the plant (Kennedy *et al.*, 1992). Only a few species exceptions can survive strict anoxia for a longer period of time during germination and growing, which are for example some flood tolerance plants, having developed mechanisms to deal with anoxia. Those mechanisms are decreased ethanol production or removal of ethanol, lactate transport in the vacuole for an extended lactate fermentation, adaptations of the mitochondria (which are usually dependent on oxygen) and the enhanced production of anaerobic stress proteins (ASP). Another observation made for high ethanol levels, is the enhanced reverse production of acetaldehyde from ethanol. Only germinating seeds seem to be a part of the plant, that is in many species usually exposed to anoxia for a short time during the imbibition, before the rupture of the testa. The seed coat is in the first hours impermeable to oxygen, resulting in increased ADH activity for alcoholic fermentation. If the radicle or shoot causes the rupture, aerobic respiration is induced and ADH levels decrease. In germinating seeds also, the lactate dehydrogenase (LDH) activity is increased in the first hours of germination, which is responsible for the formation of lactade out of pyruvate. This reaction does only transiently occur, during the transfer from the aerobic to the anaerobic metabolism. Due to the increasing cytosolic pH the LDH activity is inhibited and the ethanol synthesis starts. (Kennedy *et al.*, 1992; Chkaiban *et al.*, 2007).

Despite anoxia during germination, the respiration rate is also in dry seeds drastically reduced. In addition, the seed coat functions like an oxygen barrier and the drying process could result in a change of metabolism. Thus, slight anoxia is possible. Furthermore, in dormant seeds the activities of ADH, lactate dehydrogenase (LDH) and pyruvate phosphate dikinase (PPDK), all linked to anaerobic glycolysis and the latter one, participating in pyrophosphate-dependent reactions, are known to be higher as in germinating seeds, indicating a certain level of fermentation during dormancy (Ma *et al.*, 2017). This could mean, that the discussed mechanism is to some extend responsible for the pyruvate and acetaldehyde production in dry seeds. Pyruvate is also generated during glycolysis, which is an oxygen-independent metabolic pathway. Therefore, pyruvate is produced in the aerobic and anaerobic metabolism. In the anaerobic metabolism the pyruvate decarboxylase (PDC) is responsible for the decarboxylation of pyruvate, thus yielding CO_2 and acetaldehyde, the precursor of ethanol, as described above. ADH activity increases in most plants under anaerobic conditions for the catalysis of acetaldehyde into ethanol, which is under stress a safe energy mechanism. Another stress situation to induce ADH, is dehydration, thus, also the drying process of seeds could be an activator. Therefore, it is probable, that all three requirements for the occurrence of pyranoanthocyanidins suggested by (Rentzsch *et al.*, 2007b) are present in Arabidopsis, which are the presence of anthocyanidins and flavanols in Arabidopsis seeds (Routaboul *et al.*, 2006; Shi and Xie, 2014), the presence of the reaction partners (see discussion above) and the storage time, which is a natural characteristic of seeds.

4.6. The function of flavanol–anthocyanidin, pyranoanthocyanidin and flavanol–pyranoanthocyanidin in seeds

Due to the enhanced stability of flavanol–anthocyanidin condensations and pyranoanthocyanidins (Somers, 1971; Bakker and Timberlake, 1997; Romero and Bakker, 1999; Oliveira *et al.*, 2014), these compounds could fulfill functions as stable antioxidants or reactive oxygen species (ROS) scavengers in seeds. Since, carboxypyranoanthocyanidins seem to appear at maximum high in relatively young wine or juices, the emerging of this compound in young ripe seeds, as used in this work, supports the occurence of vitisin compounds (Rentzsch *et al.*, 2007b). In addition, it was shown in wine, that the concentration of acetaldehyde at the end of the fermentation is higher than pyruvic acid, therefore the formation of vitisin B tends to be higher at this stage of the fermentation (Morata *et al.*, 2003). Even if seeds and wine are not really comparable, it could be a hint, that in the presence of much pyruvate more vitisin A compounds are produced, thus probably in younger seeds, and that with the ongoing accumulation of acetaldehyde, more vitisin B compounds are built. This would be in concordance with the comparable low levels of the suggested (epi)gallocatechin–vitisin B compound in this work, for young dry seeds. It would be interesting, to investigate in more old seeds, if the level of vitisin B compounds rises.

Seeds are permanently confronted with ROS. Therefore, they need to be protected from oxidative damage. This is especially important for seed germination, since the germination involves an increase in ROS production, while at the same time the rupture of the radicle or shoot involves damages, like leakage of cellular solutes and harm to organelles, membranes and DNA, making protection from ROS even more crucial (Ma *et al.*, 2017). Thus, ROS production in plants, at the same time include the antioxidant system for ROS scavenging, such as the ascorbate–glutathione cycle (Foyer and Halliwell, 1976), superoxide dismutase (SOD, Alscher *et al.*, 2002) and catalase (CAT, Chelikani *et al.*, 2004) as the main ones, but also antioxidants like flavonoids. Since, seeds remain for a certain storage time in a reduced biosynthetic state, long lasting flavonoid antioxidant compounds, like pyranoanthocyanidins could be of great importance. In addition, enzymes for ROS scavenging, like SOD and CAT, are mainly active in germinating seeds and not very active in dormant seeds (Ma *et al.*, 2017), therefore flavonoids could be the main ROS scavenging mechanisms in dormant dry seeds, where enzymatic activities are decreased.

In addition, the regulation of the change from dormancy to germination could involve stable antioxidative compounds. The change from dormancy to germination is regulated by a very complex interacting network of different factors like for example abscisic acid (ABA), gibberellins (GA), reactive oxygen species (ROS) and reactive nitrogen species (RNS), resulting again in an intense change of gene expression and thus at the end metabolism, including the increase in ROS scavenging enzymes and a decrease in the fermentation enzymes like LDH and ADH (Alscher *et al.*, 2002; Chelikani *et al.*, 2004; Finch-Savage and Leubner-Metzger, 2006; Šírová *et al.*, 2011; Bykova *et al.*, 2015).

ROS production occurs in germinating seeds as well as in dry seeds. This include ROS species like hydrogen peroxide (H_2O_2), hydroxyl radicals ($OH_.$) and superoxide anions (O_2^-). During imbibition, ROS accumulates, as an important requirement for seed germination. They

have a positive function in cellular signal transduction, disease resistance, redox regulation, and programmed cell death (El-Maarouf-Bouteau and Bailly, 2008; Ma *et al.*, 2017). For these functions, ROS content needs to be in the so-called *oxidative window* (Bailly *et al.*, 2008). Lower ROS concentrations maintain seed dormancy, higher ROS concentrations lead to oxidative damage of cell components, that either inhibit germination or lead to abnormal seedlings. Stable antioxidants, like pyranoanthocyanidins could be involved in this ROS regulation.

Since, the dormant seed can be exposed to many stress factors, which it needs to survive until germination, stable ROS scavengers until germination could be provided by pyranoanthocyanidins, to protect also the dormant seed (Weidner *et al.*, 2014; Mera *et al.*, 2019).

Apart from ROS, reactive nitrogen species (RNS) seem to play a crucial role in the regulation of seed dormancy release towards germination. The most important compound seems to be nitric oxide (NO), which is present as different forms like the nitrosyl cation (NO^+), the NO radical ($NO_.$) and nitroxyl (NO^-) and which seems to be responsible for germination and dormancy release, due to high NO levels during germination and low levels during dormancy (Hendricks and Taylorson, 1974; Sarath *et al.*, 2006). Probably, NO plays a role in nitrosylation of proteins for the germination process (Sen, 2010). Besides enzymatic generation of NO, it can also be produced by non-enzymatic conversion of nitrite (Santolini *et al.*, 2017), which is accumulating in tissues under anaerobic conditions (Morard *et al.*, 2004), thus probably also in seeds, which then can be converted to NO. Since, high NO concentrations are only wanted during germination, in plants, NO can be scavenged by the S-nitrosoglutathione reductase (GSNOR, Sakamoto *et al.*, 2002) and the plant hemoglobin, called phytoglobin (Pgb, Igamberdiev *et al.*, 2005; Hill *et al.*, 2016). Since, Pgb is also anoxia-induced (Sowa *et al.*, 1998), it could play an important role in reducing NO levels in the dormant seed, as well as preventing excessive NO levels during germination in the anoxic phase. GSNOR activity was shown to be higher in dormant seeds, than in germinating seeds, thus reducing the free NO concentration in dormant seeds (Ma *et al.*, 2016), by binding NO in the S-nitrosoglutathione (GSNO) complex. The nitrite accumulation under anoxic conditions with the transformation of nitrite into NO, which is then bound in the GSNO complex, could be a mechanism to keep NO available in the GSNO complex in dormant seeds for germination, if the right conditions are present, by preventing at the same time too early germination (Ma *et al.*, 2016).

Takahama *et al.* (2013a) could show in studies with the cyanidin–catechin pigment vigna-cyanidin, that the molecule binds by hydrophobic interaction with starch in the stomach. They postulate that it is possible that nitrous acid could access such sites of hydrophobic interactions between starch and vignacyanidin in the stomach, enhancing in this way the oxidation of vignacyanidin. This could on the one hand lead to the formation of complex vignacyanidin polymers and on the other hand to the reduction of nitrous acid to nitric oxide (NO), the latter should increase the peristaltic activity of the stomach (Desai *et al.*, 1991). Nitrous acid arrives as salivary nitrite in the stomage, during the time of food stay in the stomach. This reduction mechanism is also known to happen by the action of ascorbic acid in the stomach (Iijima *et al.*, 2003) and should also be possible to happen through the antioxidant activity of the vignacyanidin compound. Thus, vignacyanidin indirectly can

control the time, which the food stays in the stomach. This reduction of nitrous acid could also be a possible non-enzymatic mechanism, to convert nitrite into NO in seeds which are high in starch. (Takahama *et al.*, 2013a).

Another interesting finding is, that the vignacyanidin complex with starch, prevents starch digestion in the stomach and intestine by amylases. The inhibition of starch digestion was also shown for quercetin by Hirota and Takahama (2017), while the inhibition strength seems to be dependent on the hydrophobicity of the flavonoid. In seeds, starch is the most important carbohydrate store, to provide energy for the developing seedling after germination. Therefore, starch needs to be protected from degradation (Nonogaki, 2008). Probably, condensed pigments of flavanols and anthocyanidins or pyranoanthocyanidins can contribute to this protection, apart from enzyme-production and pH conditions, regulating starch degradation (Edelman *et al.*, 1959; Hamabata *et al.*, 1988), while the latter factors are dependent on an acidic pH, whereas the condensed pigments and pyranoanthocyanidins do not seem to be that pH dependent (Bakker and Timberlake, 1997; Francia-Aricha *et al.*, 1997).

From complexation of anthocyanidins with copigments like flavonoids, it is known, that these complexes play an important role for UV protection. This mechanism involves an ultrafast internal conversion of the complex with a fast, non-radiative energy transfer, where the excited copigment transfers energy to the flavylium moiety, which subsequently deprotonates to water. Protection against UV radiation can also involve anthocyanidins alone, where the ion can eliminate the absorbed UV radiation into heat, by a fast proton transfer to water and a rapid decay into the ground state. Because in plant vacuoles, where most of the anthocyanidins are located, the pH is not in favor of the flavylium ion state, but for the hemiketal form, complexation of the anthocyanidins is predominant. (Francia-Aricha *et al.*, 1997; Costa *et al.*, 2014). It is suggested, that copigmentation can happen prior to the formation of condensed compounds of anthocyanidins and flavanols (Liao *et al.*, 1992; Brouillard and Dangles, 1994; Escribano-Bailón *et al.*, 1996). Therefore, copigmentation and condensation of flavanols and anthocyanidins could fulfill similar functions, suggesting, that the condensed compounds are involved in UV-protection, too, probably even improving the mechanism.

All these factors support the fact, that stable flavonoids with antioxidant activities, like pyranoanthocyanidins, could be very important in seeds. Pyranoanthocyanidins are more stable than normal anthocyanidins, due to the additional pyran ring, forming a double pyrylium. This doubles the number of resonant forms thus the possibility of the electronic de-localization, giving more resistance to antioxidant degradation, also by reducing the polarity of the molecule (Macz-Pop *et al.*, 2006a; Morata *et al.*, 2019). The stability of pyranoanthocyanidins was predominantly shown for the color stability in wine and model solution. The color expression is significantly increased by pyranoanthocyanidins at rising pH, with a maximum for oligomers of pyranoanthocynidins and flavanols. The reason could be the hydration equilibrium of the anthocyanidin at higher pH, which yields the colorless anthocyanidin hemiacetal. Through these mechanisms, anthocyanidins lose around 80% of their color intensity if the pH increases from 1 to 5, while the pyranoanthocyanidins color intensity is almost not changing at all. Pyranoanthocyanidins are more protected from the attack of water. Even if slow reactions could occur, as seen for oxovitisins, equilibria shift to colorless forms are less extreme. (Brouil-

lard and Lang, 1990; Oliveira *et al.*, 2006; Rentzsch *et al.*, 2007b). The nucleophilic water attack naturally occur at the C-2 and C-4 positions of the flavylium, the substitution of pyranoanthocyanidins at C-4 affects the distribution of charge throughout the molecule in a way, that positions 2 and 4 become less reactive to this nucleophilic attack. Apart from hydration, the molecule is also resistant towards SO_2 bleaching, pH changes, oxidative degradation and temperature (Bakker and Timberlake, 1997; Francia-Aricha *et al.*, 1997; Fulcrand *et al.*, 1998; Asenstorfer *et al.*, 2001; Vivar-Quintana *et al.*, 2002; Fossen and Andersen, 2003; Andersen *et al.*, 2004; Rentzsch *et al.*, 2007b; de Freitas and Mateus, 2011; Sánchez-Ilárduya *et al.*, 2012; Marquez *et al.*, 2013b).

In general, more studies about condensed pigments and pyranoanthocyanidins in plants and thus seeds are needed. In addition, the physiological properties and therefore, influences of these compounds on human health are of interest, requiring more studies about these compounds. The formation of more stable compounds could lead to enhanced stability for their biological activity. In this way pyranoanthocyanidins probably show the same positive health benefits as anthocyanidins but being more stable also under physiological conditions (McDougall *et al.*, 2005; Rentzsch *et al.*, 2007b).

4.7. Aglycones in seeds

The proposed (epi)gallocatechin–carboxypyranomalvidin substrate of BGLU3 suggest, that in Arabidopsis seeds a non-glycosylated flavanol–pyranoanthocyanidin molecule is present. Anthocyanidins in plants are most abundant in their glycosylated form, but aglycone forms were found in some cases in plant samples, too, as in azuki beans (Takahama *et al.*, 2013b,a; Yoshida *et al.*, 2019), beans (González-Paramás *et al.*, 2006; Macz-Pop *et al.*, 2006b), roses (Fukui *et al.*, 2002) and red onions (Fossen and Andersen, 2003), including condensed pigments of flavanols and anthocyanidins as well as anthocyanidin and pyranoanthocyanidins, like carboxypyranoanthocyanidins in red onion (Fossen and Andersen, 2003).

The aglycones are less soluble in water and have a shorter life span compared to their glycosylated forms. Therefore, it is assumed that the glycosylated forms are more abundant in nature (Macz-Pop *et al.*, 2006b,a).

In general, the condensation of flavanols and anthocyanidin aglycones should be possible (Santos-Buelga *et al.*, 1995), while Takahama *et al.* (2013b) suggest that apart from the reaction of a flavanol with an anthocyanidin aglycone, the removal of the sugar from condensed glycosides could account for those aglycones, too. In case of the BGLU3 candidate substrate compound, the mechanism is not known and further studies are needed for clarification. Nevertheless, it seems as if the flavanol–anthocyanidin condensation and the reaction to yield the pyranoanthocyanidin is performed without a sugar moiety. About the functional purpose nothing is known. In addition, the purpose of the broad extend of glycosylation of flavonoids is not completely clarified. Apart from factors like stabilization, more solubility and less toxicity, there might be much more purposes, as can be estimated for example in the proposed role for polymerization of oligomeric proanthocyanidins (Zhang *et al.*, 2017).

Concerning pyranoanthocyanidins Farr *et al.* (2018) made an interesting study on the role of glycosylation patterns, concerning the development of pyranoanthocyanidins, if bringing

together glycosylated anthocyanidins and pyruvic acid in a model solution. They could show, that the respective kind of glycosylation influences the reaction. Probably this is not just influenced by steric hindrance, thus the more sugar moieties attached to C-3 the less pyranoanthocyanidin formation, but more by the amount of sugar and the kind of sugar-linkage, because (1⟶6) linked disaccharides range before monoglucosides in the formation of new pyranoanthocyanidins. The authors suggest the greater degrees of rotation, through an additional torsion angle between C-5 and C-6 of the attached sugar, being responsible, on the one hand by providing more access of pyruvic acid or guiding pyruvic acid in the correct position and on the other hand by increasing collision between reactants. Unfortunately, Farr *et al.* (2018) did not compare aglycones and glycosides for pyranoanthocyanidin formation. Further studies like this could investigate, if pyranoanthocyanidin formation from aglycones is favored, and if therefore, a sugar moiety is removed from the anthocyanin prior to pyranoanthocyanidin formation, in order to favor the establishment of more stable compounds. Condensed flavanol–pyranoanthocyanidins are the most stable compounds, compared to flavanol–anthocyanidins, anthocyanidins and their respective glycosides (He *et al.*, 2010a). Since, glycosylation in general stabilizes flavonoid molecules, a glycosylation of flavanol–pyranoanthocyanidins could even enhance the stability. Therefore, it is likely, that the glycosylation of flavanol–pyranoanthocyanidins is favored, compared to flavanol–pyranoanthocyanidin aglycones for storage in seeds, even if sugar moieties are temporarily removed.

4.8. Future perspectives

Since, pyranoanthocyanidins are not very well known for natural plant samples, more knowledge about these compounds would be desirable. The first point is, that is is of great interest, to investigate, if more of these compounds exist in plants or probably especially seeds, and if so, what structures and characteristics they provide. The investigations of (Wu *et al.*, 2018) also show, that it is very likely, that the metabolome of Arabidopsis includes further, yet unidentified and also undiscovered minor flavonoids and flavonoid glycosides. Untargeted LC-MS-based metabolomic profiling and samples either derived under different abiotic environments or from different tissues, or as in our case by different mutants as well as including metabolic genome-wide association studies, as performed by Wu *et al.* (2018), could be an useful tool, to identify on the one hand metabolite-gene associations and on the other hand new, minor compounds. Especially artificial stress, can be very helpful, if keeping in mind, that the biosynthesis of many metabolites is only induced or increased under certain stress, conditions. It can also help, to discover novel metabolic pathways or signaling mechanism, that are involved in the very complex interactions of genes and environment to respond and adapt to environmental stress. (Wu *et al.*, 2018). It is assumed, that even in wine, where malvidin 3-glucoside and its pyruvic acid adducts are the most dominant compounds, thus most detailed investigated, much more new compounds, than known, exist, due to the initial chemical complexity of wine (Mateus *et al.*, 2002a). Besides, it seems as if many different types of pyranoanthocyanidins and condensed pigments can exist, like the very unique (until now) catechinopyranocyanidins A and B from azuki beans. In this molecule the flavanol and the anthocyanidin are connected by

a pyrano ring at position 6 and 7 of the catechin (Yoshida *et al.*, 2019). Or also the rosacyanin B from roses (Fukui *et al.*, 2002) or also an pyruvic acid–catechin–malvidin-3-glucoside, where the pyruvic acid substitutes at the C-6 and C-8 of the flavanol (Nave *et al.*, 2010). In addition, further complexation is possible, since there were anthocyanidin–flavanol oligomers found in grape marc extracts, with up to four units of catechin linked by vinyl bonds. (Asenstorfer *et al.*, 2001).

A systematic search for pyroanthocyanidins and flavanol–pyroanthocyanidins (and vice versa) or also flavanol–anthocyanidins (and vice versa) in Arabidopsis seeds would be interesting. Such a systematic investigation was performed by Sentandreu *et al.* (2010) in pomegranate, with very successful results, screening for specific known molecular weights of those compounds. Another suitable improved method for such a high throughoutput measurement would be the method developed by Lambert *et al.* (2015). They performed an uHPLC coupled to triple-quadrupole mass spectrometry (UHPLC-QqQ-MS) using the MRM mode to select compounds with specific characteristics, like neutral losses or specific fragment ions. This allows to scan also for minor compounds without sample purification or pre-concentration in a short time, thus improving selectivity and sensitivity in reduced time frames.

Better understanding of pyranoanthocyanidins and their derivatives, their occurence and their role, including their potential health benefits is also of great importance, also to understand more about the potential of the daily food, to mediate health, as an important health care factor (Martin and Li, 2017).

Apart from the role of pyranoanthocyanidins and condensed pigments, the role of flavonoid glycosylation in plants remains a topic of interest, with many open questions, needing further research in this field. Some aglycone decorations pointed out, to be key factors for the flavonoid physiological function against environmental stresses like enhanced UV-B tolerance (Tohge *et al.*, 2016; Peng *et al.*, 2017). Therefore, the interest in the flavonoid decorations increased, including interest about the biological function of different decorations as well as the underlying regulation mechanisms for the development of these decorations, which are often regulated tissue related and by stress induction. This makes the investigation of metabolic diversity and functional genomics approaches for flavonoid decorations and the respective enzymes an important topic, to investigate (Saigo *et al.*, 2020).

Examinations by genetic analysis of color mutants are very difficult, because many decorative genes do not exhibit a visible phenopye in mutant plants. This was only successful for some GTs that are involved in the first glycosylation step, like the anthocyanidin 3-*O*-glucosyl transferase (Saigo *et al.*, 2020).

One suggestion of Wu *et al.* (2018) is to unravel more detailed the role of glycosylation in stress response, for example be the prior discussed use of untargeted LC-MS-based metabolomic profiling and metabolic genome-wide association studies with samples from different abiotic environments. Because genes could also be induced due to other factors, like specific tissues, development stages or due to plant natural variations, this approach should be extended to those cases. For example, in the last years, some genes for different flavonoid decoration have been found, using metabolomic analysis of natural accessions (Ishihara *et al.*, 2016; Tohge *et al.*, 2016; Peng *et al.*, 2017). Also, as shown in this work the use of different mutant lines

can provide further insides. Sequence similarity of the coding sequences or (as shown in this work) of the amino acid sequences can give important further insights towards the function of genes (Saigo *et al.*, 2020).

Alternative approaches for the identification of most decorating enzymes or genes were performed by protein purification and screening for enzymatic activity (Saigo *et al.*, 2020). Most Arabidopsis GTs have been identified by transcriptome co-expression networks and following characterization of loss-of-function mutants or recombinant protein assays (Tohge *et al.*, 2005; Yonekura-Sakakibara *et al.*, 2007, 2008, 2012). For many GH1-type enzymes those studies were challenging, because recombinant expression was ineffective or not possible (Miyahara *et al.*, 2012; Nishizaki *et al.*, 2013; Miyahara *et al.*, 2014; Ishihara *et al.*, 2016). Even the removal of the N-terminus, which was in some cases a problematic factor, did not change the result (Miyahara *et al.*, 2014). In most cases, crude protein extracts were used, but for example for BGLU6, this was also not successful (Ishihara *et al.*, 2016). In the future, it will be of great interest, to purify BGLU1, BGLU3 and BGLU4 and to characterize their activity. In addition, more detailed subcellular localization studies will be performed, to increase further information on the activity and substrate availability.

Since, transcriptome co-expression networks can help to reveal the function of genes, which was also shown, to generate and underline further insides into the glycosylation pattern and the regulation by data of Wu *et al.* (2018) and from this work, co-expression analysis would be in addition interesting to the so far generated results for *BGLU1*, *BGLU3* and *BGLU4*. Since, it is known, that synthesis is regulated by environment and development, as well as organ- and tissue-specific, the co-expressed genes could give further insights to the function (Stracke *et al.*, 2010). In addition, nothing is known about the transcriptional regulation of GH1-type GT genes. Co-expressed transcription factors could give further insights into this. Therefore, co-expression analysis of *BGLU1*, *BGLU3* and *BGLU4* will be performed by Lennart Malte Sielmann in his master thesis. Since, the connection of metabolomics and transcriptomics can enhance information on the gene-to-metabolite network (Tohge *et al.*, 2005), RNA-Seq data of the respective mutants could maybe enhance information about the network, involved in the production of the respective product compounds of BGLU1, BGLU3 and BGLU4, and maybe also to find more information about putative donor substrates.

For distinc structural elucidation of the BGLU1 candidate product compound purification of the compound for future NMR measurements are in progress. Seed samples for purification and further structure elucidation by NMR of the candidate compounds of BGLU3 are growing.

4.9. Conclusion

Instead of hydrolase activity, BGLU1 to BGLU11 from Arabidopsis are supposed to exhibit GH1-type GT activity, based on phylogenetic analysis with GH1-type GTs from other organisms and AAGTs with known GH1-type GT activity. The functional characterization of *BGLU1*, *BGLU3* and *BGLU4* in this work revealed supposed GT activity for the encoded enzymes. The functional characterization was based on untargeted metabolic fingerprinting with uHPLC-DAD-ESI$^+$-QTOF-MS/MS of a loss-of-function T-DNA insertion, an overexpression line and the Col-0 wild type. Differential metabolite levels were investigated by fold changes

and metabolite features with higher peaks in the overexpression mutant and simultaneously lower or no peaks in the loss-of-function mutant compared to the wild type, were accounted for potential product compounds of the respective encoded protein. In contrast, candidate substrate compounds were suggested for metabolite features, showing higher peaks in the loss-of-function mutant and lower peaks in the overexpression mutant, compared to the wild type. Candidate product and substrate compounds were identified by their MS/MS spectra. This approach revealed a supposed triglycosylated kaempferol as product compound of BGLU1 glycosylation and a supposed triglycosylated quercetin as product compound of BGLU3 and BGLU4 glycosylation, respectively. In addition, for BGLU3 glycosylation of condensed flavanol–anthocyanidin compounds as well as a condensed (epi)gallocatechin–carboxypyranomalvidin is presumed. These condensed compounds are reported for the first time in Arabidopsis, while a condensed (epi)gallocatechin–carboxypyranomalvidin is for the first time reported in general. The genetic approach of a loss-of-function mutant and an overexpression mutant coupled to untargeted metabolic fingerprinting could successfully highlight the potential product compounds of all three investigated BGLUs, while also providing information about some substrates. This enables important insights into the function of three BGLUs from Arabidopsis, confirming the hypothesized GH1-type GT activity. Furthermore, the chosen approach provided an appropriate tool, to identify unknown minor flavonoid compounds (aglycones and glycosides), even for low metabolite levels, yielding new insights in flavonoid abundance and diversity in plants.

In future studies, the proposed enzymatic activity of BGLU1, BGLU3 and BGLU4 should be investigated by enzyme assays of the purified or recombinant protein, to verify the GH1-type GT activity. Structural elucidation by NMR should verify the proposed structures as well as reveal theprecise sugar atttachment sites.

References

Ablajan K (2011). A study of characteristic fragmentation of isoflavonoids by using negative ion ESI-MS^n. *J Mass Spectrom* 46(1):77–84.

Ablajan K, Abliz Z, Shang XY, He JM, Zhang RP and Shi JG (2006). Structural characterization of flavonol 3,7-di-*O*-glycosides and determination of the glycosylation position by using negative ion electrospray ionization tandem mass spectrometry. *J Mass Spectrom* 41(3):352–360.

Ahn D, Lee E, Kim B, Lee S, Lee T, Ahn M, Lim H, Cha D, Jeon H and Kim D (2014). Antioxidant and lifespan extending property of quercetin-3-*O*-dirhamnoside from *Curcuma longa* L. in *Caenorhabditis elegans*. *J Korean Soc Appl Biol Chem* 57(6):709–714.

Ahn YO, Shimizu B, Sakata K, Gantulga D, Zhou C, Zhou Z, Bevan DR and Esen A (2010). Scopolin-hydrolyzing β-glucosidases in roots of Arabidopsis. *Plant Cell Physiol.* 51(1):132–143.

Ahrazem O, Rubio-Moraga A, Trapero-Mozos A, Climent MF, Gómez-Cadenas A and Gómez-Gómez L (2015). Ectopic expression of a stress-inducible glycosyltransferase from saffron enhances salt and oxidative stress tolerance in Arabidopsis while alters anchor root formation. *Plant Sci.* 234:60–73.

Ahuja I, Kissen R and Bones AM (2012). Phytoalexins in defense against pathogens. *Trends Plant Sci.* 17(2):73–90.

Akiyama K, Matsuzaki K and Hayashi H (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435(7043):824–827.

Allen DR and McWhinney BC (2019). Quadrupole Time-of-Flight Mass Spectrometry: A Paradigm Shift in Toxicology Screening Applications. *Clin Biochem Rev* 40(3):135–146.

Alscher RG, Erturk N and Heath LS (2002). Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *J. Exp. Bot.* 53(372):1331–1341.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.* 215(3):403–410.

Amorim ACL, Hovell AMC, Pinto AC, Eberlin MN, Arruda NP, Pereira EJ, Bizzo HR, Catharino RR, Morais Filho ZB and Rezende CM (2009). Green and Roasted Arabica Coffees Differentiated by Ripeness, Process and Cup Quality via Electrospray Ionization Mass Spectrometry Fingerprinting. *J. Braz. Chem. Soc.* 20(2):313–321.

Andersen OM, Fossen T, Torskangerpoll K, Fossen A and Hauge U (2004). Anthocyanin from strawberry (Fragaria ananassa) with the novel aglycone, 5-carboxypyranopelargonidin. *Phytochemistry* 65(4):405–410.

Andres-Lacueva C, Shukitt-Hale B, Galli RL, Jauregui O, Lamuela-Raventos RM and Joseph JA (2005). Anthocyanins in aged blueberry-fed rats are found centrally and may enhance memory. *Nutr Neurosci* 8(2):111–120.

Aniszewski T, Definition, Typology and Occurrence of Alkaloids, 1–59 (Elsevier Science, 2007).

Appeldoorn M (2009). Dietary A- and B-type procyanidins – Characterization and biofunctional potential of an abundant and diverse group of phenolics. Ph.D. thesis, Wageningen University, The Netherlands.

