

Aus dem Institut für Pflanzenbau und Pflanzenzüchtung  
der Christian-Albrechts-Universität zu Kiel

**Functional analysis and mutagenesis of glucosinolate synthesis genes for  
breeding oilseed rape (*Brassica napus*) with lower glucosinolate content**

Dissertation  
zur Erlangung des Doktorgrades  
der Agrar- und Ernährungswissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von  
M.Sc. Srijan Jhingan  
aus Delhi, Indien  
Kiel, 2021

---

Dekan: Prof. Dr. Karl H. Mühling

1. Berichterstatter: Prof. Dr. Christian Jung
2. Berichterstatter: Prof. Dr. Daguang Cai

Tag der mündlichen Prüfung: 16.02.2022



## Table of Contents

1	General introduction.....	1
1.1	Oilseed rape – significance, cultivation and breeding.....	1
1.2	Current scenario of ‘omics’ resources for research in oilseed rape .....	2
1.3	Glucosinolates in Brassicaceae .....	3
1.3.1	The Biosynthesis of glucosinolates .....	3
1.3.2	Regulation of biosynthesis and transport of GSLs .....	8
1.3.3	Role of glucosinolates in plant defense mechanism.....	9
1.3.4	GSLs: a major deterrent in rapeseed meal quality .....	9
1.3.5	Breeding oilseed rape with lower glucosinolate content .....	10
1.4	Increasing genetic variation by mutagenesis .....	11
1.4.1	Random mutagenesis .....	11
1.4.2	Targeted mutagenesis: precise genome editing.....	12
1.5	Scientific hypotheses, questions and aims.....	13
2	Direct access to millions of mutations by whole-genome sequencing of an oilseed rape mutant population.....	14
2.1	Introduction.....	14
2.2	Materials and methods.....	15
2.2.1	Plant material.....	15
2.2.2	DNA isolation and whole-genome sequencing.....	15
2.2.3	Data processing and SNP detection.....	15
2.2.4	Estimation of mutation frequencies.....	16
2.2.5	Developing a web-based mutant database .....	16
2.3	Results .....	16
2.3.1	Whole-genome sequencing reveals high mutation density.....	16
2.3.2	Detecting functional mutations .....	20
2.3.3	Sequence analyses reveal patterns of mutation frequency and distribution .....	21
2.3.4	Validation of functional mutations.....	24
2.3.5	A web-based interface for screening the mutant population.....	25
2.4	Discussion.....	25
2.5	Supplementary data .....	29
2.5.1	Supplementary figures .....	29
2.5.2	Supplementary tables.....	30
2.6	References.....	31

3	Reducing glucosinolate content in oilseed rape ( <i>Brassica napus</i> ) by random mutagenesis in biosynthesis genes <i>BnMYB28</i> and <i>BnCYP79F1</i> .....	36
3.1	Introduction.....	36
3.2	Materials and methods.....	38
3.2.1	Plant material and growth conditions .....	38
3.2.2	<i>In silico</i> analyses .....	39
3.2.3	Expression analysis of candidate genes by RT-qPCR.....	39
3.2.4	DNA isolation and genotyping experiments.....	39
3.2.5	Conventional gel-based detection of EMS-induced mutations.....	39
3.2.6	Genotyping with the <i>Brassica</i> 19K SNP array .....	40
3.2.7	Glucosinolate determination .....	41
3.2.8	Statistical analyses.....	41
3.3	Results .....	42
3.3.1	Identification of <i>MYB28</i> and <i>CYP79F1</i> genes in oilseed rape.....	42
3.3.2	Expression profiles of <i>BnMYB28</i> and <i>BnCYP79F1</i> genes reveal putative functional paralogs .....	42
3.3.3	EMS-induced mutations detected in the selected <i>BnMYB28</i> and <i>BnCYP79F1</i> paralogs.....	45
3.3.4	The <i>Brassica</i> 19K SNP array for marker-assisted background selection.....	47
3.3.5	Seeds of <i>BnMYB28</i> and <i>BnCYP79F1</i> double mutants possess significantly reduced aliphatic GSLs.....	48
3.4	Discussion.....	54
3.5	Supplementary data .....	59
3.5.1	Supplementary figures .....	59
3.5.2	Supplementary tables.....	60
3.6	References.....	61
4	Functional analysis of the glucosinolate transporter gene <i>BnGTR2</i> in rapeseed.....	66
4.1	Introduction.....	66
4.2	Materials and methods.....	67
4.2.1	<i>In silico</i> analysis .....	67
4.2.2	Expression analysis of candidate genes by RT-qPCR.....	68
4.2.3	DNA isolation and genotyping experiments.....	68
4.2.4	Conventional gel-based screening of EMS-induced mutations .....	68
4.2.5	Plasmid vector construction for CRISPR-Cas9 mutagenesis .....	68
4.2.6	Plant material and growth conditions .....	69
4.2.7	Statistical analysis .....	70



4.3 Results .....	70
4.3.1 Seven paralogs of <i>BnGTR2</i> were identified in oilseed rape .....	70
4.3.2 Expression analyses reveal differentially expressed <i>BnGTR2</i> paralogs in leaves and seeds .....	72
4.3.3 EMS-induced mutagenesis of <i>BnGTR2.A06</i> and <i>BnGTR2.C03</i> .....	73
4.3.4 Genotyping BC <sub>1</sub> plants using the <i>Brassica</i> 19K SNP array .....	75
4.3.5 Development of final CRISPR-Cas9 constructs for the targeted mutagenesis of <i>BnGTR2</i> .....	76
4.4 Discussion.....	77
4.5 Supplementary data .....	81
4.5.1 Supplementary figures .....	81
4.5.2 Supplementary tables.....	82
4.6 References.....	82
5 Closing discussion.....	87
5.1 The TILLING by whole-genome sequencing platform is an unprecedented resource for functional genomics and breeding in rapeseed.....	87
5.2 Selecting candidate genes controlling glucosinolate metabolism, a complex quantitative trait in Brassicaceae.....	89
5.3 <i>BnMYB28</i> and <i>BnCYP79F1</i> loss of function mutants possess low aliphatic GSL content in the seeds.....	91
5.4 Introgression of the recurrent Peace parent masked mutation effects in backcrossed mutants.....	92
5.5 Controlling glucosinolate transport could dictate rapeseed breeding .....	93
6 Summary .....	96
7 Zusammenfassung .....	98
8 Appendix.....	100
8.1 Supplementary tables.....	100
8.2 Supplementary figures.....	116
9 References.....	143
10 Supplementary data on CD/DVD.....	161
11 Curriculum Vitae and Publications .....	162
11.1 Curriculum vitae.....	162
11.2 Oral and poster presentations.....	162
12 Publications and declaration of own contribution.....	164
13 Acknowledgements .....	165

## List of Abbreviations

%	Percent
°C	Celcius
µg	Microgram
µl	Microliter
µmol	Micromole
2-D	2-dimensional
3'	3 prime
4x	Fourfold
5'	5 prime
A	Adenine
<i>A. thaliana, At</i>	<i>Arabidopsis thaliana</i>
<i>A. tumefaciens</i>	<i>Agrobacterium tumefaciens</i>
ANOVA	Analysis of Variance
<i>B. juncea, Bj</i>	<i>Brassica juncea</i>
<i>B. napus, Bn</i>	<i>Brassica napus</i>
<i>B. oleracea, Bo</i>	<i>Brassica oleracea</i>
<i>B. rapa, Br</i>	<i>Brassica rapa</i>
bam	Binary Alignment Map
BC	Backcross
BLAST	Basic Local Alignment Search Tool
<i>GAPDH</i>	Glyceraldehyde 3-phosphate dehydrogenase
bp	Base pair
BRAD	The Brassica Database
BWA-MEM	Burrows-Wheeler Aligner-Maximal Exact Match
C	Cytosine
Cas	CRISPR-associated
cDNA	Complementary DNA
cm	Centimeter
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
CSV	Comma-Separated Values
C <sub>t</sub>	Cycle threshold
CTAB	Cetyltrimethyl ammonium bromide
DAP	Days after pollination
ddH <sub>2</sub> O	Deionized distilled water
DH	Doubled haploid
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleotide triphosphate
DW	Dry weight
<i>E. coli</i>	<i>Escherichia coli</i>
EMS	Ethyl methanesulfonate

## List of Abbreviations

F <sub>n</sub>	n <sup>th</sup> generation after cross
g	Gram
G	Guanine
GATK	Genome Analysis Tool Kit
Gb	Gigabase pairs
GC	Guanine-Cytosine
gDNA	Genomic DNA
GFF	General Feature Format
GOPOD	Glucose oxidase/oxidase
GSLs	Glucosinolates
<i>GTR</i>	<i>GLUCOSINOLATE TRANSPORTER</i>
GWAS	Genome-Wide Association Study
h	Hours
HPLC	High performance liquid chromatography
<i>IND</i>	<i>INDEHISCENT</i>
InDel	Insertions and deletions
IR	Infrared
kb	Kilobase pairs
LB	Luria-Bertani
M	Molar
Mb	Megabase pairs
mg	Milligram
min	Minutes
mm	Millimeter
mM	Millimolar
M <sub>n</sub>	n <sup>th</sup> generation after mutagenesis
MNU	N-nitroso- <i>N</i> -methylurea
MQ	Mapping Quality
ng	Nanogram
NGS	Next-generation sequencing
ORF	Open reading frame
PAA	Polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PCA	Principal component analysis
PCR	Polymerase chain reaction
PE	Paired-end
QTL	Quantitative trait loci
RNA	Ribonucleic Acid
rpm	Rotations per minute
RSM	Rapeseed meal
RT-qPCR	Real-time quantitative PCR
s	Seconds

## List of Abbreviations

sam	Sequence Alignment Map
SEM	Standard error of mean
<i>SFAR</i>	<i>SEED FATTY ACID REDUCER</i>
SNP	Single Nucleotide Polymorphism
SOC	Seed oil content
T	Thymine
TAIR	The Arabidopsis Information Resource
TALEN	Transcription activator-like effector nuclease
TASSEL	Trait Analysis by aSSociation, Evolution and Linkage
TILLING	Targeted Induced Local Lesions in Genomes
tracrRNA	Trans-activating crRNA
UTR	Untranslated region
V	Volt
vcf	Variant Call Format
VEP	Variant Effect Predictor
W	Watt
WGS	Whole-genome sequencing
ZFN	Zinc finger nuclease

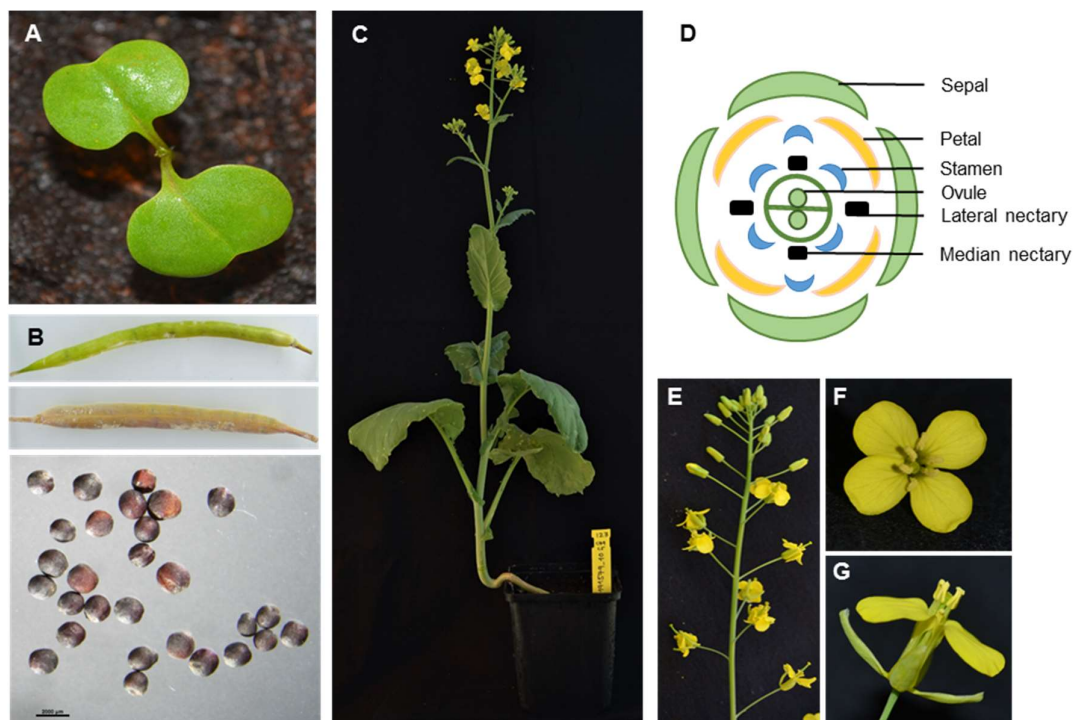
# 1 General introduction

## 1.1 Oilseed rape – significance, cultivation and breeding

Oilseed rape or rapeseed (*Brassica napus* L.) is a major oil crop in the temperate and sub-tropical regions of the world. The species belongs to the family of crucifers (Brassicaceae) and is the third-largest seed oil and second-largest protein meal source worldwide. In 2018, ~37.6 million hectares under cultivation produced 75 million tonnes of oilseed rape globally (<http://www.fao.org/faostat/>). In Europe, oilseed rape yields the highest oil per unit area compared to any other oil crop. The seeds contain 45-50% edible oil that is fit for human consumption due to a balanced and nutritious fatty acid profile (Russo et al., 2021). Moreover, it is also used for biodiesel production in Europe. After oil extraction, the leftover rapeseed meal (RSM) is conventionally utilized as animal feed due to its high protein content comprising 320-450 g/kg of dry matter (Bu et al., 2018).

*B. napus* is an allotetraploid (AACC,  $2n = 38$ ) that is a product of spontaneous interspecific hybridization between the diploid AA ( $2n = 20$ ) and CC ( $2n = 18$ ) genomes of turnip rape (*Brassica rapa* L., syn. *campestris*) and cabbage (*Brassica oleracea* L.), respectively (Chalhoub et al., 2014). It is speculated to have originated in the Middle Ages in the Mediterranean region since its wild diploid progenitors also originated here (Hu et al., 2007; Chalhoub et al., 2014). The model plant *Arabidopsis thaliana* is believed to be the common ancestor of *Brassica* species (Friedt & Snowden, 2009). Swedes or *B. napus* ssp. *napobrassica* and *B. napus* ssp. *napus* constituting the oil crop rapeseed, represent the two major sub-species of *B. napus*. In terms of growth cycles, oilseed rape is annual or biennial that is represented by three major ecotypes (i) winter-type, (ii) spring-type and (iii) semi-winter types, each with distinct vernalization requirements (Ferreira et al., 1995; Tadege et al., 2001). The winter-type oilseed rape is pre-dominantly grown in the temperate regions of Europe, where exposure to cold winters induces floral competence (Diepenbrock, 2000). Plants develop an erect stem and branched structure with glabrous waxy leaves (Figure 1A and 1C). The inflorescence is racemous, with acropetally developing flowers (Figure 1E). Flowers (Figure 1D-G) have four sepals, four characteristic pale to bright yellow-colored petals, a single superior ovary and six stamens (Gulden et al., 2008). After fertilization, the flowers develop into elongated, cylindrical pods called siliques that contain seeds, which upon maturity, develop a dark brown to black color (Figure 1B).

Oilseed rape has undergone a short evolutionary and domestication history (Rahman, 2013; Wang et al., 2014). Since the genetic diversity within *B. napus* is narrow, breeders have traditionally relied on pedigree breeding, doubled haploid (DH) production and used resynthesized oilseed rape and interspecific hybridization (Hu et al., 2007; Girke et al., 2012). Oilseed rape varieties with several improved agronomic traits such as yield, seed oil content and oil quality have been introduced using these strategies (Becker et al., 1995; Butruille et al., 1999; Rahman, 2001; Seyis et al., 2006; Xu et al., 2007; Zou et al., 2010). Presently, most modern rapeseed varieties are hybrids that have been developed via male sterility systems (Friedt et al., 2018). Major breeding targets include increasing yield and yield potential combined with winter hardiness and reducing anti-nutritive compounds characteristic to Brassicales. In terms of the oil quality, fatty acid profiles rich in oleic acid and low in linolenic acid are also targeted (Liersch et al., 2013). However, with the growing concern of climate change, breeding varieties resilient to abiotic and biotic stresses and better adapted to changing agricultural regimes are gaining significance.



**Figure 1:** Representative images of the oilseed rape plant and its parts. A) germinating plant ten days after sowing, B) siliques and mature seeds C) flowering plant D) floral diagram of *Brassica* species, and E-G) oilseed rape inflorescence.

## 1.2 Current scenario of ‘omics’ resources for research in oilseed rape

‘Omics’ based studies have always served as a strong foundation in plant and crop research. Genomics, transcriptomics, proteomics and metabolomics serve as core omics technologies that offer an integrated and comprehensive understanding of complex biological processes like cellular metabolic systems (Bino et al., 2004; Oksman-Caldentey & Saito, 2005; Saito & Matsuda, 2010), plant-pathogen or plant-microbe interactions (Kaul et al., 2016), biosynthesis of secondary metabolites (Schauer & Fernie, 2006) and response to biotic and abiotic stresses (Takeda & Matsuoka, 2008; Mosa et al., 2017). The application of multi-omics approaches in plant research can be restricted due to limited genomic resources like poorly annotated genomes, vast metabolic diversity and complicated interaction networks (Jamil et al., 2020). However, technological advancements in the past decades for whole-genome sequencing, RNA-Seq and metabolite profiling have aided plant research immensely (Mochida & Shinozaki, 2011; Fukushima et al., 2014).

With the advent of next-generation sequencing (NGS) technologies combined with their cost-effectiveness, high-quality and reliable whole-genome assemblies are now publically available from multiple plant species. The Arabidopsis Information Resource (TAIR)-<https://www.arabidopsis.org/> (Swarbreck et al., 2007) is the most extensive ‘omics’ database for the model plant *Arabidopsis thaliana*. The Brassica Database (BRAD) serves as a comprehensive web-based genomic resource for Brassicaceae plants (Cheng et al., 2011). The first whole-genome reference assembly for rapeseed was developed using the ‘Darmor-bzh’ genotype (Chalhoub et al., 2014). The assembly had an effective genome size of ~1.13 Gigabases (Gb) and was annotated with 101,040 predicted gene models. More recently, several other high-quality whole-genome assemblies (Lee et al., 2020; Rousseau-Gueutin et al., 2020; Song et al., 2020; Chen et al., 2021) have enabled convenient detection and

sequence analysis of candidate orthologs in *B. napus*. Noteworthy, the development of the ‘Genomic Variation Database of Rapeseed: BnaGVD’ now allows the analysis of genomic variation across a global collection of oilseed rape accessions via open access to large sequencing datasets (Yan et al., 2021). Furthermore, the ‘*Brassica* Expression Database (BrassicaEDB)’ is an online repository of comprehensive transcriptome data from the semi-winter rapeseed ‘ZS11’ encompassing multiple tissues and growth stages accessible via a user-friendly graphical interface (Chao et al., 2020). Such resources have significantly supported and enhanced the power of functional genomics and molecular breeding in oilseed rape.

### 1.3 Glucosinolates in Brassicaceae

#### 1.3.1 The Biosynthesis of glucosinolates

Glucosinolates (GSLs) are a diverse group of amino acid-derived, sulfur and nitrogen-containing heterogeneous secondary metabolites specific to Brassicaceae. Structurally, they consist of a thioglucose and a sulfonated aldoxime attached to a chain elongated amino acid (Halkier & Gershenzon, 2006). Depending upon amino acid precursors, GSLs are categorized as aliphatic, aromatic and indolic, originating primarily from methionine, phenylalanine/tyrosine and tryptophan, respectively (Halkier & Du, 1997). Roughly 130 distinct GSL types have been reported from 16 dicot angiosperms, most of which are edible plant types (Fahey et al., 2001; Blažević et al., 2020; Mitreiter & Gigolashvili, 2021). The model plant *Arabidopsis* alone has been detected with more than 30 GSLs (Wittstock & Burow, 2010). Economically relevant vegetable crops of *B. oleracea*: cauliflower, cabbage, broccoli, Brussels sprouts, kale, *B. rapa*: turnips and radish and the oilseed crop *B. napus* are prominent examples of GSL rich plant species (Cartea & Velasco, 2008; Verkerk et al., 2009; Sønderby et al., 2010). The GSL content in oilseed rape seeds can reach as high as 60-100  $\mu\text{mol/g}$  dry weight, of which aliphatic GSLs constitute the majority (~92%) of total GSLs (Magrath et al., 1993; Velasco et al., 2008; Yasumoto et al., 2010).

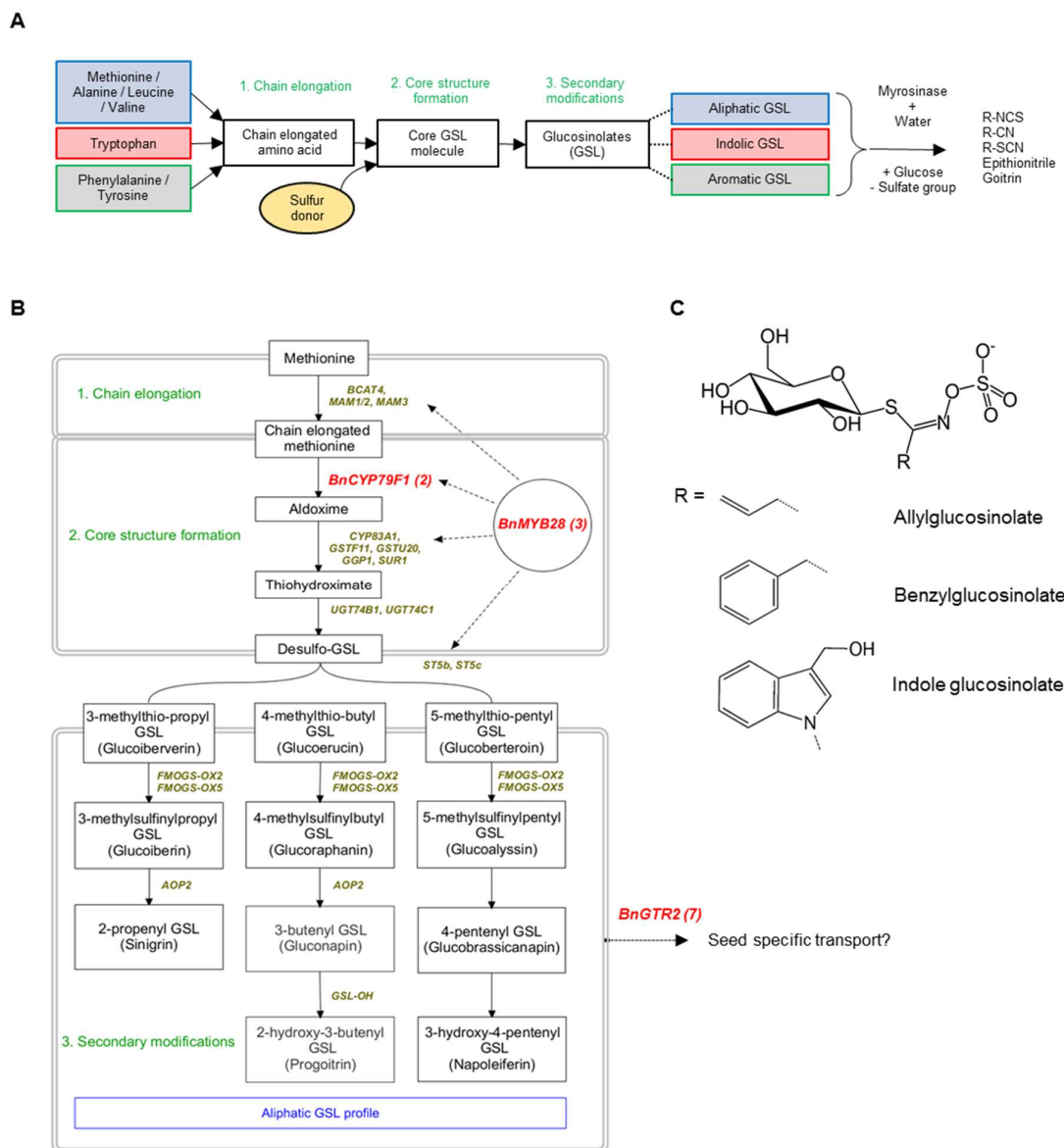
GSL biosynthesis is completed via three major biochemical steps (i) chain elongation, (ii) core structure formation and (iii) secondary side-chain modifications. The first step involves the elongation of the amino acids via the insertion of methylene ( $-\text{CH}_2-$ ) groups. The second step is marked by the addition of the sulfur group to the chain elongated amino acid. And finally, secondary modifications in each chain elongated GSL type result in distinct GSL types. The GSL metabolic pathway is complicated and diverse, where multiple gene families independently control each GSL profile (Mitreiter & Gigolashvili, 2021). Side-chain elongation of amino acid precursors is catalyzed by a branched-chain amino acid aminotransferase (BCAT), yielding the corresponding 2-oxo acid, the first major step in the synthesis of GSLs. A methylthioalkylmalate synthase (MAM) (Textor et al., 2007) then mediates the association with acetyl-CoA in singular or repeated cycles followed by oxidative isomerization and oxidative decarboxylation. An isopropylmalate isomerase (IPMI) and an isopropylmalate dehydrogenase (IPM-DH) control these processes. Consequently, the group of 2-oxo acids with variable lengths of consecutively attached  $-\text{CH}_2-$  groups then proceeds for the core structure formation. Several genes have been investigated for their control over this step. The CYP family of cytochromes P450 controls the next step. For aliphatic GSLs, CYP79F1 and CYP79F2 convert chain elongated methionine to corresponding aldoximes (Reintanz et al., 2001; Chen et al., 2003). Aldoximes are then activated to nitrile oxides or aci-nitro compounds by CYP83A1 before attachment with glutathione (GSH), the sulfur donor. The attachment of GSH is controlled by glutathione S-transferases GSTF11 and GSTU20. Subsequently, thiohydroximates are S-glucosylated by glucosyltransferases of the

UGT74 family yielding desulfo-GSLs. Noteworthy, transcriptomic analyses have revealed that almost all genes involved in methionine chain elongation and core structure pathways are co-expressed (Hirai et al., 2007). The true extent of GSL diversity and complexity is realized in the third and final step of biosynthesis. Several genes involved in the secondary modification of GSLs have been characterized. Distinct *FMO<sub>GS-OX</sub>* genes from the family of flavin monooxygenases are responsible for the *S*-oxygenation of desulfo-GSLs with variable chain lengths. *AOP2* and *AOP3* mediate a major branching point for aliphatic GSLs by converting *S*-oxygenated GSLs to alkenyl or hydroxyalkenyl GSLs. Different desulfo-GSLs (based on side-chain lengths) can finally give rise to multiple corresponding GSL types depending upon the distinct secondary modifications. Figure 2 shows a simplified overview of the GSL biosynthesis pathway based on the knowledge from Arabidopsis.

Numerous studies have shown that the concentration and composition of GSLs greatly vary not just across species but also within the plant in a tissue-specific manner. Each GSL profile has a distinct network of transcription factors and enzymes controlling their biosynthesis (Halkier & Gershenzon, 2006). Moreover, their broad diversity and complexity have been shaped by evolutionary pressures that influence the synthesis and distribution of single GSL types across different *Brassica* species (Halkier & Gershenzon, 2006; Nour-Eldin et al., 2017). Especially for oilseed rape, varieties from tropical regions have been reported to contain a higher GSL content than temperate varieties (Tripathi & Mishra, 2007). A single species may contain multiple GSL types varying in proportion and distribution within the plant (Fahey et al., 2001; Nour-Eldin & Halkier, 2009; Verkerk et al., 2009; Sønderby et al., 2010). An interplay of regulatory genes, plant hormones like jasmonate and auxin, epigenetic and environmental interactions act as cues to determine GSL content and profile across different ecotypes (Mitreiter & Gigolashvili, 2021).

Speculations on the putative function and activity of candidate orthologs in *Brassica* species are primarily sourced from the close relative and model plant Arabidopsis (Halkier & Du, 1997; Sønderby et al., 2010). Functional characterization of GSL candidate genes has been reported in multiple studies on *B. rapa* (Wiesner et al., 2014; Liu et al., 2017; Nour-Eldin et al., 2017) and *B. oleraceae* (Yin et al., 2017; Sánchez-Pujante et al., 2018; Zuluaga et al., 2019; Neequaye et al., 2021). However, knowledge transfer is somewhat limited and complicated due to the presence of multiple homoeologs in rapeseed. Associative transcriptomics, genome-wide association and QTL mapping studies have discovered many loci significantly associated with GSL content in oilseed rape (Harper et al., 2012; Qu et al., 2015; Kittipol et al., 2019; Lu et al., 2019; Wei et al., 2019; Y. Liu et al., 2020). However, comprehensive functional analyses of GSL genes have been limited so far. Table 1 summarizes major candidate genes involved in the biosynthesis, regulation and transportation of aliphatic GSLs reported and functionally characterized in Arabidopsis.





**Figure 2:** Schematic representation of the biosynthesis pathway of glucosinolates (GSLs). A) General overview and B) detailed pathway for aliphatic GSL biosynthesis. GSL biosynthesis is completed via three major biochemical steps 1) chain elongation, 2) core structure formation, and 3) secondary modifications (written in light green). GSL types originating from methionine as the amino acid precursor constitute the aliphatic GSL profile. Names of genes (based on Arabidopsis orthologs) mediating respective processes are labeled in dark green. Candidate genes considered in this study are labeled in red with the corresponding number of *B. napus* paralogs mentioned in parenthesis. Speculated gene functions are marked with dotted arrowheads. Pathways are constructed using Pathvisio 3.3.0. C) General chemical structure of glucosinolates. Glucosinolates consist of  $\beta$ -D-glucopyranose residues linked via a sulfur atom to a N-hydroximiniosulfate ester and a variable R group - modified from Halkier and Gershenzon (2006). Examples of general GSL structures are shown for aliphatic, aromatic and indolic functional groups (R).

**Table 1:** Summary of major candidate genes and their functional analyses reported from *Arabidopsis thaliana* involved in the biosynthesis, transport and regulation of aliphatic glucosinolates. All listed genes have been reported from multiple GWAS, QTL and transcriptome studies and found to be significantly associated with glucosinolate content in oilseed rape.

<i>A. thaliana</i> gene name <sup>[a]</sup>	Location of gene activity <sup>[1]</sup>		Gene function	Functional analysis by knock-out <sup>[1]</sup>	Adverse phenotypes observed in knock-out mutants (if any) <sup>[1]</sup>	No. of <i>B. napus</i> paralogs <sup>[2]</sup>	References
	Cellular level	Organ level					
<i>AtBCAT4</i> (AT3G19710)	Chloroplast and cytosol	Mature leaves, siliques	Chain elongation, transaminase activity	Reduction in methionine-derived GSLs, increased leucine-derived GSLs. Increased free amino acid availability; Methionine was maximum.	Smaller rosette leaves and inflorescence, chlorotic leaves	4	Schuster et al. (2006)
<i>AtMAM1</i> (AT5G23010)	Chloroplast and cytosol	Mature leaves, siliques	Methionine elongation to C <sub>3</sub> -C <sub>8</sub> precursors	Decreased levels of several amino acid substrates required for glucosinolate biosynthesis, no visible effect on morphology or growth rate.	-	2	Kroymann et al. (2001)
<i>AtBAT5</i> (AT4G12030)	Mitochondria and chloroplast	Mature leaves, hypocotyl	Imports 2-oxo acids into the chloroplast	~ 50% reduction in aliphatic glucosinolates, <i>MYB28</i> transcripts reduced	-	2	Gigolashvili et al. (2009)
<i>AtAPK1</i> (AT2G14750)	Chloroplast	Guard cells, vascular tissues, pollen, seeds	Secondary modifications: mediate an important branching point in sulfur assimilatory pathway	<i>apk1</i> and <i>apk2</i> double mutants: Increased auxin levels, 30% reduced indolic GSLs, aliphatic almost undetectable. No GSL reduction in single mutants.	Double mutants are semi-dwarf, 70% reduced fresh weight, delayed flowering,	4	Mugford et al. (2009)
<i>AtAPK2</i> (AT4G39940)	Chloroplast	Vascular tissues, seeds				6	
<i>AtGTR1</i> (AT3G47960)	Golgi apparatus, nuclear and plasma membrane, cytosol	Phloem cap in the inflorescence stem	Phloem loading, GSL transporter (stress induced response)	No effect in seed loading, GSL transport within leaves affected	-	3	Nour-Eldin et al. (2012)

<i>AtGTR2</i> (AT5G62680)	Plasma membrane	Siliques and inflorescence	suspected), transmembrane transport	Significantly reduced glucosinolates in seeds, increase in leaves	-	7	
<i>AtMYB28/HAG1</i> (AT5G61420)	Nucleus	Vegetative parts: mature leaves, trichomes	Transcription factors controlling the activity of most biosynthesis-related genes in the core molecule formation of aliphatic GSLs	1) Decreased transcript level of all genes in the core molecule formation. 2) Increased accumulation of aliphatic GSL in leaves but reaches a certain threshold in seeds after overexpression.	Overexpression caused growth retardation due to low methionine levels (precursor for ethylene)	3	Gigolashvili et al. (2007)
<i>AtMYB29/HAG3</i> (AT5G07690)	Nucleus	Cauline leaves	A coordinated mechanism between <i>HAG1</i> and <i>HAG3</i> speculated	<i>myb28myb29</i> double mutants show significant aliphatic GSL reduction (~50%)	-	3	Gigolashvili et al. (2008)
<i>AtCYP79F1</i> (AT1G16410)	Chloroplast, ER and nucleus	Rosette leaves, guard cells, absent in seeds	Biosynthesis (Aliphatic GSL)	Reduced short chained aliphatic GSL, a bushy mutant lacking apical dominance observed in Arabidopsis	Increased meristem numbers, crinkled leaves, bushy mutants lacking apical dominance, retarded vascularization in knock-out mutants	3	Reintanz et al. (2001)
<i>AtCYP83A1</i> (AT4G13770)	Cytosol	Mature leaves, hypocotyl	Biosynthesis (Aliphatic GSL)	Increased auxin levels, seeds with less benzoyloxyalkyl glucosinolates, increase in indolic GSL inconsistent	-	2	Hemm et al. (2003)

[1] Based on studies from Arabidopsis.

[2] Sequence analysis of orthologs and paralogs in oilseed rape are based on gene models described in the Darmar-*bzh* reference genome (Genoscope).

[a] *A. thaliana* gene names have been retrieved from The Arabidopsis Information Resource (TAIR). TAIR gene IDs are mentioned within parenthesis.

### 1.3.2 Regulation of biosynthesis and transport of GSLs

In Brassicaceae, the biosynthesis and regulation of GSLs are a major driving force influencing plant fitness. A complex network of transcription factors (TFs) interacting with environmental, both abiotic and biotic stimuli, hormonal and epigenetic factors act as the basis of GSL biosynthesis regulation (Celenza et al., 2005; Gigolashvili et al., 2007; Hirai et al., 2007; Gigolashvili et al., 2009). While numerous studies have investigated the enzymatic biosynthesis of GSLs extensively (Kliebenstein et al., 2001; Textor et al., 2007; Seo & Kim, 2017), the underlying regulatory mechanisms are far more complicated. MYB TFs constitute one of the largest families of TFs involved in a broad spectrum of primary and secondary plant biochemical processes. The R2R3-MYB TFs represent the largest subfamily (Seo & Kim, 2017). The subgroup 12 R2R3-MYB TFs have been reported as the key players in GSL biosynthesis, specifically the core molecule formation, six of which confer up-regulation in *Arabidopsis* (Gigolashvili et al., 2007; Gigolashvili et al., 2008; Frerigmann et al., 2015). Based on extensive studies in *Arabidopsis*, many MYB orthologs have been identified in other Brassicaceae crops of *B. oleracea* (Araki et al., 2013; Zuluaga et al., 2019), *B. rapa* (Kim et al., 2013; Seo et al., 2016) and *B. napus* (Lu et al., 2014; Qu et al., 2015; Seo & Kim, 2017; Wang et al., 2018; Kittipol et al., 2019). While MYB34, MYB51 and MYB122 regulate indolic GSL synthesis, aliphatic GSL synthesis is controlled by MYB28, MYB76 and MYB29, also referred to as *HIGH ALIPHATIC GLUCOSINOLATE (HAG) 1*, 2 and 3, respectively. Since these six R2R3-MYB TFs are structurally conserved, their functions are partially redundant (Mitreiter & Gigolashvili, 2021). However, MYB28 and MYB29 are crucial for aliphatic biosynthesis and MYB51 and MYB34 for indolic biosynthesis (Gigolashvili et al., 2007; Mitreiter & Gigolashvili, 2021). Additionally, a crosstalk mechanism between these TFs has also been investigated, contributing towards a balance of the two GSL profiles within the plant (Hirai et al., 2007; Gigolashvili et al., 2008; Frerigmann et al., 2015). Functional analyses have revealed that MYB28 can regulate the production of both short and long-chained GSLs, while MYB29 exclusively controls the production of short-chained GSLs (Gigolashvili et al., 2007; Sønderby et al., 2010). MYB28 (*HAG1*) and MYB29 (*HAG3*) have been demonstrated as co-expressed with multiple genes involved in aliphatic biosynthesis. Furthermore, the role of these two TFs is speculated to have a ‘master regulator’ like effect on the entire aliphatic GSL biosynthesis pathway (Hirai et al., 2007).

While it is known that GSLs are synthesized in the vegetative parts of the plant (Toroser et al., 1995; Chen et al., 2001), limited information is available about the GSL transport mechanism within and across tissues. Since GSLs are anionic thioglucosides, they cannot be transported across biological membranes via simple diffusion (Nour-Eldin & Halkier, 2009; Wittstock & Burow, 2010). Traditionally, transport mechanisms of defense compounds have been investigated via correlations with expression patterns of biosynthesis genes (Jørgensen, Nour-Eldin, et al., 2015). However, in pioneering works aimed at elucidating the source to sink translocation network between the maternal tissues and the seeds, Nour-Eldin et al. (2012) have demonstrated the role of NRT/PTR transporters in the proton-coupled seed-specific transport of GSLs in *Arabidopsis*. Two GSL transporters (*GTR*), namely *AtGTR1* and *AtGTR2*, were characterized for the long-distance transport of GSLs via the phloem without any substrate specificities for aliphatic, aromatic or indolic GSLs. Moreover, variable expression profiles for the two *GTR* genes were observed in a tissue specific manner in *Arabidopsis* (Andersen et al., 2013; 2014). This suggested a major role of GSL transporters in controlling the distribution of GSLs in space and time. More recently, *AtGTR1* and *AtGTR2* orthologs have been investigated in Brassica crops (Nour-Eldin et al., 2017). In *B. juncea* and *B. napus*, *BjGTR1* and *BjGTR2* (Nambiar et al., 2021) and *BnGTR1* and *BnGTR2* (Lu et al.,

2014; Kittipol et al., 2019), respectively, have been reported to be closely associated with seed-specific GSL content.

Agronomic and environmental factors are equally important in influencing GSL metabolism (Neugart et al., 2020). Yan and Chen (2007) have reviewed the possible factors affecting the *in vivo* degradation of GSLs. Factors like plant age, diurnal regulation of myrosinase, plant responses to temperature, light, soil fertility and sulfur availability can modulate the interaction of GSLs and myrosinase. In addition to the interplay of TFs and transporters, most plant hormones involved in the perception of abiotic and biotic stress can also promote or impede GSL synthesis. While jasmonates, abscisic acid and ethylene promote GSL synthesis (Augustine & Bisht, 2015; Chen et al., 2019), brassinosteroids inhibit GSL synthesis (Guo et al., 2013), and salicylic acid and gibberellins can do both (Kiddle et al., 1994; Schreiner et al., 2011; Mitreiter & Gigolashvili, 2021). Auxins, specifically indole-3-acetic acid has been reported with an indirect negative effect on GSL synthesis since it shares a major portion of its biosynthesis pathway with that of indolic GSLs (Malka & Cheng, 2017).

### 1.3.3 Role of glucosinolates in plant defense mechanism

Unlike phytoalexins, GSLs are anticipins which upon enzymatic cleavage by the endogenous thioglucoside glucohydrolase (also called myrosinase), yield toxic secondary by-products. The metabolic effect of GSLs, whether anti-nutritive or beneficial, is primarily dictated by the amino acid side chain structure, which further determines the corresponding products of hydrolysis (Hopkins et al., 2009). After the hydrolytic cleavage of the thioglucosidic bond by myrosinase, an aglucone is released. This aglucone is then converted to several products like isothiocyanates (ITC), thiocyanates (SCN), epithionitriles and nitriles (NI) (Figure 2A), many of which are known to confer defense against generalist herbivores and fungal and bacterial pathogens (Rask et al., 2000; Wittstock & Burow, 2010; Burow & Halkier, 2017).

The GSL-myrosinase induced defense response in Brassicaceae is a well-investigated mechanism (Thangstad et al., 1990; James & Rossiter, 1991; Halkier & Du, 1997; Kelly et al., 1998; Andréasson et al., 2001). In terms of subcellular localization, the myrosinase is localized in separate compartments called ‘myrosin cells’ isolated from the GSLs stored in ‘S-cells’ (Kissen et al., 2009; Wittstock & Burow, 2010). Upon wounding and tissue damage induced during herbivory, the hydrolysis of GSLs to secondary products ensues when the barriers between the substrates and the activating enzymes are broken (Mithen, 1992; Borgen et al., 2010; Winde & Wittstock, 2011). Moreover, the feeding mode can further influence the pattern and intensity of the defense response depending on the severity of tissue damage. In addition to the total GSL content, profile variations affect different specialist and generalist insect herbivores and pollinators differently (Burow & Halkier, 2017; Mitreiter & Gigolashvili, 2021). It is reported that the metabolism of GSLs in specialists significantly varies from that in generalists, with some specialist lepidopterans able to divert the course of GSL hydrolysis towards nitriles rather than the toxic ITCs (Jeschke et al., 2017). On the other hand, some GSLs also act as stimulants for specialist herbivores, specifically during feeding and oviposition (Halkier & Gershenzon, 2006; Hopkins et al., 2009).

### 1.3.4 GSLs: a major deterrent in rapeseed meal quality

Oilseed rape is the second major protein meal source after soybean meal (Hansen et al., 2020). With a protein content of up to 35-40%, its profile enriched with sulfur-rich amino acids makes rapeseed meal (RSM) a sustainable source of animal feed. However, the presence of anti-nutritive compounds like phenols, fibers, tannins, phytate, and GSLs can

severely undermine the utility of RSM (Mawson et al., 1993; Wittkop et al., 2009). GSLs are hydrophilic compounds and are therefore left behind in the RSM after oil extraction. After hydrolysis, GSLs yield diverse products that can adversely affect the quality of the RSM. When consumed in higher doses, dietary GSLs can cause retarded growth, reduced appetite and feed efficiency, goiter of the thyroid gland, gastrointestinal irritation, liver and kidney damage and reduced palatability due to pungency and bitter taste (Bourdon & Aumaître, 1990; Vageeshbabu & Chopra, 1997; Hopkins et al., 2009). Although mechanical treatments like heating and toasting seeds after harvest have been shown to reduce GSLs, they are not economical. Moreover, heating can reduce the digestibility and palatability of the RSM (Bourdon & Aumaître, 1990).

Aliphatic GSLs are methionine-derived thioglucosides (Mitreiter & Gigolashvili, 2021). Although leaves show more variation in GSL types, aliphatic GSLs constitute 91-94% of the total GSLs in seeds (Padilla et al., 2007; Velasco et al., 2008). Hydrolysis products of aliphatic GSLs include SCN, ITC, oxazolidinethiones and nitriles. While hydroxyalkenyl GSLs like epiprogoitrin and progoitrin are goitrogenic (Fahey et al., 2001), other aliphatic GSLs like gluconapin and sinigrin compromise on the palatability of the meal (Vageeshbabu & Chopra, 1997). Tripathi and Mishra (2007) have reviewed the effect of GSLs in animal nutrition. High SCN and nitrile content in the meal can interfere with iodine uptake and disrupt kidney and liver function, respectively. Upon ingestion in higher doses, adverse metabolic and behavioral effects have been observed in animals like fish (Kaiser et al., 2021), birds, rats and rabbits (Fenwick & Curtis, 1980; Vermorel et al., 1986; Wight et al., 1987; Tripathi et al., 2008), and higher mammals like pigs and cattle (Fenwick et al., 1983; Bourdon & Aumaître, 1990; Bischoff, 2021). Non-ruminants have been observed to be far more susceptible to anti-nutritive effects, pigs being more sensitive in comparison to fish and poultry (Bourdon & Aumaître, 1990; Tripathi & Mishra, 2007).

### 1.3.5 Breeding oilseed rape with lower glucosinolate content

In the past decades, agronomic traits like yield, seed oil content and RSM quality have significantly improved in oilseed rape. Previously, rapeseed oil was utilized as an industrial lubricant. The RSM usage as feed was limited due to the high amounts of anti-nutritive compounds that severely restricted its consumption for humans and animals. A major milestone in oilseed rape breeding was the development of ‘double low’ or ‘double zero (00)’ rapeseed varieties (also called ‘canola’). Originating from the Polish and German spring varieties ‘Bronowski’ and ‘Liho’, they were selected specifically for a low glucosinolate and low erucic acid content, respectively (Friedt et al., 2018). Introgression of these seed traits into spring and winter ecotypes has been done following extensive backcrossing programs. In 1981, the winter-type ‘Librador’ was released as the first German double zero-rapeseed variety (Hatzig et al., 2018). Breakthroughs and consistent efforts in oilseed rape breeding have significantly reduced GSL content in RSM and erucic acid in oil from 60–100  $\mu\text{mol/g}$  dry weight and 50% respectively, to what is now the benchmark for double 00-rapeseed varieties (Wittkop et al., 2009). Currently, double 00-rapeseed varieties possess <18-25  $\mu\text{mol/g}$  dry weight GSLs in the RSM and <2% erucic acid in the seed oil profile. This enabled the dramatic increase in the quality and acceptability of the oil and the protein-rich seed meal for use in the food industry, making it one of Germany’s most important vegetable oils (Matthäus & Brühl, 2003).

The winter and spring ecotypes of oilseed rape are genetically distinct (Lu et al., 2019). Therefore, introgression of traits across the two oilseed rape ecotypes has been done to introduce novel traits for improving seed yield and quality (Rahman, 2013). During oilseed

rape breeding, an intensive selection towards double zero varieties has gradually narrowed the genetic variability in modern-day cultivars (Hasan et al., 2008; Chalhoub et al., 2014). It has been observed that selection for improved seed quality in winter oilseed rape has a negative effect on seed germination (Hatzig et al., 2018). Possible pleiotropic effects and linkage drag during the inheritance of favorable alleles for reduced erucic acid and GSL content and increased seed yield are speculated to play a role here. Interestingly, based on QTL analyses, Quijada et al. (2006) have also reported that a winter-type allele conferring increased seed yield was linked to a QTL conferring high GSL content. This suggests that while breeding winter rapeseed with lower GSL content, favorable alleles contributing towards increased seed yield could have been lost. Residual portions of the original ‘Bronowski’ genotype in modern-day double zero varieties are speculated to have caused reductions in oil content, seed yield and winter hardiness (Hasan et al., 2008). Nevertheless, over the past decades, breeding rapeseed for lower GSL content has resulted in significant lowering of the GSLs in the seeds while at the same time ensuring sufficient yield and desirable oil traits ([https://www.bundessortenamt.de/apps6/bsa\\_bsl/public/de](https://www.bundessortenamt.de/apps6/bsa_bsl/public/de)).

## 1.4 Increasing genetic variation by mutagenesis

While heredity and variation are the fundamentals of biological evolution, mutations can shape this course of evolution by increasing the genetic variation in living organisms (Li et al., 2020; Lu et al., 2021). These include various DNA sequence changes spanning single point mutations (transitions and transversions) and small indels to large-scale structural changes like chromosomal inversions, duplications, deletions or translocations (Shelake et al., 2019). Artificially induced mutations are important in functional genomics and crop improvement (Fanelli et al., 2021). More specifically, mutations play a significant role in the reverse genetics approach where artificial induction of mutations can help understanding gene functions by modifying already known genes (Sessions et al., 2002) and possibly introducing novel and beneficial agronomic traits in crops. Several techniques have been applied to induce random (Henikoff et al., 2004; Till et al., 2006) and targeted mutations (Carroll, 2011; Shan et al., 2013; Fauser et al., 2014; Ma et al., 2015) in plants.

### 1.4.1 Random mutagenesis

Traditionally, high-frequency ionizing radiation (X-rays, gamma rays and fast neutrons) were used as physical agents for artificially inducing random mutations (Tanaka et al., 2010; Ulukapi & Nasircilar, 2015). However, such physical treatments may have severe modifying effects on the genome of the treated organism. Chemical mutagens like EMS (ethyl methanesulfonate) or MNU (*N*-nitroso-*N*-methylurea) are known to cause point mutations at high rates (Henikoff et al., 2004; Harloff et al., 2012) and have milder modifying effects in comparison to physical treatments. Owing to ease of handling and detoxification, EMS is a particularly favored choice of chemical mutagen in crop research (Shelake et al., 2019). Moreover, EMS treatments can be applied on larger scales, often generating thousands of mutants with heritable point mutations. Due to an alkylating effect, EMS treatments ethylate guanine (G) residues resulting in pre-dominantly G→A and C→T transitions (Till et al., 2006). These randomly induced point mutations can have a range of putative functional effects on genes in the form of premature nonsense, splice site or missense mutations, possibly conferring a loss of gene function. Targeting Induced Local Lesions in Genomes (TILLING) is an extensively used reverse genetics approach for screening random mutations in genomes induced by an EMS treatment. Screening strategies for the identification of mutants span across traditional methods like gel-based detection (Till et al., 2006; Harloff et al., 2012; Braatz et al., 2018; Karunarathna et al., 2020) to more modern methods employing

NGS based amplicon sequencing or exome capture (Tsai et al., 2011; Gilchrist et al., 2013; Krasileva et al., 2017; Sashidhar et al., 2019).

Presently, several EMS mutant resources for oilseed are publically available (Wang et al., 2008; Harloff et al., 2012; Gilchrist et al., 2013; Wells et al., 2014; Lee et al., 2018; Tang et al., 2020). The utilization of EMS mutants in breeding programs can be limited since EMS mutants possess a significant load of random and uncharacterized background mutations. This can manifest as undesirable agronomic traits in the mutagenized organism and subsequent progenies and thus severely undermines the potential of this technique. Therefore, backcrossing mutants with the EMS donor line or with elite cultivars is required before application. Nevertheless, their utilization continues to occupy a significant position in reverse genetics and functional genomics for crop improvement. This is because EMS mutagenesis is a non-transgenic approach for mutant production, so the application of mutants for breeding is currently within the legal framework of the European Union (Jung & Till, 2021).

### 1.4.2 Targeted mutagenesis: precise genome editing

Several tools facilitating targeted genome modifications are currently available. Such techniques mostly rely on the introduction of double-stranded breaks (DSBs) in the intended region, thereby triggering endogenous DNA repair mechanisms like non-homologous end-joining (NHEJ) or homology directed repair (HDR) in the targeted organism (Puchta, 2005). In this process, an error-prone repair mechanism may result in *de novo* mutations in the form of point mutations or large indels. Moreover, a replacement of entire genomic regions can be done via HDR that completes repairs using homologous genomic segments as a template within the targeted genome. These induced mutations can confer knock-outs or even chromosomal rearrangements. Currently, three major site-specific nucleases are known (i) zinc finger nucleases (ZFNs), (ii) transcription activator-like effectors (TALEs) and (iii) clustered regularly interspaced short palindromic repeats (CRISPR)-Cas-mediated RNA-guided DNA endonucleases (Chen & Gao, 2014). Application of ZFNs (Kim et al., 1996; Smith et al., 2000; Bibikova et al., 2002) was subsequently replaced with the more recent TALEs (Boch et al., 2009). TALEs have aided in the targeted mutagenesis in many crops like rice (Li et al., 2012; Shan et al., 2013), sugar cane (Kannan et al., 2018), potato (Yasumoto et al., 2019), wheat (Wang et al., 2017; Luo et al., 2019), barley (Takáč et al., 2021) and oilseed rape (Kazama et al., 2019).

Since the application of ZFNs and TALEs in design and execution is laborious, CRISPR-Cas-based genome modification technologies have gained substantial momentum. This is mainly due to the simplicity, specificity, flexibility and broad-spectrum application of the technique across multiple organisms (Chen & Gao, 2014; Carroll, 2021). One of the most extensively used CRISPR-Cas system in crop research is the type II CRISPR-Cas system (Ma et al., 2015; Arora & Narula, 2017; Braatz et al., 2017; Sashidhar et al., 2019; Karunarathna et al., 2020; Saikia et al., 2020; Luo et al., 2021) that employs the SpCas9 endonuclease isolated from the bacterium *Streptococcus pyogenes* (Doudna & Charpentier, 2014). This system utilizes a single guide RNA (sgRNA) constituted by the CRISPR RNA (crRNA) and the trans-activating crRNA called the tracrRNA (Jinek et al., 2012; Mali et al., 2013) that control the site-specific binding of the Cas9 endonuclease. In the CRISPR-Cas9 system, the crRNA consists of a 20 bp sequence (called the protospacer) for target specificity. Adjacent to the protospacer is the protospacer adjacent motif (PAM), a 3 bp 5'-NGG-3' motif required for target site binding.



Employment of different Cas endonucleases (Huang & Puchta, 2021) and further modifications to the functional components of existing CRISPR-Cas systems are now enabling improved efficiency, precision and increased flexibility of application (Fu et al., 2014; Renaud et al., 2016; Long et al., 2018; Bai et al., 2020; Moradpour & Abdulah, 2020). However, the full utilization of transgenic approaches for genome editing via the CRISPR-Cas system is currently legally restricted in many parts of the world, including the European Union (Jung & Till, 2021). Due to this, the application of genome-edited plants in breeding programs is virtually inexistent and limited to research in the EU. In light of these legal bottlenecks, the utilization of EMS mutagenized individuals presently acts as a viable long-term alternative.

## 1.5 Scientific hypotheses, questions and aims

The presence of aliphatic GSLs is a serious disadvantage to the nutritive value of rapeseed meal. As a part of the larger project ‘Improved Rapeseed as Fish Feed in Aquaculture’, the broader objective is to develop rapeseed meal as a sustainable feed source for farmed fish. In this work, I initially raised the following questions:

1. How many strongly expressed GSL biosynthesis or transporter genes can be identified as suitable candidates for functional analysis by conventional TILLING?
2. Do the candidate genes display any tissue-specific expression profiles in oilseed rape?
3. Is it possible to reduce the seed glucosinolate content without impairing the leaf glucosinolate profile?

I hypothesize that disruption of the biosynthesis genes *BnMYB28* and *BnCYP79F1* will reduce the aliphatic glucosinolate content in the whole plant. However, disruption of the GSL transporter gene, *BnGTR2* reduces glucosinolate content in a tissue-specific manner in *B. napus* seeds.

The objectives of my work are:

1. Establishing a novel TILLING by whole-genome sequencing (TbyWGS) platform to screen EMS-induced mutations on a whole-genome scale using a winter-type oilseed rape EMS population.
2. Identification of putative functional mutations for GSL biosynthesis and transporter genes in the EMS mutant population by a conventional gel-based approach.
3. Combine single mutants by crossing and backcrossing to increase phenotypic effects and reduce background mutation load in EMS mutants.
4. CRISPR-Cas9 approach to knock-out multiple paralogs of the glucosinolate transporter gene *BnGTR2*.
5. Phenotypic characterization of obtained mutants.

## 2 Direct access to millions of mutations by whole-genome sequencing of an oilseed rape mutant population

### 2.1 Introduction

Oilseed rape or rapeseed (*Brassica napus* L.) is the primary oil crop in the temperate regions and the third-largest seed oil and second-largest protein meal source worldwide (Wang et al., 2018). In 2019, ~34 million hectares yielded 70 million tons of oilseed rape globally (<http://www.fao.org/faostat/>). Seed oil content ranges between 45-50%. The oil is consumed as edible oil and for biodiesel production. Moreover, byproducts after oil extraction are conventionally utilized as animal feed. The primary aims of rapeseed breeding are to increase yield potential and stability in combination with improved seed quality. Rapeseed varieties across the world are allocated to three ecotypes. Winter-type rapeseed is primarily grown in Europe, spring-types are grown in Australia and Canada, whereas the semi-winter types are grown in China.

Oilseed rape is an allotetraploid (AACC,  $2n = 38$ ) belonging to the Brassicaceae family. It is a result of spontaneous interspecific hybridization between the diploid AA ( $2n = 20$ ) and CC ( $2n = 18$ ) genomes of turnip rape (*Brassica rapa* L., syn. *campestris*) and cabbage (*Brassica oleracea* L.), respectively. Its origin was traced back to the Mediterranean region ~7,500 years ago and the total genome size is estimated to be ~1.13 Gb (Gigabases) (Chalhoub et al., 2014). Today, several high-quality rapeseed reference genomes are available. The ‘Darmor-*bzh*’ was the first reference genome assembled using a European winter-type oilseed rape. The genome encompassed 314.2 Mb of the A sub-genome and 525.8 Mb of the C sub-genome, predicted with 101,040 gene models (Chalhoub et al., 2014). More recently, an improved long-read assembly of the ‘Darmor-*bzh*’ reference genome (Rousseau-Gueutin et al., 2020) was published. Meanwhile, ten more high-quality whole-genome assemblies of oilseed rape have been published. Song et al. (2020) sequenced eight accessions representing the three ecotypes. Winter-type and a semi-winter type oilseed rape were published by (Lee et al., 2020) and Chen et al. (2021), respectively. Genome sizes assembled in these reference genomes ranged between 921-1,033 Mb and the number of predicted gene models ranged between 89,895 and 106,059.

Due to its short history of evolution and domestication, the genetic diversity within *B. napus* is low (Rahman, 2013). Ethyl methanesulfonate (EMS) induced mutagenesis has been used to create novel allelic variation in rapeseed (Wang et al., 2008; Harloff et al., 2012; Gilchrist et al., 2013; Wells et al., 2014; Lee et al., 2018; Tang et al., 2020). In the past years, CRISPR-Cas technology to create targeted mutations has been successfully applied, and numerous mutants have been developed (Braatz et al., 2017; Karunarathna et al., 2020; Sashidhar et al., 2020; Zheng et al., 2020). Although targeted mutagenesis offers several advantages, random mutagenesis still has importance in rapeseed breeding mainly because CRISPR-Cas mutants are legally classified as genetically modified organisms in the European Union. Therefore their usage in practical breeding is presently limited (Jung & Till, 2021).

EMS mutant discovery in oilseed rape has conventionally relied on the activity of DNA mismatch specific endonucleases and polyacrylamide gel-based detection assays classically termed TILLING (Targeting Induced Local Lesions in Genomes). Mutants are screened using pooled genomic DNA from M<sub>2</sub> individuals of large mutant populations (Emrani et al., 2015; Braatz et al., 2018; Shah et al., 2018; Karunarathna et al., 2020). Because this procedure is time-consuming and laborious, amplicon sequencing-based detection methods (Gilchrist et

al., 2013; Wells et al., 2014; Sashidhar et al., 2019) gained increasing popularity owing to their efficiency and sensitivity. However, like conventional TILLING, detection of EMS mutations via amplicon sequencing approaches is restricted to single gene families with high sequence conservation.

Sequencing whole mutant populations, termed TILLING by sequencing (TbySeq), is the gold standard of mutant detection (Jung & Till, 2021). In a pioneering study, Krasileva et al. (2017) demonstrated a TILLING by exome sequencing approach to detect EMS-induced mutations in tetraploid and hexaploid wheat. The TbySeq approach has also been reported from several crops like rice (Abe et al., 2012), maize (Nie et al., 2021), tomato (Garcia et al., 2016), soybean (Lakhssassi et al., 2021), sunflower (Fanelli et al., 2021) and cotton (Fang et al., 2020). In a first TILLING by whole-genome sequencing (TbyWGS) approach for oilseed rape, a limited number of EMS mutants were whole-genome sequenced to detect EMS-induced mutations (Tang et al., 2020).

This study aimed to develop a bioinformatic resource for detecting EMS-induced mutations on a genome-wide scale. We sequenced the whole genomes of 1,988 M<sub>2</sub> plants from an EMS mutagenized winter oilseed rape population (Harloff et al., 2012). A TILLING by whole-genome sequencing (TbyWGS) pipeline was established, which allows the identification of mutations within any genomic region of interest. The sequences can be screened via a web-based resource. Thus, our TbyWGS platform constitutes a long-lasting repository of mutants that can be screened cost-effectively to detect functional mutations that possibly confer beneficial traits in oilseed rape.

## **2.2 Materials and methods**

### **2.2.1 Plant material**

The offspring of the winter oilseed rape inbred line Express617 (F<sub>11</sub>) had been treated with EMS and M<sub>2</sub> and M<sub>3</sub> generations had been produced (Harloff et al., 2012). 1,988 M<sub>2</sub> plants representing 497 M<sub>2</sub> families (4 plants/family) were grown under greenhouse conditions (22°C, 16 h light). Plants were vernalized for 12 weeks at 4°C. All M<sub>2</sub> plants were self-pollinated and M<sub>3</sub> seeds were harvested. Plant material was produced and propagated by NPZ Innovation GmbH (Holtsee, Germany).

### **2.2.2 DNA isolation and whole-genome sequencing**

Twenty 4 mm disks were taken from leaves of each of the M<sub>2</sub> plants before vernalization. DNA was isolated from bulked leaf samples of each M<sub>2</sub> family by NPZ Innovation GmbH (Holtsee, Germany). The pooled leaf samples were lyophilized (ALPHA 1-4 LDplus, Martin Christ Gefriertrocknungsanlagen GmbH, Germany) for 96 h. Genomic DNA of 497 4x DNA pools was isolated using the DNeasy Plant Mini Kit following the manufacturer's protocol (QIAGEN GmbH, Germany) and sequenced on an Illumina NovaSeq 6000 platform using 150bp paired ends reads and a 350bp insert DNA library with 20x depth of coverage (Novogene, Co., Ltd., China).

### **2.2.3 Data processing and SNP detection**

Raw data were obtained in the FASTQ format. The quality of raw reads was checked using the FastQC v0.11.5 (Andrews, 2010) and MultiQC (Ewels et al., 2016) programs. Based on the quality checks, samples representing low sequencing outputs or reads possessing poor Phred scores, deviating GC count per read, and a high number of ambiguously called bases

were removed from the analysis. A long-read genome assembly of Express617 was used as the reference genome (Lee et al., 2020). Raw reads from all 4x pools were mapped to the Express617 oilseed rape reference, using BWA-MEM (Li, 2013) with default parameters for local alignment. SAMtools was used to convert, sort, and index the resulting files to binary alignment map files. The average depth of coverage per pool was measured using the SAMtools -depth tool.

Using Picard tools AddOrReplaceReadGroups and MarkDuplicates, mapped reads were pre-processed before variant calling. Variants were called using the GATK 4.1.4.1 HaplotypeCaller program (Poplin et al., 2018). The resulting Variant Call Format (vcf) files were filtered for single nucleotide polymorphisms (SNP) characteristic to the EMS type C→T and G→A transitions (named ‘canonical’ transitions). SNPs were retained based on read depth (DP)  $\geq 10$ , mapping quality (MQ)  $\geq 30$ , and allele depth (AD) between 12.5%-60% parameters using bcftools and custom UNIX scripts. Allele depth of mutations was calculated as the percentage share of reads possessing mutations from total reads (read depth, DP) mapped at that position. These SNPs are hereafter referred to as ‘high confidence’ SNPs. The same criteria were used to create a subset of variants representing all other nucleotide substitutions referred to as ‘non-canonical’ transitions and transversions. The Ensembl Variant Effect Predictor (VEP) tool (McLaren et al., 2016) was used in the offline mode to characterize and predict SNP effects on the polypeptide level. The General Feature Format (GFF) file for the Express617 reference assembly (Lee et al., 2020) representing predicted protein models was used to characterize all SNP effects on a genome-wide scale.

## 2.2.4 Estimation of mutation frequencies

To approximate the mutation frequency per kb (kilobases) of the reference genome, the number of high confidence SNPs along the chromosomal length for each annotated chromosome (A01-A10 and C01-C09) and the non-annotated share of the reference genome was calculated. A genome-wide mutation density was calculated by dividing the number of mutations by the effective genome size (~925 Mb). Within 1 Mb (megabases) non-overlapping windows, the frequency of C→T and G→A transitions and all other transitions and transversions were then calculated from each of the 4x pools for all chromosomes. A moving average for the number of SNPs across all sequenced pools per 1 Mb non-overlapping windows was calculated for all chromosomes. The General Additive Model (GAM) in the ggplot2 package on R (Wickham, 2009) was used to smoothen the curves while plotting values. The number of putative functional mutations conferred by C→T and G→A transitions and all other mutations were then calculated.

## 2.2.5 Developing a web-based mutant database

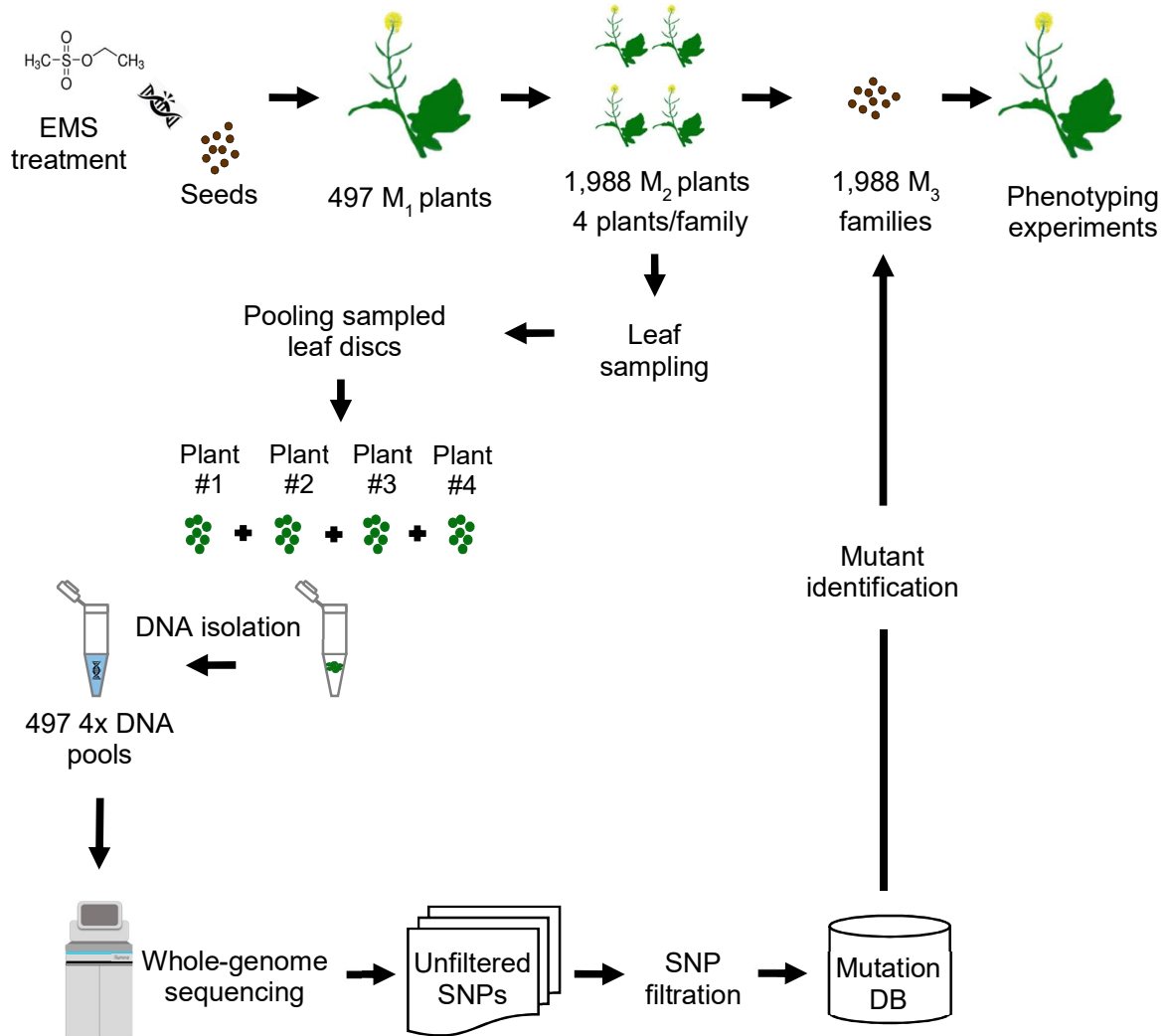
The web-accessible database was built using Django (<https://www.djangoproject.com/>), a Python-based free and open-source web framework. PostgreSQL (<http://www.postgresql.org>), an open-source object-relational database system, was used as the database engine. Webpages were built using HTML and Bootstrap (<https://getbootstrap.com/>).