Appelhagen I, Nordholt N, Seidel T, Spelt K, Koes R, Quattrochio F, Sagasser M and Weisshaar B (2015). TRANSPARENT TESTA 13 is a tonoplast P3A -ATPase required for vacuolar deposition of proanthocyanidins in Arabidopsis thaliana seeds. *Plant J.* 82(5):840–849.

Armah CN, Mackie AR, Roy C, Price K, Osbourn AE, Bowyer P and Ladha S (1999). The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol. *Biophys. J.* 76(1 Pt 1):281–290.

Asen S, Stewart R and Norris K (1975). Anthocyanin, flavonol copigments, and pH responsible for larkspur flower color. *Phytochemistry* 14(12):2677–2682.

Asenstorfer RE, Hayasaka Y and Jones GP (2001). Isolation and structures of oligomeric wine pigments by bisulfite-mediated ion-exchange chromatography. *J. Agric. Food Chem.* 49(12):5957–5963.

Babst BA, Chen HY, Wang HQ, Payyavula RS, Thomas TP, Harding SA and Tsai CJ (2014). Stress-responsive hydroxycinnamate glycosyltransferase modulates phenylpropanoid metabolism in Populus. *J. Exp. Bot.* 65(15):4191–4200.

Bailly C, El-Maarouf-Bouteau H and Corbineau F (2008). From intracellular signaling networks to cell death: the dual role of reactive oxygen species in seed physiology. *C. R. Biol.* 331(10):806–814.

Bais HP, Weir TL, Perry LG, Gilroy S and Vivanco JM (2006). The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu Rev Plant Biol* 57:233–266.

Bakker J, Bridle P, Honda T, Kuwano H, Saito N, Terahara N and Timberlake CF (1997). Identification of an Anthocyanin occurring in some Red Wines. *Phytochemistry* 44(7):1375–1382.

Bakker J, Picinelli A and Bridle P (1993). Model wine solutions: Colour and composition changes during ageing. *Vitis* 32:111–118.

Bakker J and Timberlake CF (1997). Isolation, Identification, and Characterization of New Color-Stable Anthocyanins Occurring in Some Red Wines. *J. Agric. Food Chem.* 45:35–43.

Banerjee S and Mazumdar S (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem* 2012:282574.

Barth C and Jander G (2006). Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *Plant J.* 46(4):549–562.

Baumert A, Milkowski C, Schmidt J, Nimtz M, Wray V and Strack D (2005). Formation of a complex pattern of sinapate esters in *Brassica napus* seeds, catalyzed by enzymes of a serine carboxypeptidase-like acyltransferase family? *Phytochemistry* 66(11):1334–1345.

Bednarek P, Pislewska-Bednarek M, Svatos A, Schneider B, Doubsky J, Mansurova M, Humphry M, Consonni C, Panstruga R, Sanchez-Vallet A, Molina A and Schulze-Lefert P (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* 323(5910):101–106.

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E and Huala E (2015). The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* 53(8):474–485.

Berini JL, Brockman SA, Hegeman AD, Reich PB, Muthukrishnan R, Montgomery RA and Forester JD (2018). Combinations of Abiotic Factors Differentially Alter Production of Plant Secondary Metabolites in Five Woody Plant Species in the Boreal-Temperate Transition Zone. *Front Plant Sci* 9:1257.

Blasco C and Picó Y, Liquid chromatography-mass spectrometry, 509–559 (Elsevier B.V., 2007). Chapter 14.

Bomser J, Madhavi DL, Singletary K and Smith MA (1996). In vitro anticancer activity of fruit extracts from Vaccinium species. *Planta Med.* 62(3):212–216.

Bowles D, Lim EK, Poppenberger B and Vaistij FE (2006). Glycosyltransferases of lipophilic small molecules. *Annu Rev Plant Biol* 57:567–597.

Brazier-Hicks M, Evans KM, Gershater MC, Puschmann H, Steel PG and Edwards R (2009). The C-glycosylation of flavonoids in cereals. *J. Biol. Chem.* 284(27):17926–17934.

Brouillard R and Dangles O (1994). Anthocyanin molecular interactions: the first step in the formation of new pigments during wine aging? *Food Chem.* 51:365–371.

Brouillard R and Lang J (1990). The hemiacetal-*cis*-chalcone equilibrium of malvin, a natural anthocyanin. *Can. J. Chem.* 68(5):755–761.

Burmeister WP, Cottaz S, Driguez H, Iori R, Palmieri S and Henrissat B (1997). The crystal structures of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure* 5(5):663–675.

Burmeister WP, Cottaz S, Rollin P, Vasella A and Henrissat B (2000). High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base. *J. Biol. Chem.* 275(50):39385–39393.

Bykova NV, Hu J, Ma Z and Igamberdiev AU, The Role of Reactive Oxygen and Nitrogen Species in Bioenergetics, Metabolism, and Signaling During Seed Germination, 177–195 (Springer International Publishing, Cham, 2015).

Calderón-Montaño JM, Burgos-Morón E, Pérez-Guerrero C and López-Lázaro M (2011). A review on the dietary flavonoid kaempferol. *Mini Rev Med Chem* 11(4):298–344.

Cameira dos Santos PJ, Brillouet JM, Cheynier V and Moutounet M (1996). Detection and partial characterisation of new anthocyanin-derived pigments in wine. *J. Sci. Food Agric.* 70:204–208.

Cao YY, Yang JF, Liu TY, Su ZF, Zhu FY, Chen MX, Fan T, Ye NH, Feng Z, Wang LJ, Hao GF, Zhang J and Liu YG (2017). A Phylogenetically Informed Comparison of GH1 Hydrolases between Arabidopsis and Rice Response to Stressors. *Front Plant Sci* 8:350.

Carter C, Pan S, Zouhar J, Avila EL, Girke T and Raikhel NV (2004). The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. *Plant Cell* 16(12):3285–3303.

Castagnino C and Vercauteren J (1996). Castavinol, a new series of polyphenols from Bordeaux red wines. *Tetrahedron Lett.* 37(43):7739?7742.

Chapelle A, Morreel K, Vanholme R, Le-Bris P, Morin H, Lapierre C, Boerjan W, Jouanin L and Demont-Caulet N (2012). Impact of the absence of stem-specific β-glucosidases on lignin and monolignols. *Plant Physiol.* 160(3):1204–1217.

Chelikani P, Fita I and Loewen PC (2004). Diversity of structures and properties among catalases. *Cell. Mol. Life Sci.* 61(2):192–208.

Chen AY and Chen YC (2013). A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. *Food Chem* 138(4):2099–2107.

Chen Z, Zheng W, Chen L, Li C, Liang T, Chen Z, Xu H, Han Y, Kong L, Zhao X, Wang F, Wang Z and Chen S (2019). Green Fluorescent Protein- and Discosoma sp. Red Fluorescent Protein-Tagged Organelle Marker Lines for Protein Subcellular Localization in Rice. *Front Plant Sci* 10:1421.

Cheng J, Wei G, Zhou H, Gu C, Vimolmangkang S, Liao L and Han Y (2014). Unraveling the mechanism underlying the glycosylation and methylation of anthocyanins in peach. *Plant Physiol.* 166(2):1044–1058.

Chernushevich IV, Loboda AV and Thomson BA (2001). An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* 36(8):849–865.

Chiang Y, Kresge A and Pruszynski P (1992). Keto-enol equilibria in the pyruvic acid system: determination of the keto-enol equilibrium constants of pyruvic acid and pyruvate anion and the acidity constant of pyruvate enol in aqueous solution. *J. Am. Chem. Soc.* 114(8):3103–3107.

Chien A, Edgar DB and Trela JM (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* 127(3):1550–1557.

Chkaiban L, Botondi R, Bellincontro A, de Santis D, Kefalas P and Mencarelli F (2007). Influence of postharvest water stress on lipoxygenase and alcohol dehydrogenase activities, and on the composition of some volatile compounds of Gewuerztraminer grapes dehydrated under controlled and uncontrolled thermohygrometric conditions. *Aust. J. Grape. Wine Res.* 13(3):142–149.

Chong J, Baltz R, Schmitt C, Beffa R, Fritig B and Saindrenan P (2002). Downregulation of a pathogen-responsive tobacco UDP-Glc:phenylpropanoid glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance. *Plant Cell* 14(5):1093–1107.

Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Oonanant W, Bevan DR, Esen A, Chen CJ, Opassiri R, Svasti J and Cairns JR (2008). Structural insights into rice BGlu1 beta-glucosidase oligosaccharide hydrolysis and transglycosylation. *J. Mol. Biol.* 377(4):1200–1215.

Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Svasti J and Cairns JR (2011). The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *J. Struct. Biol.* 173(1):169–179.

Clough SJ and Bent AF (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* 16(6):735–743.

Colantuoni A, Bertuglia S, Magistretti MJ and Donato L (1991). Effects of Vaccinium Myrtillus anthocyanosides on arterial vasomotion. *Arzneimittelforschung* 41(9):905–909.

Costa D, Galvão AM, Di Paolo RE, Freitas AA, Lima J, Quina FH and Maçanita AL (2014). Photochemistry of the hemiketal form of anthocyanins and its potential role in plant protection from UV-B radiation. *Tetrahedron* xxx:1–6.

Cumming G, Fidler F and Vaux DL (2007). Error bars in experimental biology. *J. Cell Biol.* 177(1):7–11.

Curtis MD and Grossniklaus U (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* 133(2):462–469.

Cuyckens F and Claeys M (2005). Determination of the glycosylation site in flavonoid mono-*O*-glycosides by collision-induced dissociation of electrospray-generated deprotonated and sodiated molecules. *J Mass Spectrom* 40(3):364–372.

Cuyckens F, Rozenberg R, de Hoffmann E and Claeys M (2001). Structure characterization of flavonoid *O*-diglycosides by positive and negative nano-electrospray ionization ion trap mass spectrometry. *J Mass Spectrom* 36(11):1203–1210.

Czjzek M, Cicek M, Zamboni V, Bevan DR, Henrissat B and Esen A (2000). The mechanism of substrate (aglycone) specificity in beta-glucosidases is revealed by crystal structures of mutant maize beta-glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. *Proc. Natl. Acad. Sci. U.S.A.* 97(25):13555–13560.

Czjzek M, Cicek M, Zamboni V, Burmeister WP, Bevan DR, Henrissat B and Esen A (2001). Crystal structure of a monocotyledon (maize ZMGlu1) beta-glucosidase and a model of its complex with p-nitrophenyl beta-D-thioglucoside. *Biochem. J.* 354(Pt 1):37–46.

Dai J and Mumper RJ (2010). Plant phenolics: extraction, analysis and their antioxidant and anticancer properties. *Molecules* 15(10):7313–7352.

Dai W, Tan J, Xie D, Li P, Lv H, Zhu Y and Lin Z (2017). Study of the glycosylated secondary metabolites in tea (*Camellia sinensis* L.) using UHPLC/Q-TOF/MS. *Agilent Technologies* 1–12.

Dallas C, Ricardo da Silva J and Laureano O (1996). Products formed in model wine solutions involving anthocyanins, procyanidin B2, and acetaldehyde. *J. Agric. Food Chem.* 44:2402–2407.

D'Auria JC (2006). Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* 9(3):331–340.

de Freitas V and Mateus N (2011). Formation of pyranoanthocyanins in red wines: a new and diverse class of anthocyanin derivatives. *Anal Bioanal Chem* 401(5):1463–1473.

de Moraes CM, Mescher MC and Tumlinson JH (2001). Caterpillar-induced nocturnal plant volatiles repel conspecific females. *Nature* 410(6828):577–580.

de Pascual-Teresa S (2014). Molecular mechanisms involved in the cardiovascular and neuroprotective effects of anthocyanins. *Arch. Biochem. Biophys.* 559:68–74.

Del Rio D, Stewart AJ, Mullen W, Burns J, Lean ME, Brighenti F and Crozier A (2004). HPLC-MSn analysis of phenolic compounds and purine alkaloids in green and black tea. *J. Agric. Food Chem.* 52(10):2807–2815.

Delgado-Vargas F, Jim?nez AR and Paredes-L?pez O (2000). Natural pigments: carotenoids, anthocyanins, and betalains– characteristics, biosynthesis, processing, and stability. *Crit Rev Food Sci Nutr* 40(3):173–289.

Desai KM, Sessa WC and Vane JR (1991). Involvement of nitric oxide in the reflex relaxation of the stomach to accommodate food or fluid. *Nature* 351(6326):477–479.

Dietz S, Herz K, Doell S, Haider S, Jandt U, Bruelheide H and Scheel D (2019). Semi?polar root exudates in natural grassland communities. *Ecology and Evolution* 1–16.

Dincheva I, Badjakov I, Kondakova V, Dobson P, McDougall G and Stewart D (2013). Iidentification of the phenolic compounds in bulgarian raspberry cultivars by LC-ESI-MS^n. *IJASR* 3(3):127–138.

Dong X, Jiang Y and Hur Y (2019). Genome-Wide Analysis of Glycoside Hydrolase Family 1 β-glucosidase Genes in Brassica rapa and Their Potential Role in Pollen Development. *Int J Mol Sci* 20(7).

Dudareva N, Klempien A, Muhlemann JK and Kaplan I (2013). Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol.* 198(1):16–32.

Dudareva N, Pichersky E and Gershenzon J (2004). Biochemistry of plant volatiles. *Plant Physiol.* 135(4):1893–1902.

Dueñas M, Fulcrand H and Cheynier V (2006). Formation of anthocyanin–flavanol adducts in model solutions. *Anal. Chim. Acta* 563:15–25.

Dueñas M, Pérez-Alonso JJ, Santos-Buelga C and Escribano-Bailón (2008). Anthocyanin composition in fig (*Ficus carica* L.). *J. Food Compost. Anal.* 21:107–115.

Duffey S and Stout M (1996). Antinutritive defenses and toxic components against insects. *Arch. Insect Biochem. Physiol.* 32:3–37.

Edelman J, Shibko S and Keys A (1959). The role of the scutellum of cereal seedlings in the synthesis and transport of sucrose. *J. Exp. Bot.* 10(2):178–189.

Edwards K, Johnstone C and Thompson C (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* 19(6):1349.

Eglinton J, Griesser M, Henschke P, Kwiatkowski M, Parker M and Herderich M, Yeast-Mediated Formation of Pigmented Polymers in Red Wine, 7–21 (American Chemical Society, Washington, DC, 2004). ACS Symposium Series 886, Chapter 2.

El-Aneed A, Cohen A and Banoub J (2009). Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Appl Spectrosc Rev* 44:210–230.

El-Maarouf-Bouteau H and Bailly C (2008). Oxidative signaling in seed germination and dormancy. *Plant Signal Behav* 3(3):175–182.

Es-Safi NE, Fulcrand H, Cheynier V and Moutounet M (1999). Studies on the Acetaldehyde-Induced Condensation of (−)-Epicatechin and Malvidin 3-*O*-Glucoside in a Model Solution System. *J. Agric. Food Chem.* 47(5):2096–2102.

Escamilla-Treviño LL, Chen W, Card ML, Shih MC, Cheng CL and Poulton JE (2006). Arabidopsis thaliana beta-Glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. *Phytochemistry* 67(15):1651–1660.

Escribano-Bailón T, Dangles O and Brouillard R (1996). Ccoupling reactions between flavylium ions and catechin. *Phytochemistry* 41(6):1583–1592.

Falcone Ferreyra ML, Rius SP and Casati P (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci* 3:222.

Farr J, Sigurdson G and de Giusti M (2018). Influence of cyanidin glycosylation patterns on carboxypyranoanthocyanin formation. *Food Chem.* 259:261–269.

Ferrer JL, Austin MB, Stewart C and Noel JP (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Biochem.* 46(3):356–370.

Fiehn O (2002). Metabolomics–the link between genotypes and phenotypes. *Plant Mol. Biol.* 48(1-2):155–171.

Finch-Savage WE and Leubner-Metzger G (2006). Seed dormancy and the control of germination. *New Phytol.* 171(3):501–523.

Fini A, Guidi L, Ferrini F, Brunetti C, Di Ferdinando M, Biricolti S, Pollastri S, Calamai L and Tattini M (2012). Drought stress has contrasting effects on antioxidant enzymes activity and phenylpropanoid biosynthesis in Fraxinus ornus leaves: an excess light stress affair? *J. Plant Physiol.* 169(10):929–939.

Flamini R (2012). Recent Applications of Mass Spectrometry in the Study of Grape and Wine Polyphenols. *ISRN Spectroscopy* 2013:1–45.

Fossen T and Andersen OM (2003). Anthocyanins from red onion, Allium cepa, with novel aglycone. *Phytochemistry* 62(8):1217–1220.

Fossen T, Rayyan S and Andersen OM (2004). Dimeric anthocyanins from strawberry (Fragaria ananassa) consisting of pelargonidin 3-glucoside covalently linked to four flavan-3-ols. *Phytochemistry* 65(10):1421–1428.

Foyer CH and Halliwell B (1976). The presence of glutathione and glutathione reductase in chloroplasts: A proposed role in ascorbic acid metabolism. *Planta* 133(1):21–25.

Francia-Aricha EM, Guerra MT, Rivas-Gonzalo JC and Santos-Buelga C (1997). New Anthocyanin Pigments Formed after Condensation with Flavanols. *J. Agric. Food Chem.* 45(6):2262–2266.

Fraser CM, Thompson MG, Shirley AM, Ralph J, Schoenherr JA, Sinlapadech T, Hall MC and Chapple C (2007). Related Arabidopsis serine carboxypeptidase-like sinapoylglucose acyltransferases display distinct but overlapping substrate specificities. *Plant Physiol.* 144(4):1986–1999.

Fraser K, Harrison S, Lane G, Otter D, Hemar Y, Quek SY and Rasmussen S (2012). HPLC-MS/MS profiling of proanthocyanidins in teas: A comparative study. *J Food Compost Anal* 26(1–2):43–51.

Frommann JF (2016). Studies towards a functional characterization oft the *Arabidopsis thaliana* genes *BGLU1* through *BGLU5*. *Master thesis: University of Bielefeld, Chair of Genome Research* .

Frydman A, Liberman R, Huhman DV, Carmeli-Weissberg M, Sapir-Mir M, Ophir R, W Sumner L and Eyal Y (2013). The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. *Plant J.* 73(1):166–178.

Fukuchi-Mizutani M, Okuhara H, Fukui Y, Nakao M, Katsumoto Y, Yonekura-Sakakibara K, Kusumi T, Hase T and Tanaka Y (2003). Biochemical and molecular characterization of a novel UDP-glucose:anthocyanin 3'-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian. *Plant Physiol.* 132(3):1652–1663.

Fukui Y, Kusumi T, Masuda K, Iwashita T and Nomoto K (2002). Structure of rosacyanin B, a novel pigment from the petals of *Rosa hybrida*. *Tetrahedron Lett.* 43:2637–2639.

Fulcrand H, Benabdeljalil C, Rigaud J, Cheynier V and Moutounet M (1998). A new class of wine pigments generated by reaction between pyruvic acid and grape anthocyanins. *Phytochemistry* 47(7):1401–1407.

Fulcrand H, dos Santos P, Sarni-Manchado P, Cheynier V and Favre-Bonvin J (1996). Structure of new anthocyanin-derived wine pigments. *J. Gem. Soc.,Perkin Trans. I* 7:735–739.

Gachon CM, Langlois-Meurinne M and Saindrenan P (2005). Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci.* 10(11):542–549.

Garabedian A, Benigni P, Ramirez C, Baker E, Liu T, Smith R and Fernandez-Lima F (2018). Towards discovery and targeted peptide biomarker detection using nanoESI-TIMS-TOF MS. *J. Am. Soc. Mass Spectrom.* 29:817–826.

Garrido-Gil P, Fernandez-Rodr?guez P, Rodr?guez-Pallares J and Labandeira-Garcia JL (2017). Laser capture microdissection protocol for gene expression analysis in the brain. *Histochem. Cell Biol.* 148(3):299–311.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA and Smith HO (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6(5):343–345.

Gođevac D, Tešević V, Veličković M, Vujisić L, Vajs V and Milosavljević S (2010). Polyphenolic compounds in seeds from some grape cultivars grown in Serbia. *J. Serb. Chem. Soc.* 75(10):1641–1652.

Gómez-Anduro G, Ceniceros-Ojeda EA, Casados-Vázquez LE, Bencivenni C, Sierra-Beltrán A, Murillo-Amador B and Tiessen A (2011). Genome-wide analysis of the beta-glucosidase gene family in maize (*Zea mays* L. var B73). *Plant Mol. Biol.* 77(1-2):159–183.

González-Manzano S, Pérez-Alonso JJ, Salinas-Moreno Y, Mateus N, Silva AMS, de Freitas V and Santos-Buelga C (2008). Flavanol–anthocyanin pigments in corn: NMR characterisation and presence in different purple corn varieties. *J. Food Compost. Anal.* 21:521–526.

González-Paramás AM, Lopes da Silva F, Martín-López P, Macz-Pop G, González-Manzano S, Alcalde-Eon C, Pérez-Alonso JJ, Escribano-Bailón MT, Rivas-Gonzalo J and Santos-Buelga C (2006). Flavanol-anthocyanin condensed pigments in plant extracts. *Food Chem.* 94:428–436.

González-Sarrías A, Combet E, Pinto P, Mena P, Dall?Asta M, Garcia-Aloy M, Rodríguez-Mateos A, Gibney E, Dumont J, Massaro M, Sánchez-Meca J, Morand C and García-Conesa M (2017). A systematic review and meta-analysis of the effects of flavanol-containing tea, cocoa and apple products on body composition and blood lipids: exploring the factors responsible for variability in their efficacy. *Nutrients* 9(7):2–28.

Gould K, McKelvie J and Markham K (2002). Do anthocyanins function as antioxidants in leaves? Imaging of H_2O_2 in red and green leaves after mechanical injury. *Plant Cell Environ.* 25(10):1261–1269.

Grassi D, Desideri G, Necozione S, di Giosia P, Barnabei R, Allegaert L, Bernaert H and Ferri C (2015). Cocoa consumption dose-dependently improves flow-mediated dilation and arterial stiffness decreasing blood pressure in healthy individuals. *J. Hypertens.* 33(2):294–303.

Griesser M, Hoffmann T, Bellido ML, Rosati C, Fink B, Kurtzer R, Aharoni A, Muñoz Blanco J and Schwab W (2008). Redirection of flavonoid biosynthesis through the down-regulation of an anthocyanidin glucosyltransferase in ripening strawberry fruit. *Plant Physiol.* 146(4):1528–1539.

Gronquist M, Bezzerides A, Attygalle A, Meinwald J, Eisner M and Eisner T (2001). Attractive and defensive functions of the ultraviolet pigments of a flower (Hypericum calycinum). *Proc. Natl. Acad. Sci. U.S.A.* 98(24):13745–13750.

Grotewold E (2006). The genetics and biochemistry of floral pigments. *Annu Rev Plant Biol* 57:761–780.

Grotewold E and Davies K (2008). Trafficking and sequestration of anthocyanins. *Nat Prod Commun* 3(8):1251–1258.

Hackman RM, Polagruto J, Zhu Q, Sun B and Fujii C H adn Keen (2008). Flavanols: digestion, absorption and bioactivity. *Phytochem Rev* 7:195–208.

Halliwell B (1996). Antioxidants in human health and disease. *Annu. Rev. Nutr.* 16:33–50.

Hamabata A, Garc?a-Maya M, Romero T and Bernal-Lugo I (1988). Kinetics of the Acidification Capacity of Aleurone Layer and Its Effect upon Solubilization of Reserve Substances from Starchy Endosperm of Wheat. *Plant Physiol.* 86(3):643–644.

Han Y, Watanabe S, Shimada H and Sakamoto A (2020). Dynamics of the leaf endoplasmic reticulum modulate β-glucosidase-mediated stress-activated ABA production from its glucosyl ester. *J. Exp. Bot.* 71(6):2058–2071.

Hancock SM, Corbett K, Fordham-Skelton AP, Gatehouse JA and Davis BG (2005). Developing promiscuous glycosidases for glycoside synthesis: residues W433 and E432 in Sulfolobus solfataricus beta-glycosidase are important glucoside- and galactoside-specificity determinants. *Chembiochem* 6(5):866–875.

Hancock SM, Vaughan MD and Withers SG (2006). Engineering of glycosidases and glycosyltransferases. *Curr Opin Chem Biol* 10(5):509–519.

Hartmann T (2007). From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry* 68(22-24):2831–2846.

Hause B, Meyer K, Viitanen PV, Chapple C and Strack D (2002). Immunolocalization of 1- O-sinapoylglucose:malate sinapoyl-transferase in *Arabidopsis thaliana*. *Planta* 215(1):26–32.

He J, Carvalho AR, Mateus N and De Freitas V (2010a). Spectral features and stability of oligomeric pyranoanthocyanin-flavanol pigments isolated from red wines. *J. Agric. Food Chem.* 58(16):9249–9258.

He J, Oliveira J, Silva AM, Mateus N and De Freitas V (2010b). Oxovitisins: a new class of neutral pyranone-anthocyanin derivatives in red wines. *J. Agric. Food Chem.* 58(15):8814–8819.

He J, Santos-Buelga C, Silva AM, Mateus N and de Freitas V (2006). Isolation and structural characterization of new anthocyanin-derived yellow pigments in aged red wines. *J. Agric. Food Chem.* 54(25):9598–9603.

Hectors K, Van Oevelen S, Geuns J, Guisez Y, Jansen MA and Prinsen E (2014). Dynamic changes in plant secondary metabolites during UV acclimation in *Arabidopsis thaliana*. *Physiol Plant* 152(2):219–230.

Hellström JK and Mattila PH (2008). HPLC determination of extractable and unextractable proanthocyanidins in plant materials. *J. Agric. Food Chem.* 56(17):7617–7624.

Hendricks SB and Taylorson RB (1974). Promotion of seed germination by nitrate, nitrite, hydroxylamine, and ammonium salts. *Plant Physiol.* 54(3):304–309.

Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP and Davies G (1996). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. U.S.A.* 93(11):5674.

Henry BS, Natural food colours, 40–79 (Springer US, Boston, MA, 1996).

Heredia F, Francia-Aricha E, Rivas-Gonzalo J, Vicario I and Santos-Buelga C (1998). Chromatographic characterization of anthocyanins from red grapes-I. pH ef- fect. *Food Chem.* 63(4):491–498.

Hill R, Hargrove M and Arredondo-Peter R (2016). Phytoglobin: a novel nomenclature for plant globins accepted by the globin community at the 2014 XVIII conference on Oxygen-Binding and Sensing Proteins. *F1000Res* 5:212.

Hillebrand S, Schwarz M and Winterhalter P (2004). Characterization of anthocyanins and pyranoanthocyanins from blood orange [*Citrus sinensis* (L.) Osbeck] juice. *J. Agric. Food Chem.* 52(24):7331–7338.

Hirota S and Takahama U (2017). Inhibition of Pancreatin-Induced Digestion of Cooked Rice Starch by Adzuki (Vigana angularis) Bean Flavonoids and the Possibility of a Decrease in the Inhibitory Effects in the Stomach. *J. Agric. Food Chem.* 65(10):2172–2179.

Hollands WJ, Voorspoels S, Jacobs G, Aaby K, Meisland A, Garcia-Villalba R, Tomas-Barberan F, Piskula MK, Mawson D, Vovk I, Needs PW and Kroon PA (2017). Development, validation and evaluation of an analytical method for the determination of monomeric and oligomeric procyanidins in apple extracts. *J Chromatogr A* 1495:46–56.

Hollman P and Arts I (2000). Flavonols, flavones and flavanols ? nature, occurrence and dietary burden. *J. Sci. Food Agric.* 80:1081–1093.

Hösel W, Tober I, Eklund SH and Conn EE (1987). Characterization of beta-glucosidases with high specificity for the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) moench seedlings. *Arch. Biochem. Biophys.* 252(1):152–162.

Hou DX (2003). Potential mechanisms of cancer chemoprevention by anthocyanins. *Curr. Mol. Med.* 3(2):149–159.

Hua Y, Ekkhara W, Sansenya S, Srisomsap C, Roytrakul S, Saburi W, Takeda R, Matsuura H, Mori H and Ketudat Cairns JR (2015). Identification of rice Os4BGlu13 as a β-glucosidase which hydrolyzes gibberellin A4 1-O-β-d-glucosyl ester, in addition to tuberonic acid glucoside and salicylic acid derivative glucosides. *Arch. Biochem. Biophys.* 583:36–46.

Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, Mueller LA, Bhattacharyya D, Bhaya D, Sobral BW, Beavis W, Meinke DW, Town CD, Somerville C and Rhee SY (2001). The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* 29(1):102–105.

Hvattum E and Ekeberg D (2003). Study of the collision-induced radical cleavage of flavonoid glycosides using negative electrospray ionization tandem quadrupole mass spectrometry. *J Mass Spectrom* 38(1):43–49.

Igamberdiev AU, Baron K, Manac'h-Little N, Stoimenova M and Hill RD (2005). The haemoglobin/nitric oxide cycle: involvement in flooding stress and effects on hormone signalling. *Ann. Bot.* 96(4):557–564.

Iijima K, Fyfe V and McColl KE (2003). Studies of nitric oxide generation from salivary nitrite in human gastric juice. *Scand. J. Gastroenterol.* 38(3):246–252.

Ishihara H, Tohge T, Viehöver P, Fernie AR, Weisshaar B and Stracke R (2016). Natural variation in flavonol accumulation in Arabidopsis is determined by the flavonol glucosyltransferase BGLU6. *J. Exp. Bot.* 67(5):1505–1517.

Ishii I, Sakaguchi K, Fujita K, Ozeki Y and Miyahara T (2017). A double knockout mutant of acyl-glucose-dependent anthocyanin glucosyltransferase genes in Delphinium grandiflorum. *J. Plant Physiol.* 216:74–78.

Islam MM, Tani C, Watanabe-Sugimoto M, Uraji M, Jahan MS, Masuda C, Nakamura Y, Mori IC and Murata Y (2009). Myrosinases, TGG1 and TGG2, redundantly function in ABA and MeJA signaling in Arabidopsis guard cells. *Plant Cell Physiol.* 50(6):1171–1175.

Iuchi S, Suzuki H, Kim YC, Iuchi A, Kuromori T, Ueguchi-Tanaka M, Asami T, Yamaguchi I, Matsuoka M, Kobayashi M and Nakajima M (2007). Multiple loss-of-function of Arabidopsis gibberellin receptor AtGID1s completely shuts down a gibberellin signal. *Plant J.* 50(6):958–966.