## 2.3 Results

### 2.3.1 Whole-genome sequencing reveals high mutation density

To gather the first results on sequencing quality and mutant detection, we first performed pilot experiments with fifty 4x pools sequenced at 10x coverage by mapping the raw reads to

the Express617 reference genome. The effective genome size of this reference assembly was 925,095,059 bp (Lee et al., 2020). After SNP detection and filtration, many SNPs were filtered into the ‘high confidence’ category but later could not be validated via Sanger sequencing (data not shown). Therefore, we decided to double the intended coverage to 20x for sequencing the 4x pools and sequenced pooled DNA from 1,988 M<sub>2</sub> plants on the Illumina NovaSeq 6000 platform (Figure 3). The raw dataset encompassing a total of 497 4x pools represented an average of 35.8 gigabases (Gb) per pool. On average, ~116 million raw reads per pool were generated. In terms of read quality, Phred scores for all paired-end reads varied well above the optimum scores (Supplementary Figure 1). Mapping raw reads to the Express617 reference genome resulted in an average coverage depth of 34.7x.



**Figure 3:** Workflow of the mutant detection pipeline. M<sub>2</sub> seeds were harvested from 497 M<sub>1</sub> plants and four plants/M<sub>2</sub> family were grown. Leaf material was sampled from 1,988 M<sub>2</sub> plants as leaf discs. Leaves from each family were pooled and DNA was isolated and sequenced on the Illumina NovaSeq 6000 platform. Raw reads were mapped to the Express617 reference genome. SNPs were called and filtered using GATK 4.1.4.1 and custom UNIX scripts. Raw SNPs were retained based on ‘canonical’ EMS mutations (C→T and G→A transitions) possessing read depth (DP) ≥10 allele depth (AD) =12.5%-60% and quality controlled for mapping quality (MQ) ≥30. Mutation effects were predicted using Ensembl Variant Effect Predictor (VEP) tool. High confidence mutations from each pool were merged into a single accessible database. After identification of desirable M<sub>2</sub> mutants, corresponding M<sub>3</sub> plants can be analyzed to verify for phenotypic effects.

Then, we calculated the number and frequency of SNPs. Variant calling revealed ~843,000 unfiltered SNPs per pool on average. Using our SNP filtering criteria (read depth  $\geq 10$ , allele depth = 12.5%-60% and mapping quality  $\geq 30$ ), on average, 156,480 SNPs per pool were C→T and G→A transitions, characteristic for EMS-induced mutations (Table 2).

**Table 2:** Summary statistics of SNPs detected after whole-genome sequencing (including regions without chromosomal annotations) of 1,988 M<sub>2</sub> rapeseed plants assembled in 497 4x DNA pools.

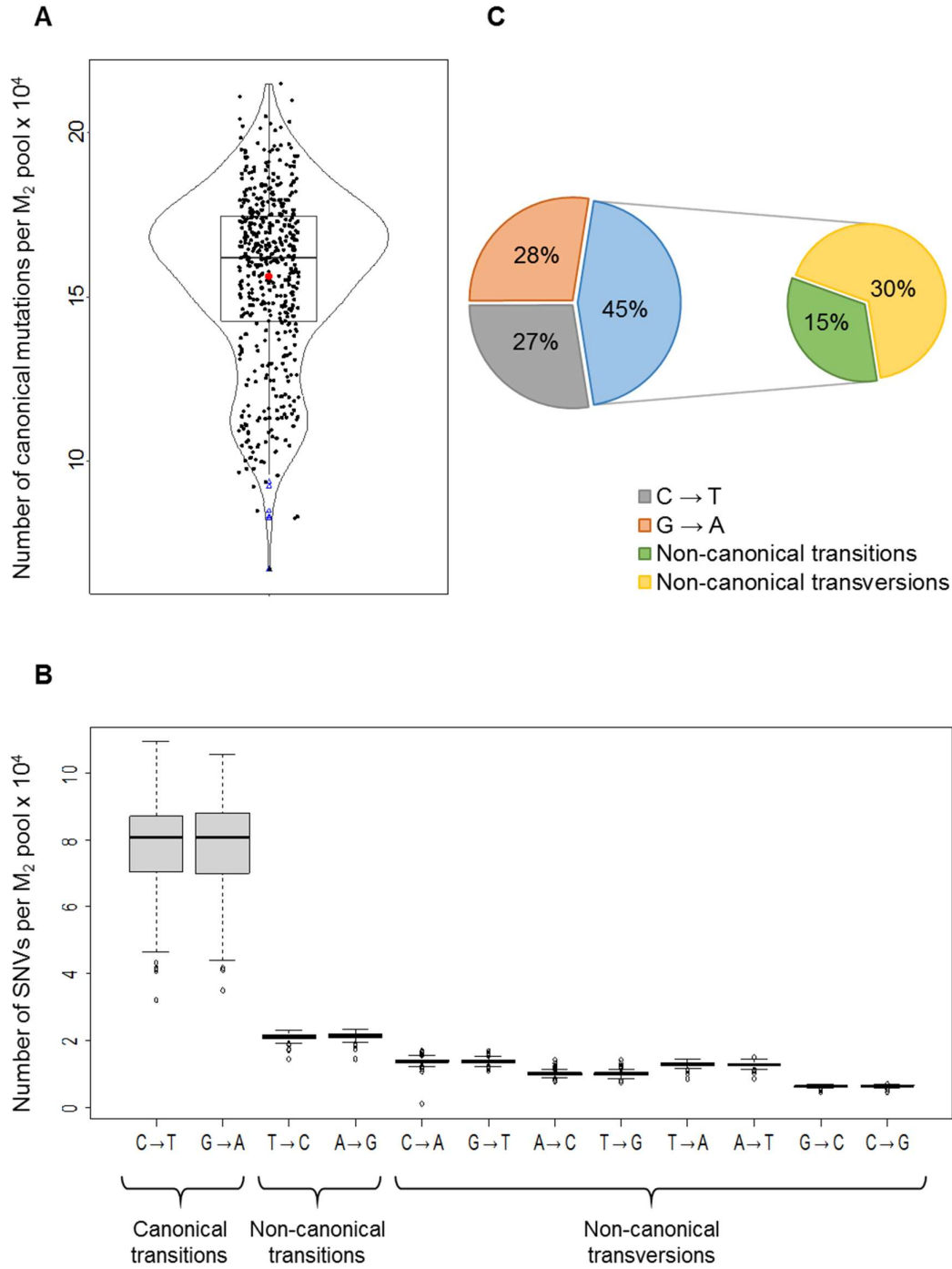
	SNP	No. SNPs <sup>[a]</sup>	Sum	% share
<b>Canonical transitions</b>	C → T	78,170	156,480	27
	G → A	78,310		28
<b>Non-canonical transitions</b>	T → C	21,073	42,247	7
	A → G	21,174		7
<b>Non-canonical transversions</b>	C → A	13,770	85,715	5
	G → T	13,776		5
	A → C	9,942		3
	T → G	9,930		3
	T → A	12,839		5
	A → T	12,754		4
	G → C	6,290		2
	C → G	6,414		2

SNPs were filtered based on DP  $\geq 10$ , AD = 12.5%-60% and MQ  $\geq 30$ . C→T and G→A transitions were named as ‘canonical’ transitions. All other substitutions were named as ‘non-canonical’ transitions or transversions.

[a] Average of 497 4x pools analyzed.

DP: Read depth, AD: Allelic depth, MQ: Mapping quality

This corresponds to an average of 39,120 EMS type mutations per single M<sub>2</sub> plant. Following the terminology used by Fanelli et al. (2021), we termed the EMS type C→T and G→A mutations as ‘canonical’ transitions and all others as ‘non-canonical’ transitions or transversions. We termed them collectively as single nucleotide variants (SNVs). We detected an average of one C→T and G→A mutation per 23.6 kb of the genomic sequence of the Express617 reference genome (925 Mb). Noteworthy, the 45% share of non-canonical transitions (15%) or transversions (30%) was almost equal to that of the canonical C→T and G→A mutations among the high confidence SNPs (Figure 4). Additionally, on average, we detected 88,725 high confidence insertions and deletions (InDels) per M<sub>2</sub> DNA pool. However, we restricted subsequent sequence analyses to SNVs only.



**Figure 4:** Analysis of SNP counts from 1,988 sequenced  $M_2$  plants constituting 497 4x pools. A) Violin plot showing the number of canonical C→T and G→A transitions. Black dots and blue triangles depict the number of filtered SNPs from each of the sequenced pools and outliers (pools with a very low number of called SNPs), respectively. The red dot represents the mean of SNP counts from all pools. B) Boxplots showing the number of individual single nucleotide variants (SNVs) observed and C) Percentage share of SNVs by type. C→T and G→A transitions were named as ‘canonical’ transitions. All other nucleotide substitutions have been named as ‘non-canonical’ transitions or transversions. All SNVs were filtered based on minimum read depth (DP)  $\geq 10$ , allele depth (AD) = 12.5%-60% and mapping quality (MQ)  $\geq 30$  parameters. All boxplots show the upper and lower quartiles separated by the median (horizontal line). Whiskers represent maximum and minimum values.

### 2.3.2 Detecting functional mutations

We questioned how many high confidence EMS mutations could have a putative effect on gene function. First, we analyzed mutations within predicted gene models from the Express617 reference. We reasoned that an SNV might alter the function of the encoded protein if it (i) introduces a premature stop codon, (ii) results in a splice variant, (iii) confers a non-synonymous amino acid substitution, or (iv) causes a change in the translation start site. On average, 12,129 mutations per M<sub>2</sub> DNA pool fulfilled these criteria (Table 3). 75% of all characterized mutations were located within annotated regions of the Express617 reference genome (chromosomes A01-A10 and C01-C09), whereas 25% were located on non-annotated scaffolds (occupying ~17% of the reference genome) that could not be confidently anchored to any chromosome (Supplementary Table 1). Out of all high confidence canonical mutations located within annotated chromosomes, 0.4%, 7.0%, 4.1%, and 4.9% were predicted as nonsense, missense, synonymous and intronic mutations, respectively (Table 3). Out of all canonical mutations located within coding regions, 2.3% were nonsense mutations, 40.4% were missense mutations, 24.1% were synonymous mutations and 28.7% were located within introns (Supplementary Figure 2). Start site loss and splice site variants were observed as the rarest type of mutations contributing <1% of the total canonical mutations (Figure 5).

**Table 3:** Summary statistics of predicted mutation effects within the gene models of the Express617 reference genome.

Mutation type	No. mutations predicted <sup>[a]</sup>					
	Canonical transitions	Non-canonical transitions	Non-canonical transversions			
	C→T / G→A	T→C / A→G	C→A / G→T	A→C / T→G	T→A / A→T	G→C / C→G
Nonsense	637	0	37	3	11	1
Splice site <sup>[1]</sup>	247	8	7	3	10	4
Missense	11,228	956	622	685	534	546
START lost	17	4	0	3	2	0
STOP retained	16	3	0	0	0	0
Synonymous	6,701	1,857	551	521	564	357
Splice region <sup>[1]</sup>	959	195	104	98	137	45
Intronic	7,969	1,958	1,134	997	1,297	595
Downstream <sup>[2]</sup>	45,108	9,074	4,915	3,907	4,872	2,215
Upstream <sup>[2]</sup>	45,310	9,193	4,943	4,002	4,937	2,291
Intergenic	42,054	10,365	6,275	4,711	6,140	2,588
Total	160,246	33,613	18,588	14,930	18,504	8,642

Mutation effects are predicted using high confidence SNVs from regions assigned to chromosomes (chrA01-A10 and chrC01-C09). The Ensembl Variant Effect Predictor (offline mode) using the general feature format (GFF) of the Express617 reference genome was used (Lee et al., 2020).

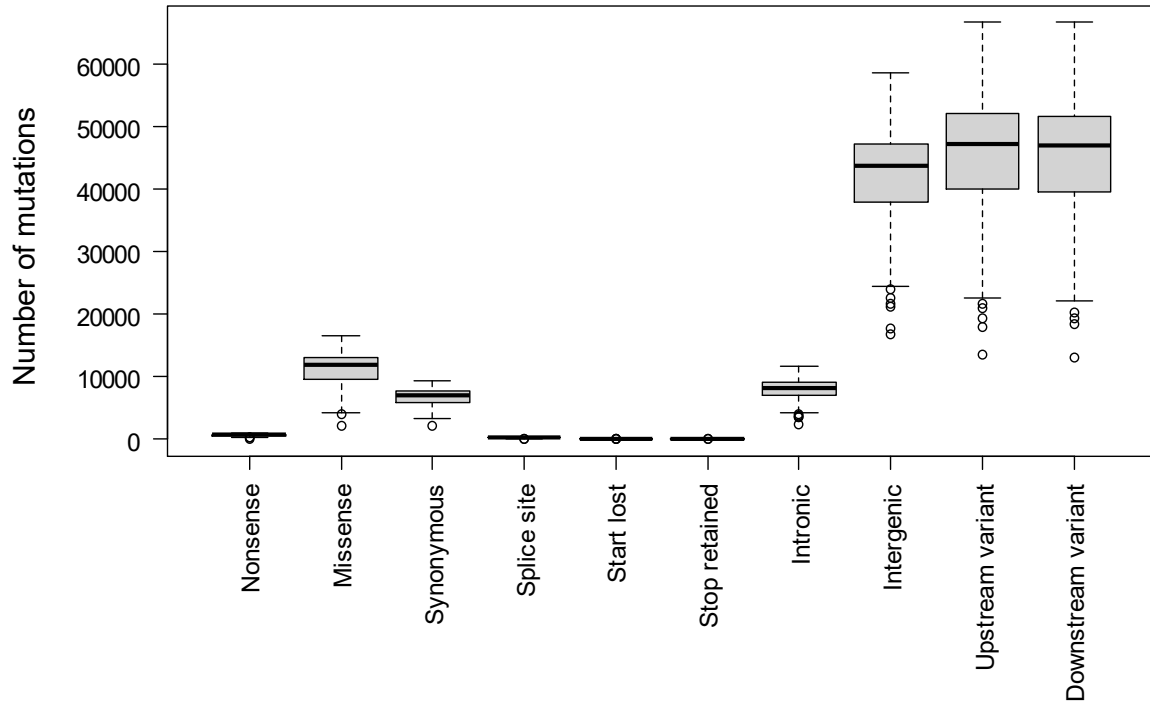
[a] Predicted mutation effects may overlap across mutation types depending on the structure of predicted gene models.

[1] Splice site variants include both acceptor and donor site mutations. Splice region variants include changes in the splice site region, within 1-3 bases of the exon or 3-8 bases of the intron.

[2] Variants located 5 kb upstream or downstream of the translation START and STOP sites, respectively.



We then analyzed the functional effects of non-canonical transitions and transversions (Supplementary Figure 3). Roughly 85% of such mutations were located either 5 kb upstream or downstream (~27% each) of gene coding regions or within intergenic regions (31%). The remaining 15% were located within regions coding for genes (Supplementary Figure 4). As predicted, 3.6% of these mutations could have drastic effects in the form of nonsense (0.05%), missense (3.5%), and splice site (0.03%), or even as START site (0.01%) mutations. On the contrary, 10.4% do not confer functional effects since they were characterized as silent mutations like synonymous (4%) or intronic (6.3%) and rare mutations where the stop codon is retained (0.003%) (Table 3).



**Figure 5:** Characterization of canonical C→T and G →A mutations with predicted effects on a genome-wide scale. Boxplots show the distribution of SNP effects from the sequenced 4x pools. The upper and lower quartiles are separated by the median (horizontal line). Whiskers represent maximum and minimum values. Circles depict outliers. Mutation effects are predicted using SNPs filtered with DP ≥10, AD =12.5%-60% and MQ ≥30 parameters. The Ensembl Variant Effect Predictor release 99 was used in offline mode. Mutation effects were characterized within all predicted gene models using the general feature format (GFF) file of the Express617 reference genome. Splice site variants include acceptor and donor site mutations. Stop retained mutations refer to unchanged stop codons despite induced mutations. Upstream and downstream variants are located within 5 kb from the translation START and STOP sites, respectively. DP: Read depth, AD: Allele depth and MQ: Mapping quality.

### 2.3.3 Sequence analyses reveal patterns of mutation frequency and distribution

Our mutation detection approach operates on the whole-genome scale. Therefore, we investigated the distribution of EMS mutations across all chromosomes, including intergenic regions and 5 kb upstream and downstream regions. To check for a possible bias in mutation frequencies for the A and C sub-genomes of oilseed rape, we calculated the number and density of high confidence canonical mutations across all annotated chromosomes (Table 4).

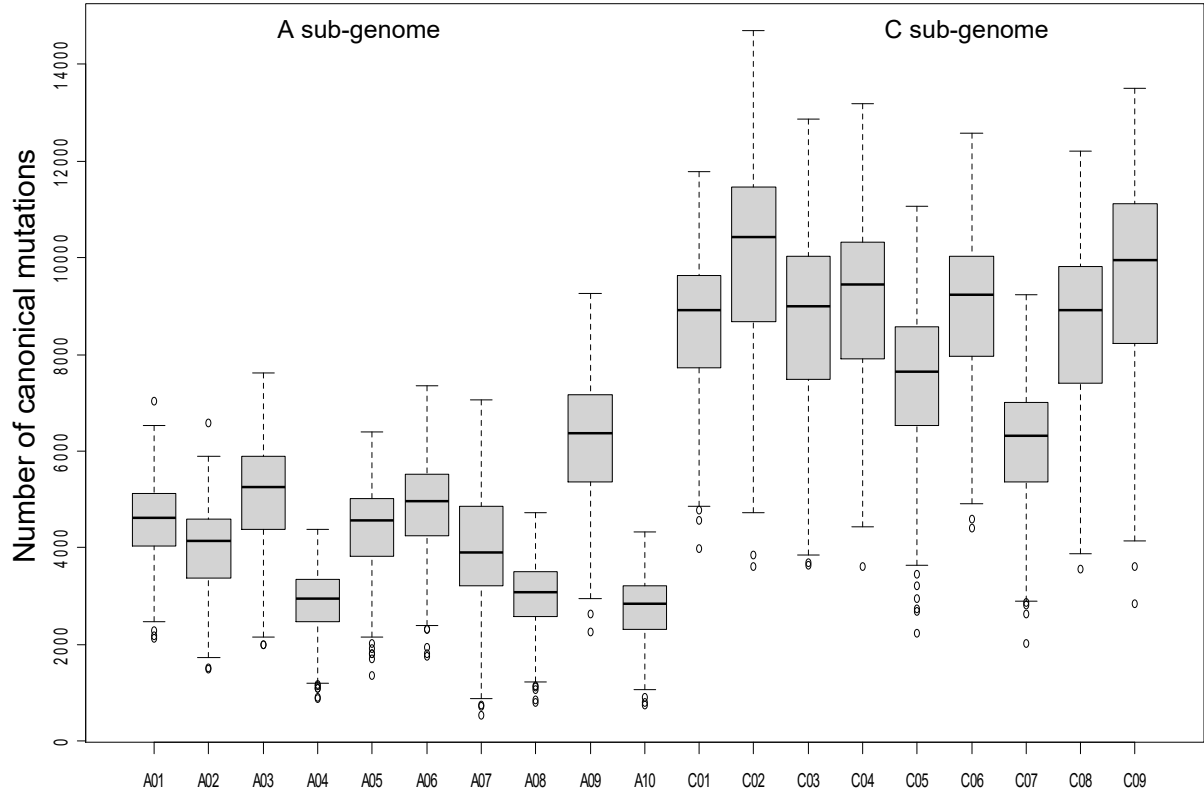
**Table 4:** Summary statistics of canonical mutations in the A and C sub-genomes of the Express617 mutant population. Mutation counts and frequencies were calculated for individual chromosomes of the two sub-genomes (chrA01-A10 and chrC01-C09) for all sequenced 4x pools.

	Chromosome	Chromosome length (bp) <sup>[1]</sup>	No. mutations/ chromosome <sup>[2]</sup>	No. mutations /Mb per 4x pool <sup>[2]</sup>	No. mutations /Mb per plant <sup>[2]</sup>	Average No. Mutations /Mb
<b>A sub-genome</b>	A01	29,969,127	4,521	151	38	35
	A02	30,174,133	3,975	132	33	
	A03	38,328,780	5,055	132	33	
	A04	22,208,790	2,852	128	32	
	A05	29,330,807	4,372	149	37	
	A06	31,862,090	4,787	150	38	
	A07	27,714,229	3,988	144	36	
	A08	22,295,061	2,963	133	33	
	A09	43,308,710	6,183	143	36	
	A10	20,498,486	2,716	133	33	
<b>C sub-genome</b>	C01	44,118,044	8,606	195	49	41
	C02	61,556,739	9,961	162	40	
	C03	62,379,756	8,626	138	35	
	C04	56,192,105	9,056	161	40	
	C05	46,536,336	7,457	160	40	
	C06	46,870,576	8,935	191	48	
	C07	39,009,375	6,106	157	39	
	C08	52,066,946	8,509	163	41	
	C09	60,209,689	9,562	159	40	

[1] Based on the assembled Express617 reference genome (Lee et al. 2020).

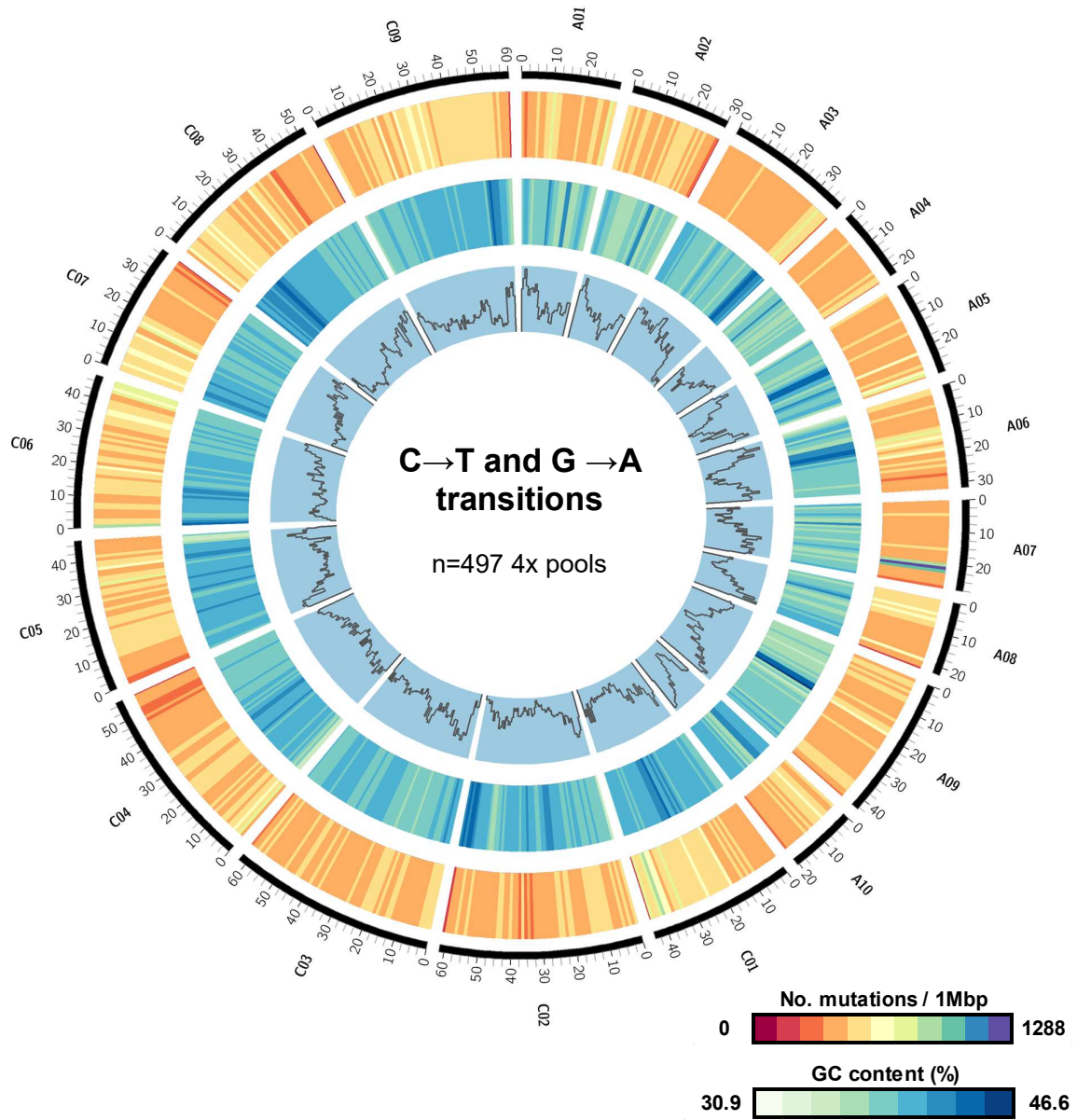
[2] Calculated as an average of 497 4x pools analyzed. For estimations per plant, frequencies of each pool were divided by four since each sequenced pool is a combination of 4 single M<sub>2</sub> plants.

As expected, the number of C sub-genome mutations (76,818) was significantly higher than A sub-genome mutations (41,412) because, in the assembled genome, the C genome exceeds the A genome by 173.25 Mb (Figure 6). Surprisingly, mutation frequencies on average were higher for the C sub-genome (41 mutations/Mb) than the A sub-genome (35 mutations/Mb).



**Figure 6:** Number of canonical transitions across the A and C sub-genomes of rapeseed detected from 497 sequenced  $M_2$  pools. Boxplots represent the average number of SNPs originating from the regions of the Express617 genome annotated with chromosomes (chrA01-A10 and chrC01-C09). The upper and lower quartiles are separated by the median (horizontal line). Whiskers represent maximum and minimum values. Circles depict outliers.

In the next step, we searched both sub-genomes for mutation hotspots by visualizing SNVs in 1 Mb non-overlapping windows (Figure 7). While canonical mutations were evenly distributed across all chromosomes in general, we found regions with significantly increased or decreased mutation densities. As a general tendency, mutation frequencies were lower towards the end of chromosomes except for A06, C03 and C06, where mutation frequencies were higher in the telomeric regions. Furthermore, we compared the distribution of canonical and non-canonical mutations for all chromosomes. For this, we used all C→T and G→A transitions and then all other mutation types that qualified as high confidence SNPs (read depth  $\geq 10$ , allele depth = 12.5%-60%, and mapping quality  $\geq 30$ ) as distinct sets. Interestingly, a similar distribution of high confidence canonical and non-canonical mutations per 1 Mb non-overlapping windows was observed across most chromosomes (Supplementary Figure 5). This suggested that the distribution of the non-canonical mutations is not random across chromosomes but follows a defined distribution pattern like that observed for the C→T and G→A transitions.



**Figure 7:** Visualization of C→T and G→A mutations at the whole genome level. The Circos plot shows four tracks (outside to inside) representing chromosomes A01-A10 and C01-C09 (black), number of mutations (dark red to violet), and GC content (%) per 1 Mb windows (white to dark blue) and regions coding for genes. Chromosomal lengths are in megabases (Mb). Mutation densities were calculated per 1 Mb windows across all chromosomes from all sequenced 4x pools. Regions coding for genes was calculated as the number of base pairs per 1 Mb annotated within gene models. Mutation densities and GC content were plotted using ‘scale\_log\_base = 0.5’ values. 1 Mb non-overlapping windows were used for calculations.

### 2.3.4 Validation of functional mutations

The functional mutations detected by the TbyWGS approach were verified in two ways. We chose four gene families controlling agronomically important traits in oilseed rape. First, we verified mutations in the four gene families previously detected by a conventional gel-based TILLING of the Express617 TILLING population (Harloff et al., 2012; Emrani et al., 2015; Karunaratna et al., 2020). *SFAR* (*SEED FATTY ACID REDUCER*) genes encode GDSL

lipases which control a wide range of primary and secondary functions in plants. Plants possessing EMS-induced functional mutations in *BnSFAR* genes had displayed an elevated seed oil content (Karunaratna et al., 2020). *REF1* and *SGT* genes encode for enzymes UDP-glucose:sinapic acid glucosyltransferase and sinapaldehyde dehydrogenase/coniferaldehyde dehydrogenase, respectively. Both genes play a crucial role in the biosynthesis of sinapine, an important anti-nutritive compound in oilseed rape. EMS-induced loss of function mutations had been detected in both genes. Double mutants showed a significant reduction of the seed sinapine content (Harloff et al., 2012; Emrani et al., 2015). Using these previously characterized mutants, we identified the 7 M<sub>2</sub> families respective to the M<sub>2</sub> mutants. To verify the mutations by our TbyWGS approach, we then screened whole genome sequences of the 4x pools harboring these M<sub>2</sub> families (Supplementary Table 2). As a result, all seven mutations could be found, proving the reliability of our TbyWGS pipeline (Supplementary Figure 6).

As a next step, we screened the WGS dataset for new mutations. We chose genes involved in the glucosinolate biosynthesis and transportation pathway. *MYB28* and *CYP79F1* genes encode for an R2R3-MYB transcription factor (Gigolashvili et al., 2008) and a cytochrome P450 enzyme (Reintanz et al., 2001), respectively. Both genes are involved in the core structure formation of aliphatic glucosinolates in oilseed rape. The *GTR* (*GLUCOSINOLATE TRANSPORTER*) genes have been characterized by a seed-specific glucosinolate transportation activity in *Brassica* species (Nour-Eldin et al., 2017). We detected 12 M<sub>2</sub> DNA pools with nonsense mutations for the selected gene families (Supplementary Figure 7). Then, we isolated genomic DNA from each of the 48 M<sub>2</sub> plants (4 plants/pool). DNA fragments were amplified by PCR using paralog-specific primers flanking the mutation sites (Supplementary Table 3). Sanger sequencing revealed a success rate of 100% because all PCR fragments were detected with the expected mutations (Supplementary Figure 8).

### 2.3.5 A web-based interface for screening the mutant population

Our TbyWGS resource provides access to 78,150,284 high confidence canonical EMS mutations. We developed a web-based resource enabling the user-friendly and convenient detection of high confidence C→T and G→A mutations with predicted effects for ease of screening. Access will be given via the institute's web page. As a first step, the Express617 reference genome assembly and the supplemented GFF file assembled by Lee et al. (2020) must be downloaded from <https://doi.org/10.5281/zenodo.3524259>. By using BLAST queries, genomic regions of interest within this reference assembly concerning the chromosome and their exact locations will be identified. Then, the WGS database can be searched based on the 'Chromosome', 'Start' and 'End' information. By screening the database, mutations within genomic regions of interest can be identified from the respective 4x pools. Finally, M<sub>3</sub> seeds corresponding to the M<sub>2</sub> pools harboring the desired mutations can be requested.

## 2.4 Discussion

This study aimed to develop a database for identifying putative functional mutations on a genome-wide scale in oilseed rape. We have established a TILLING by whole-genome sequencing (TbyWGS) protocol to detect mutations in any sequence of the rapeseed genome. Thus, M<sub>2</sub> plants with the desired mutations can be directly identified to perform functional studies in the M<sub>3</sub> offspring. This procedure avoids the laborious gel-based mutant screening and the long-term storage of high-quality M<sub>2</sub> DNA needed for conventional TILLING.