Jaakola L (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci.* 18(9):477–483.

Jackson R, Knisley D, McIntosh C and Pfeiffer P (2011). Predicting Flavonoid UGT Regioselectivity. *Adv Bioinformatics* 2011:506583.

Jefferson RA, Kavanagh TA and Bevan MW (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6(13):3901–3907.

Jeong CY, Kim JH, Lee WJ, Jin JY, Kim J, Hong SW and Lee H (2018). AtMyb56 Regulates Anthocyanin Levels via the Modulation of *AtGPT2* Expression in Response to Sucrose in Arabidopsis. *Mol. Cells* 41(4):351–361.

Jiang Y, Cai Z, Xie W, Long T, Yu H and Zhang Q (2012). Rice functional genomics research: progress and implications for crop genetic improvement. *Biotechnol. Adv.* 30(5):1059–1070.

John J and Sarada S (2012). Role of phenolics in allelopathic interactions. *Allelopathy Journal.* 29(2):215–229.

Jones P, Messner B, Nakajima J, Sch?ffner AR and Saito K (2003). UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. *J. Biol. Chem.* 278(45):43910–43918.

Jones P and Vogt T (2001). Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta* 213(2):164–174.

Katoh K and Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30(4):772–780.

Kennedy RA, Rumpho ME and Fox TC (1992). Anaerobic metabolism in plants. *Plant Physiol.* 100(1):1–6.

Kent K, Charlton K, Roodenrys S, Batterham M, Potter J, Traynor V, Gilbert H, Morgan O and Richards R (2017). Consumption of anthocyanin-rich cherry juice for 12 weeks improves memory and cognition in older adults with mild-to-moderate dementia. *Eur J Nutr* 56(1):333–341.

Kerhoas L, Aouak D, Cingöz A, Routaboul JM, Lepiniec L, Einhorn J and Birlirakis N (2006). Structural characterization of the major flavonoid glycosides from *Arabidopsis thaliana* seeds. *J. Agric. Food Chem.* 54(18):6603–6612.

Kessler A and Baldwin IT (2001). Defensive function of herbivore-induced plant volatile emissions in nature. *Science* 291(5511):2141–2144.

Kessler A and Baldwin IT (2002). Plant responses to insect herbivory: the emerging molecular analysis. *Annu Rev Plant Biol* 53:299–328.

Kessler A and Kalske A (2018). Plant secondary metabolite diversity and species interactions. *Annu. Rev. Ecol. Evol. Syst.* 49:115–138.

Ketudat Cairns J, Mahong B, Baiya S and Jeon J (2015). β-Glucosidases: multitasking, moonlighting or simply misunderstood? *Plant Sci.* 241:246–259.

Kim I, Heo J, Chang K, Lee SA, Lee M, Lim C and Lim J (2010). Overexpression and inactivation of UGT73B2 modulate tolerance to oxidative stress in Arabidopsis. *J. Plant Biol.* 53:233–239.

Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J and Bolton EE (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47(D1):D1102–D1109.

Kiran NS, Benková E, Reková A, Dubová J, Malbeck J, Palme K and Brzobohatý B (2012). Retargeting a maize β-glucosidase to the vacuole–evidence from intact plants that zeatin-O-glucoside is stored in the vacuole. *Phytochemistry* 79:67–77.

Kitamura S, Shikazono N and Tanaka A (2004). TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. *Plant J.* 37(1):104–114.

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD and Penin AA (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant J.* 88(6):1058–1070.

Köllner TG, Schnee C, Gershenzon J and Degenhardt J (2004). The variability of sesquiterpenes emitted from two Zea mays cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. *Plant Cell* 16(5):1115–1131.

König S, Feussner K, Kaever A, Landesfeind M, Thurow C, Karlovsky P, Gatz C, Polle A and Feussner I (2014). Soluble phenylpropanoids are involved in the defense response of Arabidopsis against Verticillium longisporum. *New Phytol.* 202(3):823–837.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH and Phillippy AM (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27(5):722–736.

Krikorian R, Shidler MD, Nash TA, Kalt W, Vinqvist-Tymchuk MR, Shukitt-Hale B and Joseph JA (2010). Blueberry supplementation improves memory in older adults. *J. Agric. Food Chem.* 58(7):3996–4000.

Kruve A, Herodes K and Leito I (2010). Optimization of electrospray interface and quadrupole ion trap mass spectrometer parameters in pesticide liquid chromatography/electrospray ionization mass spectrometry analysis. *Rapid Commun. Mass Spectrom.* 24:919–926.

Kuromori T, Takahashi S, Kondou Y, Shinozaki K and Matsui M (2009). Phenome analysis in plant species using loss-of-function and gain-of-function mutants. *Plant Cell Physiol.* 50(7):1215–1231.

Lairson LL, Wakarchuk WW and Withers SG (2007). Alternative donor substrates for inverting and retaining glycosyltransferases. *Chem. Commun. (Camb.)* (4):365–367.

Lambert M, Meudec E, Verbaere A, Mazerolles G, Wirth J, Masson G, Cheynier V and Sommerer N (2015). A High-Throughput UHPLC-QqQ-MS Method for Polyphenol Profiling in Rosé Wines. *Molecules* 20:7890–7914.

Lattanzio V, Kroon P, Quideau S and Treutter D, Plant Phenolics – Secondary Metabolites with Diverse Functions, 1–35 (Wiley-Blackwell, Oxford, 2008).

Le Roy J, Huss B, Creach A, Hawkins S and Neutelings G (2016). Glycosylation Is a Major Regulator of Phenylpropanoid Availability and Biological Activity in Plants. *Front Plant Sci* 7:735.

Lee JH, Kang NS, Shin SO, Shin SH, Lim SG, Suh DY, Baek IY, Park KY and Ha TJ (2009). Characterisation of anthocyanins in the black soybean (*Glycine max* L.) by HPLC-DAD-ESI/MS analysis. *Food Chem.* 112:226–231.

Lee KH, Piao HL, Kim HY, Choi SM, Jiang F, Hartung W, Hwang I, Kwak JM, Lee IJ and Hwang I (2006). Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. *Cell* 126(6):1109–1120.

Leikert JF, Räthel TR, Wohlfart P, Cheynier V, Vollmar AM and Dirsch VM (2002). Red wine polyphenols enhance endothelial nitric oxide synthase expression and subsequent nitric oxide release from endothelial cells. *Circulation* 106(13):1614–1617.

Li HJ and Deinzer ML (2007). Tandem mass spectrometry for sequencing proanthocyanidins. *Anal. Chem.* 79(4):1739–1748.

Li P, Li YJ, Zhang FJ, Zhang GZ, Jiang XY, Yu HM and Hou BK (2017). The Arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. *Plant J.* 89(1):85–103.

Li Z, Ngojeh G, DeWitt P, Zheng Z, Chen M, Lainhart B, Li V and Felpo P (2008). Synthesis of a library of glycosylated flavonols. *Tetrahedron Lett.* 49(51):7243–7245.

Liao H, Cai Y and Haslam E (1992). Polyphenol Interactions. Anthocyanins: Co-pigmentation and Colour Changes in Red Wines. *J Sci Food Agric* 59:299–305.

Lijavetzky D, Ruiz-García L, Cabezas JA, De Andrés MT, Bravo G, Ibanez A, Carreno J, Cabello F, Ibáñez J and Martínez-Zapater JM (2006). Molecular genetics of berry colour variation in table grape. *Mol. Genet. Genomics* 276(5):427–435.

Lillo C, Lea US and Ruoff P (2008). Nutrient depletion as a key factor for manipulating gene expression and product formation in different branches of the flavonoid pathway. *Plant Cell Environ.* 31(5):587–601.

Lim EK and Bowles DJ (2004). A class of plant glycosyltransferases involved in cellular homeostasis. *EMBO J.* 23(15):2915–2922.

Lim EK, Higgins GS, Li Y and Bowles DJ (2003). Regioselectivity of glucosylation of caffeic acid by a UDP-glucose:glucosyltransferase is maintained in planta. *Biochem. J.* 373(Pt 3):987–992.

Lim EK, Li Y, Parr A, Jackson R, Ashford DA and Bowles DJ (2001). Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in Arabidopsis. *J. Biol. Chem.* 276(6):4344–4349.

Lin LZ and Harnly JM (2007). A screening method for the identification of glycosylated flavonoids and other phenolic compounds using a standard analytical approach for all plant materials. *J. Agric. Food Chem.* 55(4):1084–1096.

Lin Y, Golovnina K, Chen ZX, Lee HN, Negron YL, Sultana H, Oliver B and Harbison ST (2016). Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual Drosophila melanogaster. *BMC Genomics* 17:28.

Livak KJ and Schmittgen TD (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.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM and Henrissat B (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42(Database issue):D490–495.

Lorenzen M, Racicot V, Strack D and Chapple C (1996). Sinapic acid ester metabolism in wild type and a sinapoylglucose-accumulating mutant of arabidopsis. *Plant Physiol.* 112(4):1625–1630.

Lotito SB, Actis-Goretta L, Renart ML, Caligiuri M, Rein D, Schmitz HH, Steinberg FM, Keen CL and Fraga CG (2000). Influence of oligomer chain length on the antioxidant activity of procyanidins. *Biochem. Biophys. Res. Commun.* 276(3):945–951.

Lu Y, Foo Y and Sun Y (2002). New pyranoanthocyanins from blackcurrant seeds. *Tetrahedron Lett.* 43(41):7341–7344.

Lu Y, Sun Y and Foo Y (2000). Novel pyranoanthocyanins from black currant seed. *Tetrahedron Lett.* 41:5975–5978.

Luang S, Cho JI, Mahong B, Opassiri R, Akiyama T, Phasai K, Komvongsa J, Sasaki N, Hua YL, Matsuba Y, Ozeki Y, Jeon JS and Cairns JR (2013). Rice Os9BGlu31 is a transglucosidase with the capacity to equilibrate phenylpropanoid, flavonoid, and phytohormone glycoconjugates. *J. Biol. Chem.* 288(14):10111–10123.

Ma Z, Bykova NV and Igamberdieva AU (2017). Cell signaling mechanisms and metabolic regulation of germination and dormancy in barley seeds. *Crop J.* 5(6):459–477.

Ma Z, Marsolais F, Bykova NV and Igamberdiev AU (2016). Nitric Oxide and Reactive Oxygen Species Mediate Metabolic Changes in Barley Seed Embryo during Germination. *Front Plant Sci* 7:138.

MacKenzie K and Smith M, Chapter1 - Introduction, 3–19 (Elsevier Scienc Ltd., 2002). Pergamon Materials Series, Volume 6.

Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Bélanger A, Fournel-Gigleux S, Green M, Hum DW, Iyanagi T, Lancet D, Louisot P, Magdalou J, Chowdhury JR, Ritter JK, Schachter H, Tephly TR, Tipton KF and Nebert DW (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* 7(4):255–269.

Macz-Pop GA, González-Param?ás AM, Pérez-Alonso JJ and Rivas-Gonzalo JC (2006a). New flavanol–anthocyanin condensed pigments and anthocyanin composition in guatemalan beans (*Phaseolus* spp.). *J. Agric. Food Chem.* 54(2):536–542.

Macz-Pop GA, Rivas-Gonzalo JC, Pérez-Alonso JJ and González-Paramás AM (2006b). Natural occurrence of free anthocyanin aglycones in beans (*Phaseolus vulgaris* L.). *Food Chem.* 94:448–456.

Mahajan M, Ahuja PS and Yadav SK (2011). Post-transcriptional silencing of flavonol synthase mRNA in tobacco leads to fruits with arrested seed set. *PLoS ONE* 6(12):e28315.

Majer P, Neugart S, Krumbein A, Schreiner M and Hideg E (2014). Singlet oxygen scavenging by leaf flavonoids contributes to sunlight acclimation in *Tilia platyphyllos*. *Environ. Exp. Bot.* 100:1–9.

Malviya R, Bansal V, Pal O and Sharma P (2010). High performance liquid chromatography: A short review. *J. Glob. Pharma Technol.* 2(5):22–26.

Marana SR (2006a). Molecular basis of substrate specificity in family 1 glycoside hydrolases. *IUBMB Life* 58(2):63–73.

Marana SR (2006b). Molecular basis of substrate specificity in family 1 glycoside hydrolases. *IUBMB Life* 58(2):63–73.

Marinova K, Pourcel L, Weder B, Schwarz M, Barron D, Routaboul JM, Debeaujon I and Klein M (2007). The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. *Plant Cell* 19(6):2023–2038.

Marone M, Mozzetti S, De Ritis D, Pierelli L and Scambia G (2001). Semiquantitative RT-PCR analysis to assess the expression levels of multiple transcripts from the same sample. *Biol Proced Online* 3:19–25.

Marquez A, Dueñas M, Serratosa M and Merida J (2013a). Identification by HPLC-MS of Anthocyanin Derivatives in Raisins. *J. Chem.* 2013:7 pages.

Marquez A, Dueñas M, Serratosa MP and Merida J (2012). Formation of vitisins and anthocyanin–flavanol adducts during red grape drying. *J. Agric. Food Chem.* 60(27):6866–6874.

Marquez A, Serratosa M and Merida J (2013b). Pyranoanthocyanin Derived Pigments in Wine: Structure and Formation during Winemaking. *J. Chem.* 2013:15 pages.

Marsh KJ, Wallis IR, Kulheim C, Clark R, Nicolle D, Foley WJ and Salminen JP (2020). New approaches to tannin analysis of leaves can be used to explain in vitro biological activities associated with herbivore defence. *New Phytol.* 225(1):488–498.

Martin C and Li J (2017). Medicine is not health care, food us health care: plant metabolic engineering, diet and human health. *New Phytol.* 261(3):699–719.

Masada S, Terasaka K, Oguchi Y, Okazaki S, Mizushima T and Mizukami H (2009). Functional and structural characterization of a flavonoid glucoside 1,6-glucosyltransferase from Catharanthus roseus. *Plant Cell Physiol.* 50(8):1401–1415.

Masuda K, Hori T, Tanabe K, Kano Y, Hirayama A and Nagase S (2007). Proanthocyanidin promotes free radical-scavenging activity in muscle tissues and plasma. *Appl Physiol Nutr Metab* 32(6):1097–1104.

Mateus N, de Pascual-Teresa S, Rivas-Gonzalo JC, Santos-Buelga C and de Freitas V (2002a). Structural diversity of anthocyanin-derived pigments in port wines. *Food Chem.* 76:335–342.

Mateus N, Oliveira J, González-Paramás A, Santos-Buelga C and de Freitas V (2005). Screening of Portisins (Vinylpyranoan-thocyanin Pigments) in Port Wine by LC/DAD-MS. *Food Sci Tech Int* 11(5):353–358.

Mateus N, Oliveira J, Pissarra J, González-Paramás A, Rivas-Gonzalo JC, Santos-Buelga C, Silva A and de Freitas V (2006). A new vinylpyranoanthocyanin pigment occurring in aged red wine. *Food Chem.* 97(4):689–695.

Mateus N, Silva AM, Rivas-Gonzalo JC, Santos-Buelga C and de Freitas V (2003). A new class of blue anthocyanin-derived pigments isolated from red wines. *J. Agric. Food Chem.* 51(7):1919–1923.

Mateus N, Silva AM, Santos-Buelga C, Rivas-Gonzalo JC and de Freitas V (2002b). Identification of anthocyanin-flavanol pigments in red wines by NMR and mass spectrometry. *J. Agric. Food Chem.* 50(7):2110–2116.

Mateus N, Silva AM, Vercauteren J and de Freitas V (2001). Occurrence of anthocyanin-derived pigments in red wines. *J. Agric. Food Chem.* 49(10):4836–4840.

Matsuba Y, Sasaki N, Tera M, Okamura M, Abe Y, Okamoto E, Nakamura H, Funabashi H, Takatsu M, Saito M, Matsuoka H, Nagasawa K and Ozeki Y (2010). A novel glucosylation reaction on anthocyanins catalyzed by acyl-glucose-dependent glucosyltransferase in the petals of carnation and delphinium. *Plant Cell* 22(10):3374–3389.

Mayer C, Zechel DL, Reid SP, Warren RA and Withers SG (2000). The E358S mutant of Agrobacterium sp. beta-glucosidase is a greatly improved glycosynthase. *FEBS Lett.* 466(1):40–44.

Mazzuca S, Spadafora A and Innocenti A (2006). Cell and tissue localization of β-glucosidase during the ripening of olive fruit (*Olea europaea*) by in situ activity assay. *Plant Sci.* 171:726–733.

McDougall GJ, Gordon S, Brennan R and Stewart D (2005). Anthocyanin-flavanol condensation products from black currant (Ribes nigrum L.). *J. Agric. Food Chem.* 53(20):7878–7885.

Meissner D, Albert A, B?ttcher C, Strack D and Milkowski C (2008). The role of UDP-glucose:hydroxycinnamate glucosyltrans-ferases in phenylpropanoid metabolism and the response to UV-B radiation in *Arabidopsis thaliana*. *Planta* 228(4):663–674.

Meistrowitz L (2016). Studies on functional characterisation of the genes *BGLU1* and *BGLU4* in *Arabidopsis thaliana*. *Bachelor thesis* University of Bielefeld, Chair of Genome Research.

Meng X, Li Y, Zhou T, Sun W, Shan X, Gao X and Wang L (2019). Functional Differentiation of Duplicated Flavonoid 3-O-Glycosyltransferases in the Flavonol and Anthocyanin Biosynthesis of Freesia hybrida. *Front Plant Sci* 10:1330.

Mera I, Falconí D and Córdova V (2019). Secondary metabolites in plants: main classes, phytochemical analysis and pharmacological activities. *Revista Bionatura* 4(4):1000–1009.

Miao J, Guo D, Zhang J, Huang Q, Qin G, Zhang X, Wan J, Gu H and Qu LJ (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 23(10):1233–1236.

Miki D, Itoh R and Shimamoto K (2005). RNA silencing of single and multiple members in a gene family of rice. *Plant Physiol.* 138:1903–1913.

Miller JR, Koren S and Sutton G (2010). Assembly algorithms for next-generation sequencing data. *Genomics* 95(6):315–327.

Miyahara T, Matsuba Y, Ozeki Y and Sasaki N (2011). Identification of genes in *Arabidopsis thaliana* with homology to a novel acyl-glucose dependent glucosyltransferase of carnations. *Plant Biotechnology* 28:311–315.

Miyahara T, Sakiyama R, Ozeki Y and Sasaki N (2013). Acyl-glucose-dependent glucosyltransferase catalyzes the final step of anthocyanin formation in Arabidopsis. *J. Plant Physiol.* 170(6):619–624.

Miyahara T, Takahashi M, Ozeki Y and Sasaki N (2012). Isolation of an acyl-glucose-dependent anthocyanin 7-*O*-glucosyltransferase from the monocot *Agapanthus africanus*. *J. Plant Physiol.* 169(13):1321–1326.

Miyahara T, Tani T, Takahashi M, Nishizaki Y, Ozeki Y and Sasaki N (2014). Isolation of anthocyanin 7-*O*-glucosyltransferase from Canterbury bells (*Campanula medium*). *Plant Biotechnology* 31:555–559.

Mo Y, Nagel C and Taylor LP (1992). Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proc. Natl. Acad. Sci. U.S.A.* 89(15):7213–7217.

Morant AV, Jørgensen K, Jørgensen C, Paquette SM, Sénchez-Pérez R, Møller BL and Bak S (2008). beta-Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69(9):1795–1813.

Morard P, Silvestre J, Lacoste L, Caumes E and Lamaze T (2004). Nitrate uptake and nitrite release by tomato roots in response to anoxia. *J. Plant Physiol.* 161(7):855–865.

Morata A, Gómez-Cordovés MC, Colomo B and Suárez JA (2003). Pyruvic acid and acetaldehyde production by different strains of Saccharomyces cerevisiae: relationship with Vitisin A and B formation in red wines. *J. Agric. Food Chem.* 51(25):7402–7409.

Morata A, López C, Tesfaye W, González C and Escott C, Anthocyanins as natural pigments in baverages, 383–428 (Technical University of Madrid, Madrid, Spain, 2019). Volume 14, Chapter 12.

Morazzoni P and Magistretti MJ (1990). Activity of Myrtocyan, an anthocyanoside complex from *Vaccinium myrtillus* (VMA), on platelet aggregation and adhesiveness. *Fitoterapa* 61(1):13–21.

Morikawa-Ichinose T, Miura D, Zhang L, Kim SJ and Maruyama-Nakashita A (2020). Involvement of BGLU30 in Glucosinolate Catabolism in the Arabidopsis Leaf under Dark Conditions. *Plant Cell Physiol.* .

Nabavi SM, ?amec D, Tomczyk M, Milella L, Russo D, Habtemariam S, Suntar I, Rastrelli L, Daglia M, Xiao J, Giampieri F, Battino M, Sobarzo-Sanchez E, Nabavi SF, Yousefi B, Jeandet P, Xu S and Shirooie S (2020). Flavonoid biosynthetic pathways in plants: Versatile targets for metabolic engineering. *Biotechnol. Adv.* 38:107316.

Nakabayashi R, Mori T and Saito K (2014). Alternation of flavonoid accumulation under drought stress in *Arabidopsis thaliana*. *Plant Signal Behav* 9(8).

Nakano RT, Yamada K, Bednarek P, Nishimura M and Hara-Nishimura I (2014). ER bodies in plants of the Brassicales order: biogenesis and association with innate immunity. *Front Plant Sci* 5:73.

Nakazaki A, Yamada K, Kunieda T, Sugiyama R, Hirai MY, Tamura K, Hara-Nishimura I and Shimada T (2019). Leaf Endoplasmic Reticulum Bodies Identified in Arabidopsis Rosette Leaves Are Involved in Defense against Herbivory. *Plant Physiol.* 179(4):1515–1524.

Nave F, Teixeira N, Mateus N and de Freitas V (2010). Hemisynthesis and structural characterization of flavanol-(4,8)-vitisins by mass spectrometry. *Rapid Commun. Mass Spectrom.* 24(14):1964–1970.

Negre F, Kish CM, Boatright J, Underwood B, Shibuya K, Wagner C, Clark DG and Dudareva N (2003). Regulation of methylbenzoate emission after pollination in snapdragon and petunia flowers. *Plant Cell* 15(12):2992–3006.

Neugart S, Rohn S and Schreiner M (2015). Identification of complex, naturally occurring flavonoid glycosides in Vicia faba and Pisum sativum leaves by HPLC-DAD-ESI-MS^n and the genotypic effect on their flavonoid profile. *Food Res. Int.* 76(1):114–121.

Nishizaki Y, Sasaki N, Yasunaga M, Miyahara T, Okamoto E, Okamoto M, Hirose Y and Ozeki Y (2014). Identification of the glucosyltransferase gene that supplies the *p*-hydroxybenzoyl-glucose for 7-polyacylation of anthocyanin in delphinium. *J. Exp. Bot.* 65(9):2495–2506.

Nishizaki Y, Yasunaga M, Okamoto E, Okamoto M, Hirose Y, Yamaguchi M, Ozeki Y and Sasaki N (2013). p-Hydroxybenzoyl-glucose is a zwitter donor for the biosynthesis of 7-polyacylated anthocyanin in Delphinium. *Plant Cell* 25(10):4150–4165.

Nishizawa M, Hara T, Miura T, Fujita S, Yoshigai E, Ue H, Hayashi Y, Kwon AH, Okumura T and Isaka T (2011). Supplementation with a flavanol-rich lychee fruit extract influences the inflammatory status of young athletes. *Phytother Res* 25(10):1486–1493.

Noguchi A, Horikawa M, Fukui Y, Fukuchi-Mizutani M, Iuchi-Okada A, Ishiguro M, Kiso Y, Nakayama T and Ono E (2009). Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in Lamiales. *Plant Cell* 21(5):1556–1572.

Nonogaki H, Seed Germination and Reserve Mobilization, 1–19 (John Wiley & Sons, Ltd, Chichester, 2008).

Nygaard V, R?dland EA and Hovig E (2016). Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics* 17(1):29–39.

Oba K, Conn EE, Canut H and Boudet AM (1981). Subcellular Localization of 2-(beta-d-Glucosyloxy)-Cinnamic Acids and the Related beta-glucosidase in Leaves of Melilotus alba Desr. *Plant Physiol.* 68(6):1359–1363.

Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ and Davies GJ (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* 25(6):1396–1405.

Oliveira J, Azevedo J, Silva AM, Teixeira N, Cruz L, Mateus N and de Freitas V (2010). Pyranoanthocyanin dimers: a new family of turquoise blue anthocyanin-derived pigments found in Port wine. *J. Agric. Food Chem.* 58(8):5154–5159.

Oliveira J, de Freitas V, Silva AM and Mateus N (2007). Reaction between hydroxycinnamic acids and anthocyanin-pyruvic acid adducts yielding new portisins. *J. Agric. Food Chem.* 55(15):6349–6356.

Oliveira J, Fernandes V, Miranda C, Santos-Buelga C, Silva A, de Freitas V and Mateus N (2006). Color properties of four cyanidin-pyruvic acid adducts. *J. Agric. Food Chem.* 54(18):6894–6903.

Oliveira J, Mateus N and de Freitas V (2014). Previous and recent advances in pyranoanthocyanins equilibria in aqueous solution. *Dyes Pigm.* 100(1):190–200.

Oliveira J, Mateus N, Rodriguez-Borges JE, Cabrita EJ, Silva AMS and De Freitas V (2011). Synthesis of a new pyranoanthocyanin dimer linked through a methyl-methine bridge. *Tetrahedron Lett.* 52(23):2957–2960.

Olsen KM, Slimestad R, Lea US, Brede C, L?vdal T, Ruoff P, Verheul M and Lillo C (2009). Temperature and nitrogen effects on regulators and products of the flavonoid pathway: experimental and kinetic model studies. *Plant Cell Environ.* 32(3):286–299.

Olsson L, Veit M, Weissenböck G and Bornman J (1998). Differential flavonoid response to enhanced UV-B radiation in *Brassica napus*. *Phytochemistry* 49(4):1021–1028.

Onkokesung N, Reichelt M, van Doorn A, Schuurink RC, van Loon JJ and Dicke M (2014). Modulation of flavonoid metabolites in Arabidopsis thaliana through overexpression of the MYB75 transcription factor: role of kaempferol-3,7-dirhamnoside in resistance to the specialist insect herbivore Pieris brassicae. *J. Exp. Bot.* 65(8):2203–2217.

Opassiri R, Hua Y, Wara-Aswapati O, Akiyama T, Svasti J, Esen A and Ketudat Cairns JR (2004). Beta-glucosidase, exo-beta-glucanase and pyridoxine transglucosylase activities of rice BGlu1. *Biochem. J.* 379(Pt 1):125–131.

Opassiri R, Pomthong B, Onkoksoong T, Akiyama T, Esen A and Ketudat Cairns JR (2006). Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 beta-glucosidase. *BMC Plant Biol.* 6:33.

Osmani SA, Bak S, Imberty A, Olsen CE and M?ller BL (2008). Catalytic key amino acids and UDP-sugar donor specificity of a plant glucuronosyltransferase, UGT94B1: molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. *Plant Physiol.* 148(3):1295–1308.

Ottaviani JI, Heiss C, Spencer JPE, Kelm M and Schroeter H (2018). Recommending flavanols and procyanidins for cardiovascular health: Revisited. *Mol. Aspects Med.* 61:63–75.

Palmer RG and Rodriguez de Cianzio S (1985). Conditional lethality involving nuclear and cytoplasmic chlorophyll mutants in soybeans. *Theor. Appl. Genet.* 70(4):349–354.

Panche AN, Diwan AD and Chandra SR (2016). Flavonoids: an overview. *J Nutr Sci* 5:e47.

Peng M, Shahzad R, Gul A, Subthain H, Shen S, Lei L, Zheng Z, Zhou J, Lu D, Wang S, Nishawy E, Liu X, Tohge T, Fernie AR and Luo J (2017). Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat Commun* 8(1):1975.

Perugino G, Trincone A, Rossi M and Moracci M (2004). Oligosaccharide synthesis by glycosynthases. *Trends Biotechnol.* 22(1):31–37.

Pichersky E, Noel JP and Dudareva N (2006). Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science* 311(5762):808–811.

Pinasseau L, Vallverd?-Queralt A, Verbaere A, Roques M, Meudec E, Le Cunff L, P?ros JP, Ageorges A, Sommerer N, Boulet JC, Terrier N and Cheynier V (2017). Cultivar Diversity of Grape Skin Polyphenol Composition and Changes in Response to Drought Investigated by LC-MS Based Metabolomics. *Front Plant Sci* 8:1826.

Pojer E, Mattivi F, Johnson D and Stockley C (2013). The case for anthocyanin consumption to promote human health: a review. *Compr. Rev. Food Sci. Food Saf.* 212(5):483–508.

Pollastri S and Tattini M (2011). Flavonols: old compounds for old roles. *Ann. Bot.* 108(7):1225–1233.

Poustka F, Irani NG, Feller A, Lu Y, Pourcel L, Frame K and Grotewold E (2007). A trafficking pathway for anthocyanins overlaps with the endoplasmic reticulum-to-vacuole protein-sorting route in Arabidopsis and contributes to the formation of vacuolar inclusions. *Plant Physiol.* 145(4):1323–1335.

Pucci G, Alcidi R, Tap L, Battista F, Mattace-Raso F and Schillaci G (2017). Sex- and gender-related prevalence, cardiovascular risk and therapeutic approach in metabolic syndrome: A review of the literature. *Pharmacol. Res.* 120:34–42.

Pucker B, Holtgräwe D, Stadermann KB, Frey K, Huettel B, Reinhardt R and Weisshaar B (2019). A chromosome-level sequence assembly reveals the structure of the Arabidopsis thaliana Nd-1 genome and its gene set. *PLoS ONE* 14(5):e0216233.

Quispe C, Viveros-Valdez E, Yarlewue J, Arones M, Paniagua J and Schmeda-Hirschmann G (2013). High speed centrifugal countercurrent chromatography (HSCCC) isolation and identification by LC-MSn analysis of the polar phenolics from *Vasconcellea quercifolia*. *J. Chil. Chem. Soc.* 58(3):1830–1835.

Raguso R (2008). Wake up and smell the roses: the ecology and evolution of floral scent. *Annu. Rev. Ecol. Evol. Syst.* 39:549–569.

Ramakrishna A and Ravishankar GA (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal Behav* 6(11):1720–1731.

Rasheed S, Bashir K, Kim JM, Ando M, Tanaka M and Seki M (2018). The modulation of acetic acid pathway genes in Arabidopsis improves survival under drought stress. *Sci Rep* 8(1):7831.

Raskin I, Ribnicky DM, Komarnytsky S, Ilic N, Poulev A, Borisjuk N, Brinker A, Moreno DA, Ripoll C, Yakoby N, O'Neal JM, Cornwell T, Pastor I and Fridlender B (2002). Plants and human health in the twenty-first century. *Trends Biotechnol.* 20(12):522–531.

Raychaudhuri A and Tipton PA (2002). Cloning and expression of the gene for soybean hydroxyisourate hydrolase. Localization and implications for function and mechanism. *Plant Physiol.* 130(4):2061–2068.

Remy S, Fulcrand H, Labarbe B, Cheynier V and Moutounet M (2000). First confirmation in red wine of products resulting from direct anthocyanin-tannin reactions. *J. Sci. Food Agric.* 80(6):745–751.

Rentzsch M, Quast P, Hillebrand S, Mehnert J and Winterhalter P (2007a). Isolation and identification of 5-carboxy-pyranoanthocyanins in beverages from cherry (*Prunus cerasus* L.). *Innov. Food Sci. Emerg. Technol.* 8(3):333–338.

Rentzsch M, Schwarz M and Winterhalter P (2007b). Pyranoanthocyanins – an overview on structures, occurrence, and pathways of formation. *Trends Food Sci. Technol.* 18:526–534.

Reznik H, Vergleichende Biochemie der Phenylpropane, 14–46 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1960).

Ringli C, Bigler L, Kuhn BM, Leiber RM, Diet A, Santelia D, Frey B, Pollmann S and Klein M (2008). The modified flavonol glycosylation profile in the Arabidopsis rol1 mutants results in alterations in plant growth and cell shape formation. *Plant Cell* 20(6):1470–1481.

Rivas-Gonzalo J, Bravo-Haro S and Santos-Buelga C (1995). Detection of compounds formed through the reaction of malvidin 3-monoglucoside and catechin in the presence of acetaldehyde. *J. Agric. Food Chem.* 43(6):1444–1449.

Robbins RJ, Leonczak J, Johnson JC, Li J, Kwik-Uribe C, Prior RL and Gu L (2009). Method performance and multi-laboratory assessment of a normal phase high pressure liquid chromatography-fluorescence detection method for the quantitation of flavanols and procyanidins in cocoa and chocolate containing samples. *J Chromatogr A* 1216(24):4831–4840.

Rodrìguez-Ramiro I, Ramos S, López-Oliva E, Agis-Torres A, Bravo L, Goya L and Martìn MA (2013). Cocoa polyphenols prevent inflammation in the colon of azoxymethane-treated rats and in TNF-α-stimulated Caco-2 cells. *Br. J. Nutr.* 110(2):206–215.

Roepke J (2015). Biochemical characterization of β-glucosidase-mediated catabolism of flavonol bisglycosides in *Arabidopsis thaliana*. Ph.D. thesis, The University of Guelph, Canada.

Roepke J and Bozzo GG (2015). *Arabidopsis thaliana* β-glucosidase BGLU15 attacks flavonol 3-O-β-glucoside-7-O-α-rhamnosides. *Phytochemistry* 109:14–24.

Roepke J, Gordon HOW, Neil KJA, Gidda S, Mullen RT, Freixas Coutin JA, Bray-Stone D and Bozzo GG (2017). An Apoplastic β-Glucosidase is Essential for the Degradation of Flavonol 3-O-β-Glucoside-7-O-α-Rhamnosides in Arabidopsis. *Plant Cell Physiol.* 58(6):1030–1047.

Romero C and Bakker J (1999). Interactions between grape anthocyanins and pyruvic acid, with effect of pH and acid concentration on anthocyanin composition and color in model solutions. *J. Agric. Food Chem.* 47(8):3130–3139.

Ross J, Li Y, Lim E and Bowles DJ (2001). Higher plant glycosyltransferases. *Genome Biol.* 2(2):REVIEWS3004.

Routaboul JM, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, Einhorn J and Lepiniec L (2006). Flavonoid diversity and biosynthesis in seed of Arabidopsis thaliana. *Planta* 224(1):96–107.

Ruttkies C, Schymanski EL, Wolf S, Hollender J and Neumann S (2016). MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J Cheminform* 8:3.

Saibaba S, Sathish Kumar M and Shanmuga Pandiyan P (2016). Mini review on LC/MS techniques. *WJPPS* 5(4):2381–2395.

Saigo T, Wang T, Watanabe M and Tohge T (2020). Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. *Curr. Opin. Plant Biol.* 55:93–99.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA and Arnheim N (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230(4732):1350–1354.

Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T and Fernie AR (2013). The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiol. Biochem.* 72:21–34.

Sakamoto A, Ueda M and Morikawa H (2002). Arabidopsis glutathione-dependent formaldehyde dehydrogenase is an S-nitrosoglutathione reductase. *FEBS Lett.* 515(1-3):20–24.

Salas E, Atanasova V, Poncet-Legrand F, Meudec E, Mazauric JP and Cheynier V (2004). Demonstration of the occurrence of flavanol-anthocyanin adducts in wine and in model solutions. *Anal. Chim. Acta* 513:325–332.

Sánchez-Ilárduya MB, Sánchez-Fernández C, Viloria-Bernal M, López-Márquez DM, Berrueta LA, Gallo B and Vicente F (2012). Mass spectrometry fragmentation pattern of coloured flavanol-anthocyanin and anthocyanin-flavanol derivatives in aged red wines of Rioja. *Aust. J. Grape Wine Res.* 18:203–214.

Santolini J, André F, Jeandroz S and Wendehenne D (2017). Nitric oxide synthase in plants: Where do we stand? *Nitric Oxide* 63:30–38.

Santos-Buelga C, Bravo-Haro S and Rivas-Gonzalo JC (1995). Interactions between catechin and malvidin-3-monoglucoside in model solutions. *Z Lebensm Unters Forsch* 201:269–274.

Sanz-Aparicio J, Hermoso JA, Martínez-Ripoll M, Lequerica JL and Polaina J (1998). Crystal structure of beta-glucosidase A from Bacillus polymyxa: insights into the catalytic activity in family 1 glycosyl hydrolases. *J. Mol. Biol.* 275(3):491–502.

Sarath G, Bethke PC, Jones R, Baird LM, Hou G and Mitchell RB (2006). Nitric oxide accelerates seed germination in warm-season grasses. *Planta* 223(6):1154–1164.

Sasaki N, Nishizaki Y, Ozeki Y and Miyahara T (2014). The role of acyl-glucose in anthocyanin modifications. *Molecules* 19(11):18747–18766.

Sato M, Maulik G, Ray PS, Bagchi D and Das DK (1999). Cardioprotective effects of grape seed proanthocyanidin against ischemic reperfusion injury. *J. Mol. Cell. Cardiol.* 31(6):1289–1297.

Schiestl FP (2010). The evolution of floral scent and insect chemical communication. *Ecol. Lett.* 13(5):643–656.

Schulze K, Schreiber L and Szankowski I (2005). Inhibiting effects of resveratrol and its glucoside piceid against Venturia inaequalis, the causal agent of apple scab. *J. Agric. Food Chem.* 53(2):356–362.

Schwarz M, Jerz G and Winterhalter P (2003a). Isolation and structure of Pinotin A, a new anthocyanin derivative from Pinotage wine. *Vitis* 42(2):105–106.

Schwarz M, Quast P, von Baer D and Winterhalter P (2003b). Vitisin A content in Chilean wines from Vitis vinifera Cv. Cabernet Sauvignon and contribution to the color of aged red wines. *J. Agric. Food Chem.* 51(21):6261–6267.

Schwarz M, Wabnitz TC and Winterhalter P (2003c). Pathway leading to the formation of anthocyanin-vinylphenol adducts and related pigments in red wines. *J. Agric. Food Chem.* 51(12):3682–3687.

Schwarz M, Wray V and Winterhalter P (2004). Isolation and identification of novel pyranoanthocyanins from black carrot (*Daucus carota* L.) juice. *J. Agric. Food Chem.* 52(16):5095–5101.

Sen S (2010). S-Nitrosylation process acts as a regulatory switch for seed germination in wheat. *Am. J. Plant Physiol.* 5:122–132.

Sentandreu E, Navarro JL and Sendra JM (2010). LC-DAD-ESI/MS(n) determination of direct condensation flavanol-anthocyanin adducts in pressure extracted pomegranate (Punica granatum L.) juice. *J. Agric. Food Chem.* 58(19):10560–10567.

Sentandreu E, Navarro JL and Sendra JM (2012). Identification of New Coloured Anthocyanin–Flavanol Adducts in Pressure-Extracted Pomegranate (Punica granatum L.) Juice by High-Performance Liquid Chromatography/Electrospray Ionization Mass Spectrometry. *Food Anal. Methods* 5:702–709.

Shaikh FA and Withers SG (2008). Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis. *Biochem. Cell Biol.* 86(2):169–177.

Sharma V and Strack D (1985). Vacuolar localization of 1-sinapolglucose: L-malate sinapoyltransferase in protoplasts from cotyledons of *Raphanus sativus*. *Planta* 163(4):563–568.

Shi MZ and Xie DY (2014). Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Pat Biotechnol* 8(1):47–60.

Siadjeu C, Pucker B, Vieh?ver P, Albach DC and Weisshaar B (2020). High Contiguity De Novo Genome Sequence Assembly of Trifoliate Yam (Dioscorea dumetorum) Using Long Read Sequencing. *Genes (Basel)* 11(3).

Sies H, Schewe T, Heiss C and Kelm M (2005). Cocoa polyphenols and inflammatory mediators. *Am. J. Clin. Nutr.* 81(1 Suppl):304S–312S.

Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L and Xenarios I (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41(Database issue):D344–347.

Snyder M and Gerstein M (2003). Genomics. Defining genes in the genomics era. *Science* 300(5617):258–260.

Solovchenko A and Schmitz-Eiberger M (2003). Significance of skin flavonoids for UV-B-protection in apple fruits. *J. Exp. Bot.* 54(389):1977–1984.

Somers TC (1966). Wine tannins–isolation of condensed flavonoid pigments by gel-filtration. *Nature* 209(5021):368–370.

Somers TC (1971). The polymeric nature of wine pigments. *Phytochemistry* 10:2175–2186.

Soto M, Falqué E and Domínguez H (2015). Relevance of natural phenolics from grape and derivative products in the formulation of cosmetics. *Cosmetics.* 2:259–276.

Sowa AW, Duff SM, Guy PA and Hill RD (1998). Altering hemoglobin levels changes energy status in maize cells under hypoxia. *Proc. Natl. Acad. Sci. U.S.A.* 95(17):10317–10321.

Spagnuolo C, Moccia S and Russo GL (2018). Anti-inflammatory effects of flavonoids in neurodegenerative disorders. *Eur J Med Chem* 153:105–115.

Stobiecki M, Skirycz A, Kerhoas L, Kachlicki P, Muth D, Einhorn J and Mueller-Roeber B (2006). Profiling of phenolic glycosidic conjugates in leaves of *Arabidopsis thaliana* using LC/MS. *Metabolomics* 2:197–219.

Stracke R, Jahns O, Keck M, Tohge T, Niehaus K, Fernie AR and Weisshaar B (2010). Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in Arabidopsis thaliana plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol.* 188(4):985–1000.

Sugiyama R and Hirai MY (2019). Atypical Myrosinase as a Mediator of Glucosinolate Functions in Plants. *Front Plant Sci* 10:1008.

Sun J, Lin LZ and Chen P (2012). Study of the mass spectrometric behaviors of anthocyanins in negative ionization mode and its applications for characterization of anthocyanins and non-anthocyanin polyphenols. *Rapid Commun. Mass Spectrom.* 26(9):1123–1133.

Sun YG, Wang B, Jin SH, Qu XX, Li YJ and Hou BK (2013). Ectopic expression of Arabidopsis glycosyltransferase UGT85A5 enhances salt stress tolerance in tobacco. *PLoS ONE* 8(3):e59924.

Suzuki Y, Kawazu T and Koyama H (2004). RNA isolation from siliques, dry seeds, and other tissues of Arabidopsis thaliana. *BioTechniques* 37(4):542, 544.

Swain E, Li CP and Poulton JE (1992). Tissue and Subcellular Localization of Enzymes Catabolizing (R)-Amygdalin in Mature *Prunus serotina* Seeds. *Plant Physiol.* 100(1):291–300.

Taiz L, Zeiger E, Møller L and Murphy A, Appendix 4 Secondary Metabolites, 1–18 (Oxford University Press, 2014). 6th edition.

Takahama U, Yamauchi R and Hirota S (2013a). Interactions of starch with a cyanidin-catechin pigment (vignacyanidin) isolated from Vigna angularis bean. *Food Chem* 141(3):2600–2605.

Takahama U, Yamauchi R and Hirota S (2013b). Isolation and characterization of a cyanidin-catechin pigment from adzuki bean (*Vigna angularis*). *Food Chem* 141(1):282–288.

Tamura K, Shimada T, Ono E, Tanaka Y, Nagatani A, Higashi SI, Watanabe M, Nishimura M and Hara-Nishimura I (2003). Why green fluorescent fusion proteins have not been observed in the vacuoles of higher plants. *Plant J.* 35(4):545–555.

Tanaka Y, Sasaki N and Ohmiya A (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54(4):733–749.

Tava A, Pecio L, Lo Scalzo R, Stochmal A and Pecetti L (2019). Phenolic Content and Antioxidant Activity in Trifolium Germplasm from Different Environments. *Molecules* 24(2).

Taylor J (University Science Books, Sausalito, California, 1997). 2nd edition.

Thermo Fisher (2016). Thermo Fisher Scientific real-time PCR handbook. Website. Online available at https://www.thermofisher. com/content/dam/LifeTech/Documents/PDFs/PG1503-PJ9169-CO019861-Update-qPCR-Handbook-branding-Americas-FLR.pdf; obtained on 28th May 2020.

Tholl D, Chen F, Petri J, Gershenzon J and Pichersky E (2005). Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. *Plant J.* 42(5):757–771.

Tian Q, Giusti MM, Stoner GD and Schwartz SJ (2005). Screening for anthocyanins using high-performance liquid chromatography coupled to electrospray ionization tandem mass spectrometry with precursor-ion analysis, product-ion analysis, common-neutral-loss analysis, and selected reaction monitoring. *J Chromatogr A* 1091(1-2):72–82.

Tikunov YM, Molthoff J, de Vos RC, Beekwilder J, van Houwelingen A, van der Hooft JJ, Nijenhuis-de Vries M, Labrie CW, Verkerke W, van de Geest H, Viquez Zamora M, Presa S, Rambla JL, Granell A, Hall RD and Bovy AG (2013). Non-smoky glycosyltransferase1 prevents the release of smoky aroma from tomato fruit. *Plant Cell* 25(8):3067–3078.

Timberlake CF and Bridle P (1976). Interactions between anthocyanins, phenolic compounds and acetaldehyde and their significance in red wines. *Am J Enol Vitic.* 27:97–105.

Tohge T and Fernie AR (2010). Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat Protoc* 5(6):1210–1227.

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M and Saito K (2005). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J.* 42(2):218–235.

Tohge T, Wendenburg R, Ishihara H, Nakabayashi R, Watanabe M, Sulpice R, Hoefgen R, Takayama H, Saito K, Stitt M and Fernie AR (2016). Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat Commun* 7:12399.

Trapero A, Ahrazem O, Rubio-Moraga A, Jimeno ML, Gómez MD and Gómez-Gómez L (2012). Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in *Crocus sativus. Plant Physiol.* 159(4):1335–1354.

Tsai AY, Kunieda T, Rogalski J, Foster LJ, Ellis BE and Haughn GW (2017). Identification and Characterization of Arabidopsis Seed Coat Mucilage Proteins. *Plant Physiol.* 173(2):1059–1074.

Turan Y (2008). A pseudo-beta-glucosidase in Arabidopsis thaliana: correction by site-directed mutagenesis, heterologous expression, purification, and characterization. *Biochemistry Mosc.* 73(8):912–919.

Šírová J, Sedlářová M, Piterková J, Luhová L and Petřivalský M (2011). The role of nitric oxide in the germination of plant seeds and pollen. *Plant Sci.* 181(5):560–572.

Valenta K, Burke RJ, Styler SA, Jackson DA, Melin AD and Lehman SM (2013). Colour and odour drive fruit selection and seed dispersal by mouse lemurs. *Sci Rep* 3:2424.

van der Hooft JJ, Vervoort J, Bino RJ, Beekwilder J and de Vos RC (2011). Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation. *Anal. Chem.* 83(1):409–416.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A and Speleman F (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3(7):1–12.

Veit M and Pauli GF (1999). Major flavonoids from *Arabidopsis thaliana* leaves. *J. Nat. Prod.* 62(9):1301–1303.

Venugopala KN, Rashmi V and Odhav B (2013). Review on natural coumarin lead compounds for their pharmacological activity. *Biomed Res Int* 2013:963248.

Verdan AM, Wang HC, Garc?a CR, Henry WP and Brumaghim JL (2011). Iron binding of 3-hydroxychromone, 5-hydroxychromone, and sulfonated morin: Implications for the antioxidant activity of flavonols with competing metal binding sites. *J. Inorg. Biochem.* 105(10):1314–1322.

Verdoucq L, Morini?re J, Bevan DR, Esen A, Vasella A, Henrissat B and Czjze M (2004). Structural determinants of substrate specificity in family 1 beta-glucosidases: novel insights from the crystal structure of sorghum dhurrinase-1, a plant beta-glucosidase with strict specificity, in complex with its natural substrate. *J. Biol. Chem.* 279(30):31796–31803.

Vivar-Quintana AM, Santos-Buelga C and Rivas-Gonzalo JC (2002). Anthocyanin-derived pigments and colour of red wines. *Anal. Chim. Acta* 458:147–155.

Vogt T and Jones P (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.* 5(9):380–386.

Vranová E, Coman D and Gruissem W (2012). Structure and dynamics of the isoprenoid pathway network. *Mol Plant* 5(2):318–333.

Žuvela P, Skoczylas M, Jay Liu J, Bączek T, Kaliszan R, Wong MW and Buszewski B (2019). Column Characterization and Selection Systems in Reversed-Phase High-Performance Liquid Chromatography. *Chem. Rev.* 119(6):3674–3729.

Wagner EM (2013). Monitoring gene expression: quantitative real-time rt-PCR. *Methods Mol. Biol.* 1027:19–45.

Wallace T and Giusti M (2015). Anthocyanins. *Adv. Nutr* 6:620–622.

Wang D, Xia M, Yan X, Li D, Wang L, Xu Y, Jin T and Ling W (2012). Gut microbiota metabolism of anthocyanin promotes reverse cholesterol transport in mice via repressing miRNA-10b. *Circ. Res.* 111(8):967–981.

Wang H, Cao G and Prior RL (1997). Oxygen Radical Absorbing Capacity of Anthocyanins. *J. Agric. Food Chem.* 45(2):304–309.

Wang J and Mazza G (2002). Effects of Anthocyanins and Other Phenolic Compounds on the Production of Tumor Necrosis Factor α in LPS/IFN-γ-Activated RAW 264.7 Macrophages. *J. Agric. Food Chem.* 50(15):4183–4189.

Wang LX and Huang W (2009). Enzymatic transglycosylation for glycoconjugate synthesis. *Curr Opin Chem Biol* 13(5-6):592–600.

Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Cr?semann M, Boudreau PD, Esquenazi E, Sandoval-Calder?n M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodr?guez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, M?ller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Guti?rrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC and Bandeira N (2016a). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34(8):828–837.

Wang P, Liu H, Hua H, Wang L and Song C (2011). A vacuole localized β-glucosidase contributes to drought tolerance in *Arabidopsis*. *Chinese Sci Bull* 56(33):3538–3546.

Wang R, Zhang L, Zhang Z and Tian Y (2016b). Comparison of ESI- and APCI-LC-MS/MS methods: A case study of levonorgestrel in human plasma. *J Pharm Anal* 6(6):356–362.

Wang X (2009). Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Lett.* 583(20):3303–3309.

Weidner S, Chrzanowski S, Karama? M, Kr?l A, Badowiec A, Mostek A and Amarowicz R (2014). Analysis of phenolic compounds and antioxidant abilities of extracts from germinating Vitis californica seeds submitted to cold stress conditions and recovery after the stress. *Int J Mol Sci* 15(9):16211–16225.

Wiswedel I, Hirsch D, Kropf S, Gruening M, Pfister E, Schewe T and Sies H (2004). Flavanol-rich cocoa drink lowers plasma F(2)-isoprostane concentrations in humans. *Free Radic. Biol. Med.* 37(3):411–421.

Wittstock U and Burow M (2010). Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. *Arabidopsis Book* 8:e0134.

Wu S, Tohge T, Cuadros-Inostroza A, Tong H, Tenenboim H, Kooke R, Meret M, Keurentjes JB, Nikoloski Z, Fernie AR, Willmitzer L and Brotman Y (2018). Mapping the Arabidopsis Metabolic Landscape by Untargeted Metabolomics at Different Environmental Conditions. *Mol Plant* 11(1):118–134.

Xing HL, Dong L, Wang ZP, Zhang HY, Han CY, Liu B, Wang XC and Chen QJ (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* 14:327.

Xu Z, Escamilla-Treviño L, Zeng L, Lalgondar M, Bevan D, Winkel B, Mohamed A, Cheng CL, Shih MC, Poulton J and Esen A (2004). Functional genomic analysis of Arabidopsis thaliana glycoside hydrolase family 1. *Plant Mol. Biol.* 55(3):343–367.

Xu ZY, Lee KH, Dong T, Jeong JC, Jin JB, Kanno Y, Kim DH, Kim SY, Seo M, Bressan RA, Yun DJ and Hwang I (2012). A vacuolar β-glucosidase homolog that possesses glucose-conjugated abscisic acid hydrolyzing activity plays an important role in osmotic stress responses in Arabidopsis. *Plant Cell* 24(5):2184–2199.

Yamada K, Hara-Nishimura I and Nishimura M (2011). Unique defense strategy by the endoplasmic reticulum body in plants. *Plant Cell Physiol.* 52(12):2039–2049.

Yang H, Wu JJ, Tang T, Liu KD and Dai C (2018). Author Correction: CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in *Brassica napus*. *Sci Rep* 8(1):4877.

Yin R, Han K, Heller W, Albert A, Dobrev PI, Zažìmalová E and Schäffner AR (2014). Kaempferol 3-O-rhamnoside-7-O-rhamnoside is an endogenous flavonol inhibitor of polar auxin transport in Arabidopsis shoots. *New Phytol.* 201(2):466–475.

Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada K, Matsuda F, Sugawara S, Inoue E, Kuromori T, Ito T, Shinozaki K, Wangwattana B, Yamazaki M and Saito K (2012). Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.* 69(1):154–167.

Yonekura-Sakakibara K and Hanada K (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* 66(1):182–193.

Yonekura-Sakakibara K, Nakabayashi R, Sugawara S, Tohge T, Ito T, Koyanagi M, Kitajima M, Takayama H and Saito K (2014). A flavonoid 3-O-glucoside:2"-O-glucosyltransferase responsible for terminal modification of pollen-specific flavonols in *Arabidopsis thaliana*. *Plant J.* 79(5):769–782.

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E and Saito K (2008). Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *Plant Cell* 20(8):2160–2176.

Yonekura-Sakakibara K, Tohge T, Niida R and Saito K (2007). Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. *J. Biol. Chem.* 282(20):14932–14941.

Yoshida K, Nagai N, Ichikawa Y, Goto M, Kazuma K, Oyama KI, Koga K, Hashimoto M, Iuchi S, Takaya Y and Kondo T (2019). Structure of two purple pigments, catechiopyranocyanidins A and B from the seed-coat of the small red bean, *Vigna angularis*. *Sci Rep* 9(1):1484.

Youdim KA, McDonald J, Kalt W and Joseph JA (2002). Potential role of dietary flavonoids in reducing microvascular endothelium vulnerability to oxidative and inflammatory insults (small star, filled). *J. Nutr. Biochem.* 13(5):282–288.

Zamioudis C, Hanson J and Pieterse CM (2014). β-Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in Arabidopsis roots. *New Phytol.* 204(2):368–379.

Zhang J, Subramanian S, Stacey G and Yu O (2009). Flavones and flavonols play distinct critical roles during nodulation of Medicago truncatula by Sinorhizobium meliloti. *Plant J.* 57(1):171–183.

Zhang L, Kawaguchi R, Morikawa-Ichinose T, Allahham A, Kim SJ and Maruyama-Nakashita A (2020). Sulfur Deficiency-Induced Glucosinolate Catabolism Attributed to Two β-Glucosidases, BGLU28 and BGLU30, is Required for Plant Growth Maintenance under Sulfur Deficiency. *Plant Cell Physiol.* 61(4):803–813.

Zhang L, Tai Y, Wang Y, Meng Q, Yang Y, Zhang S, Yang H, Zhang Z, Li D and Wan X (2017). The proposed biosynthesis of procyanidins by the comparative chemical analysis of five Camellia species using LC-MS. *Sci Rep* 7:46131.

Zhang Y, Butelli E and Martin C (2014). Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.* 19:81–90.

Zhao J, Huhman D, Shadle G, He XZ, Sumner LW, Tang Y and Dixon RA (2011). MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in Medicago truncatula. *Plant Cell* 23(4):1536–1555.

Zhao J, Pang Y and Dixon RA (2010a). The mysteries of proanthocyanidin transport and polymerization. *Plant Physiol.* 153(2):437–443.

Zhao Q, Duan CQ and Wang J (2010b). Anthocyanins profile of grape berries of Vitis amurensis, its hybrids and their wines. *Int J Mol Sci* 11(5):2212–2228.

Zhou C, Tokuhisa JG, Bevan DR and Esen A (2012). Properties of β-thioglucoside hydrolases (TGG1 and TGG2) from leaves of Arabidopsis thaliana. *Plant Sci.* 191-192:82–92.

Zhu QY, Holt RR, Lazarus SA, Orozco TJ and Keen CL (2002). Inhibitory effects of cocoa flavanols and procyanidin oligomers on free radical-induced erythrocyte hemolysis. *Exp. Biol. Med. (Maywood)* 227(5):321–329.

Zouhar J, Vévodová J, Marek J, Damborský J, Su XD and Brzobohatý B (2001). Insights into the functional architecture of the catalytic center of a maize beta-glucosidase Zm-p60.1. *Plant Physiol.* 127(3):973–985.