In our TbyWGS approach, we used a 4x pooling strategy instead of the original 8x pooling previously used for gel-based screening of the Express617 mutant population (Harloff et al., 2012). We found that a sequence coverage of 10x resulted in the false identification of sequencing errors or artifacts as genuine mutations. Therefore, a 20x sequence coverage was chosen, which theoretically results in five raw reads from each  $M_2$  plant in the pool. The degree of heterozygosity in the Express617 inbred line ( $F_{11}$ ) used for EMS mutagenesis is expected to be below 0.0005%. Therefore, we expect that >99.9995 % of the SNPs detected within our  $M_2$  families are caused by EMS mutagenesis, not by residual heterozygosity. Moreover, we adjusted the SNP filtering criteria to remove low-quality SNPs with poor mapping quality. According to classical Mendelian genetics, the frequency of mutant alleles in a segregating  $M_2$  generation is expected to be 50% if we assume the  $M_1$  to be hemizygous for the mutation. However, this value might be lower due to a mosaic of mutated and non-mutated cells expected for almost all  $M_1$  plants. Also, poor fitness of mutant gametophytes and low vitality of homozygous  $M_2$  plants can account for decreased mutant frequencies across the  $M_2$  offspring. In addition, we can expect a statistical bias caused by the low number of  $M_2$  plants in each pool since only four  $M_2$  plants per pool were selected. In the worst case, only one plant out of four might carry a mutant allele. The mutation might be more frequent in another case due to an increased representation of mutated plants carrying one or more mutant alleles within an  $M_2$  pool. Respecting the Mendelian segregation and biases due to mosaic  $M_1$  and fitness of  $M_2$  plants in all sequenced pools, it was reasonable to select a lower limit for the allelic depth of 12.5% to cover rare mutations in each pool. On the other hand, the upper limit from the expected Mendelian average of 50% was increased to 60% to correct for an overrepresentation of mutations in a given  $M_2$  pool. A recommended mapping quality (MQ)  $\geq 30$  was consistently ensured for all called variants for quality control. Moreover, we only used SNPs from loci that had a total coverage depth of at least ten mapped reads at that position. SNV calling was purposely based on these relaxed parameters to avoid false-negative classification of putative functional mutants in target genes based on excessively stringent filtering parameters. This might also have resulted in the calling of a high percentage of non-canonical SNVs. However, validation by Sanger sequencing of known mutants revealed that the detected mutations were all correctly called and thus made this dataset a valuable resource for identifying genotypes with mutations in the selected target genes.

Variable mutation frequencies have been reported between different rapeseed EMS mutant populations (Wang et al., 2008; Harloff et al., 2012; Gilchrist et al., 2013; Wells et al., 2014; Tang et al., 2020). This can be due to several factors like EMS concentration and permeability and the physiology and developmental stage of the treated tissue (Henry et al., 2014). Based on previous studies using conventional gel-based assays, the mutation frequency in our EMS population ranged between 1/12 kb to 1/72 kb (Harloff et al., 2012; Guo et al., 2014; Emrani et al., 2015; Braatz et al., 2018; Shah et al., 2018; Karunaratna et al., 2020). While using an amplicon sequencing approach, a mutation frequency of 1/27 kb was reported from the same TILLING platform by Sashidhar et al. (2019). These differences can be explained because mutation frequencies were estimated based on the length of the amplicons analyzed. Moreover, only protein-coding sequences were studied. In our TbyWGS approach, we estimated a mutation frequency of one high confidence canonical mutation per 23.6 kb of the Express617 reference genome. Similar to our study, Tang et al. (2020) calculated mutation frequencies across the whole rapeseed genome. Across 20 EMS mutants, mutation rates ranged from 1/23.6 kb to 1/303.9 kb. Mutation frequencies after EMS mutagenesis have also been calculated in other crops. As expected, they were lower in diploid species than in polyploid species. In diploid rice (Till et al., 2007) and tomato (Garcia et al.,

2016), they were estimated to be 1/530 kb and 1/125 kb, respectively, whereas in hexaploid wheat, substantially higher frequencies (1/30.3 kb-1/43.4 kb) were found (Krasileva et al., 2017).

The non-random distribution of mutation across the whole genome is an important result of our study. To the best of our knowledge, a lower density of EMS-induced mutations at the chromosomal ends has not been reported before. Using the representative group of 19 annotated chromosomes in the Express617 reference assembly (chromosomes A01-A10 and C01-C09), we calculated mutation frequencies from different regions of the genome (Table 4, Supplementary Figure 5). While its genome size partly explains the higher number of mutations in the C sub-genome, the striking difference in the mutation frequencies between the two sub-genomes deserves a further explanation. One reason could be different GC contents. In the assembled Express617 genome, the average GC contents of the A and C sub-genomes are 35.2% and 36.2%, respectively. Although this difference is not high, we reason that a high GC content invariably affects mutation frequencies since mutations observed were primarily C→T and G→A transitions (55%). Therefore, regions with high GC content could be potential mutation hotspots. Furthermore, it can be speculated that EMS does not target both sub-genomes randomly. At the time of EMS treatment, the vast majority of cells are in the interphase. It is known from studies with polyploid species, e.g., wheat (Fussell & Moens, 1987; Martinez-Perez et al., 2001), that sub-genomes adopt the ‘Rabl configuration’ and are located separately within the interphase nucleus. This could have an impact on their accessibility to EMS, resulting in varying mutation densities. Also, the frequency and distribution of repetitive elements alter the conformation of the chromatin fiber during the interphase. Therefore, it is tempting to speculate that EMS targets condensed and relaxed chromatin differentially.

EMS is believed to confer mainly C→T and G→A transitions (Till et al., 2006) which is reflected in our TbyWGS resource since 55% of all high confidence SNPs were canonical mutations. Using *in vitro* DNA ethylation experiments, Sega (1984) has demonstrated a range of possible modifications other than C→T and G→A transitions resulting from DNA depurination and depyrimidation. It was observed that even transversions could originate because a random nucleotide was inserted opposite apurinic or apyrimidic sites. Since we observed 45% non-canonical mutations in our TbyWGS approach, we questioned why such mutations were rarely observed in our previous studies with the Express617 mutant population. Traditionally, functional genomic studies utilizing EMS mutants have focused on identifying mutations within coding regions (Emrani et al., 2015; Braatz et al., 2018; Karunaratna et al., 2020). Moreover, these studies used conventional detection assays utilizing the CELI endonuclease, an S1 single strand-specific nuclease that cleaves heteroduplexes of mutant and wildtype DNA. It has been demonstrated that although CELI identifies all sequence variants, including InDels (Oleykowski et al., 1998), it has the highest preference for cleaving heteroduplexes arising from G→C, G→A, G→T, and C→G mutations (Oleykowski et al., 1998; Yang et al., 2000; Triques et al., 2007). Using our TbyWGS approach, we observed that 76.3% of the SNVs in gene coding regions were within this group of mutations highly preferred by CELI. Out of this, 66.7% were canonical mutations, and only 9.6% were G→C, G→T or C→G mutations. We reason that such non-canonical mutations were rarely reported from previous conventional screenings mainly due to two reasons. Firstly, mutant detections were done mainly within gene coding regions. Secondly, most non-canonical transitions and transversions within gene coding regions represented the least preferred mismatch sites for CELI cleavage (Oleykowski et al., 1998).

In conclusion, a low frequency of non-canonical mismatch heteroduplexes favored by CELI is possibly a reason for low detection rates within gene coding regions in the past.

We detected 45% non-canonical transitions and transversions. This is in line with a recent report where >50% of all filtered variants were observed to be non-canonical transitions and transversions in three out of ten whole-genome sequenced M<sub>2</sub> EMS mutagenized rapeseed plants (Tang et al., 2020). Moreover, our observation is in line with several reports utilizing conventional and sequencing-based detection methods for EMS mutants from rice (Till et al., 2007), wheat (Krasileva et al., 2017), maize (Lu et al., 2018), barley (Caldwell et al., 2004), soybean (Lakhssassi et al., 2020) and sunflower (Fanelli et al., 2021). In these studies, the share of non-canonical transitions and transversions ranged from 12.3% to 31%.

Presently, our mutant database encompasses the canonical transitions. Adding a screening option for non-canonical transitions and transversions could be interesting and a possible expansion in the future, especially regarding the regulatory up-and downstream regions of genes. Insertions and deletions due to EMS treatment are expected in our M<sub>2</sub> population, an effect that has been reported before by Sega (1984). Moreover, via an exome sequencing approach Krasileva et al. (2017) have confirmed the presence of small InDels (<20 bp) in EMS mutagenized populations of tetraploid and hexaploid wheat. While small (<10 bp) InDels can be detected in our Illumina sequenced population, screening for large InDels requires long-range sequencing technology.

In conclusion, our web-accessible whole-genome sequencing mutant platform is an unprecedented resource that can serve as a strong foundation for molecular breeding and functional genomics in oilseed rape research. The process of mutant identification will bypass cumbersome detection methods in a time and cost-effective manner. Since the resource represents whole-genome sequencing data, it enables functional characterization of EMS mutations within genes and offers a novel opportunity to analyze mutations within regulatory sequences. In this respect, investigating mutations detected upstream and downstream of genes or even within intergenic regions is now a possibility and perhaps a compelling proposition for researchers and plant breeders. Although CRISPR-Cas mediated editing of multiple gene families has gained interest in the past years, especially for the improvement of oilseed rape (Braatz et al., 2017; Karunarathna et al., 2020; Sashidhar et al., 2020; Zheng et al., 2020), legal restrictions within the European Union, have deterred application in current breeding programs. In this regard, EMS mutagenized functional mutants hold great promise for successful integration in plant breeding and crop improvement (Jung & Till, 2021).



## 2.5 Supplementary data

### 2.5.1 Supplementary figures

**Supplementary Figure 1:** Screenshot of the FastQC reports showing a graphical overview of quality checks for raw reads from all sequenced 4x pools. Mean quality scores represent Phred scores for individual bases called across the length 150 bp paired-end reads. Phred scores for all 4x pools are within the optimum levels of  $>30$  (zone marked in green). Raw reads with Phred score = 20-26 are tolerated (yellow zones), Phred score  $\leq 20$  (red zones) are rejected. Analysis was done via FastQC v0.11.5 and MultiQC.

**Supplementary Figure 2:** Share of EMS-induced mutations predicted to have putative functional effects A) on a genome-wide scale and B) specifically within coding regions. SNP effects were first predicted on a genome-wide scale for all predicted gene models on the Express617 reference genome. The share of mutation effects within coding regions was calculated using SNPs that were present within and including the translation START and STOP sites. Splice site variants include the sum of splice acceptor and donor site mutations. Upstream and downstream variants encompass 5 kb genomic regions from the START and STOP sites, respectively. The Ensembl Variant Effect Predictor tool was used for SNP effect prediction using the Express617 reference genome (Lee et al., 2020). \*Reflects the mean SNP counts from 497 IRFFA 4x pools.

**Supplementary Figure 3:** Characterization of ‘non-canonical’ type A) Transitions ( $T \rightarrow C$  and  $A \rightarrow G$ ) and B) Transversions ( $C \rightarrow A$ ,  $G \rightarrow T$ ,  $A \rightarrow C$ ,  $T \rightarrow G$ ,  $T \rightarrow A$ ,  $A \rightarrow T$ ,  $G \rightarrow C$  and  $C \rightarrow G$ ) with predicted effects on a genome-wide scale. Boxplots show the distribution of SNP effects as a mean of 497 4x pools. Mutation effects are predicted using filtered SNPs with  $DP \geq 10$ ,  $AD = 12.5\%-60\%$  and  $MQ \geq 30$ . The Ensembl Variant Effect Predictor tool was used in offline mode. Mutation effects were predicted within the gene models of the Express617 genome. All gene IDs were extracted from the general feature format (GFF) of the Express617 reference genome. Splice site variants include acceptor and donor site mutations. Upstream and downstream variants are located 5 kb from the translation START and STOP sites, respectively. DP: Read depth, AD: Allele depth and MQ: Mapping quality.

**Supplementary Figure 4:** Characterization of ‘non-canonical’ A) Transitions ( $T \rightarrow C$  and  $A \rightarrow G$ ) and B) Transversions ( $C \rightarrow A$ ,  $G \rightarrow T$ ,  $A \rightarrow C$ ,  $T \rightarrow G$ ,  $T \rightarrow A$ ,  $A \rightarrow T$ ,  $G \rightarrow C$  and  $C \rightarrow G$ ) with predicted effects exclusively within coding regions. Boxplots show the distribution of SNP effects as a mean of 497 4x pools. Mutation effects are predicted using filtered SNPs with  $DP \geq 10$ ,  $AD = 12.5\%-60\%$  and  $MQ \geq 30$ . The Ensembl Variant Effect Predictor release 99 was used in offline mode. Mutation effects were predicted within the gene models of the Express617 genome. All gene IDs were extracted from the general feature format (GFF) of the Express617 reference genome. Splice site variants include acceptor and donor site mutations. Upstream and downstream variants are located within 5 kb from the translation START and STOP sites, respectively. DP: Read depth, AD: Allele depth and MQ: Mapping quality.

**Supplementary Figure 5:** Distribution and density of filtered mutations per 1 Mb windows across all sequenced pools for chromosomes A01-A10 and C01-C09. The red and blue lines represent the running average of the frequency distribution of filtered canonical ( $C \rightarrow T$  and  $G \rightarrow A$  transitions) and non-canonical type mutations (all other nucleotide substitutions), respectively across all pools. Black dots reflect the corresponding mutation densities per 1 Mb non-overlapping windows for each of the sequenced pools. The x-axis represents the length of

chromosomes in Mb and each interval corresponds to a 1 Mb window. The y axis shows the number of SNPs located within each window. To smoothen the curves, the General Additive Model (GAM) in the ggplot2 package of R was used.

**Supplementary Figure 6:** Screenshot of the Integrative Genomics Viewer (IGV) showing read alignments from the seven selected 4x pools harboring previously detected EMS mutations. Pool IDs are named with the prefix ‘IR’ or ‘D’. (1) Tracks showing 41 bp regions with 20 bases up- and downstream of the detected mutations. (2) Coverage tracks showing read alignments with horizontal bars representing individual reads. SNPs are marked in different colors. (3) Inset windows representing the count details for each of the called nucleotides (A: Adenine, C: Cytosine, G: Guanine, T: Thymine or N: Unknown) and their share from all mapped reads at that position.

**Supplementary Figure 7:** Screenshot of the Integrative Genomics Viewer (IGV) showing read alignments from 12 fourfold pools harboring EMS mutations detected within candidate genes – *BnGTR2*, *BnMYB28* and *BnCYP79F1*. Pool IDs are named with the prefix ‘D’. (1) Tracks showing 41 bp regions with 20 bases up and downstream of the detected mutations. (2) Coverage tracks showing read alignments with horizontal bars representing individual reads. SNPs are marked in different colors. (3) Inset windows representing the count details for each of the called nucleotides (A: Adenine, C: Cytosine, G: Guanine, T: Thymine or N: Unknown) and their individual share from all mapped reads at that position.

**Supplementary Figure 8:** Sanger sequencing results from experiments validating the presence of EMS-induced mutations detected in 12 fourfold pools (Supplementary Table 3). Using genomic DNA from the four representative individuals of the selected pools, PCR amplicons encompassing detected mutants were Sanger sequenced. Genomic sequences of the candidate genes were extracted from the Express617 reference genome. Location and type of detected mutations were annotated for the candidate genes (marked in blue arrows). Sanger sequencing reads were mapped to the corresponding candidate gene sequences to check for the presence of the expected mutations. Validated mutations from M<sub>2</sub> individuals are marked with orange arrows. Fourfold pools are named with the ‘D’ prefix. The four M<sub>2</sub> individuals corresponding to each of the pools have been named with the pool IDs followed by the ‘\_1’, ‘\_2’, ‘\_3’ and ‘\_4’ suffixes. Sequence analysis was done using the CLC Main Workbench 7.

## 2.5.2 Supplementary tables

**Supplementary Table 1:** Summary of canonical transitions originating from the annotated and non-annotated regions of the Express617 genome. Regions of the Express617 reference assembly without chromosomal annotations for the A or C sub-genomes are denoted as non-annotated regions. Share of C→T and G→A transitions located within annotated and non-annotated regions was calculated as an average from 497 sequenced 4x pools.

**Supplementary Table 2:** Summary of validation experiments to confirm mutations in seven M<sub>2</sub> DNA pools harboring previously detected and characterized EMS mutations within four candidate gene families, *BnREF*, *BnSGT*, *BnSFAR4* and *BnSFAR1*. For each of the selected mutants, corresponding M<sub>2</sub> families were identified. Individual read alignments from each of the 4x pools (named with prefix ‘D’ or ‘IR’) representing the selected mutant M<sub>2</sub> families were visualized for regions harboring expected mutations.

**Supplementary Table 3:** Summary of validation experiments to confirm the presence of mutations in 12 pools harboring EMS mutations within candidate gene families *BnMYB28*,

*BnCYP79F1* and *BnGTR2*. 4x pools (named with prefix 'D') with nonsense mutations for the candidate genes were first identified. Genomic DNA was isolated separately from the leaf samples of the four individuals (#1-4) bulked in each pool. Standard PCR with locus-specific primers was used to amplify regions encompassing the detected mutations. PCR fragments were Sanger sequenced to validate mutation presence.

## 2.6 References

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., & Tamiru, M. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology*, 30(2), 174-178.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. In: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
- Braatz, J., Harloff, H.-J., Emrani, N., Elisha, C., Heepe, L., Gorb, S. N., & Jung, C. (2018). The effect of *INDEHISCENT* point mutations on silique shatter resistance in oilseed rape (*Brassica napus*). *Theoretical and Applied Genetics*, 131(4), 959-971.
- Braatz, J., Harloff, H.-J., Mascher, M., Stein, N., Himmelbach, A., & Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiology*, 174(2), 935-942.
- Caldwell, D. G., McCallum, N., Shaw, P., Muehlbauer, G. J., Marshall, D. F., & Waugh, R. (2004). A structured mutant population for forward and reverse genetics in Barley (*Hordeum vulgare* L.). *The Plant Journal*, 40(1), 143-150.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., & Samans, B. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, 345(6199), 950-953.
- Chen, X., Tong, C., Zhang, X., Song, A., Hu, M., Dong, W., Chen, F., Wang, Y., Tu, J., & Liu, S. (2021). A high-quality *Brassica napus* genome reveals expansion of transposable elements, subgenome evolution and disease resistance. *Plant Biotechnology Journal*, 19(3), 615-630.
- Emrani, N., Harloff, H.-J., Gudi, O., Kopisch-Obuch, F., & Jung, C. (2015). Reduction in sinapine content in rapeseed (*Brassica napus* L.) by induced mutations in sinapine biosynthesis genes. *Molecular Breeding*, 35(1), 37. doi:10.1007/s11032-015-0236-2
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048.
- Fanelli, V., Ngo, K. J., Thompson, V. L., Silva, B. R., Tsai, H., Sabetta, W., Montemurro, C., Comai, L., & Harmer, S. L. (2021). A TILLING by sequencing approach to identify induced mutations in sunflower genes. *Scientific Reports*, 11(1), 9885. doi:10.1038/s41598-021-89237-w
- Fang, D. D., Naoumkina, M., Thyssen, G. N., Bechere, E., Li, P., & Florane, C. B. (2020). An EMS-induced mutation in a tetratricopeptide repeat-like superfamily protein gene (*Ghir\_A12G008870*) on chromosome A12 is responsible for the *li<sub>y</sub>* short fiber phenotype in cotton. *Theoretical and Applied Genetics*, 133(1), 271-282. doi:10.1007/s00122-019-03456-4

Fussell, C. P., & Moens, P. (1987). The Rab1 orientation: a prelude to synapsis. *Meiosis*, 275-299.

Garcia, V., Bres, C., Just, D., Fernandez, L., Tai, F. W. J., Mauxion, J.-P., Le Paslier, M.-C., Bérard, A., Brunel, D., Aoki, K., Alseekh, S., Fernie, A. R., Fraser, P. D., & Rothan, C. (2016). Rapid identification of causal mutations in tomato EMS populations via mapping-by-sequencing. *Nature Protocols*, 11(12), 2401-2418. doi:10.1038/nprot.2016.143

Gigolashvili, T., Engqvist, M., Yatusевич, R., Müller, C., & Flügge, U. I. (2008). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytologist*, 177(3), 627-642. doi:10.1111/j.1469-8137.2007.02295.x

Gilchrist, E. J., Sidebottom, C. H., Koh, C. S., MacInnes, T., Sharpe, A. G., & Haughn, G. W. (2013). A mutant *Brassica napus* (Canola) population for the identification of new genetic diversity via TILLING and next generation sequencing. *PLoS One*, 8(12), e84303.

Guo, Y., Harloff, H.-J., Jung, C., & Molina, C. (2014). Mutations in single *FT*- and *TFL1*-paralogs of rapeseed (*Brassica napus* L.) and their impact on flowering time and yield components. *Frontiers in Plant Science*, 5, 282.

Harloff, H.-J., Lemcke, S., Mittasch, J., Frolov, A., Wu, J. G., Dreyer, F., Leckband, G., & Jung, C. (2012). A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. *Theoretical and Applied Genetics*, 124(5), 957-969.

Henry, I. M., Nagalakshmi, U., Lieberman, M. C., Ngo, K. J., Krasileva, K. V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., & Tai, T. H. (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *The Plant Cell*, 26(4), 1382-1397.

Jung, C., & Till, B. (2021). Mutagenesis and genome editing in crop improvement: perspectives for the global regulatory landscape. *Trends in Plant Science*.

Karunaratna, N. L., Wang, H., Harloff, H. J., Jiang, L., & Jung, C. (2020). Elevating seed oil content in a polyploid crop by induced mutations in *SEED FATTY ACID REDUCER* genes. *Plant Biotechnology Journal*, 18(11), 2251-2266.

Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R. H., Wang, X., & Borrill, P. (2017). Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences*, 114(6), E913-E921.

Lakhssassi, N., Lopes-Caitar, V. S., Knizia, D., Cullen, M. A., Badad, O., El Baze, A., Zhou, Z., Embaby, M. G., Meksem, J., & Lakhssassi, A. (2021). TILLING-by-Sequencing<sup>+</sup> Reveals the Role of Novel Fatty Acid Desaturases (GmFAD2-2s) in Increasing Soybean Seed Oleic Acid Content. *Cells*, 10(5), 1245.

Lakhssassi, N., Zhou, Z., Liu, S., Piya, S., Cullen, M. A., El Baze, A., Knizia, D., Patil, G. B., Badad, O., & Embaby, M. G. (2020). Soybean TILLING-by-Sequencing<sup>+</sup> reveals the role of novel *GmSACPD* members in the unsaturated fatty acid biosynthesis while maintaining healthy nodules. *Journal of Experimental Botany*.

- Lee, H., Chawla, H. S., Obermeier, C., Dreyer, F., Abbadi, A., & Snowdon, R. (2020). Chromosome-Scale Assembly of Winter Oilseed Rape *Brassica napus*. *Frontiers in Plant Science*, 11(496). doi:10.3389/fpls.2020.00496
- Lee, Y.-H., Park, W., Kim, K.-S., Jang, Y.-S., Lee, J.-E., Cha, Y.-L., Moon, Y.-H., Song, Y.-S., & Lee, K. (2018). EMS-induced mutation of an endoplasmic reticulum oleate desaturase gene (*FAD2-2*) results in elevated oleic acid content in rapeseed (*Brassica napus* L.). *Euphytica*, 214(2), 1-12.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Lu, X., Liu, J., Ren, W., Yang, Q., Chai, Z., Chen, R., Wang, L., Zhao, J., Lang, Z., & Wang, H. (2018). Gene-indexed mutations in maize. *Molecular Plant*, 11(3), 496-504.
- Martinez-Perez, E., Shaw, P., & Moore, G. (2001). The *Ph1* locus is needed to ensure specific somatic and meiotic centromere association. *Nature*, 411(6834), 204-207.
- Nie, S., Wang, B., Ding, H., Lin, H., Zhang, L., Li, Q., Wang, Y., Zhang, B., Liang, A., & Zheng, Q. (2021). Genome assembly of the Chinese maize elite inbred line RP125 and its EMS mutant collection provide new resources for maize genetics research and crop improvement. *The Plant Journal*.
- Nour-Eldin, H. H., Madsen, S. R., Engelen, S., Jørgensen, M. E., Olsen, C. E., Andersen, J. S., Seynnaeve, D., Verhoye, T., Fulawka, R., Denolf, P., & Halkier, B. A. (2017). Reduction of antinutritional glucosinolates in *Brassica* oilseeds by mutation of genes encoding transporters. *Nature Biotechnology*, 35, 377. doi:10.1038/nbt.3823
- <https://www.nature.com/articles/nbt.3823#supplementary-information>
- Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., & Yeung, A. T. (1998). Mutation detection using a novel plant endonuclease. *Nucleic Acids Research*, 26(20), 4597-4602.
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., & Roazen, D. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178.
- Rahman, H. (2013). Review: Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. *Canadian Journal of Plant Science*, 93(3), 363-373. doi:10.4141/cjps2012-074
- Reintanz, B., Lehnen, M., Reichelt, M., Gershenzon, J., Kowalczyk, M., Sandberg, G., Godde, M., Uhl, R., & Palme, K. (2001). bus, a bushy *Arabidopsis* *CYP79F1* knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. *The Plant Cell*, 13(2), 351-367.
- Rousseau-Gueutin, M., Belser, C., Da Silva, C., Richard, G., Istace, B., Cruaud, C., Falentin, C., Boideau, F., Boutte, J., & Delourme, R. (2020). Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience*, 9(12), gaa137.
- Sashidhar, N., Harloff, H. J., & Jung, C. (2019). Identification of phytic acid mutants in oilseed rape (*Brassica napus*) by large scale screening of mutant populations through amplicon sequencing. *New Phytologist*.

- Sashidhar, N., Harloff, H. J., Potgieter, L., & Jung, C. (2020). Gene editing of three *BnITPK* genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnology Journal*, 18(11), 2241-2250.
- Sega, G. A. (1984). A review of the genetic effects of ethyl methanesulfonate. *Mutation Research/Reviews in Genetic Toxicology*, 134(2-3), 113-142.
- Shah, S., Karunarathna, N. L., Jung, C., & Emrani, N. (2018). An *APETALA1* ortholog affects plant architecture and seed yield component in oilseed rape (*Brassica napus* L.). *BMC Plant Biology*, 18(1), 380. doi:10.1186/s12870-018-1606-9
- Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., & Zhou, R. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants*, 6(1), 34-45.
- Tang, S., Liu, D. X., Lu, S., Yu, L., Li, Y., Lin, S., Li, L., Du, Z., Liu, X., & Li, X. (2020). Development and screening of EMS mutants with altered seed oil content or fatty acid composition in *Brassica napus*. *The Plant Journal*.
- Till, B. J., Cooper, J., Tai, T. H., Colowit, P., Greene, E. A., Henikoff, S., & Comai, L. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biology*, 7(1), 1-12.
- Till, B. J., Zerr, T., Comai, L., & Henikoff, S. (2006). A protocol for TILLING and Ecotilling in plants and animals. *Nature Protocols*, 1(5), 2465.
- Triques, K., Sturbois, B., Gallais, S., Dalmais, M., Chauvin, S., Clepet, C., Aubourg, S., Rameau, C., Caboche, M., & Bendahmane, A. (2007). Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea. *The Plant Journal*, 51(6), 1116-1125.
- Wang, B., Wu, Z., Li, Z., Zhang, Q., Hu, J., Xiao, Y., Cai, D., Wu, J., King, G. J., & Li, H. (2018). Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. *Plant Biotechnology Journal*, 16(7), 1336-1348.
- Wang, N., Wang, Y., Tian, F., King, G. J., Zhang, C., Long, Y., Shi, L., & Meng, J. (2008). A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of *FAEI* point mutations by TILLING. *New Phytologist*, 180(4), 751-765.
- Wells, R., Trick, M., Soumpourou, E., Clissold, L., Morgan, C., Werner, P., Gibbard, C., Clarke, M., Jennaway, R., & Bancroft, I. (2014). The control of seed oil polyunsaturate content in the polyploid crop species *Brassica napus*. *Molecular Breeding*, 33(2), 349-362.
- Wickham, H. (2009). *Elegant graphics for data analysis* (Vol. 35).
- Yang, B., Wen, X., Kodali, N. S., Oleykowski, C. A., Miller, C. G., Kulinski, J., Besack, D., Yeung, J. A., Kowalski, D., & Yeung, A. T. (2000). Purification, cloning, and characterization of the CEL I nuclease. *Biochemistry*, 39(13), 3533-3541.
- Zheng, M., Zhang, L., Tang, M., Liu, J., Liu, H., Yang, H., Fan, S., Terzaghi, W., Wang, H., & Hua, W. (2020). Knockout of two *BnaMAX1* homologs by CRISPR/Cas9-targeted

mutagenesis improves plant architecture and increases yield in rapeseed (*Brassica napus* L.).  
*Plant Biotechnology Journal*, 18(3), 644-654.

### 3 Reducing glucosinolate content in oilseed rape (*Brassica napus*) by random mutagenesis in biosynthesis genes *BnMYB28* and *BnCYP79F1*

#### 3.1 Introduction

Oilseed rape (*Brassica napus* L.) is an essential oil crop, ranking as the third-largest source of vegetable oil in the world (<http://www.fao.org/faostat/>). In Europe, it is grown as a winter crop, sown in autumn and flowers in the following spring after exposure to cold temperatures over winter. The seeds contain 45-50% oil with a healthy lipid profile that makes it an edible oil fit for human consumption (Russo et al., 2021). Moreover, it is also utilized for biodiesel production. After oil extraction, the leftover rapeseed meal (RSM) serves as a protein-rich (40%) animal feed. However, major anti-nutritive compounds like glucosinolates in RSM adversely affect its nutritional and commercial value (Bourdon & Aumaître, 1990). Therefore, improving seed meal quality is a major goal for rapeseed breeding.

Glucosinolates (GSLs) are diverse heterogeneous secondary metabolites specific to Brassicales. They are natural organic compounds containing sulfur and nitrogen synthesized from glucose and amino acids as precursors. They comprise a thioglucose and a sulfonated oxime attached to a chain elongated amino acid. Depending upon amino acid precursors, GSLs are categorized as aliphatic, aromatic, and indolic, originating primarily from methionine, phenylalanine or tyrosine and tryptophan, respectively (Halkier & Du, 1997). Biosynthesis of the three types is independently controlled by distinct gene families (Sønderby et al., 2010). So far, around 130 different GSL types have been described from 16 dicot angiosperms (Fahey et al., 2001; Mitreiter & Gigolashvili, 2021). Apart from *B. napus*, *Brassica oleracea* (cauliflower, cabbage, broccoli, Brussels sprouts and kale), *Brassica rapa* (turnips and radish) are economically relevant vegetable crops rich in GSLs (Cartea & Velasco, 2008; Sønderby et al., 2010). Fifteen major GSL types have been identified in *B. napus* (Fahey et al., 2001), reaching levels as high as 60-100  $\mu\text{mol/g}$  dry weight in seeds. Aliphatic GSLs originating from methionine constitute a significant share of all GSLs in Brassicaceae crops, constituting up to 92% of all GSL types in rapeseed (Velasco et al., 2008).

GSLs yield toxic by-products after enzymatic cleavage by the endogenous myrosinase enzyme (Halkier & Du, 1997). Upon physical damage to tissues, the myrosinase is released from so-called 'myrosin cells' and comes into contact with GSLs stored separately in 'S-cells' (Andréasson et al., 2001). Hydrolysis of GSLs generates various products like isothiocyanates (ITC), thiocyanates (SCN), epithionitriles and nitriles (NI), many of which are known to confer defense against generalist herbivores and bacterial and fungal pathogens (Rask et al., 2000). GSLs have been shown to confer antimicrobial properties against *Xanthomonas campestris* pv. *campestris* and the necrotrophic fungus *Sclerotinia sclerotiorum* (Madloo et al., 2019). However, high consumption of GSLs through the feed can result in several adverse metabolic effects in animals. Hydroxyalkenyl GSLs like epiprogoitrin and progoitrin are goitrogenic by causing inflammation of the thyroid gland (Fahey et al., 2001). Retarded growth, reduced appetite and feed efficiency, gastrointestinal irritation, liver and kidney damage and behavioral effects have been observed in fish (Kaiser et al., 2021), poultry (Tripathi & Mishra, 2007) and higher mammals like pigs (Bourdon & Aumaître, 1990). On the contrary, other GSL types like sulforaphane and indole-3-carbinol are known for their beneficial effects on human health with anti-carcinogenic properties (Keck & Finley, 2004).



GSL biosynthesis is completed in three major steps, (i) chain elongation, (ii) core structure formation, and (iii) secondary side-chain modifications (Halkier & Du, 1997). Insertion of methylene groups results in chain elongated amino acids. The addition of the sulfur group to the chain elongated amino acid and *S*-glucosylation completes the core structure formation. Lastly, secondary modifications like benzylation, desaturation, hydroxylation, methoxylation and oxidation result in distinct GSL types (Sønderby et al., 2010; Mitreiter & Gigolashvili, 2021). Environmental effects combined with specific genetic mechanisms for GSL biosynthesis, regulation, transport and storage result in the variable GSL content and diverse profiles observed across *Brassica* species (Mitreiter & Gigolashvili, 2021). A complex network of transcription factors (TFs) influenced by abiotic and biotic stimuli, hormonal and epigenetic factors control the spatiotemporal regulation of GSL biosynthesis (Gigolashvili et al., 2007; Hirai et al., 2007). While a lot is known about the biosynthesis of GSLs, knowledge regarding the long-distance transport of GSL within the plant is limited. In *Arabidopsis*, the biosynthesis process occurs in the vegetative parts and silique walls (Sønderby et al., 2010). After synthesis, GSLs are loaded into the seeds via GSL specific transporters. In a pioneering work, Nour-Eldin et al. (2012) demonstrated the activity of nitrate/peptide transporters NRT/PTR elucidating a molecular basis for the source and sink relationship for the proton-coupled seed-specific transport of GSLs in *Arabidopsis*.

Presently, *Arabidopsis* serves as a model for studying GSL biosynthesis (Mitreiter & Gigolashvili, 2021). The most notable genes controlling aliphatic GSL biosynthesis in *Arabidopsis* are R2R3-MYB TFs. Three TFs *MYB28*, *MYB76* and *MYB29*, also referred to as *HIGH ALIPHATIC GLUCOSINOLATE (HAG) 1*, *2* and *3*, respectively, have been described (Gigolashvili et al., 2007; Hirai et al., 2007), of which *HAG1* and *HAG3* have been speculated to have a ‘master regulator’ effect by up-regulating almost all genes involved in the core structure formation (Hirai et al., 2007). In the first step of the core structure formation, the genes *CYP79F1* and *CYP79F2* convert chain elongated methionine to corresponding aldoximes (Reintanz et al., 2001; Chen et al., 2003). In subsequent steps, genes *GSTF11* and *GSTU20* encoding glutathione S-transferases control the addition of the glutathione sulfur group. *S*-glucosylation by *UGT74* genes encoding glucosyltransferases then results in desulfo-GSLs. The sulfate group in the core structure is then added by sulfotransferases SOT17 and SOT18. For secondary modifications, the flavin monooxygenase *FMO<sub>GS-OX</sub>* genes are responsible for the *S*-oxygenation of the original S-methyl group of methionine in the GSLs molecules. Other genes like *AOP2* and *AOP3* mediate a major branching point by converting *S*-oxygenated GSL to alkenyl or hydroxyalkenyl GSLs (Sønderby et al., 2010).