Supplement
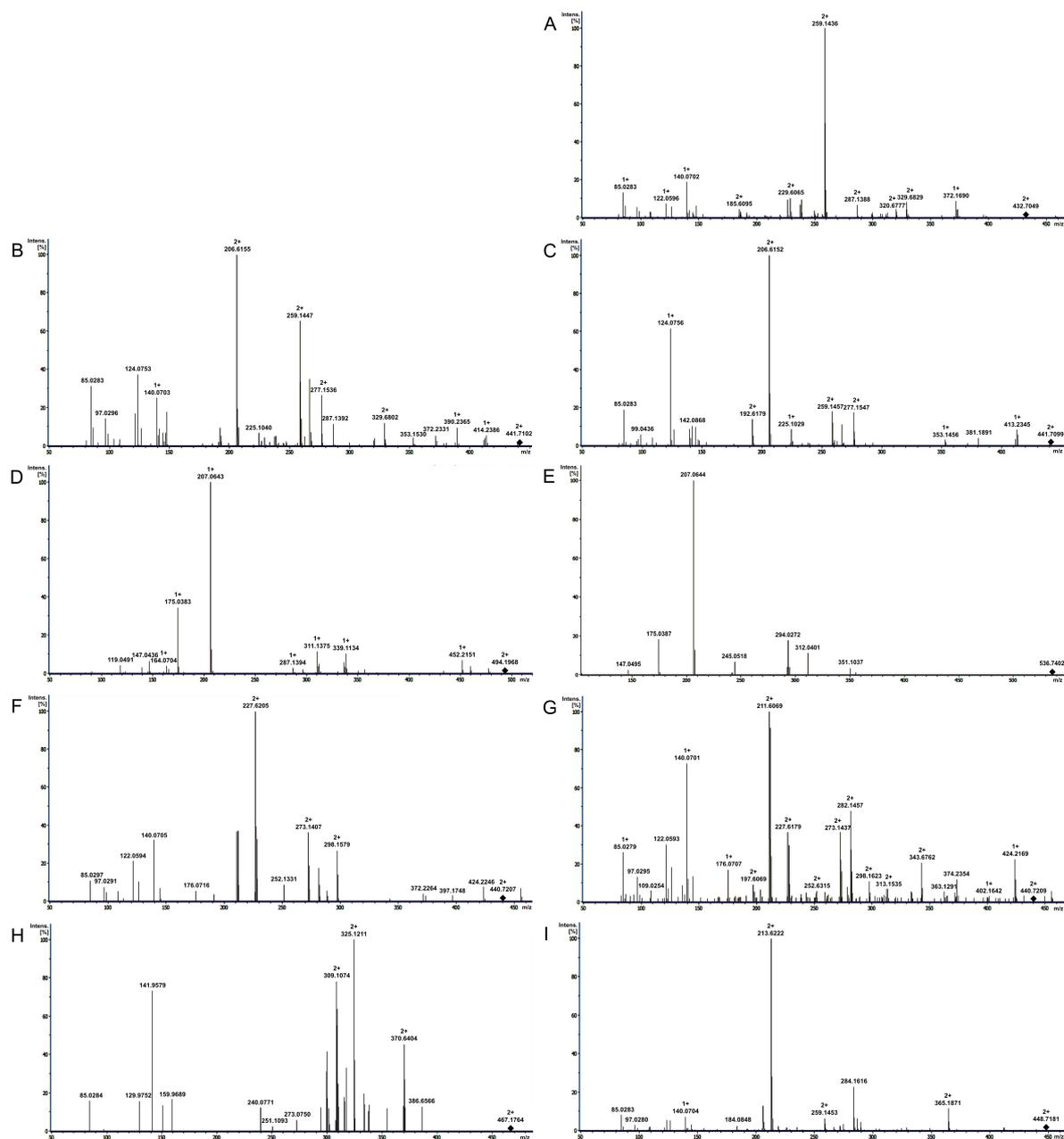

Figure S1 – *Double-charged* **BGLU3** *candidate product features.* *MS/MS spectra of all additional double-charged candidate* BGLU3 *product features are listed. The respective precursor ions are indicated by a rhombus.*
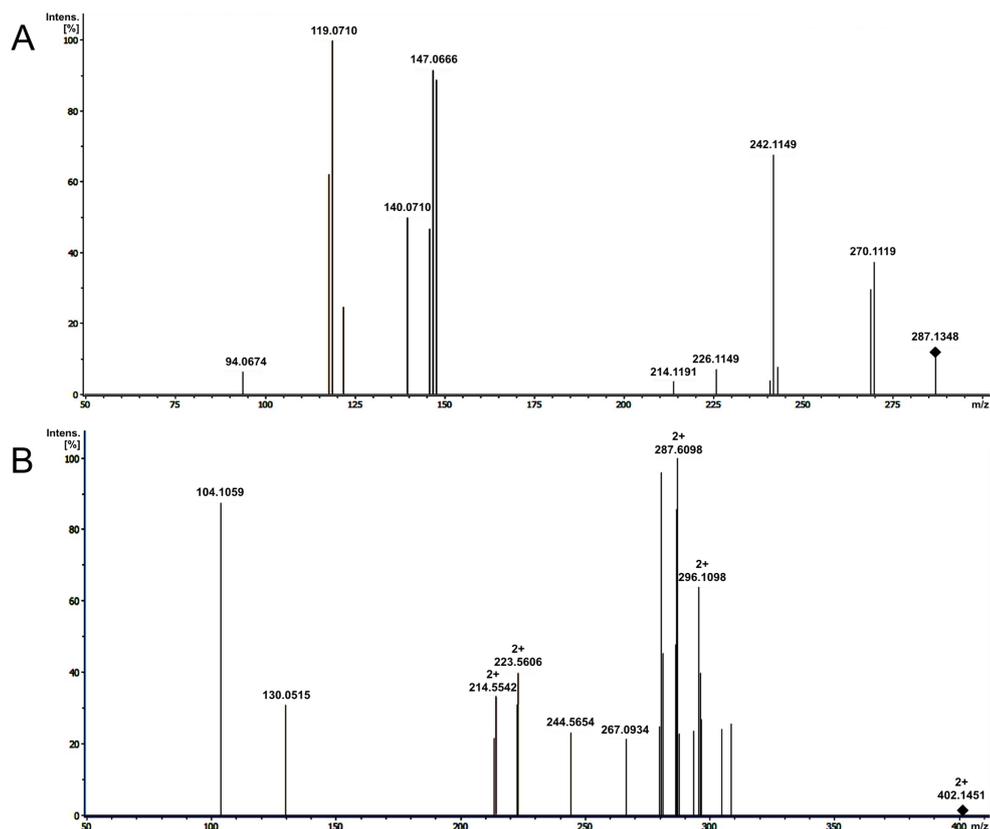
Figure S2 – *Further* **BGLU3** *candidate substrate features.* The MS/MS spectra of the two additional candidate BGLU3 substrate features 287.1382 m/z at 7.25 min (A) and 402.1451 m/z at 13.11 min (B) are shown. The respective precursor ions are indicated by a rhombus.
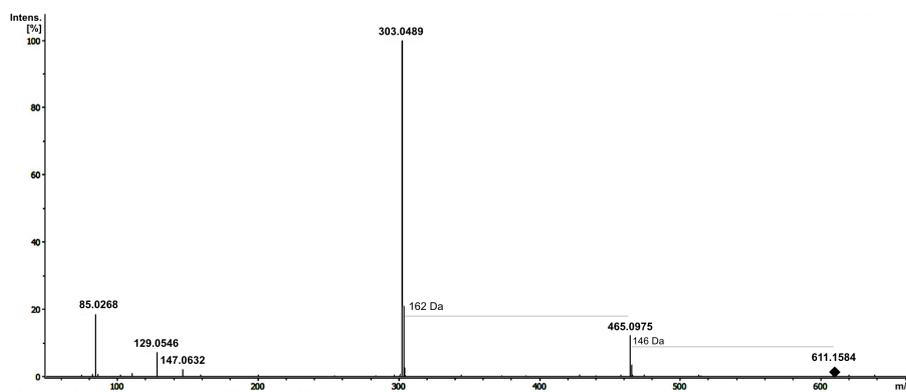
Figure S3 – *Candidate* **BGLU4** *feature 611.1584 m/z at 8.75 min.* The MS/MS spectrum of the BGLU4 candidate substrate feature 611.1584 m/z at 8.75 min shows the main fragment ions 465 m/z with a neutral loss of 146 Da and 303 m/z with a neutral loss of 162 Da. The precursor ion is indicated by a rhombus.
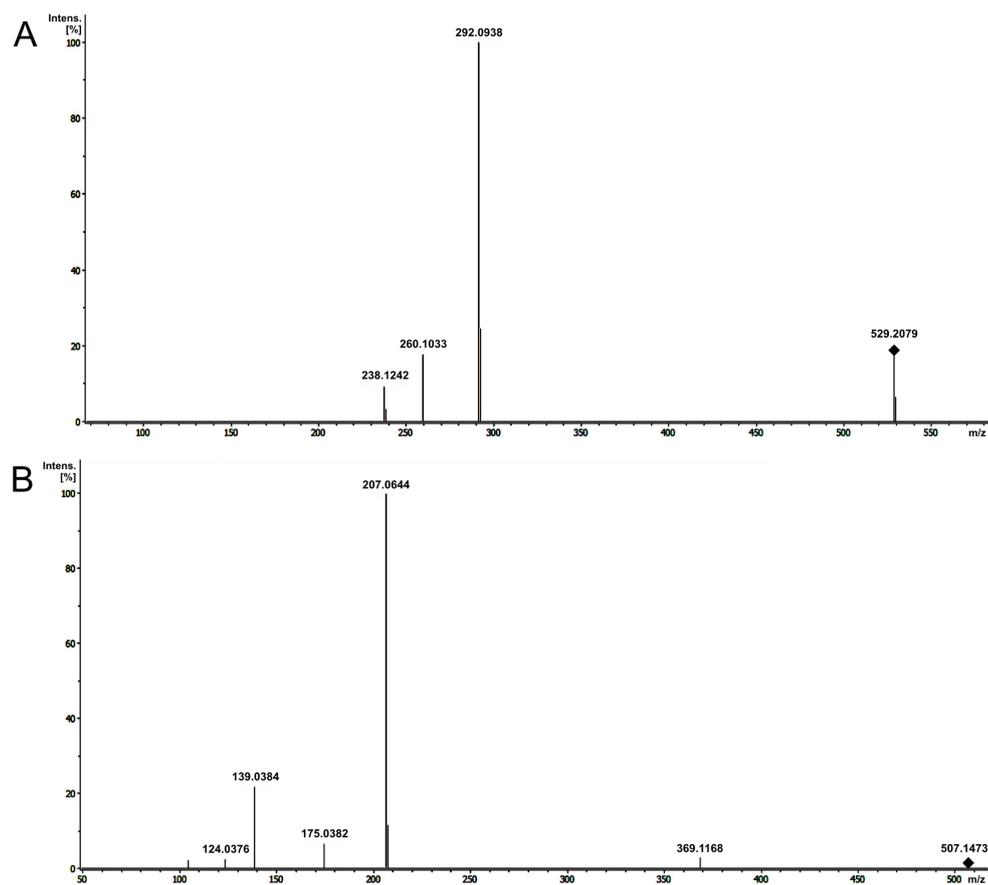
Figure S4 – ***Further* BGLU4 *candidate product features.*** *The MS/MS spectra of the two additional candidate BGLU4 product features 529.2079 m/z at 27.45 min (A) and 507.1473 m/z at 14.37 min (B) are shown.The respective precursor ions are indicated by a rhombus.*
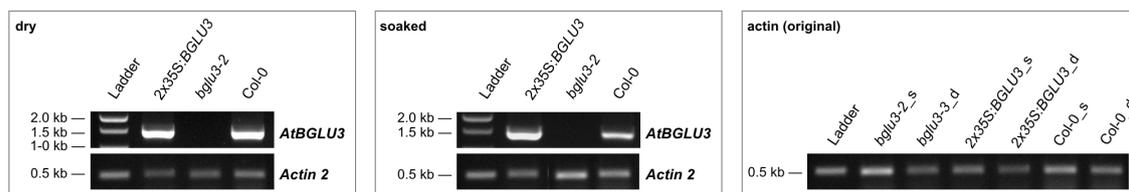
Figure S5 – ***Complete RT-PCR results of* BGLU3.** *RT-PCR results for the intact transcript of BGLU3 for dry seed samples in* 2x35S:BGLU3, bglu3-2 *and the Col-0 wild type are given left. The same RT-PCR results of BGLU3, but for soaked seed samples are given in the middle. The bands of the* Actin 2 *control are cut from the original figure for visualization. Therefore, the results for the* Actin 2 *control as originally present in the gel, are shown right.*
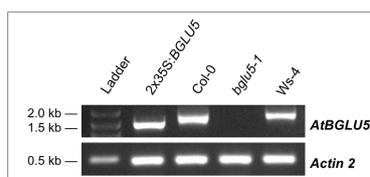
Figure S6 – ***RT-PCR results of* BGLU5.** *RT-PCR results for the intact transcript of BGLU5 in* 2x35S:BGLU5, bglu5-1 *and the Col-0 wild type as reference for* 2x35S:BGLU5 *as well as Ws-4 as reference for* bglu5-1 *are shown.*
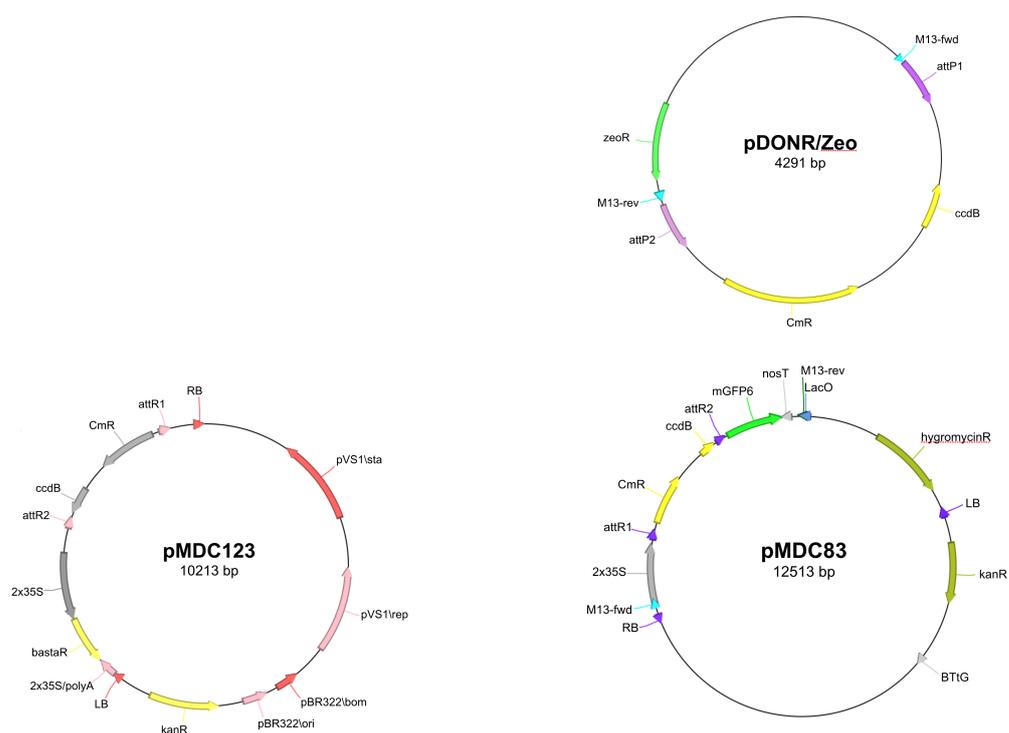
Figure S7 – **Vectors.** *The figures of all used vectors from this work are listed. Figures were constructed with ApE. 2x35S: CaMV35S promotor, LB: T-DNA left border, RB: T-DNA rigth border, Cm: chloramphenicol, kan: kanamycin, zeoR: zeocin resistance, bastaR: basta resistance.*
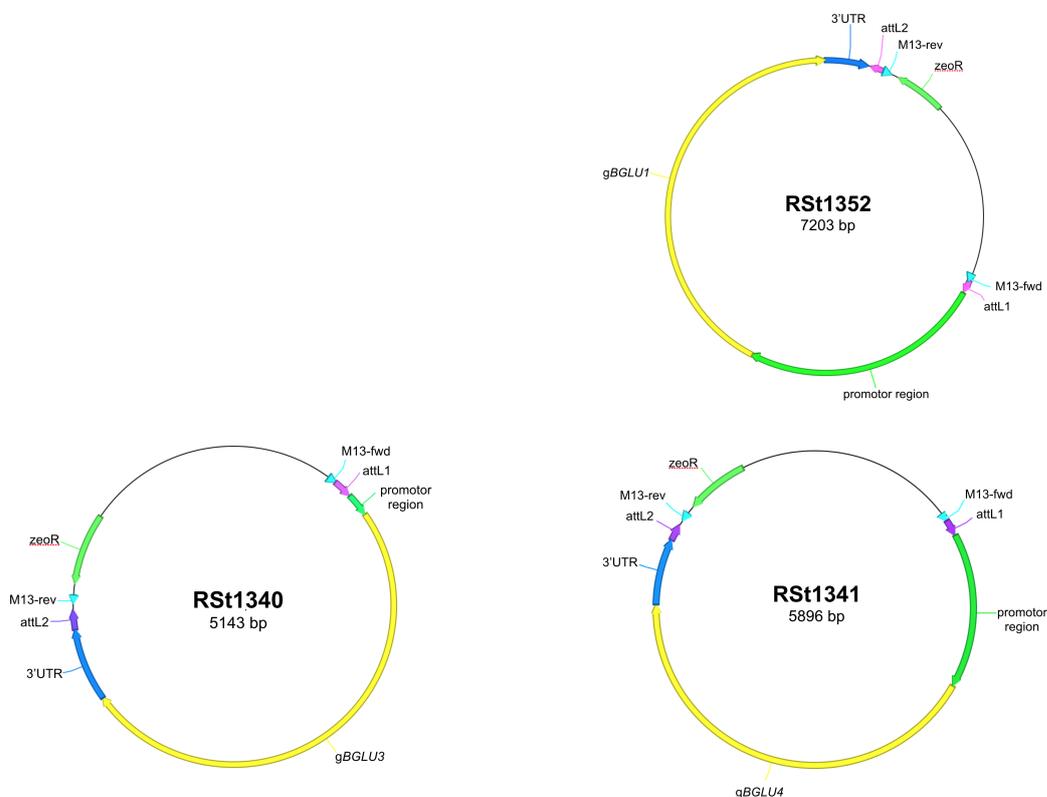
Figure S8 – **Genomic DNAs in pDONR/Zeo.** *The figures of all plasmids with genomic DNAs in pDONR/Zeo are listed. Figures were constructed with ApE. zeoR: zeocin.*
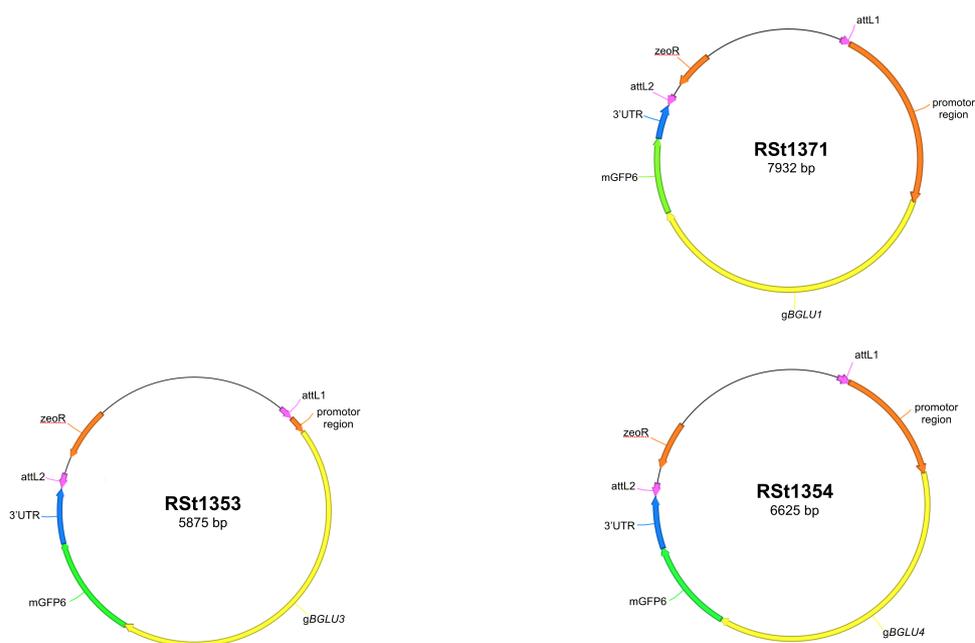
Figure S9 – ***Genomic DNAs with GFP in pDONR/Zeo.*** *The figures of all plasmids with genomic DNAs with a GFP-tag before the stop codon in pDONR/Zeo are listed. Figures were constructed with ApE. zeoR: zeocin.*
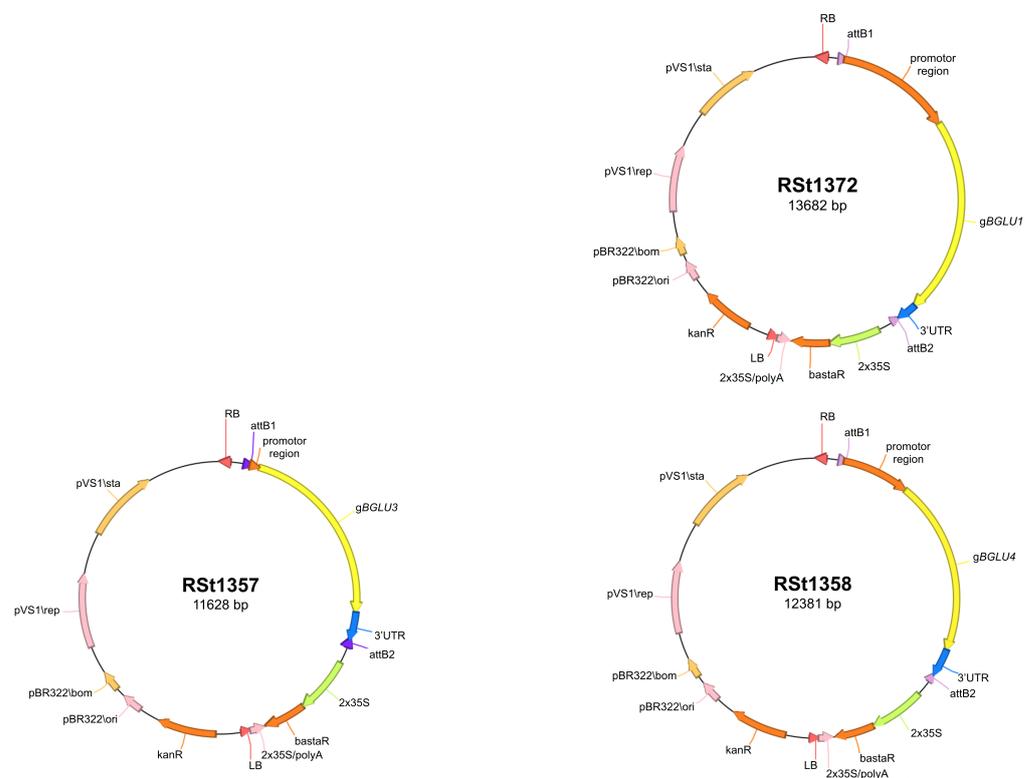
Figure S10 – ***Genomic DNAs in pMDC123.*** *The figures of all plasmids with genomic DNAs in pMDC123 are listed. Figures were constructed with ApE. 2x35S: CaMV35S promotor, LB: T-DNA left border, RB: T-DNA rigth border, kan: kanamycin, bastaR: basta resistance.*
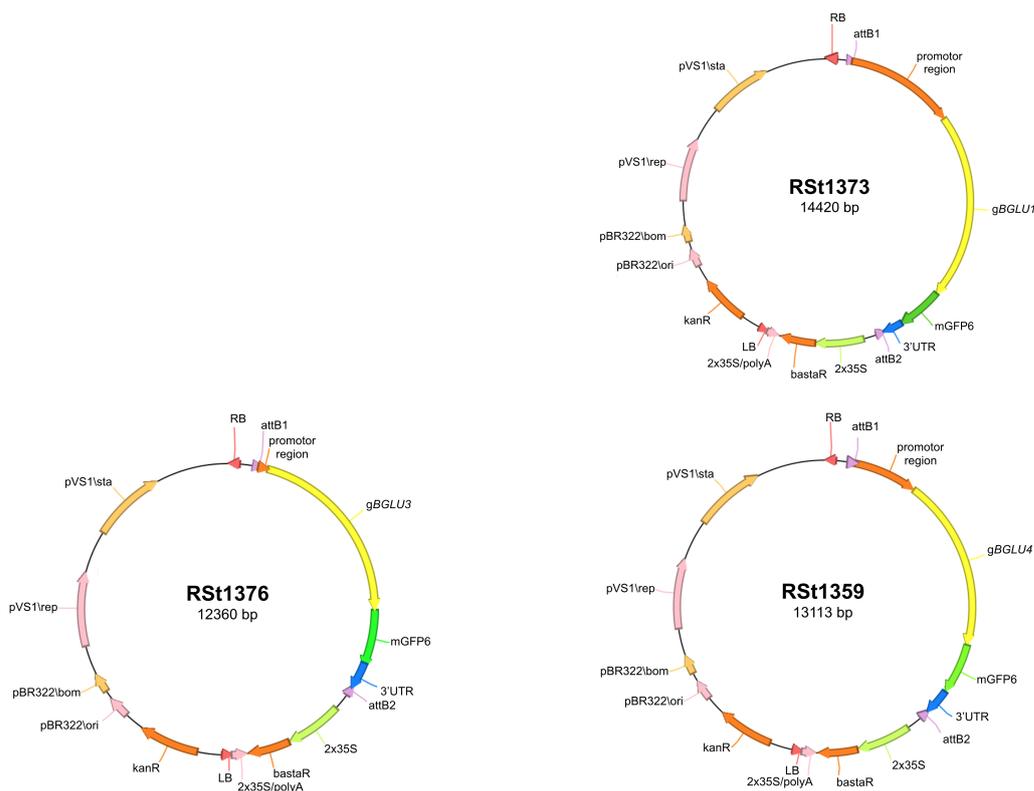
Figure S11 – *Genomic DNAs with GFP in pMDC123.* *The figures of all plasmids with genomic DNAs with a GFP-tag before the stop codon in pMDC123 are listed. Figures were constructed with ApE. 2x35S: CaMV35S promotor, LB: T-DNA left border, RB: T-DNA rigth border, kan: kanamycin, bastaR: basta resistance.*

Table S1 – *Primers.* *All used primers are listed, giving their name, strand orientation, sequence and their used purpose in this work.*

Name	Strand	Sequence (5' $\rightarrow$ 3')	Purpose
3144	fwd & rev	GTGGATTGATGTGATATCTCC	T-DNA insertion, sequencing *bglu3-2* and *bglu4-2* genotyping *bglu4-2* RB of pAC106 and pAC161
unil	fwd	CGTTGTAAAACGACGGCCAGT	sequencing GATEWAY[TM]
revl	rev	CAGGAAACAGCTATGACCATG	sequencing GATEWAY[TM]
B097	rev	AATACGCAAACCGCCTCTC	sequencing *bglu1-1*
B230	rev	GCCTTTTCAGAAATGGATAAA TAGCCTTGCTTCC	sequencing *bglu3-2*

Table S1 – continued

Name	Strand	Sequence (5' → 3')	Purpose
JFF001	rev	CATCATTCTGTGGTTGAGCCATCC	T-DNA insertion, sequencing *bglu4-2*, sequencing *BGLU4*
JFF002	fwd	CTCCACGACAAGTCTACTTACTTCC	Sequencing *BGLU1*
JFF003	rev	TGAGAAATTCGGACAGTCCAGTCG	Sequencing *BGLU1*
JFF004	fwd	TGGCCGAAACTGGCTTACATAC	Sequencing *BGLU1*
JFF005	rev	CTGAGGATGATCGTAGTGGAACAGTGCAGTG	Sequencing *BGLU3*
JFF006	rev	TGCTATATCTCCGTTACCTTGGTC	Sequencing *BGLU4*
JFF010	rev	ACGAAGAAGCTGAAGAAGAAAAGTTGCTCTGC	RT-PCR *BGLU3*
JFF022	fwd	AATTACTAAACGGATACAAGAG	Sequencing *BGLU1*
JFF024	fwd	ACACGAGAGGTTACTTCA	Sequencing *BGLU3*
JFF025	rev	AACCAAGAAAAGTGGTG	Sequencing *BGLU3*
JFF026	fwd	GGTTTTGGGTTGTACA	Sequencing *BGLU4*
JFF030	fwd	TGCAGTTCAAAGAGCCAAAG	Sequencing *BGLU3*
JFF031	rev	CATTTCATCGGGATAGTCACC	Sequencing *BGLU3*
JFF032	fwd	GCCGGGATATCTGCTTATCA	qRT-PCR *BGLU1*
JFF033	rev	CAAGCTATGTCTCCATTATCC	qRT-PCR *BGLU1*

Table S1 – continued

Name	Strand	Sequence (5' → 3')	Purpose
		ATTT	
JFF035	fwd	<u>GGGGACAAGTTTGTACAAAAA</u> <u>AGCAGGCTT</u>CACACAACCATA CACAACTGC	attB1-*BGLU1*
JFF036	rev	<u>GGGGACCACTTTGTACAAGAA</u> <u>AGCTGGGTC</u>AGTTGGTCATCG CCTTTCG	attB2-*BGLU1*, Sequencing *BGLU1*
JFF037	fwd	<u>GGGGACAAGTTTGTACAAAAA</u> <u>AGCAGGCTT</u>CTGTGTGTTTCT TTTGAGTCCAC	attB1-*BGLU3*
JFF038	rev	<u>GGGGACCACTTTGTACAAGAA</u> <u>AGCTGGGTC</u>CTTTGGCGATTC ATAGTTATGG	attB2-*BGLU3*
JFF039	fwd	<u>GGGGACAAGTTTGTACAAAAA</u> <u>AGCAGGCTT</u>CCCAAACAAAGC CCAATAATGTTGC	attB1-*BGLU4*
JFF041	rev	<u>GGGGACCACTTTGTACAAGAA</u> <u>AGCTGGGTC</u>TCTTGGACTAGG ATTCTCAC	attB2-*BGLU4*
JFF042	fwd	TGCTAGTGTAGTAGAATTGC	Gibson backbone *BGLU1*
JFF043	rev	TCTGGAAGAAGAGAAGTTG	Gibson backbone *BGLU1*
JFF044	fwd	GCAACTTCTCTTCTTCCAGAA TGAGTAAAGGAGAAGAACTTT TC	Gibson GFP insert *BGLU1*
JFF045	rev	GCAATTCTACTACACTAGCAT TAGTGGTGGTGGTGGTG	Gibson GFP insert *BGLU1*

Table S1 – continued

Name	Strand	Sequence (5' → 3')	Purpose
JFF046	fwd	TAGTAATGTTGCATTTACTTA ATTCTC	Gibson backbone *BGLU3*
JFF047	rev	CGAAGAAGCTGAAGAAGAAAA G	Gibson backbone *BGLU3*
JFF048	fwd	TTTCTTCTTCAGCTTCTTCGA TGAGTAAAGGAGAAGAACTTT TC	Gibson GFP insert *BGLU3*
JFF049	rev	AAGTAAATGCAACATTACTAT TAGTGGTGGTGGTGGTG	Gibson GFP insert *BGLU3*
JFF050	fwd	ACTCCGTATTGTTTATGAG	Gibson backbone *BGLU4*
JFF051	rev	GAAGGAAGAAGAGTATTT GC	Gibson backbone *BGLU4*
JFF052	fwd	GCAAATACTCTTCTTCCTTCA TGAGTAAAGGAGAAGAACTTT TC	Gibson GFP insert *BGLU4*
JFF053	rev	TCTCATAAACAATACGGAGTT TAGTGGTGGTGGTGGTG	Gibson GFP insert *BGLU4*
JFF056	rev	TTCCGGGAAATCGCTCCTG	Sequencing *BGLU1*
JFF059	fwd	GGTCTAGGCTTATACCGAATG G	qRT-PCR *BGLU3*
JFF072	fwd	ACTACCCTGAGGGATTCG	qRT-PCR *BGLU4*
JFF073	rev	GTGGCTTACTAACTCTTGG	qRT-PCR *BGLU4*
F067	rev	CAGCAGAGCGCAGATACCAAA TACTG	T-DNA insertion, sequencing *bglu1-1*

Table S1 – continued

Name	Strand	Sequence (5' → 3')	Purpose
F068	rev	CGCCACCTCTGACTTGAGCGTCG	sequencing *bglu1-1*
F306	fwd	TTCATATCCACGGTAGGACTCAAG	Sequencing *BGLU4*
F307	rev	GGTTTCCGGACAGAGAAGGTTTGG	Sequencing *BGLU4*
KT04	rev	TTAGTGGTGGTGGTGGTGGTGTTTG	Sequencing GFP
P35S	fwd	ACAATCCCACTATCCTTC	Sequencing RSt1376
RS469	fwd	TCCGCTCTTTCTTTCCAAGCTCAT	*ACTIN 2* primer for cDNA synthesis quality control
RS470	rev	TCCAGCACAATACCGGTTGTACG	*ACTIN 2* primer for cDNA synthesis quality control
RS629	rev	TGGTTCACGTAGTGGGCCATCG	Genotyping *bglu4-2*
RS1229	fw	ATGGAACAGTTTTTTGCTCTGTTTACCATTTTTC	Genotyping *BGLU5*
RS1230	rev	CTAAGAGGAAGAAGAGAAGTTGCTCTGC	Genotyping *BGLU5*
RS1237	rev	CGTGTGCCAGGTGCCCACGGAATAGT	Genotyping *bglu5-1*
RS1239		GGAGAAAGCAAAAGCAAGAAAAATG	Sequencing *BGLU4*