The GSL biosynthesis pathway has been thoroughly investigated in *Arabidopsis*. However, the transfer of knowledge to *B. napus* is limited and complicated due to its polyploid genome, where multiple genes with functional redundancies may exist. Several genes associated with GSLs in rapeseed have been revealed through associative transcriptomics, genome-wide association, and QTL mapping studies (Harper et al., 2012; Kittipol et al., 2019; Lu et al., 2019; Wei et al., 2019; S. Liu et al., 2020). These studies demonstrated the significant association of biosynthesis genes *BnMYB28* and *BnCYP79F1* and the transporter gene *BnGTR2* with GSL content in rapeseed. Recent studies have characterized the functions of *GLUCOSINOLATE TRANSPORTER GTR1* and *GTR2* responsible for the accumulation of GSL in seeds of *Brassica* crops (Nour-Eldin et al., 2017; Nambiar et al., 2021).

A major breeding goal for rapeseed is to reduce anti-nutritive compounds that severely restrict the application of the rapeseed meal (RSM) as feed for animals. A milestone in oilseed rape breeding was the development of ‘double zero (00)’ rapeseed varieties with low

erucic acid and glucosinolate content (Wittkop et al., 2009). Alleles conferring low seed GSL content were introgressed from the Polish spring variety ‘Bronowski’ (Kondra & Stefansson, 1970) to develop modern rapeseed cultivars with improved seed meal traits. In modern varieties, the GSL content of the RSM has been reduced to <18-25  $\mu\text{mol}$  per gram dry weight (Wittkop et al., 2009).

Rapeseed is an early allotetraploid (AACC,  $2n = 38$ ) product of spontaneous interspecific hybridization between *B. rapa* L., syn. *campestris* (AA,  $2n = 20$ ) and *B. oleracea* L. (CC,  $2n = 18$ ). It is believed to have originated in the Mediterranean region ~7,500 years ago, diverging from its close relative *Arabidopsis* ~20.4 million years ago (Chalhoub et al., 2014). The genome has a size of ~1.13 gigabases (Gb) and 101,040 gene models have been predicted encompassing the two sub-genomes A (314.2 Mb) and C (525.8 Mb) with ten and nine chromosomes, respectively (Chalhoub et al., 2014).

Our study aimed to reduce aliphatic GSLs in rapeseed since they have the highest share of all GSLs in the seeds. Past studies have revealed the association of major candidate genes involved in the biosynthesis, regulation and transport of GSL in rapeseed (Kittipol et al., 2019; Lu et al., 2019; Y. Liu et al., 2020). However, the functional effects of the genes have not been demonstrated. Using an EMS mutagenized winter oilseed rape population (Harloff et al., 2012), we detected loss of function mutations in *BnMYB28* and *BnCYP79F1* genes involved in the core structure biosynthesis of aliphatic GSLs. We found a significant reduction in the aliphatic GSL content, especially progoitrin levels in mutants of *BnMYB28* and *BnCYP79F1*.

## 3.2 Materials and methods

### 3.2.1 Plant material and growth conditions

The winter-type oilseed rape cultivar Express617 was used for expression studies. Plants were grown in 11 cm pots under greenhouse conditions (16 h light, ~25°C). After three weeks, plants were vernalized (16 h light, 4°C) for eight weeks and then transferred to greenhouse conditions (16 h light, ~25°C) and randomized twice a week. Flowers were hand-pollinated and marked with pollination dates. Leaves and seeds were sampled at 15, 25, 35 and 45 days after pollination (DAP). 50-100 mg of fresh weight tissues were collected from five biological replicates at the four developmental stages. Tissues were shock frozen in liquid nitrogen and stored at -70°C.

Seeds of the  $M_2$  and  $M_3$  generations were produced after the EMS treatment of an advanced inbred line ( $F_{11}$ ) of the winter rapeseed Express617 (Harloff et al., 2012).  $M_3$  seeds were obtained from NPZ Innovation GmbH (Holtsee, Germany).  $M_3$  plants were grown in 11 cm pots under greenhouse conditions (16 h light, 20-25°C) for three weeks. Leaf samples were collected from each plant for DNA isolation and genotyping before vernalization for eight weeks (16 h light, 4°C). Simultaneously, additional plants from the winter-type Express617 and the Canadian spring variety Peace were grown as wildtype crossing parents.  $M_3$  plants selected after genotyping were crossed in three ways. (i) Direct  $M_3 \times M_3$  crosses of single homozygous mutants were done to generate  $F_2$  populations segregating for homozygous wildtypes, single and double mutants, (ii) double mutants were backcrossed with Peace to produce segregating  $BC_1F_2$  families and (iii) homozygous single mutants were first backcrossed with Peace and then combined at the  $BC_1$  generation to subsequently produce segregating  $BC_1F_2$  progenies. All plants were grown in 11 cm pots under greenhouse conditions (16 h light, 20-25°C). Plants selected for crossing experiments were hand-

pollinated after emasculation. The inflorescences of plants chosen for selfings were isolated with plastic bags before anthesis.

### 3.2.2 *In silico* analyses

Genomic DNA and polypeptide sequences of aliphatic glucosinolate biosynthesis genes *AtMYB28* and *AtCYP79F1* were retrieved from The Arabidopsis Information Resource (TAIR - <https://www.arabidopsis.org/>). Using the Darmor-*bzh* oilseed rape reference genome (<http://www.genoscope.cns.fr/brassicanapus/>), amino acid sequences of the Arabidopsis orthologs were used as BLAST queries. Chromosomal locations and gene sequences of hits with the lowest e-values and >80% sequence similarity were retrieved as paralogs of *BnMYB28* and *BnCYP79F1*. Gene annotations (exons, 5' and 3' untranslated regions and open reading frames) for the paralogs were made using the CLC Main workbench 7 (QIAGEN® Aarhus A/S, Aarhus C, Denmark). Sequence alignments were made for the genomic DNA, cDNA and amino acid sequences of the retrieved genes. Conserved and functional domain analyses were done using the NCBI Conserved Domain Database.

### 3.2.3 Expression analysis of candidate genes by RT-qPCR

Frozen leaf and seed samples were pulverized in 2 ml reaction tubes with three 3 mm steel balls using the GenoGrinder2010 (SPEX® SamplePrep LLC, USA) at 1,200 strokes/min in three 1 min intervals. RNA was isolated using the peqGold Plant RNA Kit (PEQLAB Biotechnologie GmbH, Erlangen, Germany) following the manufacturer's instructions. RNA quality was assessed with the NanoDrop2000 spectrophotometer (ThermoFisher Scientific, Waltham, United States) and agarose gel electrophoresis (1.5% agarose, 100 V, 30 min). The RNase-free DNase kit (ThermoFisher Scientific, Waltham, United States) was used to treat samples with DNase I to remove contaminating gDNA. cDNA was synthesized with 1 µg RNA using the First Strand cDNA Kit (ThermoFisher Scientific, Waltham, United States). The cDNA was diluted 1:20 with ddH<sub>2</sub>O and 2 µl template was used to perform RT-qPCR on the Bio-Rad CFX96 Real-Time System (Bio-Rad Laboratories GmbH, Munich, Germany) using paralog-specific primers (Supplementary Table 4). Relative expression was calculated according to the  $\Delta\Delta C_q$  method for each paralog normalized against the two reference genes *BnGAPDH* and *BnACTIN2*. Relative expression levels of each candidate paralogs were determined as a mean of five biological replicates represented by three technical replicates each.

### 3.2.4 DNA isolation and genotyping experiments

For DNA isolation, leaf samples were collected and lyophilized for 72 h (Martin Christ Gefriertrocknungsanlagen GmbH, Germany). Freeze-dried samples were pulverized using the GenoGrinder2010 (SPEX® SamplePrep LLC, USA) at 1,200 strokes/min. Genomic DNA was isolated using the standard CTAB method (Saghai-Marooof et al., 1984). To identify desired genotypes for mutant analyses and crossing experiments, standard PCR with primers flanking detected EMS mutations were used for amplification (Supplementary Table 5). PCR products were checked for specificity using agarose gel electrophoresis (1%, 100 V, 12-30 min). Plants were genotyped by Sanger sequencing the PCR fragments.

### 3.2.5 Conventional gel-based detection of EMS-induced mutations

3,840 M<sub>2</sub> plants from the EMS mutagenized winter oilseed rape Express617 population (Harloff et al., 2012) were screened to detect EMS-induced mutations in *BnMYB28* and *BnCYP79F1* paralogs. Paralog-specific primers were designed using the Darmor-*bzh*

reference genome (Supplementary Table 5). Amplicons were evaluated for specificity using agarose gel electrophoresis (1% agarose, 100 V, 10 min) and Sanger sequencing for validation. Using the protocol described by Till et al. (2006), M<sub>2</sub> DNA pools were amplified using 5' end infrared (IR) labeled probes DY-681 and DY-781 (Biomers, Ulm, Germany) for forward and reverse primers (100 pmol/μl), respectively. The resulting amplicons were processed for heteroduplex formation, followed by treatment with CELI nuclease (15 min at 45°C). 5 μl 50 mM EDTA (pH 8.0) was added to terminate the digestion reaction. After digestion, samples were purified on Sephadex G-50 Fine columns (GE Healthcare, Chicago, USA). 4 μl formamide dye (96% deionized formamide, 5 ml 0.25 M EDTA, 0.01% bromophenol blue) was added to each sample. Samples were concentrated to ~20% of the original volume after incubation at 95°C for 30 min. Concentrated samples in 0.65 μl aliquots were separated on polyacrylamide gels using the LI-COR 4300 DNA Analyzer (LI-COR Biosciences, Lincoln, USA) using standard parameters (1,500 V, 40 mA and 40 W for 4 h 15 min).

The GelBuddy imaging software (Zerr & Henikoff, 2005) was used to analyze the gel images and identify single M<sub>2</sub> mutants. Standard PCR was done to amplify regions harboring the expected single mutations using the gDNA isolated from single M<sub>2</sub> plants. Amplicons were Sanger sequenced to validate the detected EMS-induced point mutations. Mutation effects conferred by SNPs were then characterized on the polypeptide level. Mutation frequencies (F) were estimated following the formula given by Harloff et al. (2012):

$$F [1 \text{ per kb}] = 1 / \frac{(\text{amplicon size} - 100) \times \text{number of M1 plants}}{\text{Number of screened mutations} \times 1000}$$

### 3.2.6 Genotyping with the *Brassica* 19K SNP array

BC<sub>1</sub> plants originating from backcrosses with the spring-type oilseed rape Peace were genotyped using the *Brassica* 19K Infinium SNP chip (SGS-INSTITUT FRESENIUS GmbH, TraitGenetics Section, Gatersleben, Germany). For this, genomic DNA from selected mutant plants was isolated using the standard CTAB method and normalized to 50-200 ng/μl. Raw data were obtained in an Excel file with a list of all markers and corresponding SNPs from each plant. All data analyses were done using the method described by Karunarathna et al. (2021). Raw marker data were first converted to the Hapmap (Haplotype Map) format. The software 'TASSEL5 - Trait Analysis by aSSociation, Evolution and Linkage' (Bradbury et al., 2007) converted the Hapmap data into the ABH genotype format. Parental alleles from Express617 and Peace wildtypes were assigned the 'A' and 'B' alleles, respectively. Heterozygous alleles were designated as 'H'. The degree of relatedness within the sampled individuals was calculated using principal component analyses (PCA). The ABH genotype files were exported in a CSV (comma-separated values) format and filtered for polymorphic markers across the Express617 and Peace genomes. The Peace genome share was then calculated for each genotyped BC<sub>1</sub> individual using the formula:

$$\frac{(\text{No. heterozygous alleles} \times 1) + (\text{No. homozygous alleles} \times 2)}{\text{Total number of alleles}} \times 100$$

### 3.2.7 Glucosinolate determination

Glucosinolate (GSL) determination in leaves and mature seeds was done in two ways. Quantitative measurements were done with an enzymatic determination method using myrosinase/thioglucosidase from *Sinapis alba* (Sigma-Aldrich CAS-No. 9025-38-1) and a D-Glucose Assay Kit (glucose oxidase/peroxidase; GOPOD Format) (Megazyme International, Ireland). A qualitative assessment of GSL types was done using high-performance liquid chromatography (HPLC). 5-7g fresh leaves (~15 days after pollination) and 200-600 mg mature seeds were sampled in 50 ml and 2 ml sample tubes, respectively. Sampled leaves were lyophilized for 72 h (Martin Christ Gefriertrocknungsanlagen GmbH, Germany). Freeze-dried leaf samples and mature seeds were pulverized using the GenoGrinder2010 (SPEX® SamplePrep LLC, USA) at 1,400 strokes/min in 4-5 1 min intervals. ~ 200 mg milled samples were used to prepare crude extracts following the protocol of Fiebig and Arens (1992) and then stored at -20°C.

For quantitative measurements, 4 ml crude extracts were run through 0.5 ml DEAE-Sephadex A-25 columns (GE Healthcare, Chicago, USA). The bound GSL was digested on the column for 18 h at room temperature with ~0.8 units of myrosinase/thioglucosidase. After digestion, columns were washed twice with 0.5 ml deionized distilled water (ddH<sub>2</sub>O) for glucose elution. The eluate was shock frozen and freeze-dried for ~72 h. The residue was dissolved in 100 µl ddH<sub>2</sub>O and a 5-40 µl aliquot was used for analysis in duplicates using the D-Glucose Assay Kit (Megazyme International, Ireland) following the manufacturer's instructions. For qualitative analyses, GSL in 1 ml crude extracts were bound on 250 µl DEAE-Sephadex A-25 columns and digested on the column for 18 h at room temperature with 18 U sulfatase H1 enzyme (Merck KGaA, Darmstadt, Germany). Desulfoglucosinolates were eluted two times with 1 ml ddH<sub>2</sub>O. 10-50 µl aliquots were separated on a 250 x 4.6 mm Lichrosorb column (Merck KGaA, Darmstadt, Germany) according to the method of Fiebig and Arens (1992). GSLs were measured by their absorbance at 229 nm and identified using commercially available GSL standards (PhytoLab GmbH, Vestenbergsgreuth, Germany). Quantification was done with individual calibration curves for each standard (Supplementary Table 6).

### 3.2.8 Statistical analyses

For expression studies, significantly expressed paralogs were identified by performing an ANOVA test ( $p < 0.05$ ) for the relative expression levels of each paralog. Mean relative expressions from five biological replicates were compared across the four sampling points 15, 25, 35 and 45 DAP in seeds and leaves. The LSD test ( $\alpha \leq 0.05$ ) was performed to generate statistical groups using the 'Agricolae' package in R. The standard error of mean was calculated across the five biological replicates with three technical replicates each.

In GSL determination experiments, the total GSL content using the GOPOD assay and the contents of individual GSL types using HPLC were evaluated for statistical significance across the analyzed samples. An ANOVA test ( $p < 0.05$ ) was performed for the analysis of the variance along with the LSD test ( $\alpha \leq 0.05$ ) for statistical grouping using the 'Agricolae' package in R. The standard error of mean was calculated across the five biological replicates.

### 3.3 Results

#### 3.3.1 Identification of *MYB28* and *CYP79F1* genes in oilseed rape

For the identification of possible paralogs in rapeseed, the Darmor-*bzh* reference genome (<https://www.genoscope.cns.fr/brassicanaapus/>) was searched for *MYB28* and *CYP79F1* genes using polypeptide sequences from *A. thaliana* genes *AtMYB28* (AT5G61420) and *AtCYP79F1* (AT1G16410) as queries. Based on the lowest e-values and highest sequence similarities (>80%), three and two paralogs were detected for *BnMYB28* and *BnCYP79F1*, respectively (Table 5, Supplementary Figure 9). Furthermore, we aligned the polypeptide sequences of the candidate paralogs to identify conserved functional domains characteristic to the two gene families (Supplementary Figure 10). The protein sequences of identified paralogs shared more than 80% similarity with the Arabidopsis orthologs. In line with previous reports, the highly conserved DNA binding R2-R3 domain-specific to the subgroup 12 MYB TFs and the nuclear localization signal (NLS) with the 'LKKRL' amino acid residues were present across all three *BnMYB28* paralogs (Dubos et al., 2010). The paralog *BnMYB28.C09* was annotated in a truncated form in the Darmor-*bzh* reference genome but harbored all the conserved domains required for gene function. Both *BnCYP79F1* paralogs possessed the five conserved domains characteristic to the family of cytochrome P450 enzymes, including a 'heme' group speculated to act as the catalytic domain (Reintanz et al., 2001).

**Table 5:** Features of *BnMYB28* and *BnCYP79F1* paralogs with homology to the Arabidopsis genes *AtMYB28* (AT5G61420) and *AtCYP79F1* (AT1G16410) in oilseed rape.

<i>B. napus</i> homoeologs <sup>[a]</sup>	<i>B. napus</i> gene name	Chromosome	Gene length (bp)	Coding region (bp)	Polypeptide length	Shared protein identity with Arabidopsis
<i>BnaA03g40190D</i>	<i>BnMYB28.A03</i>	A03	2,018	987	328	85.2%
<i>BnaCnnng43220D</i>	<i>BnMYB28.Cnn</i>	Cnn <sup>[b]</sup>	2,072	1,011	337	87.1%
<i>BnaC09g05300D</i>	<i>BnMYB28.C09</i>	C09	1,072	420	140	82.8%
<i>BnaC05g12520D</i>	<i>BnCYP79F1.C05</i>	C05	2,397	1,623	540	86.1%
<i>BnaA06g11010D</i>	<i>BnCYP79F1.A06</i>	A06	2,380	1,623	540	85.8%

[a] Sequence analysis of orthologs and paralogs in oilseed rape are based on gene models described in the Darmor-*bzh* reference genome (Genoscope).

[b] Not anchored to a specific *B. napus* chromosome.

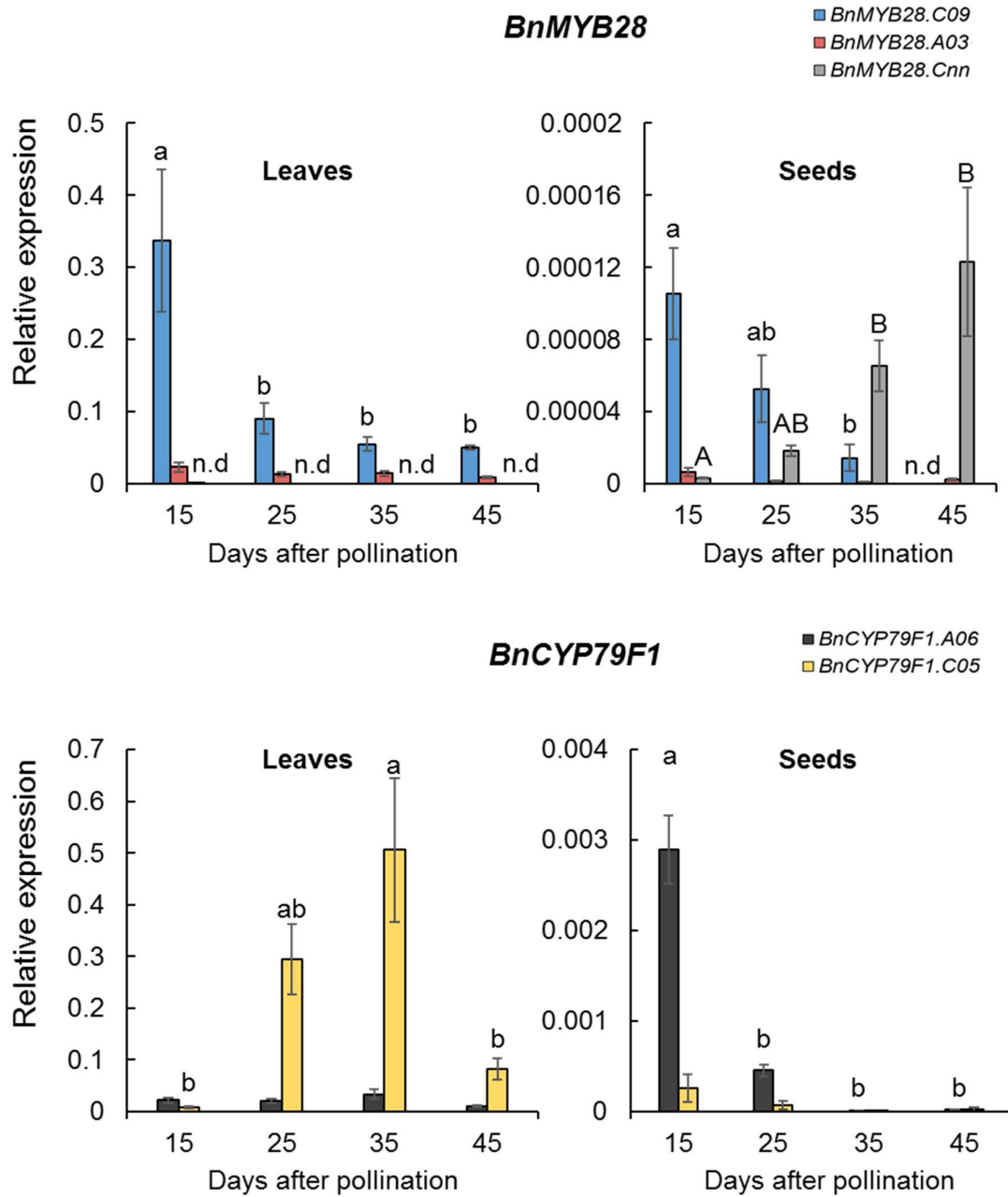
#### 3.3.2 Expression profiles of *BnMYB28* and *BnCYP79F1* genes reveal putative functional paralogs

We aimed to select the significantly expressed paralogs for knock-out studies. Therefore, we investigated the expression profiles of *BnMYB28* and *BnCYP79F1* genes in the German winter-type rapeseed Express617 by RT-qPCR. The relative expression of candidate genes was analyzed in leaves and seeds at growth stages 15, 25, 35 and 45 days after pollination (DAP).

The two *BnMYB28* paralogs *BnMYB28.C09* and *BnMYB28.A03* showed a thousandfold higher relative expression in leaves than seeds across all growth stages (Figure 8). *BnMYB28.C09* was the most highly expressed paralog in the leaves. However, expression sharply declined after 15 DAP and remained low during later stages of seed development (25-

45 DAP). The expression of *BnMYB28.A03* was consistently lower across all growth stages in the leaves, accounting for ~17% of the expression levels of *BnMYB28.C09*. In leaves, the expression of *BnMYB28.Cnn* was not detectable, whereas measurable expression levels were observed in the seeds at later stages of seed maturity. Conclusively, due to significantly higher expression levels, paralogs *BnMYB28.C09* and *BnMYB28.A03* were selected for further studies as putative functional paralogs of *BnMYB28*.

Similar to *BnMYB28*, *BnCYP79F1* paralogs were significantly more expressed in the leaves than seeds, attaining a hundredfold higher expression level. There was also a significant difference between the two *BnCYP79F1* paralogs in leaves as the expression of *BnCYP79F1.C05* was more than tenfold higher than *BnCYP79F1.A06* (Figure 8). Whereas in seeds, the opposite was observed. In the early stages of seed development (15 DAP), *BnCYP79F1.A06* was tenfold higher expressed than *BnCYP79F1.C05*, followed by a drastic decrease as the plants matured. In leaves, the expression of *BnCYP79F1.C05* sharply increased between 25 to 35 DAP and then decreased as the plants approached maturity. Interestingly, *BnMYB28* and *BnCYP79F1* genes displayed opposite expression patterns in the leaves. While *BnMYB28* was highly expressed at early stages (15 DAP) followed by a sharp decline, *BnCYP79F1* expression increased during later stages (25-35 DAP).

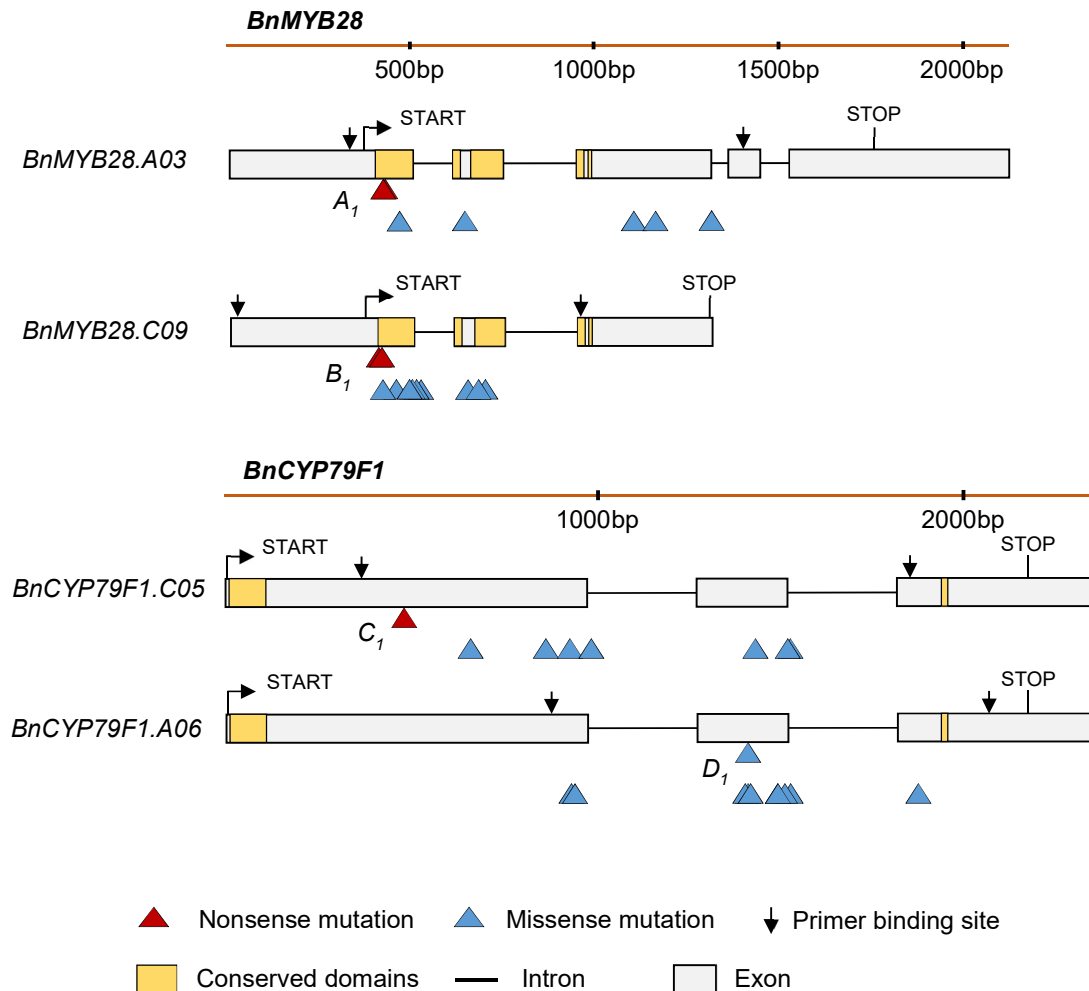


**Figure 8:** Relative expression of three *BnMYB28* and two *BnCYP79F1* paralogs in the winter-type oilseed rape Express617. Plants were grown in the greenhouse (20-25°C and 16 h light) after vernalization (4°C, 16 h light for 8 weeks). Leaves and seeds were sampled from five biological replicates 15, 25, 35 and 45 days after pollination. RT-qPCR was performed with three independent samples of each plant (three technical replicates) and the relative expression was calculated after normalization with two reference genes, *BnGAPDH* and *BnACTIN2*. Error bars represent the standard error of the mean of five biological replicates, with three technical replicates each. Statistical significance was calculated with an ANOVA ( $p < 0.05$ , linear model, grouping: Tukey test) using the R package ‘Agricolae’. Alphabets over bars represent statistical groups. n.d: not detectable.



### 3.3.3 EMS-induced mutations detected in the selected *BnMYB28* and *BnCYP79F1* paralogs

The M<sub>2</sub> population was screened for EMS-induced mutations in four genes (*BnMYB28.C09*, *BnMYB28.A03*, *BnCYP79F1.C05* and *BnCYP79F1.A06*) using a conventional polyacrylamide gel-based assay.



**Figure 9:** Location of EMS-induced nonsense and missense mutations detected in *BnMYB28* and *BnCYP79F1* genes. Allele identities are given next to the mutation site for the mutations used for further studies (refer Table 6 for all allele codes). Regions coding for functional and conserved domains characteristic to the gene families are marked in orange boxes. START and STOP represent the translation start and stop sites, respectively. Primer binding sites (downward arrows) represent the extent of the amplicons within which EMS-induced mutations were screened. For *BnCYP79F1*, the 5' untranslated regions are not defined on the Darmor-*bzh* reference genome.

35 and 43 EMS-induced mutations were detected in *BnMYB28* and *BnCYP79F1* paralogs, respectively (Supplementary Table 7), which could be categorized as 6 nonsense, 50 missense and 22 silent mutations (Figure 9). Splice site mutations could be found for any of the paralogs. Mutation frequencies ranged between 1/31.5 kb - 1/67.0 kb across the two gene families. An average mutation frequency of one EMS-induced mutation per 43.6 kb was estimated (Supplementary Table 8), which is in the range calculated in former studies on this EMS population (Guo et al., 2014; Emrani et al., 2015; Braatz et al., 2018; Shah et al., 2018;

Sashidhar et al., 2019; Karunarathna et al., 2020). All premature nonsense mutations detected for paralogs *BnMYB28.C09* and *BnMYB28.A03* were located within the conserved DNA binding R2 domain. Additionally, seven missense mutations were also detected within the R2 domain. We found one nonsense mutation for *BnCYP79F1.C05*, while none could be detected for *BnCYP79F1.A06*. For further studies, we selected all M<sub>3</sub> plants with nonsense mutations plus a *BnCYP79F1.A06* missense mutant possessing a slight modification to the protein structure. For ease of understanding, we assigned unique one-letter allele codes to wildtype and mutant alleles (Table 6).

**Table 6:** Allele codes assigned to EMS mutants and wildtype plants selected as crossing parents in this study. Single mutants were selected as crossing parents to combine single mutations for enhanced phenotypic effects. For each of the analyzed paralogs, mutants and wildtype parents are assigned unique allele codes.

	Gene name	Mutation position on gDNA <sup>[1]</sup>	Allele code	cDNA change <sup>[1]</sup>	AA change	Mutation type	M <sub>3</sub> seed code <sup>[2]</sup>
M <sub>3</sub> single mutants	<i>BnMYB28.C09</i>	G51A	<i>A<sub>l</sub></i>	G51A	W17*	Nonsense	190623
	<i>BnMYB28.A03</i>	G50A	<i>B<sub>l</sub></i>	G50A	W17*	Nonsense	190625
	<i>BnCYP79F1.C05</i>	C424T	<i>C<sub>l</sub></i>	C424T	E142*	Nonsense	190628
	<i>BnCYP79F1.A06</i>	G1379A	<i>D<sub>l</sub></i>	G1090A	E364K	Missense	190630
Wildtype Express617	<i>BnMYB28.C09</i>	n.a	<i>A<sub>e</sub></i>			n.a	
	<i>BnMYB28.A03</i>		<i>B<sub>e</sub></i>				
	<i>BnCYP79F1.C05</i>		<i>C<sub>e</sub></i>				
	<i>BnCYP79F1.A06</i>		<i>D<sub>e</sub></i>				
Wildtype Peace	<i>BnMYB28.C09</i>	n.a	<i>A<sub>p</sub></i>			n.a	
	<i>BnMYB28.A03</i>		<i>B<sub>p</sub></i>				
	<i>BnCYP79F1.C05</i>		<i>C<sub>p</sub></i>				
	<i>BnCYP79F1.A06</i>		<i>D<sub>p</sub></i>				

[1] Position with reference to the translation start site.

[2] Seed codes correspond to the M<sub>3</sub> seeds harvested from the selected M<sub>2</sub> mutants.

Non-mutated wildtype alleles from Express617 and Peace are represented with the ‘e’ and ‘p’ suffixes in subscript, respectively.

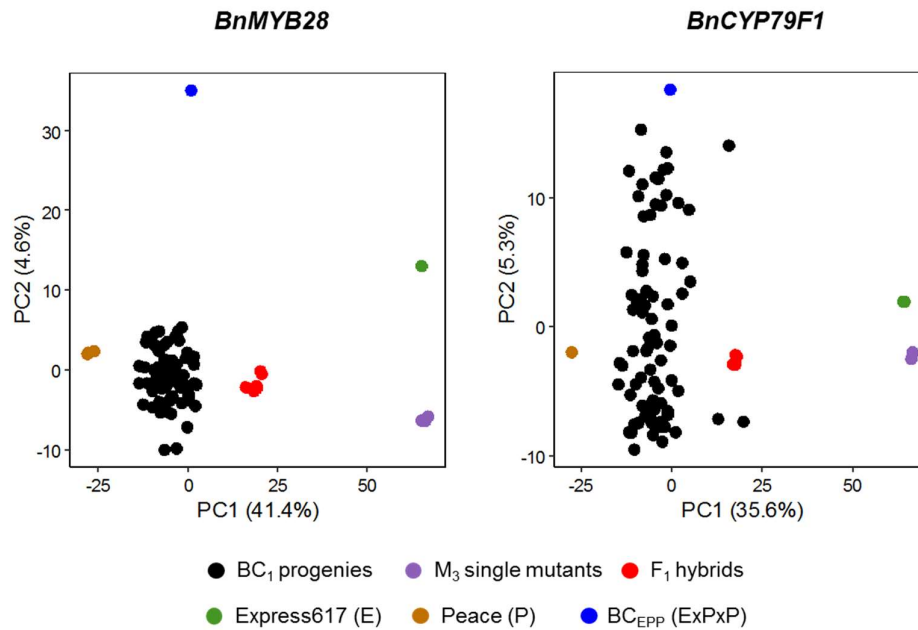
\*Premature stop codon, n.a: not applicable.