Table S1 – continued

Name	Strand	Sequence (5' → 3')	Purpose
RS1242	fwd	TTGTTAGGGAACTCTTCTACT GAAAC	Genotyping *bglu5-1*
RS1251	fwd	CATTTATGCCGGTTTTGTTTC ACTTG	Genotyping *BGLU5*
RS1259	fwd	ATGGAACAGATTTTGGCTCTG TTTGCCATTTT	Genotyping, sequencing *BGLU4*, genotyping, T-DNA insertion, sequencing *bglu4-2*
RS1260	rev	CTAGAAGGAAGAAGAGTATTT GCTCTGCA	RT-PCR, genotyping, sequencing *BGLU4*
RS1261	fwd	CAACCCAAAGAGCCCAAGATT TCTA	RT-PCR, sequencing *BGLU4*
RS1267	fwd	AGGAAGAAGAGTATTTGCTCT GCA	Sequencing *BGLU4*
RS1302	fwd	ACAGCAGGAGCGATTTCCCGG AAG	RT-PCR, genotyping, *BGLU1*, T-DNA insertion, sequencing *bglu1-1*
RS1303	rev	TGTAAGCCAGTTTCGGCCATG AG	Genotyping, sequencing *BGLU1*
RS1319	fwd	GAGATGCAGCGACAAGAACGA CTTCC	RT-PCR, genotyping *BGLU3*
RS1320	rev	AGTTCCCTGATGAGCAGTTTC TACC	qRT-PCR, genotyping *BGLU3*, T-DNA insertion, sequencing *bglu3-2*
RS1389	fwd	GAATTTCAAGGCTACCAAATT CAG	Sequencing *BGLU1*
RS1390	rev	CCAAATCTTGGTTCATTGTTT	RT-PCR, sequencing

Table S1 – continued

Name	Strand	Sequence (5' → 3')	Purpose
		TACC	*BGLU1*
RS1392	rev	GAGAACAGCAGCAATATAAGC	Sequencing *BGLU3*
RS1393	fwd	GGGGACAAGTTTGTACAAAAA AGCAGGCTTACGATCTCACCC GTCGTATATACC	Sequencing *BGLU1*
RS1474	rev	TGGCAGTCTTGATCCAATGGT TCTTTTC	Sequencing *BGLU1*
RS1484		GACCCAGCTTTCTTGTAC	Sequencing RSt1352
RS1503	fwd	AAAGGGCAAGCAGTTCAAACA CAG	Sequencing *BGLU3*
RS1845	rev	GGGACTTACCATGTTTGACAA GTT	Genotyping, T-DNA insertion, sequencing *bglu1-1*, sequencing RSt1352
RS1846	fwd	ATAATAACGCTGCGGACATCT ACATTTT	Genotyping *bglu3-2*, T-DNA insertion, sequencing *bglu1-1*
RS1847	fwd	TTATTCACCAAGCAAACTCTT GAA	Genotyping, T-DNA insertion sequencing *bglu3-2*, sequencing *BGLU3*
RS1848	fwd	ATATTGACCATCATACTCATT GC	Genotyping *bglu1-1* sequencing *bglu3-2*
RS1484	fwd	GACCCAGCTTTCTTGTAC	Sequencing RSt1371
SeLB	rev	GTAACATCAGAGATTTTGAGA CAC	Sequencing RSt1352, RSt1371
T35S	rev	TCTGGGAACTACTCACAC	T-DNA insertion, sequencing *bglu3-2*, sequencing

Table S1 – continued

Name	Strand	Sequence (5' → 3')	Purpose
			BGLU3

Table S2 – $\boldsymbol{C_t}$ ***values.*** *All C_t values from qPCR results are listed, showing the respective biological replicate with the three technical replicates for each line and for each investigated gene expression (*BGLU1, BGLU3 *and* BGLU4*). WT: wild type, OX: overexpression, KO: T-DNA insertion line.*

Sample	Line$_{Rep}$	C_t	Samples	Line$_{Rep}$	C_t	Samples	Line$_{Rep}$	C_t
BGLU1 **Replicate 1**	WT$_1$	27.89	*BGLU3* **Replicate 1**	WT$_1$	23.97	*BGLU4* **Replicate 1**	WT$_1$	23.79
	WT$_2$	27.95		WT$_2$	23.74		WT$_2$	24.13
	WT$_3$	27.95		WT$_3$	23.81		WT$_3$	24.12
	OX$_1$	19.43		OX$_1$	21.01		OX$_1$	18.40
	OX$_2$	19.60		OX$_2$	21.19		OX$_2$	18.89
	OX$_3$	19.39		OX$_3$	21.00		OX$_3$	18.72
	KO$_1$	31.69		KO$_1$	33.70		KO$_1$	33.80
	KO$_2$	32.48		KO$_2$	33.71		KO$_2$	33.37
	KO$_3$	32.17		KO$_3$	34.58		KO$_3$	32.19
BGLU1 **Replicate 2**	WT$_1$	26.55	*BGLU3* **Replicate 2**	WT$_1$	25.68	*BGLU4* **Replicate 2**	WT$_1$	23.20
	WT$_2$	26.51		WT$_2$	25.30		WT$_2$	22.93
	WT$_3$	26.60		WT$_3$	25.45		WT$_3$	22.89
	OX$_1$	21.76		OX$_1$	20.05		OX$_1$	21.00
	OX$_2$	21.80		OX$_2$	20.15		OX$_2$	21.14
	OX$_3$	21.74		OX$_3$	20.11		OX$_3$	21.12
	KO$_1$	31.10		KO$_1$	34.83		KO$_1$	0.00
	KO$_2$	31.89		KO$_2$	34.00		KO$_2$	0.00
	KO$_3$	31.36		KO$_3$	0.00		KO$_3$	0.00
	WT$_2$	28.51		WT$_2$	26.48		WT$_2$	21.51
	WT$_1$	28.32		WT$_1$	26.54		WT$_1$	21.31
	WT$_3$	28.75		WT$_3$	26.45		WT$_3$	21.76

Table S2 – continued

Sample	Line$_{Rep}$	C$_t$	Samples	Line$_{Rep}$	C$_t$	Samples	Line$_{Rep}$	C$_t$
***BGLU1* Replicate 3**	OX$_1$	22.33	***BGLU3* Replicate 3**	OX$_1$	20.89	***BGLU4* Replicate 3**	OX$_1$	16.95
	OX$_2$	22.41		OX$_2$	20.93		OX$_2$	17.29
	OX$_3$	22.45		OX$_3$	20.85		OX$_3$	16.99
	KO$_1$	30.58		KO$_1$	36.04		KO$_1$	31.42
	KO$_2$	31.18		KO$_2$	33.67		KO$_2$	32.47
	KO$_3$	31.41		KO$_3$	34.15		KO$_3$	31.73
***Pex 4* for *BGLU1* Replicate 1**	WT$_1$	25.22	***Pex 4* for *BGLU3* Replicate 1**	WT$_1$	23.62	***Pex 4* for *BGLU4* Replicate 1**	WT$_1$	23.09
	WT$_2$	25.54		WT$_2$	23.51		WT$_2$	23.15
	WT$_3$	25.29		WT$_3$	23.76		WT$_3$	23.10
	OX$_1$	25.05		OX$_1$	23.78		OX$_1$	23.55
	OX$_2$	25.17		OX$_2$	23.88		OX$_2$	23.86
	OX$_3$	24.98		OX$_3$	23.59		OX$_3$	23.51
	KO$_1$	24.96		KO$_1$	23.74		KO$_1$	24.17
	KO$_2$	25.04		KO$_2$	23.81		KO$_2$	24.41
	KO$_3$	24.87		KO$_3$	23.94		KO$_3$	24.29
***Pex 4* for *BGLU1* Replicate 2**	WT$_1$	24.65	***Pex 4* for *BGLU3* Replicate 2**	WT$_1$	24.16	***Pex 4* for *BGLU4* Replicate 2**	WT$_1$	23.53
	WT$_2$	24.76		WT$_2$	24.34		WT$_2$	22.80
	WT$_3$	24.69		WT$_3$	24.20		WT$_3$	22.67
	OX$_1$	24.63		OX$_1$	23.75		OX$_1$	25.68
	OX$_2$	24.71		OX$_2$	23.79		OX$_2$	25.89
	OX$_3$	24.62		OX$_3$	23.73		OX$_3$	25.74
	KO$_1$	24.48		KO$_1$	23.56		KO$_1$	24.33
	KO$_2$	24.57		KO$_2$	23.85		KO$_2$	24.50
	KO$_3$	24.84		KO$_3$	23.51		KO$_3$	24.66
	WT$_1$	24.70		WT$_1$	24.21		WT$_1$	22.15
	WT$_2$	24.64		WT$_2$	24.08		WT$_2$	21.78
	WT$_3$	24.96		WT$_3$	24.58		WT$_3$	21.94

Table S2 – continued

Sample	Line$_{Rep}$	C$_t$	Samples	Line$_{Rep}$	C$_t$	Samples	Line$_{Rep}$	C$_t$
Pex 4 for *BGLU1* Replicate 3	OX$_1$	24.84	*Pex 4* for *BGLU3* Replicate 3	OX$_1$	24.14	*Pex 4* for *BGLU4* Replicate 3	OX$_1$	22.07
	OX$_2$	24.65		OX$_2$	23.88		OX$_2$	22.54
	OX$_3$	24.60		OX$_3$	23.96		OX$_3$	22.04
	KO$_1$	24.07		KO$_1$	23.99		KO$_1$	23.04
	KO$_2$	24.11		KO$_2$	24.03		KO$_2$	23.14
	KO$_3$	24.16		KO$_3$	24.12		KO$_3$	22.99
EF1α for *BGLU1* Replicate 1	WT$_1$	19.39	*EF1α* for *BGLU3* Replicate 1	WT$_1$	20.76	*EF1α* for *BGLU4* Replicate 1	WT$_1$	18.00
	WT$_2$	19.46		WT$_2$	20.55		WT$_2$	17.97
	WT$_3$	19.36		WT$_3$	20.81		WT$_3$	17.95
	OX$_1$	19.13		OX$_1$	21.19		OX$_1$	17.92
	OX$_2$	18.98		OX$_2$	21.02		OX$_2$	17.75
	OX$_3$	19.05		OX$_3$	20.78		OX$_3$	17.73
	KO$_1$	19.10		KO$_1$	20.99		KO$_1$	18.06
	KO$_2$	19.29		KO$_2$	20.84		KO$_2$	18.23
	KO$_3$	19.01		KO$_3$	21.04		KO$_3$	18.02
EF1α for *BGLU1* Replicate 2	WT$_1$	19.33	*EF1α* for *BGLU3* Replicate 2	WT$_1$	19.90	*EF1α* for *BGLU4* Replicate 2	WT$_1$	17.22
	WT$_2$	19.32		WT$_2$	19.92		WT$_2$	17.52
	WT$_3$	19.43		WT$_3$	19.77		WT$_3$	17.53
	OX$_1$	19.12		OX$_1$	19.73		OX$_1$	20.59
	OX$_2$	19.07		OX$_2$	19.47		OX$_2$	20.95
	OX$_3$	19.19		OX$_3$	20.02		OX$_3$	21.10
	KO$_1$	18.82		KO$_1$	20.18		KO$_1$	18.99
	KO$_2$	19.09		KO$_2$	20.44		KO$_2$	19.13
	KO$_3$	19.01		KO$_3$	20.24		KO$_3$	18.94
	WT$_1$	19.85		WT$_1$	20.09		WT$_1$	15.48
	WT$_2$	20.12		WT$_2$	19.54		WT$_2$	15.45
	WT$_3$	20.03		WT$_3$	19.51		WT$_3$	15.36

Table S2 – continued

Sample	Line$_{Rep}$	C$_t$	Samples	Line$_{Rep}$	C$_t$	Samples	Line$_{Rep}$	C$_t$
EF1α BGLU1 Replicate 3			**EF1α BGLU3 Replicate 3**			**EF1α BGLU4 Replicate 3**		
	OX$_1$	18.98		OX$_1$	20.55		OX$_1$	16.21
	OX$_2$	18.80		OX$_2$	20.52		OX$_2$	15.81
	OX$_3$	18.83		OX$_3$	20.36		OX$_3$	16.04
	KO$_1$	17.98		KO$_1$	19.99		KO$_1$	17.39
	KO$_2$	17.67		KO$_2$	20.08		KO$_2$	17.23
	KO$_3$	17.97		KO$_3$	19.99		KO$_3$	17.06

Table S3 – *Fingerprint intensitybglu. The normalized intensity values from all candidate features for all blgu samples from the metabolic fingerprinting are listed.*

Feature	bglu 1	bglu 2	bglu 3	bglu 4
10.82min: 741.2221m/z	0	0	0	0
11.75min: 866.4164m/z	0	0	0	0
10.51min: 773.2116m/z	48.4960702986	0	0	64.0896473122
11.76min: 433.7128m/z	0	0	0	0
9.45min: 441.7099m/z	0	0	0	0
9.96min: 441.7102m/z	0	0	0	0
11.05min: 448.7181m/z	0	0	0	0
11.76min: 460.1686m/z	0	0	0	0
11.87min: 411.7178m/z	0	0	0	0
12.45min: 440.7207m/z	0	0	0	0
12.90min: 440.7209m/z	129.6152273595	0	150.7160121114	137.533602122
14.62min: 432.7049m/z	0	0	0	0
12.89min: 467.1764m/z	0	0	0	0
19.11min: 494.1968m/z	0	0	0	0
17.07min: 536.7408m/z	0	0	0	0
11.63min: 757.2152m/z	0	0	0	0
13.74min: 943.3804m/z	0	0	0	0
13.11min: 704.3635m/z	3164.0697560654	2802.2995146799	3022.9069102734	2821.8406251508
13.11min: 330.6915m/z	10831.8552259275	10779.385743145	10611.6877169856	10317.4219407867
15.38min: 781.3270m/z	1473.2724295093	1453.0346968138	1437.5055601646	1413.5117085709
13.21min: 660.3742m/z	431.0640724485	380.3582702855	443.6864716374	362.6690298598
13.11min: 402.1451m/z	452.1076502433	473.7036228646	438.4535974815	427.3907249753
17.37min: 619.2742m/z	1410.3428058874	1543.5266932888	1483.2396467691	1412.2476129632
18.19min: 575.2851m/z	275.3750402433	263.1370731239	249.1650184823	177.2262042047
7.25min: 287.1382m/z	146.8661831111	159.2944292622	150.7104133726	123.754959997
19.68min: 633.2899m/z	466.5263616219	425.5208365022	406.9726708711	402.614451064
17.03min: 795.3419m/z	110.7824681408	98.9580926425	118.5862306615	154.9781215088
11.71min: 757.2149m/z	0	0	0	0
27.45min: 529.2079m/z	0	0	0	0
14.37min: 507.1473m/z	0	0	69.13269115	0
8.75min: 611.1584m/z	786.5855232	814.5742586	774.1789584	846.3438981
11.34min: 611.1583m/z	223.594619	277.4040061	249.7351479	288.3299289

Table S4 – Fingerprint intensity2x35S:BGLU. *The normalized intensity values from all candidate features for all 2x35S:BGLU samples from the metabolic fingerprinting are listed.*

Feature	2x35S:BGLU 1	2x35S:BGLU 2	2x35S:BGLU 3	2x35S:BGLU 4
10.82min: 741.2221m/z	80.46356559	90.30071763	0	100.5087952
11.75min: 866.4164m/z	724.8221050298	501.6734964438	851.6081014691	637.112940115
10.51min: 773.2116m/z	542.1311514669	551.8676162857	469.7562818721	516.668310554
11.76min: 433.7128m/z	5793.8729342211	5817.2977131672	7757.5418971511	6745.8204380434
9.45min: 441.7099m/z	172.8873569642	187.7929337004	221.2136774002	214.7507018456
9.96min: 441.7102m/z	228.0288728642	221.3895312484	248.6501986728	248.9496071806
11.05min: 448.7181m/z	166.8292618486	157.0072068646	246.6059088525	188.9575880585
11.76min: 460.1686m/z	386.9939958173	451.8809295613	544.3190632132	506.765597225
11.87min: 411.7178m/z	652.3016199909	581.0471312872	706.0331474221	673.8450976963
12.45min: 440.7207m/z	157.2370273539	0	210.0238804884	196.5573447992
12.90min: 440.7209m/z	1987.3413051913	1816.0901813638	2379.01537986	1979.5063353542
14.62min: 432.7049m/z	922.349533048	986.6156196071	1119.4100679333	1045.5422910252
12.89min: 467.1764m/z	117.1472310692	0	191.1948953018	151.64969133
19.11min: 494.1968m/z	293.8922297355	324.8563436148	460.0728037761	357.7643059707
17.07min: 536.7408m/z	89.8537942817	0	109.4233024886	93.3848896498
11.63min: 757.2152m/z	107.434939153	118.7258247986	121.6890414106	93.9606287972
13.74min: 943.3804m/z	167.2190329692	102.6637064498	101.6765200109	0
13.11min: 704.3635m/z	367.0679388097	403.6945745131	522.6926287979	500.0870231191
13.11min: 330.6915m/z	1889.647984844	2149.5129880981	2214.1810638207	2184.6997673592
15.38min: 781.3270m/z	384.937608287	419.7566928628	391.3201092922	345.9040795419
13.21min: 660.3742m/z	0	98.1127729175	120.1827225948	118.1416729728
13.11min: 402.1451m/z	94.2943702211	98.915878342	82.2019696178	0
17.37min: 619.2742m/z	722.4493618112	675.0105236308	902.0697817709	772.9873788163
18.19min: 575.2851m/z	130.6770196376	145.0944690888	119.967534193	121.9415513431
7.25min: 287.1382m/z	78.8441312781	0	0	89.9304547679
19.68min: 633.2899m/z	187.3128470318	208.8075385417	241.5489814026	219.3566150222
17.03min: 795.3419m/z	0	0	0	0
11.71min: 757.2149m/z	135.8199225	160.0864651	136.4937619	128.5884601
27.45min: 529.2079m/z	83.00106378	95.25806186	95.06002	0
14.37min: 507.1473m/z	94.37252834	97.90411913	92.2012602	101.817866
8.75min: 611.1584m/z	462.1915831	473.9750092	418.781996	393.494292
11.34min: 611.1583m/z	149.8482714	160.7479794	155.9654619	135.981764

Table S5 – Fingerprint intensity wild type *The normalized intensity values from all candidate features for all wild type samples from the metabolic fingerprinting are listed.*

Feature	Col-0 1	Col-0 2	Col-0 3	Col-0 4
10.82min: 741.2221m/z	0	0	0	45.23697059
11.75min: 866.4164m/z	905.8171678546	445.7656967554	482.9795922223	511.3669549315
10.51min: 773.2116m/z	408.0441757629	504.4640328754	442.0085679908	421.9936701451
11.76min: 433.7128m/z	5903.7619909728	5094.959343984	5056.0337312132	5781.5802766537
9.45min: 441.7099m/z	164.6105856813	157.5306218816	162.3888040728	197.576789952
9.96min: 441.7102m/z	217.62071119883	200.8547323933	194.0890126023	183.1028145852
11.05min: 448.7181m/z	166.0013848349	147.5861251104	164.332847844	192.79897284
11.76min: 460.1686m/z	304.5393416511	398.1714166456	413.7475330274	421.7126220805
11.87min: 411.7178m/z	574.8119399672	540.9332969732	488.0635879304	548.6058235925
12.45min: 440.7207m/z	167.6914316401	129.5950596457	173.7530297718	124.3637689242
12.90min: 440.7209m/z	1548.8326303501	1305.2904540671	1525.4977708524	1569.7939690954
14.62min: 432.7049m/z	617.4498789248	123.4651042987	546.0809507106	0
12.89min: 467.1764m/z	0	82.5537904721	91.960510593	102.3014957931
19.11min: 494.1968m/z	254.8303300776	234.5289235108	181.0799647623	252.8027347954
17.07min: 536.7408m/z	0	71.6399749035	0	0
11.63min: 757.2152m/z	0	98.6261547135	0	0
13.74min: 943.3804m/z	317.8284792759	77.2003706612	206.0513554438	0
13.11min: 704.3635m/z	511.9117739865	305.7380079165	503.7641629098	500.9681765126
13.11min: 330.6915m/z	2107.6090154804	1826.441604759	2280.7701920032	2256.9564889122
15.38min: 781.3270m/z	688.5667056346	249.969327479	519.7637964599	535.9586606505
13.21min: 660.3742m/z	83.9976538005	76.531478541	123.2121312665	113.5434184079
13.11min: 402.1451m/z	123.7254286831	83.6422969346	114.240741351	118.180711486
17.37min: 619.2742m/z	1016.9931771444	823.513623589	950.1090801799	1019.6423811438
18.19min: 575.2851m/z	139.0098061339	114.3360490488	162.0897455011	164.5536422716
7.25min: 287.1382m/z	0	0	85.6802806021	0
19.68min: 633.2899m/z	238.6939495574	144.4219448906	257.0408418052	238.8908555599
17.03min: 795.3419m/z	75.8298060951	0	75.9608770429	0
11.71min: 757.2149m/z	0	0	0	0
27.45min: 529.2079m/z	56.53074944	59.69904844	0	0
14.37min: 507.1473m/z	77.80208624	0	64.92125946	77.23147019
8.75min: 611.1584m/z	572.1500415	623.5356332	538.477097	512.6071425
11.34min: 611.1583m/z	204.8304001	185.4856044	205.3791306	170.1994105

A. Sequences

A.1. Sequence of transgene in *2x35S:BGLU1*

The *BGLU1* cDNA sequence without stop is given in bold. The promoter region and the 3'UTR region is given in italic. The underlined sequence is the T-DNA left (LB) and right border (RB).

<u>GTGGCAGGATATATTGTGGTGTAAAC</u>AAATTGACGCTTAGACAACTTAATAACACATTG
CGGACGTTTTTAATGATCGAATACTAACGTCTCTACCAGATATCAGCTTGCATGCCGGT
CGATCTAGTAACATAGATGACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATAT
TTTGTTTTCTATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATC
TCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAATTCAACAGA
AATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTG
CCAAATGTTTGAACGATCTGCTTGACTCTAGGGGTCATCAGATTTCGGTGACGGGCAGG
ACCGGACGGGGCGGCACCGGCAGGCTGAAGTCCAGCTGCCAGAAACCCACGTCATGCC
AGTTCCCGTGCTTGAAGCCGGCCGCCCGCAGCATGCCGCGGGGGGCATATCCGAGCGC
CTCGTGCATGCGCACGCTCGGGTCGTTGGGCAGCCCGATGACAGCGACCACGCTCTTGA
AGCCCTGTGCCTCCAGGGACTTCAGCAGGTGGGTGTAGAGCGTGGAGCCCAGTCCCGT
CCGCTGGTGGCGGGGGGAGACGTACACGGTCGACTCGGCCGTCCAGTCGTAGGCGTTG
CGTGCCTTCCAGGGGCCCGCGTAGGCGATGCCGGCGACCTCGCCGTCCACCTCGGCGAC
GAGCCAGGGATAGCGCTCCCGCAGACGGACGAGGTCGTCCGTCCACTCCTGCGGTTCCT
GCGGCTCGGTACGGAAGTTGACCGTGCTTGTCTCGATGTAGTGGTTGACGATGGTGCA
GACCGCCGGCATGTCCGCCTCGGTGGCACGGCGGATGTCGGCCGGGCGTCGTTCTGGG
CTCATGGTAGATCCCCCTCGATCGAGTTGAGAGTGAATATGAGACTCTAATTGGATACC
GAGGGGAATTTATGGAACGTCAGTGGAGCATTTTTGACAAGAAATATTTGCTAGCTGA
TAGTGACCTTAGGCGACTTTTGAACGCGCAATAATGGTTTCTGACGTATGTGCTTAGCT
CATTAAACTCCAGAAACCCGCGGCTCAGTGGCTCCTTCAACGTTGCGGTTCTGTCAGTT
CCAAACGTAAAACGGCTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCTTAA
TTCTCCGCTCATGATCAGATTGTCGTTTCCCGCCTTCGGTTTGGGCGCGCCG*GTCTCA*
GAAGACCAGAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGG
ATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAGATGGCT
TCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGAATGCCTCTACCGAC
AGTGGTCCCAAAGATGGACCCCCACCCACGAGGAACATCGTGGAAAAAGAAGACGTTCC
AACCACGTCTTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACTCTCG
TCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAGAGGGCTATTGAGACTTTTC
AACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCA
TCGAAAGGACAGTAGAAAAGGAAGATGGCTTCTACAAATGCCATCATTGCGATAAAGGA
AAGGCTATCGTTCAAGAATGCCTCTACCGACAGTGGTCCCAAAGATGGACCCCCACCCAC
GAGGAACATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGAT
GTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTT
*CCTCTATATAAGGAAGTTC*ATTTCATTTGGAGAGGACCTCGAGAAAGAGGATCCACCTG
AGGCCTCCGTTCCATGGGCTAGAAGCTTCTCCTCCTCTGCTAACGTAAGCCTCTCTGTT
TTTTTTCTCTGTTTCTTTTGAAATGAATCCAATTAGTGATGATAATCTGTGTTTGATGT
ATCATTGATTTAACATCTTGACAATGAATCGTGATCGGAAGTGATAAAGTTATGGGTCA
ACGGTTTCAAAGAGAGAGAAAGACTTTTAGAGTCAACTCTCGACTCTTTCTTAATTATG
TTATTGCTATTTGTCTCTTTTCTTGAAGTCTGAACAATTCTTGGGATTGTTTTGCAGGT
TCTAGCTTCTCCAACCACAGGAATTCATCGCCCGGGACTGGGTTATCACAAGTTTGTA
CAAAAAAGCAGGCTCC**ATGGAAGATGTTTTGACTCTCATTACCATGATTGTGT**
TGCTTCTTCTAGCTTTCCATGGATTTGGAAAATGCAGTAGCGATCTTTACAG
CAGGAGCGATTTCCCGGAAGGCTTCGTTTTCGGAGCCGGGATATCTGCTTA

TCAGTGGGAAGGAGCTGTTGATGAAGATGGGAGGAAACCTAGCGTCTGGGA
TACTTTCCTTCACTGTCGTAAAATGGATAATGGAGACATAGCTTGTGATGGA
TATCACAAGTATAAGGAAGATGTGCAGCTCATGGCCGAAACTGGCTTACAT
ACATTCAGATTCTCCATCTCTTGGTCTAGACTCATCTCTAATGGAAGAGGTT
CCATTAACCCGAAAGGTCTACAGTTCTACAAGAATTTCATTCAAGAACTTGT
CAAACATGGAATTGAGCCACATGTTACACTACATCACTACGATTTTCCTCAA
TATCTCGAGGATGACTATGGAGGCTGGACCAACCGCAAAATCATCAAAGAC
TTTACCGCTTATGCAGATGTTTGCTTTAGAGAGTTTGGGAACCACGTCAAAT
TCTGGACCACGATCAACGAGGCTAATATATTTACTATTGGAGGTTACAACG
ATGGGAATTCACCGCCTGGTCGTTGCTCCTTTCCGGGCAGAAACTGCACGTT
AGGGAACTCTTCCACTGAAACATATATCGTAGGCCATAACTTGCTGCTTGCT
CACGCCTCTGTTTCAAGACTATATAAGCAAAGTACAAGGATATACAAGGA
GGTTCTGTTGGATTTAGTTTATTTGCAATGAATTTTACTCCTTCTACAAACT
CCAAGGATGATGAAATCGCAACGAAAGAGCCAACGATTTCTACCTCGGAT
GGATGCTTGAGCCTCTTATATATGGAGACTATCCTGATGTGATGAAAAGAA
CCATTGGATCAAGACTGCCAGTTTTCTCGAAGGAAGAATCAGAACAAGTGA
AAGGCTCATCTGACTTCATAGGAGTCATTCACTATCTCACGGCTTTGGTCAC
AAACATCGATATCAACCCTTCACTTTCAGGAATTCCAGATTTTAACTCAGAC
ATGGGTGAATCTATTAATATTTTATCTATGAGGTCTGTTGTTTCTCCATGGG
CTATGGAAGGCATCCTAGAGTATATAAAGCAGAGCTATGGCAATCCTCCAG
TCTACATTCTTGAGAATGGTAAAACAATGAACCAAGATTTGGAGCTGCAAC
AAAAGGACACACCAAGGATTGAGTACTTAGATGCTTACATTGGTGCGGTGC
TCAAAGCTGTTAGGAATGGATCAGACACGAGAGGCTACTTCGTATGGTCAT
TTATGGATTTGTACGAATTACTAAACGGATACAAGAGTAGTTTTGGATTGTA
CTCTGTCAATTTCAGTGATCCCCATCGCAAGAGATCTCCCAAACTCTCTGCT
CACTGGTACTCTGGTTTTCTCAAGGGCAAACCCACATTTCTTGGTTCCCAAG
GCATCACACAATTGCATAGCAACTTCTCTTCTTCCAGA*TACCCAGCTTTCTTGT*
ACAAAGTGGTGATAACTTCCGATCGATTCTAGACTAGTTCTAGAGTCCGCAAAAATCAC
CAGTCTCTCTCTACAAATCTATCTCTCTATTTTTCTCCAGAATAATGTGTGAGTAGTT
CCCAGATAAGGGAATTAGGGTTCTTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAA
ACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCTAA
*AACCAAAATCCAGTGACC*GGGCGGCCGCCACCGCGGTGGAGGGGGATCAGATTGTCGT
TTCCCGCCTTCAGTTTAAACTATCAGTGTTT<u>GACAGGATATATTGGCGGGTAAACC</u>

A.2. Sequence of transgene in *2x35S:BGLU3*

The *BGLU3* cDNA sequence without stop is given in bold. The promoter region and the 3'UTR region is given in italic. The underlined sequence is the T-DNA left (LB) and right border (RB).