Due to high gene redundancies in polyploid oilseed rape, single mutations rarely have a phenotypic effect. Therefore, single mutants of paralogs *BnMYB28.C09* and *BnMYB28.A03* were crossed to produce *BnMYB28* double mutants. Similarly, single mutants of paralogs *BnCYP79F1.C05* and *BnCYP79F1.A06* were crossed to produce *BnCYP79F1* double mutants. M<sub>3</sub> plants were genotyped by generating PCR fragments encompassing expected mutations (Supplementary Table 5) and then validated by Sanger sequencing of the PCR fragments (Supplementary Table 9). M<sub>3</sub> plants possessing the expected mutations were selected as parents for three types of handcrosses (Supplementary Table 10). (i) M<sub>3</sub> single mutants of *BnMYB28* and *BnCYP79F1* paralogs were crossed with each other (referred to as ‘M<sub>3</sub>xM<sub>3</sub>’). F<sub>1</sub> offspring were selfed to generate the F<sub>2</sub> populations 200527 and 200529, representing *BnMYB28* and *BnCYP79F1* double mutants, respectively (Supplementary Figure

11A). (ii) Heterozygous *BnMYB28* double mutants were backcrossed with Peace to produce the BC<sub>1</sub>F<sub>2</sub> population 210465 (Supplementary Figure 11B). (iii) Homozygous single mutants (*BnCYP79F1.C05* and *BnCYP79F1.A06*) were backcrossed with the Peace wildtype to produce BC<sub>1</sub> populations 200504 and 200508. Heterozygous BC<sub>1</sub> plants were then crossed with each other to produce the BC<sub>1</sub>F<sub>2</sub> population 210462 (Supplementary Figure 12).

### 3.3.4 The *Brassica* 19K SNP array for marker-assisted background selection

The aim was to select homozygous mutants with a low mutation load in early backcross generations, as was recently published by Karunaratna et al. (2021). 159 BC<sub>1</sub> plants from ten backcross populations (named 'BC<sub>A</sub>-BC<sub>J</sub>') were selected for genotyping with the SNP array (Supplementary Table 12). The six populations BC<sub>A</sub>-BC<sub>F</sub> were produced after backcrossing heterozygous *BnMYB28* double mutants with Peace. The remaining four populations, BC<sub>G</sub>-BC<sub>J</sub> were generated after backcrossing *BnCYP79F1.C05* and *BnCYP79F1.A06* single mutants with Peace. All BC<sub>1</sub> plants, M<sub>3</sub> single mutants, F<sub>1</sub> hybrids, and the two parents were genotyped using the *Brassica* 19K SNP chip, with 18,567 SNP markers. Additionally, the progeny of a backcross between the non-mutated Express617 and Peace wildtypes (named 'BC<sub>EPP</sub> (ExPxP)') was used as a control. Genotyping experiments were conducted separately for populations BC<sub>A</sub>-BC<sub>F</sub> and BC<sub>G</sub>-BC<sub>J</sub> with 8,961 and 8,334 polymorphic markers (for the Express617 and Peace genomes) in the two experiments, respectively (Supplementary Table 11).



**Figure 10:** Principal component analysis showing relatedness between parents, mutant lines, and their offspring. Homozygous M<sub>3</sub> plants carrying mutations in either *BnMYB28* (genotypes *A<sub>1</sub>A<sub>1</sub>B<sub>e</sub>B<sub>e</sub>* and *A<sub>e</sub>A<sub>e</sub>B<sub>1</sub>B<sub>1</sub>*) or *BnCYP79F1* (genotypes *C<sub>1</sub>C<sub>1</sub>D<sub>e</sub>D<sub>e</sub>* and *C<sub>e</sub>C<sub>e</sub>D<sub>1</sub>D<sub>1</sub>*) genes were crossed to produce F<sub>1</sub> hybrids (genotypes *A<sub>1</sub>A<sub>e</sub>B<sub>1</sub>B<sub>e</sub>* and *C<sub>1</sub>C<sub>e</sub>D<sub>1</sub>D<sub>e</sub>*). F<sub>1</sub> hybrids were backcrossed with the spring variety Peace to produce BC<sub>1</sub> progenies. 63 *BnMYB28* BC<sub>1</sub> plants (from populations BC<sub>A</sub>-BC<sub>F</sub>) and 86 *BnCYP79F1* BC<sub>1</sub> plants (from populations BC<sub>G</sub>-BC<sub>J</sub>) were genotyped with a *Brassica* 19K SNP array. Non-mutagenized Peace (P) and Express617 (E) plants and their backcrossed offspring named 'BC<sub>EPP</sub> (ExPxP)' were used as controls. Principal components were calculated by TASSEL and visualized using the 'ggplot2' package (Wickham, 2009) on R.

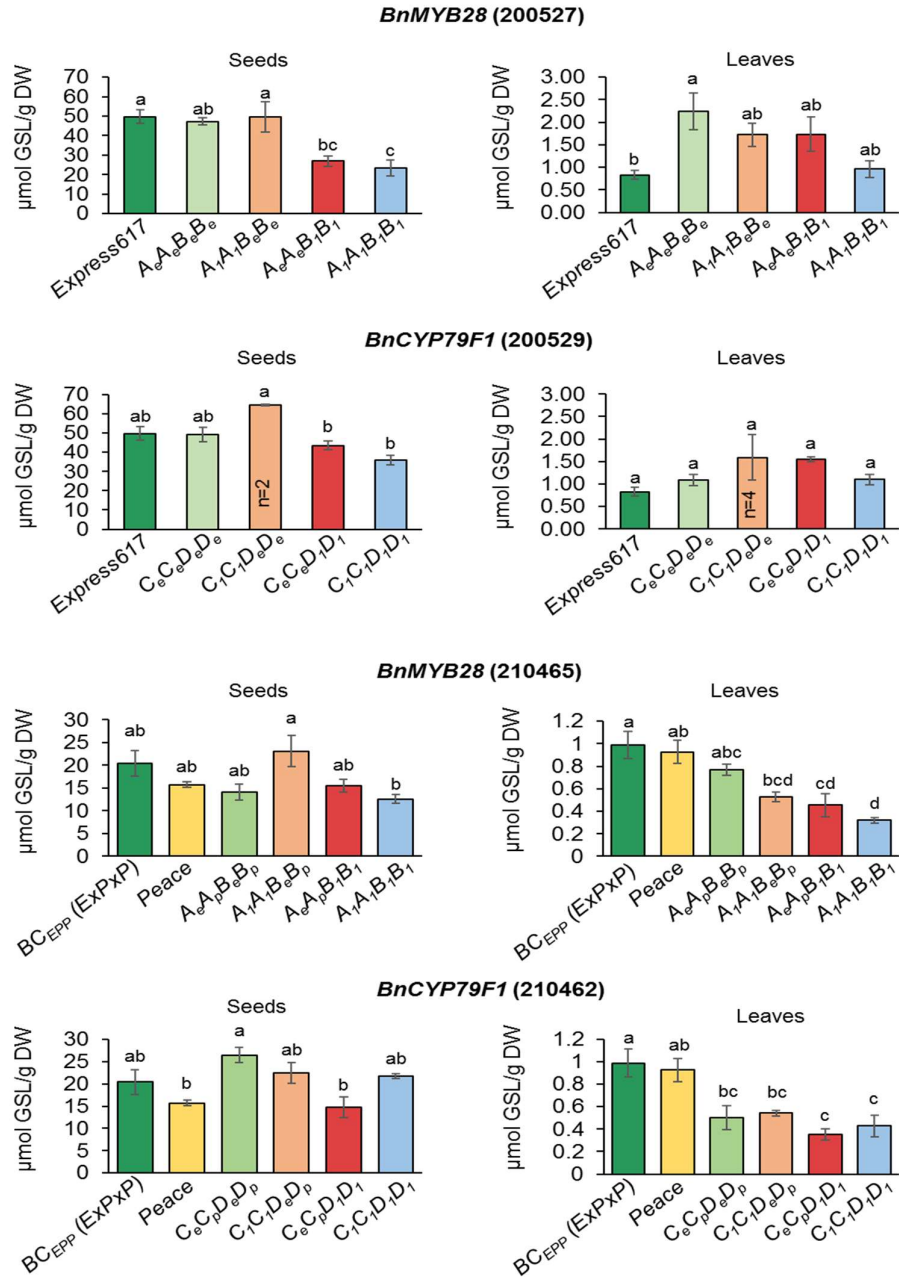
First, principal component (PC) analyses were performed separately for populations BC<sub>A</sub>-BC<sub>F</sub> and BC<sub>G</sub>-BC<sub>J</sub> to visualize the population structure of the backcrossed mutants with respect to their parents. As expected, the parents (Express617 and Peace), the M<sub>3</sub> single mutants and the F<sub>1</sub> hybrids clustered separately from the BC<sub>1</sub> progenies. The BC<sub>1</sub> plants carrying the EMS-induced mutations were observed within the same cluster at an intermediate region between the recurrent Peace parent and the F<sub>1</sub> hybrids, farthest from the wild type Express617. This indicated that the backcrossed plants carried a higher Peace background. In PC analyses for both populations, the non-mutagenized Express617 plants clustered in the neighborhood of the M<sub>3</sub> single mutants. Furthermore, the 'BC<sub>EPP</sub> (ExPxP)' controls were observed closer to the BC<sub>1</sub> mutants. A considerable genetic diversity arising possibly due to the high EMS mutation load was apparent. The variation occurring due to the EMS treatment is observable in the PC2, explaining 4.6% and 5.3% of the genetic variation for populations BC<sub>A</sub>-BC<sub>F</sub> and BC<sub>G</sub>-BC<sub>J</sub>, respectively (Figure 10).

Then, the share of parent-specific alleles in the backcrossed generations was analyzed. As expected, all F<sub>1</sub> hybrids possessed ~50% Peace and Express617 genome shares. A high percentage of the backcross parent genome was reasoned with a low mutation load. In backcrossed populations BC<sub>A</sub>-BC<sub>F</sub> segregating for *BnMYB28.C09* and *BnMYB28.A03* mutations, the Peace genome share ranged from 69.3% - 85.5%. In populations BC<sub>G</sub>-BC<sub>J</sub> segregating for *BnCYP79F1.C05* and *BnCYP79F1.A06* mutations, the Peace genome share was 52.2% - 78.5%. Therefore, plants with moderately reduced mutation loads attributed to higher Peace genome shares of >75% could be selected, which is expected as the average in the BC<sub>1</sub> generation (Supplementary Table 12). Selected BC<sub>1</sub> plants were chosen as crossing parents for combining mutations. BC<sub>1</sub> plants of *BnMYB28* double mutants were selfed to produce the BC<sub>1</sub>F<sub>2</sub> population (seed code 210465). BC<sub>1</sub> single mutants of *BnCYP79F1.C05* (seed code 200504) and *BnCYP79F1.A06* (seed code 200508) were then crossed to subsequently produce the BC<sub>1</sub>F<sub>2</sub> population 210462.

### 3.3.5 Seeds of *BnMYB28* and *BnCYP79F1* double mutants possess significantly reduced aliphatic GSLs

The effect of EMS-induced mutations in *BnMYB28* and *BnCYP79F1* genes was investigated by quantitative and qualitative GSL analysis. Segregating individuals from four F<sub>2</sub> populations (seed codes 200527, 200529, 210465 and 210462) were analyzed (Supplementary Table 10) along with non-mutagenized Express617, Peace and 'BC<sub>EPP</sub> (ExPxP)' plants as controls. F<sub>2</sub> populations 200527 and 200529 originated from M<sub>3</sub>xM<sub>3</sub> crosses of *BnMYB28* and *BnCYP79F1* single mutants, respectively. Whereas BC<sub>1</sub>F<sub>2</sub> populations 210465 and 210462 were derived from double mutants of *BnMYB28* and *BnCYP79F1* backcrossed with Peace, respectively.

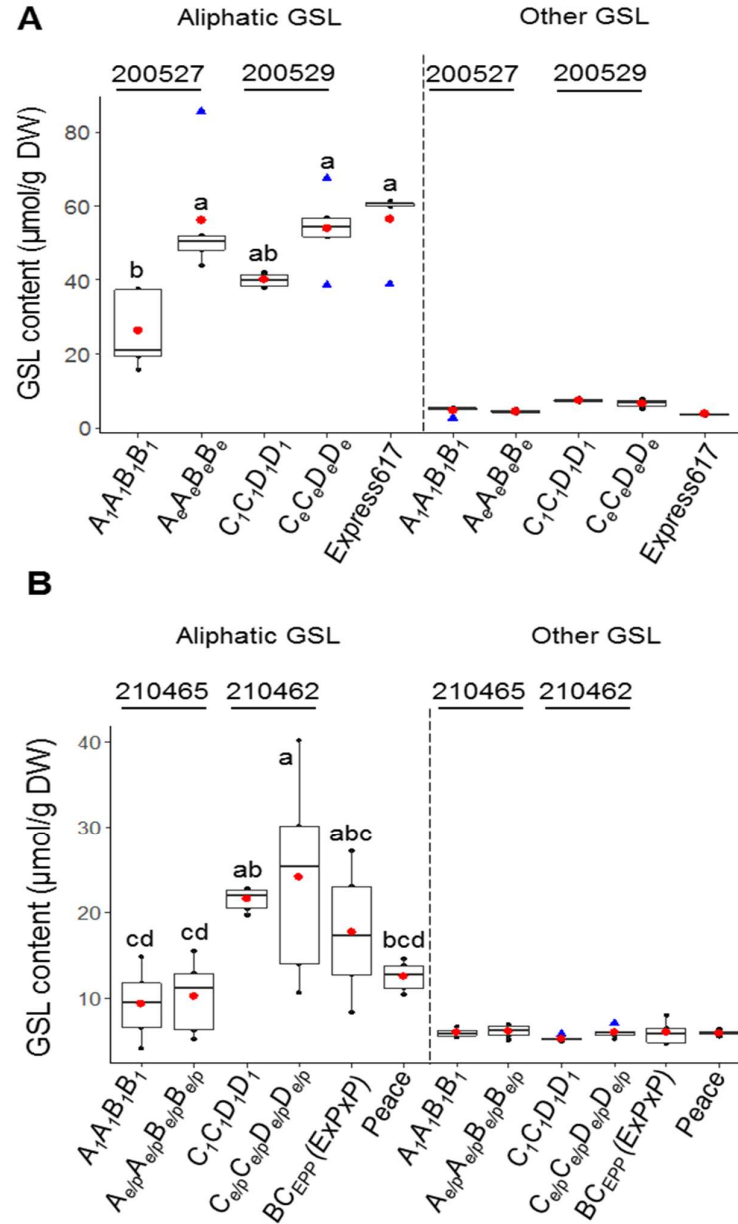
First, F<sub>2</sub> plants from the four populations were genotyped by Sanger sequencing of the PCR amplicons encompassing expected mutations (Supplementary Table 5). Four genotypes per population were selected for phenotypic studies. In F<sub>2</sub> populations 200527 and 200529, plants homozygous for the wildtype alleles (*A<sub>e</sub>A<sub>e</sub>B<sub>e</sub>B<sub>e</sub>* and *C<sub>e</sub>C<sub>e</sub>D<sub>e</sub>D<sub>e</sub>*), homozygous for one mutant allele (*A<sub>1</sub>A<sub>1</sub>B<sub>e</sub>B<sub>e</sub>* or *A<sub>e</sub>A<sub>e</sub>B<sub>1</sub>B<sub>1</sub>* and *C<sub>1</sub>C<sub>1</sub>D<sub>e</sub>D<sub>e</sub>* or *C<sub>e</sub>C<sub>e</sub>D<sub>1</sub>D<sub>1</sub>*) and homozygous for two mutant alleles (*A<sub>1</sub>A<sub>1</sub>B<sub>1</sub>B<sub>1</sub>* and *C<sub>1</sub>C<sub>1</sub>D<sub>1</sub>D<sub>1</sub>*) were selected for phenotyping. Similarly, plants segregating for wildtype alleles and homozygous single and double mutant alleles were also selected from backcrossed BC<sub>1</sub>F<sub>2</sub> populations 210465 and 210462. For GSL determination, the total content and composition were determined in seeds and leaves of plants from the selected genotypes.



**Figure 11:** Seed and leaf glucosinolate contents in populations segregating for *BnMYB28* and *BnCYP79F1* mutations. F<sub>2</sub> populations 200527 and 200529 segregating for *BnMYB28* and *BnCYP79F1* mutations, respectively, originated from M<sub>3</sub>xM<sub>3</sub> crosses. Homozygous F<sub>2</sub> double mutants ( $A_1A_1B_1B_1$  and  $C_1C_1D_1D_1$ ) were analyzed together with homozygous single mutants ( $A_1A_1B_eB_e$ ,  $A_eA_eB_1B_1$ ,  $C_1C_1D_eD_e$  and  $C_eC_eD_1D_1$ ), non-mutagenized Express617 and F<sub>2</sub> plants homozygous for the wildtype alleles ( $A_eA_eB_eB_e$  and  $C_eC_eD_eD_e$ ). The BC<sub>1</sub>F<sub>2</sub> populations 210465 and 210462 originate from backcrosses with Peace. Homozygous double mutants ( $A_1A_1B_1B_1$  and  $C_1C_1D_1D_1$ ) were analyzed with homozygous single mutants, non-mutagenized Peace, BC<sub>1</sub>F<sub>2</sub> plants homozygous for the wildtype alleles and the backcrossed progeny of non-mutagenized Express617 and Peace 'BC<sub>EPP</sub> (ExPxP)' as the controls. Plants were grown in the greenhouse. Leaf samples were taken at 15 days after pollination and mature seeds (BBCH89) were used for glucosinolate determination. Error bars represent the standard error from five plants (n=5) per genotype with two exceptions mentioned in the figure. An ANOVA ( $p < 0.05$ ) was performed and Tukey test ( $p < 0.05$ ) was done for grouping. Different alphabets (a-d) above error bars represent groups based on significance. All genotypes are as per designated allele codes given in Table 6.

In general, a significantly higher GSL content was observed in seeds than in leaves (Figure 11). In seeds of *BnMYB28* double mutants (F<sub>2</sub> population 200527), the GSL content was significantly ( $p < 0.05$ ) reduced to 23.41  $\mu\text{mol/g DW}$  compared to F<sub>2</sub> plants homozygous for the wildtype alleles (47.28  $\mu\text{mol/g DW}$ ) and the non-mutagenized Express617 control plants (49.73  $\mu\text{mol/g DW}$ ), which corresponds to a significant GSL reduction by 50.5% and 52.9%, respectively (Figure 11). No significant differences were observed between the leaves of *BnMYB28* double mutants (0.96  $\mu\text{mol/g DW}$ ) and the Express617 control (0.86  $\mu\text{mol/g DW}$ ). Also, the *BnCYP79F1* double mutants (F<sub>2</sub> population 200529) showed reduced seed GSL contents by 27.9% and 26.9% compared to the Express617 controls and the F<sub>2</sub> plants homozygous for the wild type alleles, respectively. However, the difference was not statistically significant ( $p < 0.05$ ). GSL contents in leaves varied between 0.9-1.5  $\mu\text{mol/g DW}$  in the *BnCYP79F1* double mutants and 0.7-1.5  $\mu\text{mol/g DW}$  in the wildtype F<sub>2</sub> plants without statistically significant differences between the genotypes.

Then, individual seed GSLs were measured in the same plants as used for total GSL measurements. Single GSLs were identified by retention time and co-chromatography with commercial standards and quantified using individual calibration curves (Supplementary Table 6). The estimation of total GSLs by summing up the major compounds identified by HPLC yielded generally higher values than the enzymatic method. However, an ANOVA test ( $p < 0.05$ ) did not reveal any significant differences in the estimated GSL content by the two methods for the analyzed mutants (Supplementary Figure 13). This strongly supports our data evaluation for single GSL identification and their estimated concentrations.



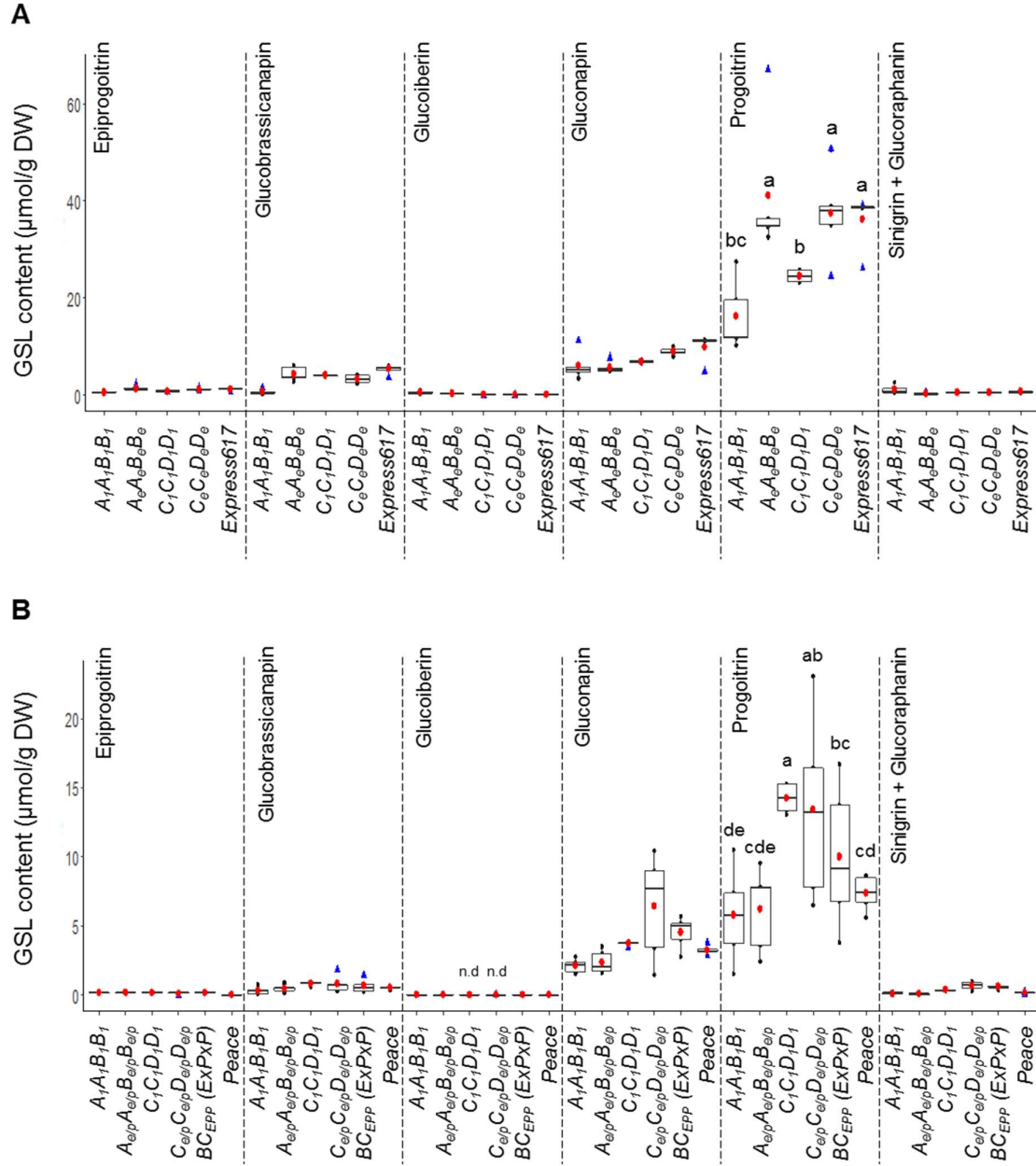
**Figure 12:** Analysis of major glucosinolate types in mature seeds of *BnMYB28* and *BnCYP79F1* mutants. A) Aliphatic and other GSL types measured from homozygous *BnMYB28* (genotype *A<sub>1</sub>A<sub>1</sub>B<sub>1</sub>B<sub>1</sub>*, seed code 200527) and *BnCYP79F1* (genotype *C<sub>1</sub>C<sub>1</sub>D<sub>1</sub>D<sub>1</sub>*, seed code 200529) originating from M<sub>3</sub>xM<sub>3</sub> crosses. F<sub>2</sub> plants homozygous for the wild type alleles (*A<sub>e</sub>A<sub>e</sub>B<sub>e</sub>B<sub>e</sub>* and *C<sub>e</sub>C<sub>e</sub>D<sub>e</sub>D<sub>e</sub>*) and non-mutated Express617 were used as controls. B) Aliphatic and other GSL types measured from homozygous BC<sub>1</sub>F<sub>2</sub> *BnMYB28* (genotype *A<sub>1</sub>A<sub>1</sub>B<sub>1</sub>B<sub>1</sub>*, seed code 210465) and *BnCYP79F1* (genotype *C<sub>1</sub>C<sub>1</sub>D<sub>1</sub>D<sub>1</sub>*, seed code 210462) mutants originating from backcrosses with the spring oilseed rape Peace. BC<sub>1</sub>F<sub>2</sub> plants homozygous for the wildtype alleles (*A<sub>e/p</sub>A<sub>e/p</sub>B<sub>e/p</sub>B<sub>e/p</sub>* and *C<sub>e/p</sub>C<sub>e/p</sub>D<sub>e/p</sub>D<sub>e/p</sub>*) and the backcrossed progeny of non-mutagenized wildtype Peace and Express617 ‘BC<sub>EPP</sub> (ExPxP)’ were used as controls. Plants were grown in the greenhouse. Mature seeds were harvested at BBCH89. GSL profiles for aliphatic and other GSL types (phenolic and indolic) were analyzed using HPLC. GSL content was calculated as μmol/g dry weight (DW). Individual and mean values are marked in black and red dots, respectively. Blue triangles represent outliers. Error bars represent the standard error of mean from five biological replicates. An ANOVA test ( $p < 0.05$ ) was performed and significant differences between groups were calculated by a Tukey test ( $p < 0.05$ ). Different alphabets (a-d) above error bars represent groups based on significance. All genotypes are as per designated allele codes given in Table 6.

We detected seven aliphatic GSLs (epiprogoitrin, glucobrassicinapin, glucoiberin, gluconapin, glucoraphanin, progoitrin and sinigrin) in varying quantities and six other GSLs (4-hydroxyglucobrassicin, 4-methoxyglucobrassicin, glucobrassicin, gluconasturtiin, glucotropaeolin and neoglucobrassicin) in trace amounts in the seeds. In line with previous reports on rapeseed (Velasco et al., 2008), the aliphatic GSL comprised the major share (94%) of the seed GSL in all genotypes studied here (Figure 12).

In *BnMYB28* F<sub>2</sub> double mutants (population 200527), the progoitrin concentrations in seeds were 55.3% lower than in the Express617 controls (reduced from 36.32  $\mu\text{mol/g DW}$  to 16.20  $\mu\text{mol/g DW}$ ) (Figure 13A, Supplementary Table 13). The next abundant aliphatic compounds in the seeds were gluconapin and glucobrassicinapin. While glucobrassicinapin levels were drastically reduced by 87% from 5.26  $\mu\text{mol/g DW}$  in the Express617 controls DW to 0.64  $\mu\text{mol/g DW}$  in the double mutants, the gluconapin content was not significantly reduced. The minor aliphatic compound epiprogoitrin, whose synthesis starts from gluconapin, was reduced in the double mutants by 51% (0.57  $\mu\text{mol/g DW}$ ) compared to the Express617 controls (1.17  $\mu\text{mol/g DW}$ ). The remaining seed GSLs analyzed did not exceed 3  $\mu\text{mol/g DW}$ . Out of the 28  $\mu\text{mol/g DW}$  reduction observed in the total seed GSL content of double mutants, the three major aliphatic GSLs accounted for 86% (24  $\mu\text{mol/g DW}$ ) of the total reduction. In the *BnCYP79F1* F<sub>2</sub> double mutants, progoitrin was 32.4% lower than in the Express617 controls, reducing from 36.32  $\mu\text{mol/g DW}$  to 24.5  $\mu\text{mol/g DW}$  (Figure 13A, Supplementary Table 14). The glucobrassicinapin content was not altered in *BnCYP79F1* double mutants. However, the 30.4% decrease in gluconapin content suggests that the mutations in the *BnCYP79F1* mutants might have a bigger effect on the synthesis of short-chained 4C aliphatic than on the 5C aliphatic GSLs.

A difference in the GSL content and composition between the spring-type Peace and the winter-type Express617 was observed. The seed GSL content in Peace (15.7  $\mu\text{mol/g DW}$ ) was strikingly lower compared to Express617 (49.7  $\mu\text{mol/g DW}$ ) (Figure 11 and Figure 12). Interestingly, the leaf GSL content was roughly the same (Figure 11). In general, plants from backcrossed populations 210465 and 210462 showed lower seed GSL than their counterpart non-backcrossed populations 200527 and 200529. This reflects the influence of the recurrent low-GSL Peace genotype in backcrossed generations. In the seeds of the two backcrossed populations, no significant changes for the GSL content were observed across the tested genotypes compared to the wildtype controls (Figure 11). Although detected in much lower quantities in the leaves than seeds, GSL content was reduced significantly in the leaves of backcrossed *BnMYB28* and *BnCYP79F1* double mutants. When compared to the Peace (0.99  $\mu\text{mol/g DW}$ ) and BC<sub>EPP</sub> (0.93  $\mu\text{mol/g DW}$ ) control plants, GSL content in leaves of *BnMYB28* double mutants (0.32  $\mu\text{mol/g DW}$ ) was reduced by 65.5% and 67.7%, respectively. Similarly, GSL content in leaves of *BnCYP79F1* double mutants (0.43  $\mu\text{mol/g DW}$ ) was significantly reduced by 53.8% compared to the Peace wildtype and 56.7% compared to the BC<sub>EPP</sub> control plants. Strikingly, the leaf GSL content for all individuals of BC<sub>1</sub>F<sub>2</sub> population 210462 was generally lower than the control plants (Figure 11).





**Figure 13:** Analysis of aliphatic glucosinolates in mature seeds of *BnMYB28* and *BnCYP79F1* mutants. A) Individual aliphatic GSLs in homozygous F<sub>2</sub> *BnMYB28* (genotype  $A_1A_1B_1B_1$ , seed code 200527) and *BnCYP79F1* (genotype  $C_1C_1D_1D_1$ , seed code 200529) double mutants originating from M<sub>3</sub>xM<sub>3</sub> crosses. F<sub>2</sub> plants homozygous for the wild type alleles ( $A_eA_eB_eB_e$  and  $C_eC_eD_eD_e$ ) and non-mutated Express617 were used as controls. B) Individual aliphatic GSLs identified and measured from homozygous *BnMYB28* (genotype  $A_1A_1B_1B_1$ , seed code 210465) and *BnCYP79F1* (genotype  $C_1C_1D_1D_1$ , seed code 210462) BC<sub>1</sub>F<sub>2</sub> mutants originating from backcrosses with the spring oilseed rape Peace. BC<sub>1</sub>F<sub>2</sub> plants homozygous for the wildtype alleles ( $A_{elp}A_{elp}B_{elp}B_{elp}$  and  $C_{elp}C_{elp}D_{elp}D_{elp}$ ) and the backcrossed progeny of non-mutagenized wildtype Peace and Express617 ‘BC<sub>EPP</sub> (ExPxP)’ were used as controls. Plants were grown in the greenhouse. Mature seeds were harvested at BBCH89. Aliphatic GSLs were identified and quantified using HPLC. The content was calculated as μmol/g dry weight (DW). Individual and mean values are marked in black and red dots, respectively. Blue triangles represent outliers. Error bars represent the standard error of mean from biological replicates. An ANOVA test ( $p < 0.05$ ) was performed and significant differences between groups were calculated by a Tukey test ( $p < 0.05$ ). Different alphabets (a-e) above error bars represent groups based on significance. All genotypes are as per designated allele codes given in Table 6. n.d.: not detectable.

In qualitative analyses on plants from backcrossed populations 210465 and 210462, no significant reductions in the total GSL content or the aliphatic profile were observed (Figure 12B). The impact of the recurrent Peace genome was also observed on the composition of aliphatic GSLs. In general, backcrossed populations were detected with lower progoitrin levels than non-backcrossed populations (Figure 13). In the non-mutagenized controls and the F<sub>2</sub> plants homozygous for the wildtype alleles from non-backcrossed populations 200527 and 200529, progoitrin accounted for 68%-78% of all major identified GSLs (Figure 13A). However, in backcrossed populations 210465 and 210462, the progoitrin share was reduced to 56%-61% in the non-mutagenized controls and BC<sub>1</sub>F<sub>2</sub> plants homozygous for the wildtype alleles (Figure 13B).

Noteworthy, an increased variation in the GSL content within biological replicates of the same genotype was observed for the backcrossed mutants. While GSL content across biological replicates of the Peace wildtypes did not vary significantly (14.3–17.5 µmol/g DW), seed GSL content varied between 9.8–15.7 µmol/g DW for *BnMYB28* double mutants and 20.6–23.5 µmol/g DW for *BnCYP79F1* double mutants. Moreover, wildtype plants segregating within the same BC<sub>1</sub>F<sub>2</sub> populations showed the most significant variation for GSL content ranging between 8.8–30.3 µmol/g DW. Consequently, a statistically significant change in the GSL content or composition could not be ascertained for backcrossed double mutants of *BnMYB28* or *BnCYP79F1*.