<u>GTGGCAGGATATATTGTGGTGTAAAC</u>AAATTGACGCTTAGACAACTTAATAACACATTG
CGGACGTTTTTAATGATCGAATACTAACGTCTCTACCAGATATCAGCTTGCATGCCGGT
CGATCTAGTAACATAGATGACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATAT
TTTGTTTTCTATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATC
TCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAATTCAACAGA
AATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTG
CCAAATGTTTGAACGATCTGCTTGACTCTAGGGGTCATCAGATTTCGGTGACGGGCAGG
ACCGGACGGGGCGGCACCGGCAGGCTGAAGTCCAGCTGCCAGAAACCCACGTCATGCC
AGTTCCCGTGCTTGAAGCCGGCCGCCCGCAGCATGCCGCGGGGGGCATATCCGAGCGC

CTCGTGCATGCGCACGCTCGGGTCGTTGGGCAGCCCGATGACAGCGACCACGCTCTTGA
AGCCCTGTGCCTCCAGGGACTTCAGCAGGTGGGTGTAGAGCGTGGAGCCCAGTCCCGT
CCGCTGGTGGCGGGGGGAGACGTACACGGTCGACTCGGCCGTCCAGTCGTAGGCGTTG
CGTGCCTTCCAGGGGCCCGCGTAGGCGATGCCGGCGACCTCGCCGTCCACCTCGGCGAC
GAGCCAGGGATAGCGCTCCCGCAGACGGACGAGGTCGTCCGTCCACTCCTGCGGTTCCT
GCGGCTCGGTACGGAAGTTGACCGTGCTTGTCTCGATGTAGTGGTTGACGATGGTGCA
GACCGCCGGCATGTCCGCCTCGGTGGCACGGCGGATGTCGGCCGGGCGTCGTTCTGGG
CTCATGGTAGATCCCCCTCGATCGAGTTGAGAGTGAATATGAGACTCTAATTGGATACC
GAGGGGAATTTATGGAACGTCAGTGGAGCATTTTTGACAAGAAATATTTGCTAGCTGA
TAGTGACCTTAGGCGACTTTTGAACGCGCAATAATGGTTTCTGACGTATGTGCTTAGCT
CATTAAACTCCAGAAACCCGCGGCTCAGTGGCTCCTTCAACGTTGCGGTTCTGTCAGTT
CCAAACGTAAAACGGCTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCTTAA
TTCTCCGCTCATGATCAGATTGTCGTTTCCCGCCTTCGGTTTGGGCGCGCCCG*GTCTCA*
GAAGACCAGAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGG
ATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAGATGGCT
TCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGAATGCCTCTACCGAC
AGTGGTCCCAAAGATGGACCCCCACCCACGAGGAACATCGTGGAAAAGAAGACGTTCC
AACCACGTCTTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACTCTCG
TCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAGAGGGCTATTGAGACTTTTC
AACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCA
TCGAAAGGACAGTAGAAAAGGAAGATGGCTTCTACAAATGCCATCATTGCGATAAAGGA
AAGGCTATCGTTCAAGAATGCCTCTACCGACAGTGGTCCCAAAGATGGACCCCCACCCAC
GAGGAACATCGTGGAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGAT
GTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTT
*CCTCTATATAAGGAAGTTC*ATTTCATTTGGAGAGGACCTCGAGAAAGAGGATCCACCTG
AGGCCTCCGTTCCATGGGCTAGAAGCTTCTCCTCCTCTGCTAACGTAAGCCTCTCTGT
TTTTTTTCTCTGTTTCTTTTGAAATGAATCCAATTAGTGATGATAATCTGTGTTTGAT
GTATCATTGATTTAACATCTTGACAATGAATCGTGATCGGAAGTGATAAAGTTATGGG
TCAACGGTTTCAAAGAGAGAGAAAGACTTTTAGAGTCAACTCTCGACTCTTTCTTAAT
TATGTTATTGCTATTTGTCTCTTTTCTTGAAGTCTGAACAATTCTTGGGATTGTTTTG
CAGGTTCTAGCTTCTCCAACCACAGGAATTCATCGCCCGGGACTGGGTTATCACAAGT
TTGTACAAAAAAGCAGGCTCC**ATGGAACTGACTTTGTCTCTGTTAACCATTTTT**
CTGCTTTTCTTTGCATTGTCTGGGAGATGCAGCGACAAGAACGACTTCCCCG
AGGGCTTCATTTTTGGATCCGCCACTTCTGCTTATCAGTGGGAAGGAGCTTT
TGATGAAGATGGAAGAAAACCTAGCGTCTGGGATACTTTCCTTCACACTCG
TAACCTAAGTAACGGAGACATAACAAGTGATGGGTACCATAAGTACAAGGA
GGATGTGAAGCTCATGGTGGAAACTGGCTTAGATGCCTTCAGATTCTCCAT
TTCTTGGTCTAGGCTTATACCGAATGGAAGAGGTCCTGTTAATCCCAAGGG
TCTACAGTTCTATAAGAACTTCATTCAAGAACTTGTAAGCCATGGAATCGAA
CCACATGTGACACTGTTCCACTACGATCATCCTCAGTATCTTGAAGATGAAT
ATGGAGGATGGATCAACCGCAGAATCATCCAAGACTTTACAGCTTATGCAA
ATGTTTGCTTCAGAGAGTTTGGGCATCACGTCAAATTCTGGACCACGATCAA
CGAGGCTAATATATTCACTATTGGAGGATACAATGATGGGATCACACCGCC
TGGTCGTTGCTCCTCACCGGGTAGAAACTGCTCATCAGGAACTCTTCAACT
GAACCATATATCGTAGGCCATAACTTGCTTCTTGCGCACGCCTCTGCTTCAA
GACTGTATAAGCAAAAGTACAAGGATATGCAAGGAGGTTCTGTGGGATTTA
GCTTATTCTCCCTAGGGTTTACACCTTCTACAAGCTCCAAGGATGATGACAT
TGCAGTTCAAAGAGCCAAAGATTTCTACTTCGGCTGGATGCTTGAGCCTTTT
ATATTCGGTGACTATCCCGATGAAATGAAAAGAACTGTTGGATCAAGACTG
CCAGTTTTCTCAAAGGAAGAATCAGAACAAGTTAAAGGCTCATCTGATTTC
ATAGGAATCATTCACTATCTTGCAGCTTCGGTCACAAGCATCAAGATCAAAC

CTTCTATTTCAGGAAACCCGGATTTTTATTCAGACATGGGTGTATCTATGAC
ATTCCTTGGGAATTTTTCAGCTTTCGAGTATGCTGTTGCTCCATGGGCTATG
GAAAGTGTCCTTGAGTATATAAAGCAAAGCTATGGCAATCCTCCTATCTACA
TTCTTGAAAATGGTACACCGATGAAGCAAGATTTGCAGCTGCAACAAAAGG
ACACACCAAGGATTGAGTATTTACATGCTTATATTGCTGCTGTTCTCAAATC
TATTAGGAATGGATCAGACACGAGAGGTTACTTCATATGGTCATTTATGGA
TTTATCGAGTTAGTGAAGGGATATGAGTTTAGTTTTGGATTATACTCTGTG
AATTTCAGTGATCCTCATCGTACGAGATCTCCAAAACTCTCTGCTCATTGGT
ACTCTGCTTTTCTCAAGGGCAACACCACTTTTCTTGGTTCTCAAGGCATCAT
GCAAATGCAGAGCAACTTTTCTTCTTCAGCTTCTTCGTACCCAGCTTTCTTGTA
CAAAGTGGTGATAACTTCCGATCGATTCTAGACTAGT*TCTAGAGTCCGCAAAAATCACC*
AGTCTCTCTCTACAAATCTATCTCTCTCTATTTTTCTCCAGAATAATGTGTGAGTAGTTC
CCAGATAAGGGAATTAGGGTTCTTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAAA
CCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCTAAA
*ACCAAAATCCAGTGACC*GGGCGGCCGCCACCGCGGTGGAGGGGGATCAGATTGTCGTT
TCCCGCCTTCAGTTTAAACTATCAGTGTTT<u>GACAGGATATATTGGCGGGTAAACC</u>

A.3. Sequence of transgene in *2x35S:BGLU4*

The *BGLU4* cDNA sequence without stop is given in bold. The promoter region and the 3'UTR region is given in italic. The underlined sequence is the T-DNA left (LB) and right border (RB).

<u>GTGGCAGGATATATTGTGGTGTAAAC</u>AAATTGACGCTTAGACAACTTAATAACACATTG
CGGACGTTTTTAATGATCGAATACTAACGTCTCTACCAGATATCAGCTTGCATGCCGGT
CGATCTAGTAACATAGATGACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATAT
TTTGTTTTCTATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATC
TCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAATTCAACAGA
AATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTG
CCAAATGTTTGAACGATCTGCTTGACTCTAGGGGTCATCAGATTTCGGTGACGGGCAGG
ACCGGACGGGGCGGCACCGGCAGGCTGAAGTCCAGCTGCCAGAAACCCACGTCATGCC
AGTTCCCGTGCTTGAAGCCGGCCGCCCGCAGCATGCCGCGGGGGGCATATCCGAGCGC
CTCGTGCATGCGCACGCTCGGGTCGTTGGGCAGCCCGATGACAGCGACCACGCTCTTGA
AGCCCTGTGCCTCCAGGGACTTCAGCAGGTGGGTGTAGAGCGTGGAGCCCAGTCCCGT
CCGCTGGTGGCGGGGGGAGACGTACACGGTCGACTCGGCCGTCCAGTCGTAGGCGTTG
CGTGCCTTCCAGGGGCCCGCGTAGGCGATGCCGGCGACCTCGCCGTCCACCTCGGCGAC
GAGCCAGGGATAGCGCTCCCGCAGACGGACGAGGTCGTCCGTCCACTCCTGCGGTTCCT
GCGGCTCGGTACGGAAGTTGACCGTGCTTGTCTCGATGTAGTGGTTGACGATGGTGCA
GACCGCCGGCATGTCCGCCTCGGTGGCACGGCGGATGTCGGCCGGGCGTCGTTCTGGG
CTCATGGTAGATCCCCCTCGATCGAGTTGAGAGTGAATATGAGACTCTAATTGGATACC
GAGGGGAATTTATGGAACGTCAGTGGAGCATTTTTGACAAGAAATATTTGCTAGCTGA
TAGTGACCTTAGGCGACTTTTGAACGCGCAATAATGGTTTCTGACGTATGTGCTTAGCT
CATTAAACTCCAGAAACCCGCGGCTCAGTGGCTCCTTCAACGTTGCGGTTCTGTCAGTT
CCAAACGTAAAACGGCTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCTTAA
TTCTCCGCTCATGATCAGATTGTCGTTTCCCGCCTTCGGTTTGGGCGCGCCCGGTCTCA
GAAGACCAGAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCG
GATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAGATGGC
TTCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGAATGCCTCTACCG
ACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAACATCGTGGAAAAAGAAGACGTT

CCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACTCT
CGTCTACTCCAAGAATATCAAAGATACA*GTCTCAGAAGACCAGAGGGCTATTGAGACTT*
TTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACT
TCATCGAAAGGACAGTAGAAAAGGAAGATGGCTTCTACAAATGCCATCATTGCGATAAA
GGAAAGGCTATCGTTCAAGAATGCCTCTACCGACAGTGGTCCCAAAGATGGACCCCCAC
CCACGAGGAACATCGTGGAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGAT
TGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGAC
*CCTTCCTCTATATAAGGAAGTTC*ATTTCATTTGGAGAGGACCTCGAGAAAGAGGATCCA
CCTGAGGCCTCCGTTCCATGGGCTAGAAGCTTCTCCTCCTCTGCTAACGTAAGCCTCTC
TGTTTTTTTTCTCTGTTTCTTTTGAAATGAATCCAATTAGTGATGATAATCTGTGTTTG
ATGTATCATTGATTTAACATCTTGACAATGAATCGTGATCGGAAGTGATAAAGTTATGG
GTCAACGGTTTCAAAGAGAGAGAAAGACTTTTAGAGTCAACTCTCGACTCTTTCTTAAT
TATGTTATTGCTATTTGTCTCTTTTCTTGAAGTCTGAACAATTCTTGGGATTGTTTTGC
AGGTTCTAGCTTCTCCAACCACAGGAATTCATCGCCCGGGACTGGGTTATCACAAGTTT
GTACAAAAAAGCAGGCTCA**ATGGAACAGATTTTGGCTCTGTTTGCCATTTTTCT**
TGCTTTTGCTTTCTCCGGGAAATGCAGCGACGTTTTCAGCAGATCTGACTAC
CCTGAGGGATTCGTCTTTGGAGCCGGCACTTCTGCTTATCAGTGGGAAGGA
GCTGCTGCAGAAGACGGAAGGAAACCTAGCCTGTGGGACACTCTTTGTCAC
TCTCGTGACCAAGGTAACGGAGATATAGCATGTGATGGGTATCACAAGTAC
AAGGATGATGTGAAGTTGATGGTGGACACTAACTTAGATGCTTTCAGATTC
TCCATCTCTTGGTCTAGGCTAATACCTAATGGAAGAGGGCCTGTTAACCAAA
AGGGTCTACAGTTCTACAAGAACCTCATCCAAGAGTTAGTAAGCCACGGAA
TTGAACCACATGTTACACTCTACCACTATGATCATCCTCAGTCTCTGGAGGA
TGAATATGGAGGATGGCTCAACCACAGAATGATCAAAGACTTTACTACTTA
TGCAGATGTTTGCTTCAGAGAGTTTGGGAACCATGTCAAACTATGGACCAC
GATCAACGAGGCTAATATATTCTCTATAGGAGGTTACAACGATGGGGATAC
ACCGCCTGGTCGTTGTTCCAAACCGAGCAAAAACTGTTCTTCAGGGAACTCT
TCCATTGAACCATATATTGTAGGCCATAATTTGTTGCTTGCGCACGCCTCTG
TTTCAAGACGATACAAACAAAAATACAAGGATAAGCAGGGAGGTTCCATAG
GATTTAGCTTATTCATATTAGGGCTTATTCCTACTACAAGCTCCAAGGATGA
TGCCACTGCAACCCAAAGAGCCCAAGATTTCTACGTCGGCTGGTTCCTTAGA
CCTCTTTTATTTGGAGACTATCCTGATACAATGAAAAGAACCATTGGATCAA
GACTGCCAGTTTTCTCTGAGAAAGAATCAGAACAAGTTAAAGGCTCATGTG
ACTTTGTAGGAGTCATACACTATCATGCGGCTTCTGTCACTAACATCAAATC
CAAACCTTCTCTGTCCGGAAACCCGGATTTCTACTCATACATGGGTGTATCT
TTTCATTATTTTGGAAAAGTTTAGATTTTCAGTATGCTAATACTCCATGGG
CTATGGAAGTTGTGCTGGAGTATATAAAGCAGAGCTATGGAAATCCTCCTG
TCTACATTCTTGAGAGTGGTACACCGATGAAGCAAGATTCGCAGCTGAAAC
AGAAGGATATACCAAGGGTGGAATACTTGCATGCCTACATTGGTGGTGTGC
TCAAATCCATAAGGAATGGATCAGACACGAGAGGCTACTTCGTATGGTCAT
TTATGGATTTATACGAGCTTCTAGGGGGATATGAGGTTGGTTTTGGGTTGT
ACACTGTCAATTTCAGCGATCCTCATCGCAAGAGATCTCCCAAACTCTCTGC
TTATTGGTACTCTGATTTTCTCAAGGGTGAATCTGCCTTTCTTGATTCCCAAGGCATC
AAGGAATTGCAGAGCAAATACTCTTCTTCCTTCGACCCAGCTTTCTTGTACAAAGTGGT
GATAACTTCCGATCGATTCTAGACTAGT*TCTAGAGTCCGCAAAAATCACCAGTCTCTCT*
CTACAAATCTATCTCTCTATTTTTCTCCAGAATAATGTGTGAGTAGTTCCCAGATAAG
GGAATTAGGGTTCTTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAAACCCTTAGTA
TGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTCCTAAAACCAAAATC
*CAGTGACC*GGGCGGCCGCCACCGCGGTGGAGGGGGATCAGATTGTCGTTTCCCGCCTT
CAGTTTAAACTATCAGTGTT<u>TGACAGGATATATTGGCGGGTAAACC</u>

A.4. Sequence of pAC161

The sequence of pAC161, the vector for the generation of the GABI-KAT T-DNA insertion mutants *bglu1-1* and *bglu3-2* is given below. The sequence is derived from the GenBank entry AJ537514.1. The sequence between the T-DNA left (LB, double-underlined) and right (RB, underlined) border is marked in bold.

GAATTCGTAATCATGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCA
CAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTG
CCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT
CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGC
GGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGA
CTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAA
GGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACAT
GTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCT
GGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACG
CTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT
TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACC
GGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCT
CACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT
GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTA
TCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGC
CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTT
CTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTAT
CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTG
ATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCA
GCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTC
TACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACC
AATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATC
CATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTT
ACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGC
TCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAG
TGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA
GCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCT
CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAA
AAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCG
CAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCAT
GCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGG
GATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGT
TCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCA
CCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGG
GAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAA
TGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAA
GTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAA
ATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTG

AAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAG
CGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGC
GGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGA
GTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATAC
CGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGA
TCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCT
GCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGT
AAAACGACGGCCAGTGCCAAGCTTGCATGTCGACATGCATGCATGTCGACA
TGCATGCATGTCGACTCTAGAGGATCCCCAATTCCCATGGAGTCAAAGATT
CAAATAGAGGACCTAACAGAACTCGCCGTAAAGACTGGCGAACAGTTCATA
CAGAGTCTCTTACGACTCAATGACAAGAAGAAAATCTTCGTCAACATGGTG
GAGCACGACACGCTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAA
GACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTC
CTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAA
AGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCG
TTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGA
GGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGG
ATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCC
TTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACA
GGGTACCCGGGGATCAGATTGTCGTTTCCCGCCTTCGGTTTAAACTATCAGT
GTTTGACAGGATATATTGGCGGGTAAACCTAAGAGAAAAGAGCGTTTATTAGAA
TAATCGGATATTTAAAAGGGCGTGAAAAGGTTTATCCGTTCGTCCATTTGTATGTGCAT
GCCAACCACAGGGTTCCCCTCGGGAGTGCTTGGCATTCCGTGCGATAATGACTTCTGTT
CAACCACCCAAACGTCGGAAAGCCTGACGACGGAGCAGCATTCCAAAAAGATCCCTTGG
CTCGTCTGGGTCGGCTAGAAGGTCGAGTGGGCTGCTGTGGCTTGATCCCTCAACGCGG
TCGCGGACGTAGCGCAGCGCCGAAAAATCCTCGATCGCAAATCCGACGCTGTCGAAAAG
CGTGATCTGCTTGTCGCTCTTTCGGCCGACGTCCTGGCCAGTCATCACGCGCCAAAGTT
CCGTCACAGGATGATCTGGCGCGAGTTGCTGGATCTCGCCTTCAATCCGGGTCTGTGGC
GGGAACTCCACGAAAATATCCGAACGCAGCAAGATATCGCGGTGCATCTCGGTCTTGCC
TGGGCAGTCGCCGCCGACGCCGTTGATGTGGACGCCGGGCCCGATCATATTGTCGCTCA
GGATCGTGGCGTTGTGCTTGTCGGCCGTTGCTGTCGTAATGATATCGGCACCTTCGACC
GCCTGTTCCGCAGAGGTGCAGGCCTCGATCTGAAACCCGAACCGCTGGAGATTGCGGG
AGCAGCGAGCAGTAGCCTCGGGGTCGATGTCGTAAAGTCGTATCCGATCGACGCCGAT
CAGCGCCTTGAAGGCCAAAGCCTGGAACTCACTTTGGGCACCGTTGCCGATCAGCGCCA
TCGTGCGCGAATCTTTACGGGCCAGATACTTTGCCGCGATCGCGGAGGTCGCGGCCGTT
CGCAAGGCCGTCAGGATTGTCATTTCCGACAGCAGCAGCGGATAGCCGCTATCGACATC
GGAGAGCACGCCGAACGCGGTTACCTAGAGCGGCCGCCACCGCGGTGCCTTGATGTGG
GCGCCGGCGGTCGAGTGGCGACGGCGCGGCTTGTCCGCGCCCTGGTAGATTGCCTGGC
CGTAGGCCAGCCATTTTTGAGCGGCCAGCGGCCGCGATAGGCCGACGCGAAGCGGCGG
GGCGTAGGGAGCGCAGCGACCGAAGGGTAGGCGCTTTTTGCAGCTCTTCGGCTGTGCG
CTGGCCAGACAGTTATGCACAGGCCAGGCGGGTTTTAAGAGTTTTAATAAGTTTTAAAG
AGTTTTAGGCGGAAAAATCGCCTTTTTTCTCTTTTATATCAGTCACTTACATGTGTGAC
CGGTTCCCAATGTACGGCTTTGGGTTCCCAATGTACGGGTTCCGGTTCCCAATGTACGG
CTTTGGGTTCCCAATGTACGTGCTATCCACAGGAAAGAGACCTTTTCGACCTTTTTCCC
CTGCTAGGGCAATTTGCCCTAGCATCTGCTCCGTACATTAGGAACCGGCGGATGCTTCG
CCCTCGATCAGGTTGCGGTAGCGCATGACTAGGATCGGGCCAGCCTGCCCCGCCTCCTC
CTTCAAATCGTACTCCGGCAGGTCATTTGACCCGATCAGCTTGCGCACGGTGAAACAGA
ACTTCTTGAACTCTCCGGCGCTGCCACTGCGTTCGTAGATCGTCTTGAACAACCATCTG
GCTTCTGCCTTGCCTGCGGCGCGGCGTGCCAGGCGGTAGAGAAAACGGCCGATGCCGG
GATCGATCAAAAAGTAATCGGGGTGAACCGTCAGCACGTCCGGGTTCTTGCCTTCTGTG
ATCTCGCGGTACATCCAATCAGCTAGCTCGATCTCGATGTACTCCGGCCGCCCGGTTTC

GCTCTTTACGATCTTGTAGCGGCTAATCAAGGCTTCACCCTCGGATACCGTCACCAGGC
GGCCGTTCTTGGCCTTCTTCGTACGCTGCATGGCAACGTGCGTGGTGTTTAACCGAATG
CAGGTTTCTACCAGGTCGTCTTTCTGCTTTCCGCCATCGGCTCGCCGGCAGAACTTGAG
TACGTCCGCAACGTGTGGACGGAACACGCGGCCGGGCTTGTCTCCCTTCCCTTCCCGGT
ATCGGTTCATGGATTCGGTTAGATGGGAAACCGCCATCAGTACCAGGTCGTAATCCCAC
ACACTGGCCATGCCGGCCGGCCCTGCGGAAACCTCTACGTGCCCGTCTGGAAGCTCGT
AGCGGATCACCTCGCCAGCTCGTCGGTCACGCTTCGACAGACGGAAAACGGCCACGTC
CATGATGCTGCGACTATCGCGGGTGCCCACGTCATAGAGCATCGGAACGAAAAAATCT
GGTTGCTCGTCGCCCTTGGGCGGCTTCCTAATCGACGGCGCACCGGCTGCCGGCGGTT
GCCGGGATTCTTTGCGGATTCGATCAGCGGCCGCTTGCCACGATTCACCGGGGCGTGC
TTCTGCCTCGATGCGTTGCCGCTGGGCGGCCTGCGCGGCCTTCAACTTCTCCACCAGG
TCATCACCCAGCGCCGCGCCGATTTGTACCGGGCCGGATGGTTTGCGACCGCTCACGC
CGATTCCTCGGGCTTGGGGGTTCCAGTGCCATTGCAGGGCCGGCAGACAACCCAGCCG
CTTACGCCTGGCCAACCGCCCGTTCCTCCACACATGGGGCATTCCACGGCGTCGGTGCC
TGGTTGTTCTTGATTTTCCATGCCGCCTCCTTTAGCCGCTAAAATTCATCTACTCATTT
ATTCATTTGCTCATTTACTCTGGTAGCTGCGCGATGTATTCAGATAGCAGCTCGGTAAT
GGTCTTGCCTTGGCGTACCGCGTACATCTTCAGCTTGGTGTGATCCTCCGCCGGCAAC
TGAAAGTTGACCCGCTTCATGGCTGGCGTGTCTGCCAGGCTGGCCAACGTTGCAGCCT
TGCTGCTGCGTGCGCTCGGACGGCCGGCACTTAGCGTGTTTGTGCTTTTGCTCATTTT
CTCTTTACCTCATTAACTCAAATGAGTTTTGATTTAATTTCAGCGGCCAGCGCCTGGAC
CTCGCGGGCAGCGTCGCCCTCGGGTTCTGATTCAAGAACGGTTGTGCCGGCGGCGGCA
GTGCCTGGGTAGCTCACGCGCTGCGTGATACGGGACTCAAGAATGGGCAGCTCGTACC
CGGCCAGCGCCTCGGCAACCTCACCGCCGATGCGCGTGCCTTTGATCGCCCGCGACAC
GACAAAGGCCGCTTGTAGCCTTCCATCCGTGACCTCAATGCGCTGCTTAACCAGCTCCA
CCAGGTCGGCGGTGGCCCATATGTCGTAAGGGCTTGGCTGCACCGGAATCAGCACGAA
GTCGGCTGCCTTGATCGCGGACACAGCCAAGTCCGCCGCCTGGGGCGCTCCGTCGATC
ACTACGAAGTCGCGCCGGCCGATGGCCTTCACGTCGCGGTCAATCGTCGGGCGGTCGA
TGCCGACAACGGTTAGCGGTTGATCTTCCCGCACGGCCGCCCAATCGCGGGCACTGCC
CTGGGGATCGGAATCGACTAACAGAACATCGGCCCCGGCGAGTTGCAGGGCGCGGGCT
AGATGGGTTGCGATGGTCGTCTTGCCTGACCCGCCTTTCTGGTTAAGTACAGCGATAA
CCTTCATGCGTTCCCCTTGCGTATTTGTTTATTTACTCATCGCATCATATACGCAGCGA
CCGCATGACGCAAGCTGTTTTACTCAAATACACATCACCTTTTTAGACGGCGGCGCTCG
GTTTCTTCAGCGGCCAAGCTGGCCGGCCAGGCCGCCAGCTTGGCATCAGACAAACCGG
CCAGGATTTCATGCAGCCGCACGGTTGAGACGTGCGCGGCGGCTCGAACACGTACCC
GGCCGCGATCATCTCCGCCTCGATCTCTTCGGTAATGAAAAACGGTTCGTCCTGGCCG
TCCTGGTGCGGTTTCATGCTTGTTCCTCTTGGCGTTCATTCTCGGCGGCCGCCAGGGC
GTCGGCCTCGGTCAATGCGTCCTCACGGAAGGCACCGCGCCGCCTGGCCTCGGTGGGC
GTCACTTCCTCGCTGCGCTCAAGTGCGCGGTACAGGGTCGAGCGATGCACGCCAAGCA
GTGCAGCCGCCTCTTTCACGGTGCGGCCTTCCTGGTCGATCAGCTCGCGGGCGTGCGC
GATCTGTGCCGGGGTGAGGGTAGGGCGGGGGCCAAACTTCACGCCTCGGGCCTTGGCG
GCCTCGCGCCCGCTCCGGGTGCGGTCGATGATTAGGGAACGCTCGAACTCGGCAATGC
CGGCGAACACGGTCAACACCATGCGGCCGGCCGGCGTGGTGGTGTCGGCCCACGGCTC
TGCCAGGCTACGCAGGCCCGCGCCGGCCTCCTGGATGCGCTCGGCAATGTCCAGTAGG
TCGCGGGTGCTGCGGGCCAGGCGGTCTAGCCTGGTCACTGTCACAACGTCGCCAGGGC
GTAGGTGGTCAAGCATCCTGGCCAGCTCCGGGCGGTCGCGCCTGGTGCCGGTGATCTT
CTCGGAAAACAGCTTGGTGCAGCCGGCCGCGTGCAGTTCGGCCCGTTGGTTGGTCAAG
TCCTGGTCGTCGGTGCTGACGCGGGCATAGCCCAGCAGGCCAGCGGCGGCGCTCTTGT
TCATGGCGTAATGTCTCCGGTTCTAGTCGCAAGTATTCTACTTTATGCGACTAAAACAC
GCGACAAGAAAACGCCAGGAAAAGGGCAGGGCGGCAGCCTGTCGCGTAACTTAGGACT
TGTGCGACATGTCGTTTTCAGAAGACGGCTGCACTGAACGTCAGAAGCCGACTGCACT
ATAGCAGCGGAGGGGTTGGATCGACCTCGAGGGGGGGCCCGGTACCGTGAACGTCGGC

TCGATTGTACCTGCGTTCAAATACTTTGCGATCGTGTTGCGCGCCTGCCCGGTGCGTC
GGCTGATCTCACGGATCGACTGCTTCTCTCGCAACGCCATCCGACGGATGATGTTTAAA
AGTCCCATGTGGATCACTCCGTTGCCCCGTCGCTCACCGTGTTGGGGGGAAGGTGCAC
ATGGCTCAGTTCTCAATGGAAATTATCTGCCTAACCGGCTCAGTTCTGCGTAGAAACCA
ACATGCAAGCTCCACCGGGTGCAAAGCGGCAGCGGC**GGCAGGATATATTCAATTGT**
AAATGGCTTCATGTCCGGGAAATCTACATGGATCAGCAATGAGTATGATGG
TCAATATGGAGAAAAGAAAGAGTAATTACCAATTTTTTTTCAATTCAAAAA
TGTAGATGTCCGCAGCGTTATTATAAAATGAAAGTACATTTTGATAAAACG
ACAAATTACGATCCGTCGTATTTATAGGCGAAAGCAATAAACAAATTATTCT
AATTCGGAAATCTTTATTTCGACGTGTCTACATTCACGTCCAAATGGGGGCT
TAGATGAGAAACTTCACGATCGATGCCTTGATTTCGCCATTCCCAGATACCC
ATTTCATCTTCAGATTGGTCTGAGATTATGCGAAAATATACACTCATATACA
TAAATACTGACAGTTTGAGCTACCAATTCAGTGTAGCCCATTACCTTACATA
ATTCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTGTCGGCCAC
TATACGATAGTTGCGCAAATTTTCAAAGTCCTGGCCTAACATCACACCTCTG
TCGGCGGCGGGTCCCATTTGTGATAAATCCACCATCACAATAGATAGTCTAA
TGGACGAAAAAGGCGAATATTTCGATGCTGAGATTCGACGCAATTAATTCG
AGAAAAATCCCGTGATTGATGCTGTTGAGTTACCAATAATATGGGCAGCGA
AGGCCATTTAATTATAAGATCCGATTTGGTGTATCGAGATTGGTTATGAAAT
TCAGATGCTAGTGTAATGTATTGGTAATTTGGGAAGATATAATAGGAAGCA
AGGCTATTTATCCATTTCTGAAAAGGCGAAATGGCGTCACCGCGAGCGTCA
CGCGCATTCCGTTCTTGCTGTAAAGCGTTGTTTGGTACACTTTTGACTAGCG
AGGCTTGGCGTGTCAGCGTATCTATTCAAAAGTCGTTAATGGCTGCGGATC
AAGAAAAGTTGGAATAGAAACAGAATACCCGCGAAATTCAGGCCCGGTTG
CCATGTCCTACACGCCGAAATAAACGACCAAATTAGTAGAAAAATAAAAAC
TGACTCGGATACTTACGTCACGTCTTGCGCACTGATTTGAAAAATCTCAATA
TAAACAAAGACGGCCACAAGAAAAAACCAAACACCGATATTCATTAATCT
TATCTAGTTTCTCAAAAAAATTCATATCTTCCACACGTGGATCGATCCGTCG
TACCAGCTTTGCAATTCATACAGAAGTGAGAAAAATGGCTTCTATGATATCC
TCTTCAGCTGTGACTACAGTCAGCCGTGCTTCTACGGTGCAATCGGCCGCG
GTGGCTCCATTCGGCGGCCTCAAATCCATGACTGGATTCCCAGTTAAGAAG
GTCAACACTGACATTACTTCCATTACAAGCAATGGTGGAAGAGTAAAGTGC
ATGGTGACGGTGTTCGGCATTCTGAATCTCACCGAGGACTCCTTCTTCGATG
AGAGCCGGCGGCTAGACCCCGCCGGCGCTGTCACCGCGGCGATCGAAATGC
TGCGAGTCGGATCAGACGTCGTGGATGTCGGACCGGCCGCCAGCCATCCGG
ACGCGAGGCCTGTATCGCCGGCCGATGAGATCAGACGTATTGCGCCGCTCT
TAGACGCCCTGTCCGATCAGATGCACCGTGTTTCAATCGACAGCTTCCAACC
GGAAACCCAGCGCTATGCGCTCAAGCGCGGCGTGGGCTACCTGAACGATAT
CCAAGGATTTCCTGACCCTGCGCTCTATCCCGATATTGCTGAGGCGGACTGC
AGGCTGGTGGTTATGCACTCAGCGCAGCGGGATGGCATCGCCACCCGCACC
GGTCACCTTCGACCCGAAGACGCGCTCGACGAGATTGTGCGGTTCTTCGAG
GCGCGGGTTTCCGCCTTGCGACGGAGCGGGTCGCTGCCGACCGGCTCATC
CTCGATCCGGGGATGGGATTTTTCTTGAGCCCCGCACCGGAAACATCGCTG
CACGTGCTGTCGAACCTTCAAAAGCTGAAGTCGGCGTTGGGGCTTCCGCTA
TTGGTCTCGGTGTCGCGGAAATCCTTCTTGGGCGCCACCGTTGGCCTTCCTG
TAAAGGATCTGGGTCCAGCGAGCCTTGCGGCGGAACTTCACGCGATCGGCA
ATGGCGCTGACTACGTCCGCACCCACGCGCCTGGAGATCTGCGAAGCGCAA
TCACCATCTCGGAAACCCTCGCGAAATTTCGCAGTCGCGACGCCAGAGACC
GAGGGTTAGATCATGCCTAGCTTGGCTGCAGGTCGACCTGCAGGCATGCCC
GCTGAAATCACCAGTCTCTCTCTACAAATCTATCTCTCTATAATAATGTG
TGAGTAGTTCCCAGATAAGGGAATTAGGGTTCTTATAGGGTTTCGCTCATG

**TGTTGAGCATATAAGAAACCCTTAGTATGTATTTGTATTTGTAAAATACTTC
TATCAATAAAATTTCTAATTCCTAAAACCAAAATCCAGGGGTACCGAGCTC**

A.5. Sequence of pROK2

The sequence of pROK2, the vector for the generation of the SALK T-DNA insertion mutant *bglu4-2* is given below. The sequence is derived from the SIGnAL T-DNA insertion database (http://signal.salk.edu/pBIN-pROK2.txt-new). The sequence between the T-DNA left (LB, double-underlined) and right (RB, underlined) border is marked in bold.

CCGGGCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGCGCCA
GAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGCCGGCGTTGTGGATACCTCG
CGGNtcttaccttAAAACTTGGCCCTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGG
GGCCGACTCACCCGGCGCGGCGTTGACAGATGAGGGGCAGGCTCGATTTCGGCCGGCG
ACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTTACGCGAGTTTCCC
ACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGGGC
GCGACTACTGACAGATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACA
GATGAGGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCAT
TTGCAAGGGTTTCCGCCCGTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTA
AACCAATATTTATAAACCTTGTTTTTAACCAGGGCTGCGCCCTGTGCGCGTGACCGCGC
ACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCCTCCCGGCCCGCTAACGCGGGCCTCC
CATCCCCCCAGGGGCTGCGCCCCTCGGCCGCGAACGGCCTCACCCCAAAAATGGCAGCG
CTGGCAGTCCTTGCCATTGCCGGGATCGGGGCAGTAACGGGATGGGCGATCAGCCCGA
GCGCGACGCCCGGAAGCATTGACGTGCCGCAGGTGCTGGCATCGACATTCAGCGACCA
GGTGCCGGGCAGTGAGGGCGGCGGCCTGGGTGGCGGCCTGCCCTTCACTTCGGCCGTC
GGGGCATTCACGGACTTCATGGCGGGGCCGGCAATTTTTACCTTGGGCATTCTTGGCA
TAGTGGTCGCGGGTGCCGTGCTCGTGTTCGGGGGTGCGATAAACCCAGCGAACCATTT
GAGGTGATAGGTAAGATTATACCGAGGTATGAAAACGAGAATTGGACCTTTACAGAAT
TACTCTATGAAGCGCCATATTTAAAAAGCTACCAAGACGAAGAGGATGAAGAGGATGA
GGAGGCAGATTGCCTTGAATATATTGACAATACTGATAAGATAATATATCTTTTATATA
GAAGATATCGCCGTATGTAAGGATTTCAGGGGGCAAGGCATAGGCAGCGCGCTTATCA
ATATATCTATAGAATGGGCAAAGCATAAAAACTTGCATGGACTAATGCTTGAAACCCA
GGACAATAACCTTATAGCTTGTAAATTCTATCATAATTGGGTAATGACTCCAACTTATT
GATAGTGTTTTATGTTCAGATAATGCCCGATGACTTTGTCATGCAGCTCCACCGATTTT
GAGAACGACAGCGACTTCCGTCCCAGCCGTGCCAGGTGCTGCCTCAGATTCAGGTTAT
GCCGCTCAATTCGCTGCGTATATCGCTTGCTGATTACGTGCAGCTTTCCCTTCAGGCGG
GATTCATACAGCGGCCAGCCATCCGTCATCCATATCACCACGTCAAAGGGTGACAGCA
GGCTCATAAGACGCCCCAGCGTCGCCATAGTGCGTTCACCGAATACGTGCGCAACAAC
CGTCTTCCGGAGACTGTCATACGCGTAAAACAGCCAGCGCTGGCGCGATTTActtGCCCC
GACATAGCCCCACTGTTCGTCCATTTCCGCGCAGACGATGACGTCACTGCCCGGCTGTA
TGCGCGAGGTTACCGACTGCGGCCTGAGTTTTTTAAGTGACGTAAAATCGTGTTGAGG
CCAACGCCCATAATGCGGGCTGTTGCCCGGCATCCAACGCCATTCATGGCCATATCAAT
GATTTTCTGGTGCGTACCGGGTTGAGAAGCGGTGTAAGTGAACTGCAGTTGCCATGTT
TtcttTACGGCAGTGAGAGCAGAGATAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACG
CACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGACACAGAAGCCACTGGA
GCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCACCCATAATTG
TGGTTTCAAAATCGGCTCCGTCGATACTATGTTATACGCCAACTTTGAAAACAACTTTG
AAAAAGCTGTTTTCTGGTATTTAAGGTTTTAGAATGCAAGGAACAGTGAATTGGAGTT
CGTCTTGTTATAATTAGCTTCTTGGGGTATCTTTAAATACTGTAGAAAAGAGGAAGGA
AATAATAAATGGCTAAAATGAGAATATCACCGGAATTGAAAAAACTGATCGAAAAATA

CCGCTGCGTAAAAGATACGGAAGGAATGTCTCCTGCTAAGGTATATAAGCTGGTGGGA
GAAAATGAAAACCTATATTTAAAAATGACGGACAGCCGGTATAAAGGGACCACCTATG
ATGTGGAACGGGAAAAGGACATGATGCTATGGCTGGAAGGAAAGCTGCCTGTTCCAAA
GGTCCTGCACTTTGAACGGCATGATGGCTGGAGCAATCTGCTCATGAGTGAGGCCGAT
GGCGTCCTTTGCTCGGAAGAGTATGAAGATGAACAAAGCCCTGAAAAGATTATCGAGC
TGTATGCGGAGTGCATCAGGCTCTTTCACTCCATCGACATATCGGATTGTCCCTATACG
AATAGCTTAGACAGCCGCTTAGCCGAATTGGATTACTTACTGAATAACGATCTGGCCG
ATGTGGATTGCGAAAACTGGGAAGAAGACACTCCATTTAAAGATCCGCGCGAGCTGTA
TGATTTTTTAAAGACGGAAAAGCCCGAAGAGGAACTTGTCTTTTCCCACGGCGACCTG
GGAGACAGCAACATCTTTGTGAAAGATGGCAAAGTAAGTGGCTTTATTGATCTTGGGA
GAAGCGGCAGGGCGGACAAGTGGTATGACATTGCCTTCTGCGTCCGGTCGATCAGGGA
GGATATCGGGGAAGAACAGTATGTCGAGCTATTTTTTGACTTACTGGGGATCAAGCCT
GATTGGGAGAAAATAAAATATTATATTTTACTGGATGAATTGTTTTAGTACCTAGATG
TGGCGCAACGATGCCGGCGACAAGCAGGAGCGCACCGACTTCTTCCGCATCAAGTGTT
TTGGCTCTCAGGCCGAGGCCCACGGCAAGTATTTGGGCAAGGGGTCGCTGGTATTCGT
GCAGGGCAAGATTCGGAATACCAAGTACGAGAAGGACGGCCAGACGGTCTACGGGACC
GACTTCATTGCCGATAAGGTGGATTATCTGGACACCAAGGCACCAGGCGGGTCAAATC
AGGAATAAGGGCACATTGCCCCGGCGTGAGTCGGGGCAATCCCGCAAGGAGGGTGAAT
GAATCGGACGTTTGACCGGAAGGCATACAGGCAAGAACTGATCGACGCGGGGTTTTCC
GCCGAGGATGCCGAAACCATCGCAAGCCGCACCGTCATGCGTGCGCCCCGCGAAACCT
TCCAGTCCGTCGGCTCGATGGTCCAGCAAGCTACGGCCAAGATCGAGCGCGACAGCGT
GCAACTGGCTCCCCCTGCCCTGCCCGCGCCATCGGCCGCCGTGGAGCGTTCGCGTCGT
CTCGAACAGGAGGCGGCAGGTTTGGCGAAGTCGATGACCATCGACACGCGAGGAACTA
TGACGACCAAGAAGCGAAAAACCGCCGGCGAGGACCTGGCAAAACAGGTCAGCGAGGC
CAAGCAGGCCGCGTTGCTGAAACACACGAAGCAGCAGATCAAGGAAATGCAGCTTTCC
TTGTTCGATATTGCGCCGTGGCCGGACACGATGCGAGCGATGCCAAACGACACGGCCC
GCTCTGCCCTGTTCACCACGCGCAACAAGAAATCCCGCGCGAGGCGCTGCAAAACAA
GGTCATTTTCCACGTCAACAAGGACGTGAAGATCACCTACACCGGCGTCGAGCTGCGG
GCCGACGATGACGAACTGGTGTGGCAGCAGGTGTTGGAGTACGCGAAGCGCACCCCTA
TCGGCGAGCCGATCACCTTCACGTTCTACGAGCTTTGCCAGGACCTGGGCTGGTCGAT
CAATGGCCGGTATTACACGAAGGCCGAGGAATGCCTGTCGCGCCTACAGGCGACGGCG
ATGGGCTTCACGTCCGACCGCGTTGGGCACCTGGAATCGGTGTCGCTGCTGCACCGCT
TCCGCGTCCTGGACCGTGGCAAGAAAACGTCCCGTTGCCAGGTCCTGATCGACGAGGA
AATCGTCGTGCTGTTTGCTGGCGACCACTACACGAAATTCATATGGGAGAAGTACCGC
AAGCTGTCGCCGACGGCCCGACNcttGGATGTTCGACTATTTCAGCTCGCACCGGGAGCC
GTACCCGCTCAAGCTGGAAACCTTCCGCCTCATGTGCGGATCGGATTCCACCCGCGTG
AAGAAGTGGCGCGAGCAGGTCGGCGAAGCCTGCGAAGAGTTGCGAGGCAGCGGCCTG
GTGGAACACGCCTGGGTCAATGATGACCTGGTGCATTGCAAACGCTAGGGCCTTGTGG
GGTCAGTTCCGGCTGGGGGTTCAGCAGCCAGCGCTTTACTGGCATTTCAGGAACAAGC
GGGCACTGCTCGACGCACTTGCTTCGCTCAGTATCGCTCGGGACGCACGGCGCGCTCT
ACGAACTGCCGATAAACAGAGGATTAAAATTGACAATTGTGATTAAGGCTCAGATTCG
ACGGCTTGGAGCGGCCGACGTGCAGGATTTCCGCGAGATCCGATTGTCGGCCCTGAAG
AAAGCTCCAGAGATGTTCGGGTCCGTTACGAGCACGAGGAGAAAAAGCCCATGGAGG
CGTTCGCTGAACGGTTGCGAGATGCCGTGGCATTCGGCGCCTACATCGACGGCGAGAT
CATTGGGCTGTCGGTCTTCAAACAGGAGGACGGCCCCAAGGACGCTCACAAGGCGCAT
CTGTCCGGCGTTTTCGTGGAGCCCGAACAGCGAGGCCGAGGGGTCGCCGGTATGCTGC
TGCGGGCGTTGCCGGCGGGTTTATTGCTCGTGATGATCGTCCGACAGATTCCAACGGG
AATCTGGTGGATGCGCATCTTCATCCTCGGCGCACTTAATATTTCGCTATTCTGGAGCT
TGTTGTTTATTTCGGTCTACCGCCTGCCGGGCGGGGTCGCGGCGACGGTAGGCGCTGT
GCAGCCGCTGATGGTCGTGTTCATCTCTGCCGCTCTGCTAGGTAGCCCGATACGATTG
ATGGCGGTCCTGGGGGCTATTTGCGGAACTGCGGGCGTGGCGCTGTTGGTGTTGACAC

CAAACGCAGCGCTAGATCCTGTCGGCGTCGCAGCGGGCCTGGCGGGGGCGGTTTCCAT
GGCGTTCGGAACCGTGCTGACCCGCAAGTGGCAACCTCCCGTGCCTCTGCTCACCTTTA
CCGCCTGGCAACTGGCGGCCGGAGGACTTCTGCTCGTTCCAGTAGCTTTAGTGTTTGA
TCCGCCAATCCCGATGCCTACAGGAACCAATGTTCTCGGCCTGGCGTGGCTCGGCCTG
ATCGGAGCGGGTTTAACCTACTTCCTTTGGTTCCGGGGGATCTCGCGACTCGAACCTA
CAGTTGTTTCCTTACTGGGCTTTCTCAGCCCCAGATCTGGGGTCGATCAGCCGGGGAT
GCATCAGGCgcttCGACAGTCGGAACTTCGGGTCCCCGACCTGTACCATTCGGTGAGCAA
TGGATAGGGGAGTTGATATCGTCAACGTTCACTTCTAAAGAAATAGCGCCACTCAGCT
TCCTCAGCGGCTTTATCCAGCGATTTCCTATTATGTCGGCATAGTTCTCAAGATCGACA
GCCTGTCACGGTTAAGCGAGAAATGAATAAGAAGGCTGATAATTCGGATCTCTGCGAG
GGAGATGATATTTGATCACAGGCAGCAACGCTCTGTCATCGTTACAATCAACATGCTA
CCCTCCGCGAGATCATCCGTGTTTCAAACCCGGCAGCTTAGTTGCCGTTCTTCCGAATA
GCATCGGTAACATGAGCAAAGTCTGCCGCCTTACAACGGCTCTCCCGCTGACGCCGTCC
CGGA<u>CTGATGGGCTGCCTGT</u>ATtCGAGTGGTGATTTTGTGCCGAGCTGCCGG
TCGGGGAGCTGTTGGCTGGCTGGTGGCAGGATATATTGTGGTGTAAACAAA
TTGACGCTTAGACAACTTAATAACACATTGCGGACGTTTTTAATGTACTGGG
GTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCG
CCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCA
GGCGAAAATCCTGTTTGATGGTGGTTCCGAAATCGGCAAAATCCCTTATAA
ATCAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAA
GAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGT
CTATCAGGGCGATGGCCCACTACGTGAACCATaCACCCAAATCAAGTTTTTT
GGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCG
ATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCCATTCAGGCTGCGCAACTGTTGGGAAGG
GCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATG
TGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACG
TTGTAAAACGACGGCCAGTGAATTCCCGccttATCTAGTAACATAGATGACAC
CGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTCTATCGCG
TATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATCTCATAAATA
ACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAATTCAACAGAAA
TTATATGATAATCATCGCAAGACCGGCAACAGGATTCAATCTTAAGAAACTT
TATTGCCAAATGTTTGAACGATCGGGGAAATTCGAGCTCGGTACCCGGGGA
TCCTCTAGAGTCCCCCGTGTacttTCTCTCCAAATGAAATGAACTTCCTTATAT
AGAGGAAGGGTCTTGCGAAGGATAGTGGGATTGTGCGTCATCCCTTACGTC
AGTGGAGATATCACATCAATCCACTTGCTTTGAAGACGTGGTTGGAACGTC
TTCTTTTTCCACGATGCTCCTCGTGGGTGGGGGTCCATCTTTGGGACCACTG
TCGGCAGAGGCATCTTCAACGATGGCCTTTCCTTTATCGCAATGATGGCATT
TGTAGGAGCCACCTTCCTTTTCCACTATCTTCACAATAAAGTGACAGATAGC
TGGGCAATGGAATCCGAGGAGGTTTCCGGATATTACCCTTTGTTGAAAAGT
CTCAATTGCCCTTTGGTCTTCTGAGACTGTATCTTTGATATTTTTGGAGTAG
ACAAGTGTGTCGTGCTCCACCATGTTGACGAAGATTTTCTTCTTGTCATTGA
GTCGTAAGAGACTCTGTATGAACTGTTCGCCAGTCTTTACGGCGAGTTCTGT
TAGGTCCTCTATTTGAATCTTTGACTCCATGGCCTTTGATTCAGTGGGAACT
ACCTTTTTAGAGACTCCAATCTCTATTACTTGCCTTGGTTTGTGAAGCAAGC
CTTGAATCGTCCATACTGGAATAGTACTTCTGATCTTGAGAAATATATCTTT
CTCTGTGTTCTTGATGCAGTTAGTCCTGAATCTTTTGACTGCATCTTTAACC
TTCTTGGGAAGGTATTTGATTTCCTGGAGATTATTGCTCGGGTAGATCGTCT
TGATGAGACCTGCTGCGTAAGCCTCTCTAACCATCTGTGGGTTAGCATTCTT
TCTGAAATTGAAAAGGCTAATCTGGGGACCTGCAGGCATGCAAGCTTGGCG
TAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGncttCTCACAA

TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA
GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
GAGAGGCGGTTTGCGTATTGGGCCAAAGACAAAAGGGCGACATTCAACCGA
TTGAGGGAGGGAAGGTAAATATTGACGGAAATTATTCATTAAAGGTGAATT
ATCACCGTCACCGACTTGAGCCATTTGGGAATTAGAGCCAGCAAAATCACC
AGTAGCACCATTACCATTAGCAAGGCCGGAAACGTCACCAATGAAACCATC
GATAGCAGCACCGTAATCAGTAGCGACAGAATCAAGTTTGCCTTTAGCGTC
AGACTGTAGCGCGTTTTCATCGGCATTTTCGGTCATAGCCCCCTTATTAGCG
TTTGCCATCTTTTCATAATCAAAATCACCGGAACCAGAGCCACCACCGGAAC
CGCCTCCCTCAGAGCCGCCACCCTCAGAACCGCCACCCTCAGAGCCACCACC
CTCAGAGCCGCCACCAGAACCACCACCAGAGCCGCCGCCAGCATTGACAGG
AGGCCCGATCTAGTAACATAGATGACACCGCGCGCGATAATTTATCCTAGT
TTGCGCGCTATATTTTGTTTTCTATCGCGTATTAAATGTATAATTGCGGGAC
TCTAATCATAAAAACCCATCTCATAAATAACGTCATGCATTACATGTTAATT
ATTACATGCTTAACGTAATTCAACAGAAATTATATGATAATCATCGCAAGAC
CGGCAACAGGATTCAATCTTAAGAAACTTTATTGCCAAATGTTTGAACGATC
GGGGATCATCCGGGTCTGTGGCGGGAACTCCACGAAAATATCCGAACGCAG
CAAGATATCGCGGTGCATCTCGGTCTTGCCTGGGCAGTCGCCGCCGACGCC
GTTGATGTGGACGCCGGGCCCGATCATATTGTCGCTCAGGATCGTGGCGTT
GTGCTTGTCGGCCGTTGCTGTCGTAATGATATCGGCACCTTCGACCGCCTGT
TCCGCAGAGATCCCGTGGGCGAAGAACTCCAGCATGAGATCCCCGCGCTGG
AGGATCATCCAGCCGGCGTCCCGGAAAACGATTCCGAAGCCCAACCTTTCA
TAGAAGGCGGCGGTGGAATCGAAATCTCGTGATGGCAGGTTGGGCGTCGCT
TGGTCGGTCATTTCGAACCCCAGAGTCCCGCTCAGAAGAACTCGTCAAGAA
GGCGATAGAAGGCGATGCGCTGCGAATCGanacntancgnNTGGAGCGGCGATA
CCGTAAAGCACGAGGAAGCGGTCAGCCCATTCGCCGCCAAGCTCTTCAGCA
ATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACACCCAGC
CGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTC
GGCAAGCAGGCATCGCCATGGGTCACGACGAGATCATCGCCGTCGGGCATG
CGCGCCTTGAGCCTGGCGAACAGTTCGGCTGGCGCGAGCCCCTGATGCTCT
TCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTC
GCTCGATGCGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAA
GCGTATGCAGCCGCCGCATTGCATCAGCCATGATGGATACTTTCTCGGCAG
GAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCCCAATAGCA
GCCAGTCCCTTCCCGCTTCAGTGACAACGTCGAGCACAGCTGCGCAAGGAA
CGCCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCCTGCAGTTCATT
CAGGGCACCGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGCCCTGCGC
TGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCCA
GTCATAGCCGAATAGCCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAA
TCCATCTTGTTCAATCATGCGAAACGATCCAGATCCGGTGCAGATTATTTGG
anacntancgnNTATTGAGAGTGAATATGAGACTCTAATTGGATACCGAGGGGAA
TTTATGGAACGTCAGTGGAGCATTTTTGACAAGAAATATTTGCTAGCTGATA
GTGACCTTAGGCGACTTTTGAACGCGCAATAATGGTTTCTGACGTATGTGC
TTAGCTCATTAAACTCCAGAAACCCGCGGCTGAGTGGCTCCTTCAACGTTGC
GGTTCTGTCAGTTCCAAACGTAAAACGGCTTGTCCCGCGTCATCGGCGGGG
GTCATAACGTGACTCCCTTAATTCTCCGCTCATGATCAGATTGTCGTTTCCC
GCCTTCAGTTTAAACTATCAGTGTTTGACAGGATATATTGGCGGGTAAACCT
AAGAGAAAAGAGCGTTTATTAGAATAATCGGATATTTAAAAGGGCGTGAAA
AGGTTTATCCGTTCGTCCATT<u>TGTATGTGCATGCCAACCACAGG</u>GTTCCCCAG
ATCTGGCGCCGGCCAGCGAGACGAtgtgcttGCAAGATTGGCCGCCGCCCGAAACGATCCG

ACAGCGCGCCCAGCACAGGTGCGCAGGCAAATTGCACCAACGCATACAGCGCCAGCAGA
ATGCCATAGTGGGCGGTGACGTCGTTCGAGTGAACCAGATCGCGCAGGAGGCCCGGCA
GCACCGGCATAATCAGGCCGATGCCGACAGCGTCGAGCGCGACAGTGCTCAGAATTAC
GATCAGGGGTATGTTGGGTTTCACGTCTGGCCTCCGGACCAGCCTCCGCTGGTCCGATT
GAACGCGCGGATTCTTTATCACTGATAAGTTGGTGGACATATTATGTTTATCAGTGATA
AAGTGTCAAGCATGACAAAGTTGCAGCCGAATACAGTGATCCGTGCCGCCCTGGACCTG
TTGAACGAGGTCGGCGTAGACGGTCTGACGACACGCAAACTGGCGGAACGGTTGGGGG
TTCAGCAGCCGGCGCTTTACTGGCACTTCAGGAACAAGCGGGCGCTGCTCGACGCACTG
GCCGAAGCCATGCTGGCGGAGAATCATACGCATTCGGTGCCGAGAGCCGACGACGACT
GGCGCTCATTTCTGATCGGGAATGCCCGCAGCTTCAGGCAGGCGCTGCTCGCCTACCGC
GATGGCGCGCGCATCCATGCCGGCACGCGACCGGGCGCACCGCAGATGGAAACGGCCG
ACGCGCAGCTTCGCTTCCTCTGCGAGGCGGGTTTTTCGGCCGGGGACGCCGTCAATGCG
CTGATGACAATCAGCTACTTCACTGTTGGGGCCGTGCTTGAGGAGCAGGCCGGCGACA
GCGATGCCGGCGAGCGCGGCGGCACCGTTGAACAGGCTCCGCTCTCGCCGCTGTTGCG
GGCCGCGATAGACGCCTTCGACGAAGCCGGTCCGGACGCAGCGTTCGAGCAGGGACTC
GCGGTGATTGTCGATGGATTGGCGAAAAGGAGGCTCGTTGTCAGGAACGTTGAAGGAC
CGAGAAAGGGTGACGATTGATCAGGACCGCTGCCGGAGCGCAACCCACTCACTACAGC
AGAGCCATGTAGACAACATCCCCTCCCCTTTCCACCGCGTCAGACGCCCGTAGCAGCC
CGCTACGGGCTTTTTCATGCCCTGCCCTAGCGTCCAAGCCTCACGGCCGCGCTCGGCCT
CTCTGGCGGCCTTCTGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTC
GTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA
ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAAC
CGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCA
CAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGG
ATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCCGCTGCATAACCCTGC
TTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCTTTTTCGCACGATATACAGG
ATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCA
GGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCG
CACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGT
AACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCCA
CCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCGG
CGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACG
GGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCC
GCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTCGGT
GATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCA
AGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAA
CGGCCGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGC
GACTTCGCGGAGCTGGTGAAGTACATCACCGACGAGCAAGGCAAGACCGAGCGCCTTT
GCGACGCTCA

Acknowledgements

At first, I would like to deeply thank Prof. Dr. Bernd Weisshaar for the opportunity to work on this wonderful PhD thesis under his doctorate supervision.

My sincere gratitude goes to my second supervisor Dr. Ralf Stracke, for his support and his subject-specific advice. His prior work on BGLU6 and further interest in GH1-type GTs introduced me to this absorbing topic of plant metabolism and flavonoid glycosylation. In the last three years I had the opportunity to dive deep into the world of glucose and flavonoids, enhancing deeply my knowledge in this field.

Also, with deep gratitude, I thank Dr. Rabea Schweiger for this wonderful and fruitful cooperation and for sharing her profound knowledge about metabolite fingerprinting and compound identification. I enjoyed every time, working with her on the identification of the candidate substrates. Especially, her optimism and enthusiasm were very supportive during all stages of the PhD process.

Another time with deep gratitude I thank Melanie Kuhlmann who supported me during my whole work with an outstanding technical knowledge and an outstanding generous, patient and joyful willingness to help in any situation.

A special thank goes to Lennart Malte Sielmann for his enormous help with the complementation lines, for the last-minute genotyping of double mutants and also his ongoing co-expression analysis. His dedication coupled to good mood is highly contagious.
A big big special thank goes to Andrea Voigt for her outstanding help in the green house. Without Andrea, sample growth for HPLC and NMR measurement would not have been possible. Also Dr. Boas Pucker provided precious support for my PhD thesis on the one hand for the MinION sequencing, on the other hand for many fruitful discussions. Additionally, I thank the sequencing core facility of the CeBiTec, especially Prisca Viehöver for the reliable generation of sequencing data. Special thanks to the Organic and Bioorganic Chemistry department and especially to Dr. Gervais Happi Mouthe for supporting me with the compound purification for NMR measurements. Furthermore, I deeply thank Christopher Acatay for his support in the harvesting and freeze-drying marathon as well as lots of genotyping PCRs. This thank deserves also Janina Mercedes Grabowski.
A heartfelt thanks is also provided to my wonderful PhD mates Mareike Busche, Jens Theine, Bianca Frommer and Hanna Schilbert, making joyful and less joyful times more delightful. Besides, I sincerely give thanks to all members of the Genetics and Genomics of Plants group for the good time and enjoyable working atmosphere.

Finally, I would like to thank my beloved family, friends and Giampiero supporting me until the last minute behind the scene.