### 3.4 Discussion

Major anti-nutritive compounds like glucosinolates (GSL) in the rapeseed meal (RSM) pose a challenge for utilization as animal feed. Therefore, a major breeding target for rapeseed is reducing the seed glucosinolate content (SGC) to an acceptable limit of <18 µmol/g dry weight. This study aimed to reduce the aliphatic GSL content in seeds by knocking out genes *BnMYB28* and *BnCYP79F1* involved in the biosynthesis of aliphatic GSLs in rapeseed. We demonstrate that independent knock-out mutants of the two genes possessed significantly reduced total and aliphatic GSLs, primarily progoitrin, in the seeds.

We targeted the aliphatic GSL biosynthesis pathway since the aliphatic profile comprises the majority of up to 92% of all GSLs in rapeseed (Velasco et al., 2008). In our study, the aliphatic GSLs comprised ~94% of all GSLs in the seeds. Moreover, major GSLs like progoitrin that have adverse metabolic effects in animals belong to the aliphatic profile (Fenwick & Curtis, 1980; Fahey et al., 2001; Tripathi & Mishra, 2007). We reasoned that functional mutations in genes involved in the secondary modification of GSLs might only confer an altered GSL profile and not a significant reduction in the overall content. Therefore, we selected candidate genes *BnMYB28* and *BnCYP79F1* due to their prominent role in the core structure formation of aliphatic GSLs (Reintanz et al., 2001; Gigolashvili et al., 2007). In Arabidopsis, a transcriptome study confirmed the role of sub-group 12 R2R3-MYB TFs in up-regulating almost all genes involved in the core structure formation of aliphatic GSLs. For aliphatic GSL biosynthesis, *MYB28* was found to co-express with almost all genes of the chain elongation and core structure formation (Hirai et al., 2007). In associative transcriptomics and QTL mapping studies, former studies have found *MYB28* and *CYP79F1* to be strongly associated with a high aliphatic GSL content in rapeseed (Feng et al., 2012; Li et al., 2014). More recently, Kittipol et al. (2019) and S. Liu et al. (2020) have also identified *MYB28* as a significant gene controlling aliphatic GSL content in rapeseed using transcriptome and genome-wide association studies, respectively.

It has been shown that GSL biosynthesis occurs in vegetative parts, especially in rosette leaves and silique walls (Toroser et al., 1995). Using histochemical analyses in Arabidopsis, Reintanz et al. (2001) have demonstrated that the activity of the biosynthesis gene *CYP79F1* is restricted to the silique walls and almost undetectable in seeds. In Arabidopsis, a reduction in leaf GSL content was followed by an increasing GSL content in the seeds towards maturity, suggesting the role of a GSL transport mechanism rather than biosynthesis in seeds (Brown et al., 2003). Moreover, using *in silico* microarray analyses in Arabidopsis seeds, Nour-Eldin and Halkier (2009) have reported the insignificant expression of genes involved in the chain elongation and GSL core structure formation steps including *MYB28* and *CYP79F1* genes. Thus, in these studies, the negligible expression levels of biosynthesis genes suggested the inability of seeds for *de novo* GSL biosynthesis.

Our expression analyses for *BnMYB28* and *BnCYP79F1* paralogs encompass seed setting and loading phases between 15-45 days after pollination (DAP). In our study, the expression profiles observed for the two biosynthesis genes complement previous studies since expression levels were significantly higher in leaves than seeds. *BnMYB28* paralogs were expressed at almost a thousandfold higher level in the leaves. Relative expression levels were higher during the early stages of growth (15 DAP), with a gradual decrease as the plant approached maturity (45 DAP). This was expected since GSL biosynthesis is high at the onset of the generative phase (Brown et al., 2003; Gigolashvili et al., 2007). *BnMYB28.C09* showed the most significant expression levels over other paralogs in the leaves. In previous mapping studies, QTL significantly associated with high aliphatic GSL content in leaves and seeds of rapeseed were linked to *BnMYB28* on chromosome C09 (Howell et al., 2003; Harper et al., 2012; Li et al., 2014; Qu et al., 2015). Although in trace amounts, *BnMYB28.Cnn* showed higher relative expression than other paralogs in seeds towards maturity. On the contrary, its expression was undetectable in leaves. Since GSL biosynthesis is suggested to be absent in seeds (Gijzen et al., 1989; Nour-Eldin et al., 2012), we reason that the expression level of *BnMYB28.Cnn* is too low to affect the seed GSL content.

Most genes involved in aliphatic GSL biosynthesis, including *CYP79F1*, are under the transcriptional control of *MYB28* (Hirai et al., 2007). This was evident since a significantly high expression of *BnMYB28.C09* at 15 DAP was followed by a significant increase in *BnCYP79F1.C05* expression levels in the leaves at 25 DAP. Based on these data, we identified paralogs *BnMYB28.C09* and *BnCYP79F1.C05* with the most significant expression levels in leaves, the primary site for GSL biosynthesis. Therefore, we selected them as the most promising candidates for functional analyses. However, due to the high functional redundancy in the polyploid rapeseed genome, we cannot wholly rely on single mutants of highly expressed paralogs for significant phenotypic effects. Therefore, we also considered genes *BnMYB28.A03* and *BnCYP79F1.A06* for pyramiding functional mutants for enhanced phenotypic effects.

Former studies demonstrating EMS-induced random mutagenesis in Brassicaceae crops have reported a wide range of mutation frequencies between 1/12 kb to 1/447 kb (Wang et al., 2008; Himelblau et al., 2009; Stephenson et al., 2010; Harloff et al., 2012; Gilchrist et al., 2013; Tang et al., 2020). In this work, we screened the EMS mutant population of the winter rapeseed Express617 developed by Harloff et al. (2012). Past studies on this resource have estimated varying mutation frequencies of 1/12 kb – 1/72 kb (Guo et al., 2014; Emrani et al., 2015; Braatz et al., 2018; Shah et al., 2018; Sashidhar et al., 2019; Karunarathna et al., 2020). This variation is expected since frequency estimations depend on factors like the length of amplicons screened, the GC content within amplified fragments and the number of pools

screened for mutant detection. We estimated an average mutation frequency of 1/52.4 kb for *BnMYB28* and 1/34.7 kb for *BnCYP79F1*, which is well within the range expected from this mutant population.

The functionality of R2R3-MYB TFs is determined by the DNA binding R2 and R3 domains and a nuclear localization signal (Stracke et al., 2001; Gigolashvili et al., 2007). We detected EMS-induced premature nonsense mutations within the conserved R2 domain in the first exons of *BnMYB28.C09* and *BnMYB28.A03*. As a result, consequent transcripts are expected to lack the downstream R3 DNA binding and the NLS domains. For both single mutants, >95% of the resultant protein sequence is expected to be truncated. Therefore, we anticipate a complete loss of function of the *BnMYB28* TF in our double mutants.

We only detected missense mutations for the gene *BnCYP79F1.A06*. A lack of nonsense mutations detected for this gene was due to two reasons. The designed paralog-specific primers encompassed only 38% of the total cDNA sequence. Moreover, the amplicon possessed only five possible amino acid motifs with the possibility of converting to stop codons after the EMS-induced C→T or G→A transitions. Since we observed a nearly undetectable gene expression for *BnCYP79F1.A06* in the leaves compared to *BnCYP79F1.C05*, we speculate that its role in the biosynthesis process is less critical. Moreover, associative transcriptomic studies have reported *BnCYP79F1.C05* to be significantly associated with the aliphatic GSL content in rapeseed (Kittipol et al., 2019). Since a loss of 73% of the consequent polypeptide sequence is expected from the EMS-induced nonsense mutation in the *BnCYP79F1.C05* mutant, we anticipated a knock-out of this paralog.

A high mutation load is a serious drawback for utilizing EMS mutagenized plants. Moreover, residual mutations in the background can adversely affect subsequent generations (Jung & Till, 2021). The EMS mutant population used in this study has been estimated to carry ~46,000 mutations per M<sub>2</sub> plant on average (Karunarathna et al., 2020). Since these unknown mutations can confer adverse phenotypes like poor plant vigor and significant yield penalties, the direct application of EMS mutants in traditional breeding practices is severely limited. Therefore, backcrossing with non-mutagenized parents, preferably the EMS donor line or elite cultivars, is suggested. In rice, SNP arrays have been used to select backcrossed progenies for a higher share of the recurrent parent (Singh et al., 2015; Thomson et al., 2017). A similar approach was recently demonstrated for rapeseed using a *Brassica* 19K SNP array (Karunarathna et al., 2021). A combined strategy for EMS mutation load reduction and speed breeding for rapeseed was developed. Karunarathna et al. (2021) have demonstrated that the introgression of the spring genotype from Peace to the winter-type Express617 genome eliminated the vernalization requirement in the BC<sub>1</sub> generation. Using this strategy, we selected individuals possessing >75% of the recurrent Peace genome. Consequently, we were able to reduce the generation time by almost 50% in our study, and seeds could be harvested 3-4 months after sowing.

Based on functional studies from *Arabidopsis*, we expected that a knock-out of the two genes would severely influence the biosynthesis of short-chained aliphatic GSLs (Reintanz et al., 2001; Gigolashvili et al., 2007), especially progoitrin that accounts for ~80% of all GSLs in the seeds (Halkier & Du, 1997). Since progoitrin is also the most abundant seed GSL type in rapeseed (Velasco et al., 2008), we anticipated the most significant changes in its levels in the mutants. In this regard, a more significant effect from *BnMYB28* mutants was expected for its prominent regulatory control over the aliphatic GSL biosynthesis (Gigolashvili et al., 2007;

Hirai et al., 2007). Therefore, a knock-out of the *BnMYB28* is expected to downregulate several genes involved in the biosynthesis process. To validate this, expressional analyses of major downstream targets like *MAM3* (Textor et al., 2007) involved in chain elongation and *CYP79F1*, *CYP79F2* (Chen et al., 2003) and *CYP83A1* (Hemm et al., 2003) involved in the core structure formation can be done in the generated *BnMYB28* double mutants.

With the sole knock-out of the *BnCYP79F1.C05* paralog in the *BnCYP79F1* double mutants, we detected significant reductions in the short-chained C4 aliphatic GSL progoitrin but not the C5 glucobrassicinapin. We could only use a missense mutation conferring a minor change of a glutamic acid exchanged with lysine residue for the paralog *BnCYP79F1.A06*. Therefore, a weaker phenotypic effect from the selected functional mutations in *BnCYP79F1* can be explained by a putative compensation effect of *BnCYP79F1.A06*. Since *CYP79F1* can metabolize short and long-chained aliphatic GSLs in *Arabidopsis* (Chen et al., 2003), a possible sub-functionalization in the *B. napus* paralogs could explain distinct functions for *BnCYP79F1.C05* and *BnCYP79F1.A06*. In this regard, we speculate that *BnCYP79F1.A06* has a more significant impact on the C5 GSL metabolism than C4 aliphatic GSLs. Therefore, functional analyses of *BnCYP79F1* mutants encompassing nonsense mutations for both paralogs are needed to confirm this speculation.

The domestication history of the winter and spring-type rapeseed is distinct. Re-sequencing rapeseed diversity panels have revealed that both ecotypes display independent selection patterns for local adaptation, especially flowering time (Wu et al., 2019) and GSL content (Lu et al., 2019). Therefore, introgression of the spring-type Peace into the winter-type Express617 genotype can strongly influence the GSL biosynthesis process in the progeny. This was evident since the GSL content in backcrossed plants showed an increased variation even across biological replicates of the same genotype. This striking variation in the GSL content in the backcrossed plants deserves a further explanation.

Although we genotyped and selected mutant alleles for *BnMYB28* and *BnCYP79F1* each generation, we selected plants for a higher recurrent Peace genome share (>75%) at the BC<sub>1</sub> generation using the SNP array. As segregating BC<sub>1</sub>F<sub>2</sub> populations were developed from BC<sub>1</sub> plants, the winter and spring-type alleles are expected to still segregate in the background. We reason that for the same mutations that conferred significant phenotypic effects without backcrosses, a varying genomic background of the diverse winter and spring-type genomes has a distinct impact on the overall phenotype. In this regard, the underlying genetic mechanisms governing GSL content and profile in Peace can thus mask the effects of the EMS-induced mutations for *BnMYB28* and *BnCYP79F1* in the backcrossed mutants. Therefore, it would be interesting to identify underlying allelic variations influencing the GSL content and profile in the winter and spring ecotypes through studies on GSL biosynthesis, transportation and regulation. While significant variations in the biosynthesis genes are expected, shedding light on ecotype-specific biosynthesis regulation and transport mechanisms would be particularly interesting. Since the EMS donor line was Express617, comparing the effect of the same mutations in plants backcrossed with Express617 should address our speculations. Alternatively, it is also possible that since the GSL content in Peace is already low, a further reduction from the knock-out of our candidates did not have an additional effect. Even the most significant GSL reductions achieved in double mutants of non-backcrossed mutants could only reach levels still above the Peace wildtype controls.

EMS mutagenesis is a powerful and robust method for introducing novel variations in plant genomes. EMS-induced mutations can be instrumental in creating allelic variations that could

confer agronomically beneficial traits. Currently, several publically available EMS mutant resources have aided functional genomics, molecular breeding and crop improvement for rapeseed (Wang et al., 2008; Harloff et al., 2012; Gilchrist et al., 2013; Tang et al., 2020). Novel EMS mutants possessing agronomically beneficial traits like increased seed oil content and improved plant architecture and oil quality have been identified from these resources. EMS-induced random mutagenesis is a cost-effective and high throughput approach for generating non-transgenic functional mutants on a large scale. Since EMS mutants are non-transgenic, their utilization and integration into breeding programs are legally permitted in the European Union (Jung & Till, 2021).

In conclusion, our study demonstrates the function of two major genes involved in the biosynthesis of aliphatic GSLs, the most abundant GSL class in rapeseed. Our results provide the first functional analysis by knock-out for the biosynthesis genes *BnMYB28* and *BnCYP79F1* in rapeseed. Furthermore, we discovered a shift in the GSL biosynthesis content and composition arising from the introgression of the distant winter and spring genotypes. In this regard, understanding ecotype-specific regulatory and transport mechanisms in the complex GSL biosynthesis process could be crucial for achieving further GSL reduction. Our study provides a strong and promising basis for breeding rapeseed with improved meal quality in the future.

### 3.5 Supplementary data

#### 3.5.1 Supplementary figures

**Supplementary Figure 9:** Gene structures of three *BnMYB28* and two *BnCYP79F1* paralogs identified in rapeseed. START and STOP refer to the translation start and stop sites, respectively. Conserved domains characteristic of respective gene families are marked in yellow boxes. All gene models are based on the Darmor-*bzh* rapeseed reference genome. Paralog *BnMYB28.C09* was truncated but retained all conserved domains required for gene function. Information on the transcription start site for the *BnCYP79F1* paralogs is unavailable on the reference genome.

**Supplementary Figure 10:** Conserved domains observed in *MYB28* and *CYP79F1* paralogs of *B. napus*. The R2 and R3 DNA binding domains are highly conserved and characteristic of the MYB family transcription factors, followed by a nuclear localization signal (NLS) (Dubos et al., 2010). Five conserved domains have been reported from the family of cytochrome P450 enzymes. The Heme group is speculated to act as a catalytic domain vital for enzyme function (Reintanz et al., 2001).

**Supplementary Figure 11:** Crossing schemes and pedigrees of *BnMYB28* and *BnCYP79F1* mutant plants used in this study. Single mutants of the same gene family were crossed with each other to produce double mutants. A) M<sub>3</sub> single mutants were crossed to generate F<sub>1</sub> hybrids (genotypes *A<sub>1</sub>A<sub>e</sub>B<sub>1</sub>B<sub>e</sub>* or *C<sub>1</sub>C<sub>e</sub>D<sub>1</sub>D<sub>e</sub>*). The F<sub>1</sub> hybrids were then selfed to generate F<sub>2</sub> populations 200527 and 200529 representing *BnMYB28* and *BnCYP79F1* mutants, respectively. B) M<sub>3</sub> single mutants of *BnMYB28* paralogs were crossed to produce heterozygous double mutants of *BnMYB28* (genotype *A<sub>1</sub>A<sub>e</sub>B<sub>1</sub>B<sub>e</sub>*). The F<sub>1</sub> hybrids were backcrossed with the non-mutagenized spring oilseed rape Peace to generate backcross generation (BC<sub>1</sub>). BC<sub>1</sub> plants were selfed to generate the segregating BC<sub>1</sub>F<sub>2</sub> population (seed code 210465). Plants from the BC<sub>1</sub> generation were first genotyped for the presence of expected EMS-induced mutations and then genotyped with the *Brassica* 19K SNP array. Leaves and seeds from segregating F<sub>2</sub> plants were used for glucosinolate determination. Genotypes mentioned were selected each generation for further experiments and follow the single letter acronyms as in Table 6. ‘e’ and ‘p’ in subscript refer to non-mutagenized alleles from Express617 and Peace, respectively.

**Supplementary Figure 12:** Crossing schemes and pedigrees of backcrossed *BnCYP79F1* mutant plants used in this study. M<sub>3</sub> single homozygous mutants of *BnCYP79F1* paralogs (genotypes *C<sub>1</sub>C<sub>1</sub>D<sub>e</sub>D<sub>e</sub>* and *C<sub>e</sub>C<sub>e</sub>D<sub>1</sub>D<sub>1</sub>*) were first backcrossed with Peace to generate corresponding BC<sub>1</sub> progenies. Plants from the BC<sub>1</sub> generation were first genotyped for the presence of expected EMS-induced mutations and then genotyped with the *Brassica* 19K SNP array. Single mutants of the BC<sub>1</sub> generation were then combined to produce the segregating BC<sub>1</sub>F<sub>2</sub> population 210462. The BC<sub>1</sub>F<sub>2</sub> population displayed a digenic segregation (1:15 for double mutants *C<sub>1</sub>C<sub>1</sub>D<sub>1</sub>D<sub>1</sub>*). Genotypes mentioned were selected each generation for further experiments and follow the single letter acronyms as in table. ‘e’ and ‘p’ in subscript refer to non-mutagenized alleles from Express617 and Peace, respectively.

**Supplementary Figure 13:** Comparison of glucosinolate measurements by HPLC and the GOPOD enzymatic assay. Total GSL content determined by HPLC analyses is the sum of all major identified peaks identified from chromatograms (epiprogoitrin, glucobrassicinapin, glucoiberin, gluconapin, glucoraphanin, progoitrin, sinigrin, 4-hydroxyglucobrassicin, 4-methoxyglucobrassicin, glucobrassicin, gluconasturtiin, glucotropaeolin and

neoglucobrassicin). Values were calculated after calibrating against GSL standards (Supplementary Table 6). Absolute GSL content was measured using the GOPOD enzymatic assay. *BnMYB28* and *BnCYP79F1* double mutants (genotypes  $A_1A_1B_1B_1$  and  $C_1C_1D_1D_1$ ) and wildtype plants from the same segregating population originating either from  $M_3 \times M_3$  crosses ( $A_eA_eB_eB_e$  and  $C_eC_eD_eD_e$ ) or from backcrosses with Peace ( $A_{e/p}A_{e/p}B_{e/p}B_{e/p}$  and  $C_{e/p}C_{e/p}D_{e/p}D_{e/p}$ ) were compared with non-mutagenized controls. For each of the analyzed genotypes, five biological replicates were analyzed.

### 3.5.2 Supplementary tables

**Supplementary Table 4:** Primers developed in this study for RT-qPCR analyses.

**Supplementary Table 5:** List of primers used for generating amplicons for mutant screening. One amplicon per paralog was used for mutant detection.

**Supplementary Table 6:** HPLC performance of commercial glucosinolate standards.

**Supplementary Table 7:** All EMS-induced mutations detected in this study for two *BnMYB28*, two *BnCYP79F1* and two *BnGTR2* paralogs. Mutation positions are relative to the translation start site.

**Supplementary Table 8:** Summary statistics of functional effects conferred by EMS-induced mutations in *BnMYB28* and *BnCYP79F1* paralogs. One amplicon per paralog was used for screening EMS-induced mutations.

**Supplementary Table 9:** Mutant genotyping in the  $M_3$  generation to select crossing parents and combination of single mutations.

**Supplementary Table 10:** Genotypes of parents used for hand crosses and the offspring derived thereof made in this study to combine single EMS mutants of *BnMYB28* and *BnCYP79F1* and genotypes of selected crossing parents and their progenies used for phenotyping experiments.

**Supplementary Table 11:** Summary statistics of marker information and samples genotyped with the *Brassica* 19K SNP array.  $BC_1$  individuals were first genotyped for the presence of expected EMS-induced mutations and then genotyped with the SNP array. Marker information was calculated as an average of all genotyped individuals.

**Supplementary Table 12:** Summary statistics of the background Peace genome share for  $BC_1$  double mutants (*BnMYB28*) and single mutants (*BnCYP79F1*) genotyped with the *Brassica* 19K SNP array.

**Supplementary Table 13:** Summary statistics of individual glucosinolates in seeds of *BnMYB28* EMS mutants and controls. Analyses were done using high-performance liquid chromatography.

**Supplementary Table 14:** Summary statistics of individual glucosinolates in seeds of *BnCYP79F1* EMS mutants and controls. Analyses were done using high-performance liquid chromatography.



### 3.6 References

- Andréasson, E., Jørgensen, L. B., Höglund, A.-S., Rask, L., & Meijer, J. (2001). Different Myrosinase and Idioblast Distribution in *Arabidopsis* and *Brassica napus*. *Plant Physiology*, 127(4), 1750-1763. doi:10.1104/pp.010334
- Bourdon, D., & Aumaître, A. (1990). Low-glucosinolate rapeseeds and rapeseed meals: effect of technological treatments on chemical composition, digestible energy content and feeding value for growing pigs. *Animal Feed Science and Technology*, 30(3), 175-191. doi:[https://doi.org/10.1016/0377-8401\(90\)90014-Y](https://doi.org/10.1016/0377-8401(90)90014-Y)
- Braatz, J., Harloff, H.-J., Emrani, N., Elisha, C., Heepe, L., Gorb, S. N., & Jung, C. (2018). The effect of *INDEHISCENT* point mutations on silique shatter resistance in oilseed rape (*Brassica napus*). *Theoretical and Applied Genetics*, 131(4), 959-971.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633-2635.
- Brown, P. D., Tokuhisa, J. G., Reichelt, M., & Gershenzon, J. (2003). Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochemistry*, 62(3), 471-481.
- Cartea, M. E., & Velasco, P. (2008). Glucosinolates in Brassica foods: bioavailability in food and significance for human health. *Phytochemistry Reviews*, 7(2), 213-229.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., & Samans, B. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, 345(6199), 950-953.
- Chen, S., Glawischnig, E., Jørgensen, K., Naur, P., Jørgensen, B., Olsen, C. E., Hansen, C. H., Rasmussen, H., Pickett, J. A., & Halkier, B. A. (2003). *CYP79F1* and *CYP79F2* have distinct functions in the biosynthesis of aliphatic glucosinolates in *Arabidopsis*. *The Plant Journal*, 33(5), 923-937.
- Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). *MYB* transcription factors in *Arabidopsis*. *Trends in Plant Science*, 15(10), 573-581.
- Emrani, N., Harloff, H.-J., Gudi, O., Kopisch-Obuch, F., & Jung, C. (2015). Reduction in sinapine content in rapeseed (*Brassica napus* L.) by induced mutations in sinapine biosynthesis genes. *Molecular Breeding*, 35(1), 37. doi:10.1007/s11032-015-0236-2
- Fahey, J. W., Zalcmann, A. T., & Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, 56(1), 5-51.
- Feng, J., Long, Y., Shi, L., Shi, J., Barker, G., & Meng, J. (2012). Characterization of metabolite quantitative trait loci and metabolic networks that control glucosinolate concentration in the seeds and leaves of *Brassica napus*. *New Phytologist*, 193(1), 96-108.
- Fenwick, G. R., & Curtis, R. F. (1980). Rapeseed meal and its use in poultry diets. A review. *Animal Feed Science and Technology*, 5(4), 255-298. doi:[https://doi.org/10.1016/0377-8401\(80\)90016-4](https://doi.org/10.1016/0377-8401(80)90016-4)

Fiebig, H., & Arens, M. (1992). Glucosinolate (HPLC-Methode) -Gemeinschaftsarbeiten der DGF, 128. Mitteilung: Deutsche Einheitsmethoden zur Untersuchung von Fetten, Fettprodukten, Tensiden und verwandten Stoffen, 98. Mitt.: Analyse von Fettrohstoffen XII\*. *12. Fett Wissenschaft Technologie*, 94, 199–203.

Gigolashvili, T., Yatusovich, R., Berger, B., Müller, C., & Flügge, U. I. (2007). The R2R3-MYB transcription factor *HAG1/MYB28* is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*, 51(2), 247-261.

Gilchrist, E. J., Sidebottom, C. H., Koh, C. S., MacInnes, T., Sharpe, A. G., & Haughn, G. W. (2013). A mutant *Brassica napus* (Canola) population for the identification of new genetic diversity via TILLING and next generation sequencing. *PLoS One*, 8(12), e84303.

Guo, Y., Harloff, H.-J., Jung, C., & Molina, C. (2014). Mutations in single *FT*- and *TFL1*-paralogs of rapeseed (*Brassica napus* L.) and their impact on flowering time and yield components. *Frontiers in Plant Science*, 5, 282.

Halkier, B. A., & Du, L. (1997). The biosynthesis of glucosinolates. *Trends in Plant Science*, 2(11), 425-431.

Harloff, H.-J., Lemcke, S., Mittasch, J., Frolov, A., Wu, J. G., Dreyer, F., Leckband, G., & Jung, C. (2012). A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. *Theoretical and Applied Genetics*, 124(5), 957-969.

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., & Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nature Biotechnology*, 30(8), 798-802.

Hemm, M. R., Ruegger, M. O., & Chapple, C. (2003). The *Arabidopsis* *ref2* Mutant Is Defective in the Gene Encoding CYP83A1 and Shows Both Phenylpropanoid and Glucosinolate Phenotypes. *The Plant Cell*, 15(1), 179-194. doi:10.1105/tpc.006544

Himmelblau, E., Gilchrist, E. J., Buono, K., Bizzell, C., Mentzer, L., Vogelzang, R., Osborn, T., Amasino, R. M., Parkin, I. A., & Haughn, G. W. (2009). Forward and reverse genetics of rapid-cycling *Brassica oleracea*. *Theoretical and Applied Genetics*, 118(5), 953-961.

Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., Araki, R., Sakurai, N., Suzuki, H., & Aoki, K. (2007). Omics-based identification of *Arabidopsis* *Myb* transcription factors regulating aliphatic glucosinolate biosynthesis. *Proceedings of the National Academy of Sciences*, 104(15), 6478-6483.

Howell, P., Sharpe, A., & Lydiat, D. (2003). Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*Brassica napus*). *Genome*, 46(3), 454-460.

Jung, C., & Till, B. (2021). Mutagenesis and genome editing in crop improvement: perspectives for the global regulatory landscape. *Trends in Plant Science*.

Kaiser, F., Harloff, H. J., Tressel, R. P., Kock, T., & Schulz, C. (2021). Effects of highly purified rapeseed protein isolate as fishmeal alternative on nutrient digestibility and growth performance in diets fed to rainbow trout (*Oncorhynchus mykiss*). *Aquaculture Nutrition*.

- Karunarathna, N. L., Patiranage, D. S. R., Harloff, H.-J., Sashidhar, N., & Jung, C. (2021). Genomic background selection to reduce the mutation load after random mutagenesis. *Scientific Reports*, 11(1), 19404. doi:10.1038/s41598-021-98934-5
- Karunarathna, N. L., Wang, H., Harloff, H. J., Jiang, L., & Jung, C. (2020). Elevating seed oil content in a polyploid crop by induced mutations in *SEED FATTY ACID REDUCER* genes. *Plant Biotechnology Journal*, 18(11), 2251-2266.
- Keck, A.-S., & Finley, J. W. (2004). Cruciferous vegetables: cancer protective mechanisms of glucosinolate hydrolysis products and selenium. *Integrative Cancer Therapies*, 3(1), 5-12.
- Kittipol, V., He, Z., Wang, L., Doheny-Adams, T., Langer, S., & Bancroft, I. (2019). Genetic architecture of glucosinolate variation in *Brassica napus*. *Journal of Plant Physiology*, 240, 152988. doi:<https://doi.org/10.1016/j.jplph.2019.06.001>
- Kondra, Z., & Stefansson, B. (1970). Inheritance of the major glucosinolates of rapeseed (*Brassica napus*) meal. *Canadian Journal of Plant Science*, 50(6), 643-647.
- Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., Harper, A. L., Trick, M., Liu, S., & Gao, G. (2014). Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Research*, 21(4), 355-367.
- Liu, S., Huang, H., Yi, X., Zhang, Y., Yang, Q., Zhang, C., Fan, C., & Zhou, Y. (2020). Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome-wide association study. *Plant Biotechnology Journal*, 18(6), 1472-1484.
- Liu, Y., Zhou, X., Yan, M., Wang, P., Wang, H., Xin, Q., Yang, L., Hong, D., & Yang, G. (2020). Fine mapping and candidate gene analysis of a seed glucosinolate content QTL, *qGSL-C2*, in rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics*, 133(2), 479-490. doi:10.1007/s00122-019-03479-x
- Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., Zhang, C., Chen, Z., Xiao, Z., & Jian, H. (2019). Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nature Communications*, 10(1), 1154.
- Madloo, P., Lema, M., Francisco, M., & Soengas, P. (2019). Role of Major Glucosinolates in the Defense of Kale Against *Sclerotinia sclerotiorum* and *Xanthomonas campestris* pv. *campestris*. *Phytopathology*, 109(7), 1246-1256.
- Mitreiter, S., & Gigolashvili, T. (2021). Regulation of glucosinolate biosynthesis. *Journal of Experimental Botany*, 72(1), 70-91.
- Nambiar, D. M., Kumari, J., Augustine, R., Kumar, P., Bajpai, P. K., & Bisht, N. C. (2021). GTR1 and GTR2 transporters differentially regulate tissue-specific glucosinolate contents and defence responses in the oilseed crop *Brassica juncea*. *Plant, Cell & Environment*.
- Nour-Eldin, H. H., Andersen, T. G., Burow, M., Madsen, S. R., Jørgensen, M. E., Olsen, C. E., Dreyer, I., Hedrich, R., Geiger, D., & Halkier, B. A. (2012). NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature*, 488(7412), 531.

Nour-Eldin, H. H., & Halkier, B. A. (2009). Piecing together the transport pathway of aliphatic glucosinolates. *Phytochemistry Reviews*, 8(1), 53-67.

Nour-Eldin, H. H., Madsen, S. R., Engelen, S., Jørgensen, M. E., Olsen, C. E., Andersen, J. S., Seynnaeve, D., Verhoye, T., Fulawka, R., Denolf, P., & Halkier, B. A. (2017). Reduction of antinutritional glucosinolates in *Brassica* oilseeds by mutation of genes encoding transporters. *Nature Biotechnology*, 35, 377. doi:10.1038/nbt.3823

<https://www.nature.com/articles/nbt.3823#supplementary-information>

Qu, C.-M., Li, S.-M., Duan, X.-J., Fan, J.-H., Jia, L.-D., Zhao, H.-Y., Lu, K., Li, J.-N., Xu, X.-F., & Wang, R. (2015). Identification of candidate genes for seed glucosinolate content using association mapping in *Brassica napus* L. *Genes*, 6(4), 1215-1229.

Rask, L., Andréasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., & Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Molecular Biology*, 42(1), 93-114.

Reintanz, B., Lehnen, M., Reichelt, M., Gershenzon, J., Kowalczyk, M., Sandberg, G., Godde, M., Uhl, R., & Palme, K. (2001). bus, a bushy *Arabidopsis CYP79F1* knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. *The Plant Cell*, 13(2), 351-367.

Russo, M., Yan, F., Stier, A., Klasen, L., & Honermeier, B. (2021). Erucic acid concentration of rapeseed (*Brassica napus* L.) oils on the German food retail market. *Food Science & Nutrition*.

Saghai-Maroo, M. A., Soliman, K. M., Jorgensen, R. A., & Allard, R. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences*, 81(24), 8014-8018.

Sashidhar, N., Harloff, H. J., & Jung, C. (2019). Identification of phytic acid mutants in oilseed rape (*Brassica napus*) by large scale screening of mutant populations through amplicon sequencing. *New Phytologist*.

Shah, S., Karunarathna, N. L., Jung, C., & Emrani, N. (2018). An *APETALA1* ortholog affects plant architecture and seed yield component in oilseed rape (*Brassica napus* L.). *BMC Plant Biology*, 18(1), 380. doi:10.1186/s12870-018-1606-9

Singh, N., Jayaswal, P. K., Panda, K., Mandal, P., Kumar, V., Singh, B., Mishra, S., Singh, Y., Singh, R., & Rai, V. (2015). Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. *Scientific Reports*, 5(1), 1-9.

Sønderby, I. E., Geu-Flores, F., & Halkier, B. A. (2010). Biosynthesis of glucosinolates—gene discovery and beyond. *Trends in Plant Science*, 15(5), 283-290.

Stephenson, P., Baker, D., Girin, T., Perez, A., Amoah, S., King, G. J., & Østergaard, L. (2010). A rich TILLING resource for studying gene function in *Brassica rapa*. *BMC Plant Biology*, 10(1), 62.

- Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, 4(5), 447-456.
- Tang, S., Liu, D. X., Lu, S., Yu, L., Li, Y., Lin, S., Li, L., Du, Z., Liu, X., & Li, X. (2020). Development and screening of EMS mutants with altered seed oil content or fatty acid composition in *Brassica napus*. *The Plant Journal*.
- Textor, S., De Kraker, J.-W., Hause, B., Gershenzon, J., & Tokuhsa, J. G. (2007). MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. *Plant Physiology*, 144(1), 60-71.
- Thomson, M. J., Singh, N., Dwiyantri, M. S., Wang, D. R., Wright, M. H., Perez, F. A., DeClerck, G., Chin, J. H., Malitic-Layaoen, G. A., Juanillas, V. M., Dilla-Ermita, C. J., Mauleon, R., Kretzschmar, T., & McCouch, S. R. (2017). Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. *Rice*, 10(1), 40. doi:10.1186/s12284-017-0181-2
- Till, B. J., Zerr, T., Comai, L., & Henikoff, S. (2006). A protocol for TILLING and Ecotilling in plants and animals. *Nature Protocols*, 1(5), 2465.
- Toroser, D., Wood, C., Griffiths, H., & Thomas, D. (1995). Glucosinolate biosynthesis in oilseed rape (*Brassica napus* L.): studies with 35SO<sub>2</sub>– 4 and glucosinolate precursors using oilseed rape pods and seeds. *Journal of Experimental Botany*, 46(7), 787-794.
- Tripathi, M., & Mishra, A. (2007). Glucosinolates in animal nutrition: A review. *Animal Feed Science and Technology*, 132(1-2), 1-27.
- Velasco, P., Soengas, P., Vilar, M., Cartea, M. E., & del Rio, M. (2008). Comparison of glucosinolate profiles in leaf and seed tissues of different *Brassica napus* crops. *Journal of the American Society for Horticultural Science*, 133(4), 551-558.
- Wang, N., Wang, Y., Tian, F., King, G. J., Zhang, C., Long, Y., Shi, L., & Meng, J. (2008). A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of *FAEI* point mutations by TILLING. *New Phytologist*, 180(4), 751-765.
- Wei, D., Cui, Y., Mei, J., Qian, L., Lu, K., Wang, Z.-M., Li, J., Tang, Q., & Qian, W. (2019). Genome-wide identification of loci affecting seed glucosinolate contents in *Brassica napus* L. *Journal of Integrative Plant Biology*, 61(5), 611-623. doi:10.1111/jipb.12717
- Wickham, H. (2009). *Elegant graphics for data analysis* (Vol. 35).
- Wittkop, B., Snowdon, R., & Friedt, W. (2009). Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. *Euphytica*, 170(1), 131-140.
- Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., & Wang, H. (2019). Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Molecular Plant*, 12(1), 30-43.
- Zerr, T., & Henikoff, S. (2005). Automated band mapping in electrophoretic gel images using background information. *Nucleic Acids Research*, 33(9), 2806-2812.

## 4 Functional analysis of the glucosinolate transporter gene *BnGTR2* in rapeseed

### 4.1 Introduction

Rapeseed (*Brassica napus*) is a major oil crop cultivated for its healthy edible seed oil and the leftover rapeseed meal (RSM) that serves as animal feed. However, the presence of major anti-nutritive compounds like glucosinolates (GSLs) adversely affects its utilization as animal feed (Tripathi & Mishra, 2007). Therefore, an important breeding goal for rapeseed is to reduce the GSL content in the seeds. The development of ‘double low’ or ‘double zero (00)’ rapeseed varieties was an achievement in oilseed rape breeding for improved seed quality. Their origin is traced back to the Polish spring variety ‘Bronowski’ and the German spring variety ‘Liho’, which were selected for a low glucosinolate and low erucic acid content, respectively (Friedt et al., 2018). Breeding efforts in the past decades have significantly reduced GSLs in rapeseed from >100 µmol/g dry weight to the acceptable limits of <18-25 µmol/g dry weight in modern rapeseed cultivars (Wittkop et al., 2009).

GSLs are diverse secondary metabolites synthesized from a β-D-glucopyranose residue attached to a variable R group of amino acid precursors through a sulfonated oxime (Halkier & Gershenzon, 2006). Three major GSL types exist depending upon the associated amino acid residues. Aliphatic GSLs are derived from methionine, aromatic GSLs are derived from phenylalanine or tyrosine, and indolic GSLs are derived from tryptophan (Mitreiter & Gigolashvili, 2021). The breakdown products of GSLs after metabolism by endogenous plant myrosinases include several toxic compounds like isothiocyanates, thiocyanates, epithionitriles and nitriles that confer defense against insect herbivores and pathogens, (Rask et al., 2000). Over 130 different GSL types have been described from 16 dicot angiosperms, most of which are edible Brassicaceae crops (Fahey et al., 2001; Mitreiter & Gigolashvili, 2021). Out of the three GSL types, the methionine-derived aliphatic GSLs are the most abundant type in rapeseed, accounting for ~92% of all GSLs (Velasco et al., 2008). The short-chained GSL progoitrin comprises the majority of all aliphatic GSLs in the seeds and severely limits the meal quality due to its anti-thyroid goitrogenic effects in animals. Moreover, other aliphatic GSLs like sinigrin are pungent and bitter-tasting compounds that reduce the palatability of the RSM (Tripathi & Mishra, 2007; Ishida et al., 2014).

The GSL biosynthesis and its regulatory mechanisms have been intensively studied in *Arabidopsis* (Mitreiter & Gigolashvili, 2021). Although *Arabidopsis* is a close relative of rapeseed, the transfer of knowledge to rapeseed is complicated. This is because, for the same *Arabidopsis* ortholog, multiple homoeologs can exist in rapeseed. In this regard, the publically available ‘Darmor-*bzh*’ reference genome (Genoscope-<https://www.genoscope.cns.fr/brassicanapus/>) has significantly supported genomic studies for rapeseed by enabling gene identification and sequence analyses of candidate orthologs. The reference assembly has an effective genome size of ~1.13 gigabases representing 101,040 gene models encompassed by ten chromosomes of the A sub-genome and nine chromosomes of the C sub-genome (Chalhoub et al., 2014).

In *Arabidopsis*, several candidate gene families have been identified and characterized with functional effects on GSL biosynthesis (Reintanz et al., 2001; Chen et al., 2003; Gigolashvili et al., 2007; Hirai et al., 2007; Textor et al., 2007; Frerigmann et al., 2015). In Brassicaceae crops, genes associated with high GSL content have also been identified and functionally

analyzed in *Brassica rapa* (Kim et al., 2013; Wiesner et al., 2014; Seo et al., 2016), *Brassica oleracea* (Araki et al., 2013; Yin et al., 2017; Sánchez-Pujante et al., 2018; Zuluaga et al., 2019) and *Brassica juncea* (Bisht et al., 2009; Augustine et al., 2013; Yang et al., 2021). In *Brassica napus*, associative transcriptomics, genome-wide association, and QTL mapping studies have revealed major gene families strongly associated with the seed GSL content (Harper et al., 2012; Kittipol et al., 2019; Wei et al., 2019; S. Liu et al., 2020). In these studies, the subgroup 12 R2R3-MYB transcription factors controlling the core structure formation of GSLs were identified to be strongly associated with the GSL content. Gene networks controlling the biosynthesis and regulation of aliphatic, aromatic and indolic GSLs are known to function independently and several environmental factors can influence the GSL content and profile (Mitreiter & Gigolashvili, 2021). Thus, GSL metabolism in Brassicaceae is vast and diverse and is quantitatively controlled by multiple gene families (Halkier & Gershenzon, 2006; Chalhoub et al., 2014). Investigations on underlying gene families have been particularly cumbersome in rapeseed due to its complex polyploid genome. As a result, comprehensive functional analyses for GSL genes are mostly missing in rapeseed.

In contrast to biosynthesis, knowledge regarding the molecular basis for the seed-specific transportation of GSLs is limited. Since GSL biosynthesis occurs primarily in the vegetative parts (Andersen et al., 2013), the long-distance transport of GSLs to seeds is expected to be controlled by transporters (Nour-Eldin & Halkier, 2009). GSL transport via nitrate/peptide transporters NRT/PTR was first demonstrated in Arabidopsis by Nour-Eldin et al. (2012). In pioneering studies for *Brassica* crops *B. rapa* and *B. juncea*, the seed-specific proton-coupled transport of GSLs was described through two *GLUCOSINOLATE TRANSPORTER* genes *GTR1* and *GTR2* (Nour-Eldin et al., 2017). Out of the two, *GTR2* had a more significant role in the seed-specific accumulation of GSLs. More recently, *GTR1* and *GTR2* were functionally characterized using knock-out mutants in *B. juncea* (Nambiar et al., 2021). In this study, the *GTR2* knock-out mutants showed a significant reduction in seed GSLs and a higher accumulation in vegetative parts, resulting in increased resistance against lepidopteran insect pests. By whole-genome resequencing 588 diverse *B. napus* accessions, Lu et al. (2019) have identified selection patterns associated with the improvement of rapeseed via GSL reduction in the seeds. The study reported a strong association of the *BnGTR2* gene with the seed GSL content.

Since GSLs play a prominent role in plant defense mechanisms, a reduction in the whole plant is likely to increase susceptibility against herbivory and insect pests (Rask et al., 2000; Sønderby et al., 2010). Therefore, seed-specific reduction of GSLs without compromising on the content in the vegetative parts is a bigger objective for breeding rapeseed with improved seed meal quality. This study aimed to reduce the seed GSL content through a functional analysis by knock-out for *BnGTR2* in rapeseed. In this chapter, I describe the production of material that can provide a strong foundation for studying the loss of function effects of *BnGTR2* on the seed-specific accumulation of GSLs in rapeseed.

## 4.2 Materials and methods

### 4.2.1 *In silico* analysis

I retrieved the genomic DNA and polypeptide sequences of the GSL transporter *GTR2* from The Arabidopsis Information Resource (TAIR). All *in silico* analyses were done as described in Chapter 3. Conserved and functional domain identification was made following previously published studies and the NCBI Conserved Domain Database (CDD). For phylogenetic

analyses, a maximum likelihood tree (GTR and WAG used as nucleotide and amino acid substitution models, respectively; 1000 bootstraps) was constructed on the CLC Main Workbench 7 using polypeptide sequences.

#### 4.2.2 Expression analysis of candidate genes by RT-qPCR

The winter-type rapeseed Express617 was used for expression studies. Plant material production, tissue sampling and expression profile analyses were done according to the methods described in Chapter 3. After RNA isolation, 1 µg RNA was used for cDNA synthesis. Synthesized cDNA was diluted 1:20 with ddH<sub>2</sub>O and paralog-specific primers were used to perform RT-qPCR for each paralog of *BnGTR2* (Supplementary Table 4). Five biological replicates, each with three technical replicates, were analyzed for the data points 15, 25, 35 and 45 days after pollination. RT-qPCR was run using the Bio-Rad CFX96 Real-Time System (Bio-Rad Laboratories GmbH, Munich, Germany). The  $\Delta\Delta C_q$  method was used to calculate the relative expression of each paralog normalized against the two reference genes *BnGAPDH* and *BnACTIN2*. Relative expression levels of each candidate paralog were calculated as a mean of five biological replicates represented by three technical replicates each.

#### 4.2.3 DNA isolation and genotyping experiments

Leaf genomic DNA was isolated following the standard CTAB method of Saghai-Maroo et al. (1984), as described in Chapter 3. For genotyping experiments, standard PCR was done to generate paralog-specific amplicons using primers in Supplementary Table 5. PCR fragments were Sanger sequenced for validation. Plants backcrossed with the spring-type oilseed rape Peace were genotyped with a *Brassica* 19K SNP array (SGS-INSTITUT FRESENIUS GmbH, TraitGenetics Section, Gatersleben, Germany). Data analyses for the SNP array genotyping were done as per the methods described in Chapter 3.

#### 4.2.4 Conventional gel-based screening of EMS-induced mutations

EMS-induced mutations were identified from the EMS mutagenized rapeseed population using the method described in Chapter 3. Paralog-specific primers were used to generate amplicons for mutation detection (Supplementary Table 5). For this, a conventional polyacrylamide gel-based approach was used following the protocol described by Till et al. (2006). Single mutants were validated using genomic DNA from single M<sub>2</sub> plants by Sanger sequencing PCR fragments encompassing the detected mutations.

#### 4.2.5 Plasmid vector construction for CRISPR-Cas9 mutagenesis

I searched for 5'-NGG-3' protospacer adjacent motifs (PAM) in exonic regions of *BnGTR2* paralogs and selected 20 bp target oligonucleotide sequences (>40-50% GC content) upstream of the PAM as target sites. Target sequences were searched using a BLAST query on the Darmor-*bzh* rapeseed reference (<https://www.genoscope.cns.fr/brassicanapus/>) to assess sequence specificity and putative off-target effects. The binary vector set of pChimera and pCas9-TPC was obtained from Prof. Dr. Holger Puchta (Karlsruher Institut für Technologie, Karlsruhe). Final Cas9 vectors for each target oligonucleotide were developed using the protocol of Fauser et al. (2014) with minor modifications described by Braatz et al. (2017) (Supplementary Figure 15A and 15B).



Roughly 500 ng pChimera vector was digested with 1  $\mu$ l *BbsI/BpiI* (10 U/ $\mu$ l) restriction enzyme and 2  $\mu$ l 10x Buffer G (ThermoFisher Scientific, Waltham, United States). The reaction mixture was completed to a final reaction volume of 20  $\mu$ l by adding ddH<sub>2</sub>O and incubated at 37°C for one hour. The 20 bp synthesized target oligonucleotides (100 pmol/ $\mu$ l) were ordered from Eurofins Genomics (Ebersberg, Germany). For oligonucleotide annealing, 1  $\mu$ l of each forward and reverse target oligonucleotides were added to 48  $\mu$ l ddH<sub>2</sub>O and incubated at 95°C for 5 minutes and kept at room temperature for 20 minutes. For ligation, 2  $\mu$ l of digested pChimera vector, 3  $\mu$ l of annealed oligonucleotides, 1  $\mu$ l of T4 ligase (5 U/ $\mu$ l) and 5  $\mu$ l 2x Quick Ligase buffer (ThermoFisher Scientific, Waltham, United States) were incubated for 1 h at room temperature.

The ligated pChimera vectors were then transformed into *Escherichia coli* cells of the strain XL1 Blue (DNA Cloning Service, Hamburg). Transformed cells were inoculated with 950  $\mu$ l of 37°C pre-heated LB liquid medium and agitated on a rotatory shaker (37°C, 220 rpm, 1 h). 100  $\mu$ l inoculum was incubated overnight at 37°C on LB solid media with ampicillin (100 mg/ml). Single positive colonies were picked and colony PCR was done to confirm target integration using primers flanking insertion sites. Colonies with confirmed target integration were picked and grown overnight in 3 ml LB liquid media with 3  $\mu$ l ampicillin (100 mg/ml) at 37°C while stirring on the rotatory shaker at 200 rpm. The plasmid was isolated from the inoculum using the NucleoSpin® Plasmid Isolation Kit (Macherey-Nagel, Germany).

Before integrating the sgRNA-target oligonucleotide complex, 17  $\mu$ l pCas9-TPC vector was digested with 1  $\mu$ l *BcuI* (10 U/ $\mu$ l) and 2  $\mu$ l Buffer Tango (ThermoFisher Scientific, Waltham, United States) for 2 hours at 37°C. The digested vector was dephosphorylated with 2  $\mu$ l of calf intestinal phosphatase (CIP) (10,000 U/ml) and 5  $\mu$ l CutSmart buffer (New England Biolabs, USA) at 37°C for 30 minutes. The digested pCas9-TPC vector was purified using the NucleoSpin® PCR clean-up kit (Macherey-Nagel, Germany). The sgRNA-target oligonucleotide complex was then cleaved out of the pChimera vector after treatment with *BcuI* (10 U/ $\mu$ l) at 37°C for 2 h. The cleaved fragment was then ligated into the pCas9-TPC vector possessing the gene encoding the Cas9 endonuclease to develop the final Cas9 vectors.

The final Cas9 vectors were then transformed into competent *Agrobacterium tumefaciens* cells (strain GV3101 pMP90RK) using the protocol of Höfgen and Willmitzer (1988). After transformation, 200  $\mu$ l of the inoculum was grown on solid LB media with 100 mg/ml spectinomycin. Single positive colonies were picked and a colony PCR was performed to confirm the successful integration of the final Cas9 vectors using vector-specific primers flanking the insertion sites. Simultaneously, single colonies were picked and inoculated in liquid LB media with 3  $\mu$ l each of spectinomycin (100 mg/ml), rifampicin (50 mg/ml) and gentamycin (50 mg/ml). The inoculum was agitated at 200 rpm on a rotatory shaker at 28°C, overnight. The *Agrobacterium* culture was aliquoted in 2 ml tubes, diluted in equal amounts with 100% glycerol, and stored at -70°C for plant transformation experiments.

#### 4.2.6 Plant material and growth conditions

M<sub>3</sub> seeds were obtained from NPZ Innovation GmbH (Holtsee, Germany). M<sub>3</sub> plants were grown in 11 cm pots under greenhouse conditions (16 h light, 20-25°C) for three weeks. Leaf samples were collected from each plant for DNA isolation and genotyping. Plants were then vernalized for eight weeks (16 h light, 4°C). Plants from the winter-type Express617 and the spring-type Peace were grown as wildtype crossing parents. I made direct M<sub>3</sub>xM<sub>3</sub> crosses of single homozygous mutants. Subsequently, I backcrossed F<sub>1</sub> double mutants with the wildtype Peace and Express617 genotypes to yield the respective BC<sub>1</sub>F<sub>2</sub> populations. All

plants were grown in 11 cm pots under greenhouse conditions (16 h light, 20-25°C). Plants selected for crossing experiments were hand-pollinated after emasculation and plants chosen for selfings were isolated in plastic bags.

#### 4.2.7 Statistical analysis

In expression studies, significantly expressed paralogs were identified by performing an ANOVA test ( $p < 0.05$ ) for the relative expression levels of each paralog. Mean relative expressions from five biological replicates were compared across the sampling points 15-45 days after pollination in seeds and leaves. I also performed the LSD test ( $\alpha \leq 0.05$ ) to generate statistical groups using the ‘Agricolae’ package in R. The standard error of mean was calculated across the five biological replicates with three technical replicates each.

### 4.3 Results

#### 4.3.1 Seven paralogs of *BnGTR2* were identified in oilseed rape

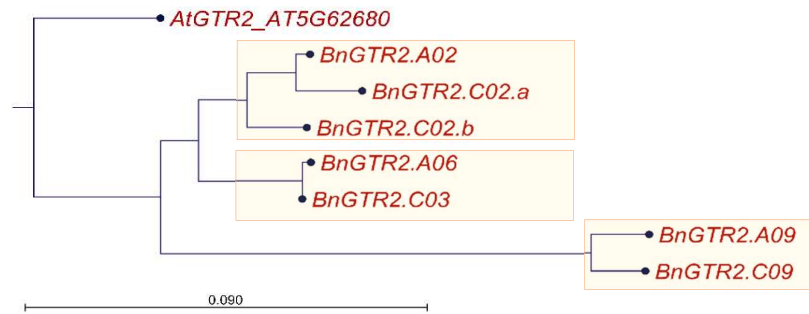
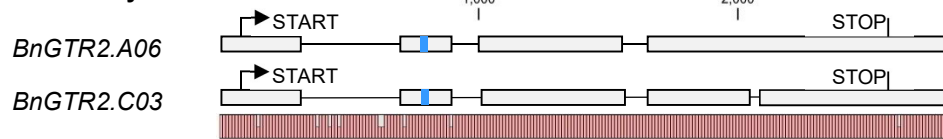
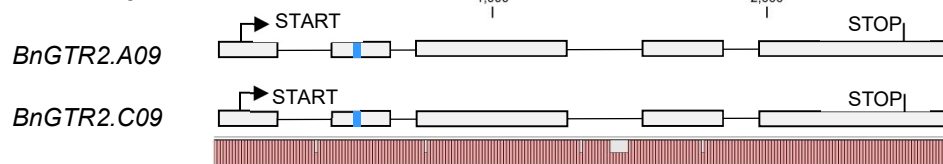
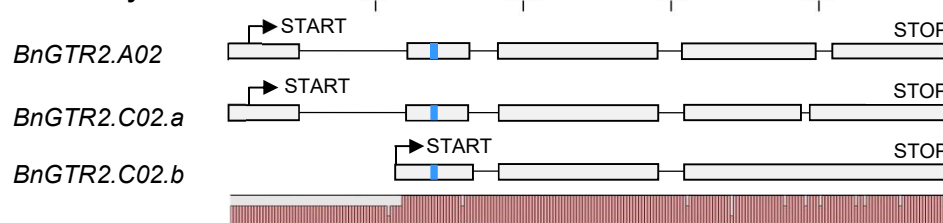
Using the *Arabidopsis thaliana* gene *AtGTR2* (AT5G62680), I searched for possible paralogs in the Darmor-*bzh* oilseed rape reference genome (<https://www.genoscope.cns.fr/brassicanapus/>). Sequences of seven paralogs were retrieved based on the lowest e-values and >80% sequence similarity with the Arabidopsis ortholog (Table 7).

**Table 7:** Features of *BnGTR2* paralogs identified in oilseed rape.

<i>B. napus</i> paralogs <sup>[a]</sup>	<i>B. napus</i> gene name	Chromosomal Location	Gene length (bp)	Coding region (bp)	Polypeptide length	Shared protein identity with Arabidopsis
<i>BnaA06g22160D</i>	<i>BnGTR2.A06</i>	A06	2,770	1,839	612	88.2%
<i>BnaC03g51560D</i>	<i>BnGTR2.C03</i>	C03	2,814	1,809	602	87.7%
<i>BnaA09g06190D</i>	<i>BnGTR2.A09</i>	A09	2,606	1,725	574	86.9%
<i>BnaC09g05810D</i>	<i>BnGTR2.C09</i>	C09	2,668	1,722	573	86.9%
<i>BnaC02g42280D</i>	<i>BnGTR2.C02.a</i>	C02	2,451	1,809	602	86.9%
<i>BnaC02g42260D</i>	<i>BnGTR2.C02.b</i>	C02	1,880	1,716	571	88.7%
<i>BnaA02g33530D</i>	<i>BnGTR2.A02</i>	A02	2,452	1,809	599	88.3%

[a] Sequence analysis of the paralogs in oilseed rape are based on gene models described in the Darmor-*bzh* reference genome (Genoscope).

In phylogenetic analyses with polypeptide sequences, the Neighbor Joining tree revealed three clades for the seven paralogs (Figure 14A). Based on the phylogenetic clades, I divided the seven paralogs into three sub-families and named them sub-family 1 (*BnGTR2.A06* and *BnGTR2.C03*), sub-family 2 (*BnGTR2.A09* and *BnGTR2.C09*) and sub-family 3 (*BnGTR2.A02*, *BnGTR2.C02.a* and *BnGTR2.C02.b*). An ‘ETFEK’ conserved domain characteristic of the proton-dependent oligopeptide transporter (POT/PTR) family (Jørgensen, Olsen, et al., 2015) was found to be highly conserved across all seven paralogs of *BnGTR2* (Figure 14B). This domain is speculated to control the proton-coupled transport of GSLs.

**A****B****Sub-family 1****Sub-family 2****Sub-family 3**

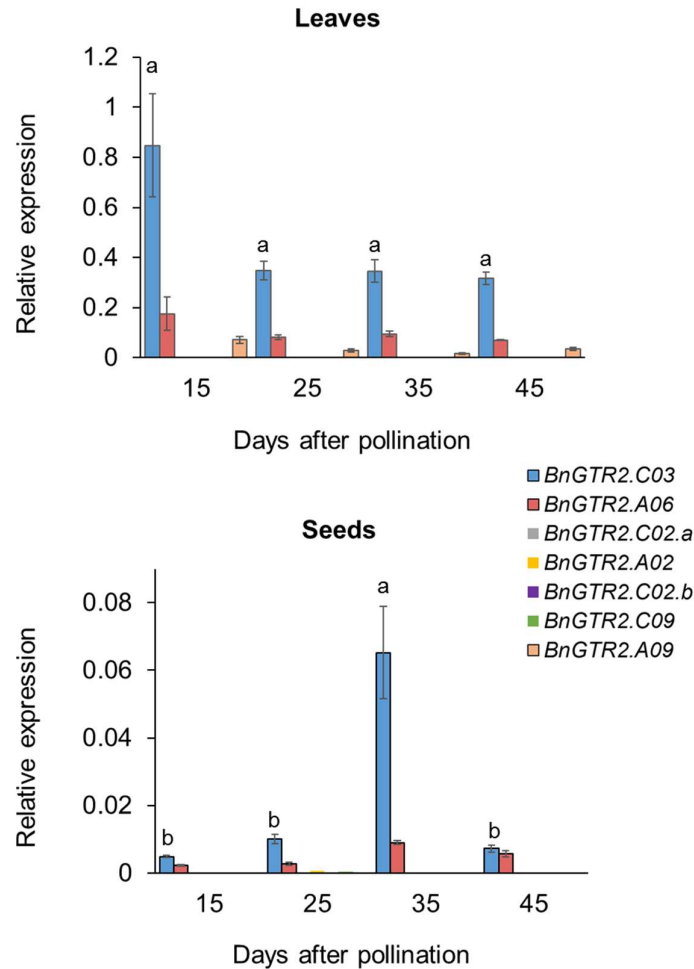
Exon — Intron ■ 'ETFEK' conserved domain

**Figure 14:** Sequence analyses of *BnGTR2* paralogs in rapeseed. A) Phylogenetic tree representing the three clades of *BnGTR2* paralogs (marked in yellow boxes). The maximum likelihood tree (GTR and WAG used as nucleotide and amino acid substitution models, respectively; 1000 bootstraps) was constructed on the CLC Main Workbench 7 using the protein sequence alignment of the seven paralogs. The protein sequence of the *GTR2* ortholog from Arabidopsis (prefixed 'At' followed by the TAIR ID) was also used. B) Gene structures of seven *BnGTR2* paralogs grouped into the three sub-families in rapeseed. Structures are based on the gene models predicted in the Darmor-*bzh* reference genome. START and STOP refer to the translation start and stop sites, respectively. For gene *BnGTR2.C02.b*, information regarding the 5' untranslated region was unavailable on the Darmor-*bzh* reference genome. The 'ETFEK' motif (blue bars) was conserved across all seven paralogs. Vertical bars below gene models represent the sequence conservation from 0-100% across the paralogs within each sub-family. All *in silico* analyses were done on the CLC Main Workbench 7.

### 4.3.2 Expression analyses reveal differentially expressed *BnGTR2* paralogs in leaves and seeds

The expression analyses covered the seed setting and loading stages between 15-45 days after pollination (DAP). The highest relative expressions were observed for the paralogs of sub-family 1 (*BnGTR2.C03* and *BnGTR2.A06*) across all sampling points in both leaves and seeds (Figure 15). Moreover, the relative expression levels were tenfold higher in leaves than in seeds for both paralogs. Although not statistically significant, expression levels for *BnGTR2.C03* decreased by almost half after 15 DAP and stabilized to a constant level between 25-45 DAP. Both paralogs showed a similar expression pattern towards plant maturity with decreasing levels towards 45 DAP. *BnGTR2.C03* was the most significantly expressed gene in the leaves. This was followed by *BnGTR2.A06*, which was expressed almost tenfold lower than *BnGTR2.C03* in the leaves. Noteworthy, the sixfold increase in expression levels of *BnGTR2.C03* at 35 DAP was striking in seeds. Sub-family 2 paralogs showed more than a thousandfold weaker expression in the leaves than sub-family 1 paralogs. However, within the sub-family, *BnGTR2.A02* showed a fourfold expression over *BnGTR2.C02.a* while *BnGTR2.C02.b* was undetectable in both tissues. Both paralogs of sub-family 3 were barely detectable in seeds. *BnGTR2.A09* was observed in detectable levels in the leaves. However, this was at a level tenfold lower than the highest expressed paralog *BnGTR2.C03*.

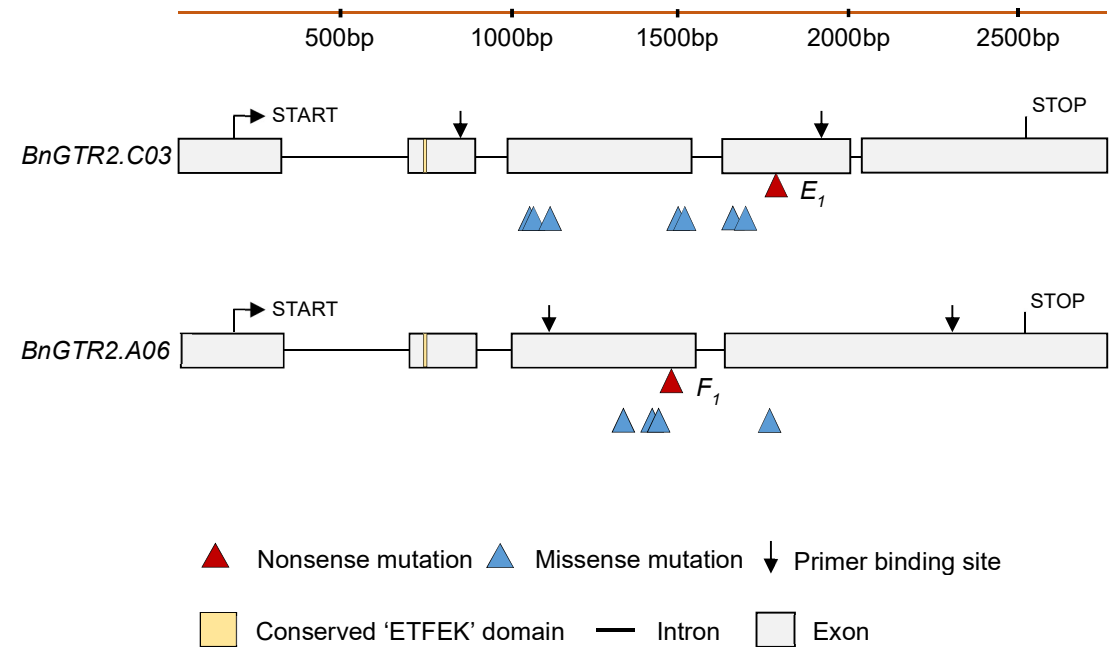
Consequently, I selected both paralogs of sub-family 1 (*BnGTR2.C03* and *BnGTR2.A06*) as promising candidates for functional analyses due to significantly higher expression levels in the leaves.



**Figure 15:** Relative expression of *BnGTR2* paralogs in leaves and seeds of the winter-type oilseed rape Express617. Plants were grown in the greenhouse (20-25°C and 16 h light) after vernalization (4°C, 16 h light for 8 weeks). Leaves and seeds were sampled from five plants 15, 25, 35 and 45 days after pollination (five biological replicates). RT-qPCR was performed with three independent samples of each plant (three technical replicates). The relative expression was calculated after normalizing with two reference genes, *BnGAPDH* and *BnACTIN2*. Error bars represent the standard error of the mean of five biological replicates, with three technical replicates each. Statistical significance was calculated with an ANOVA test ( $p < 0.05$ , linear model, grouping: Tukey test) using the R package ‘Agricolae’. Alphabets over bars represent statistical groups.

#### 4.3.3 EMS-induced mutagenesis of *BnGTR2.A06* and *BnGTR2.C03*

I used the Express617 winter oilseed rape EMS mutant population (Harloff et al., 2012) for mutation screening. In total, 27 EMS-induced mutations were detected for the two selected paralogs *BnGTR2.C03* and *BnGTR2.A06* (Supplementary Table 7). I detected one nonsense mutation each and 11 missense mutations across the two paralogs (Figure 16). Moreover, 14 silent mutations, including synonymous and intronic mutations were detected. Based on the number of mutations detected from both paralogs, an average EMS-induced mutation frequency of 1/35.1 kb was estimated (Supplementary Table 15).



**Figure 16:** The positions of EMS-induced nonsense and missense mutations detected in *BnGTR2.C03* and *BnGTR2.A06*. One amplicon per paralog was screened for EMS mutations. Allele identities are given next to the mutation site for those mutations used for further studies (refer to Table 8 for all allele codes). The ‘ETFEK’ domain characteristic of the proton-dependent oligopeptide transporter (POT/PTR) family is marked by yellow bars. START and STOP represent the translation start and stop sites, respectively. Primer binding sites (downward arrows) represent the extent of the amplicons within which EMS-induced mutations were screened.

To combine single mutants for enhanced phenotypic effects, I selected single M<sub>3</sub> mutants possessing EMS-induced premature stop codon mutations for *BnGTR2.C03* and *BnGTR2.A06*. In genotyping experiments, I first confirmed the presence of the detected mutations in the M<sub>3</sub> generation by Sanger sequencing PCR fragments encompassing the expected mutations using paralog-specific primers (Supplementary Table 5). M<sub>3</sub> plants confirmed with homozygous mutations were selected as crossing parents. Unique single letter allele codes were assigned to mutant and wildtype alleles (Table 8).

**Table 8:** Allele codes assigned to *BnGTR2* EMS mutants and wildtype plants selected as crossing parents in this study.

	Gene name	Mutation position on gDNA <sup>[1]</sup>	Allele code	cDNA change <sup>[1]</sup>	AA change	Mutation type	M <sub>3</sub> seed code <sup>[2]</sup>
M <sub>3</sub> single mutants	<i>BnGTR2.C03</i>	G 1618 A	<i>E<sub>l</sub></i>	G 1064 A	W 354 *	Nonsense	190849
	<i>BnGTR2.A06</i>	C 1225 T	<i>F<sub>l</sub></i>	C 805 T	Q 267 *	Nonsense	190850
Wildtype Express617	<i>BnGTR2.C03</i>	n.a	<i>E<sub>e</sub></i>			n.a	
	<i>BnGTR2.A06</i>	n.a	<i>F<sub>e</sub></i>				
Wildtype Peace	<i>BnGTR2.C03</i>	n.a	<i>E<sub>p</sub></i>			n.a	
	<i>BnGTR2.A06</i>	n.a	<i>F<sub>p</sub></i>				

[1] Position with reference to the translation start site.

[2] Seed codes correspond to the M<sub>3</sub> seeds harvested from the selected M<sub>2</sub> mutants.

Non-mutated wildtype alleles from Express617 and Peace are represented with the ‘e’ and ‘p’ suffixes, respectively.

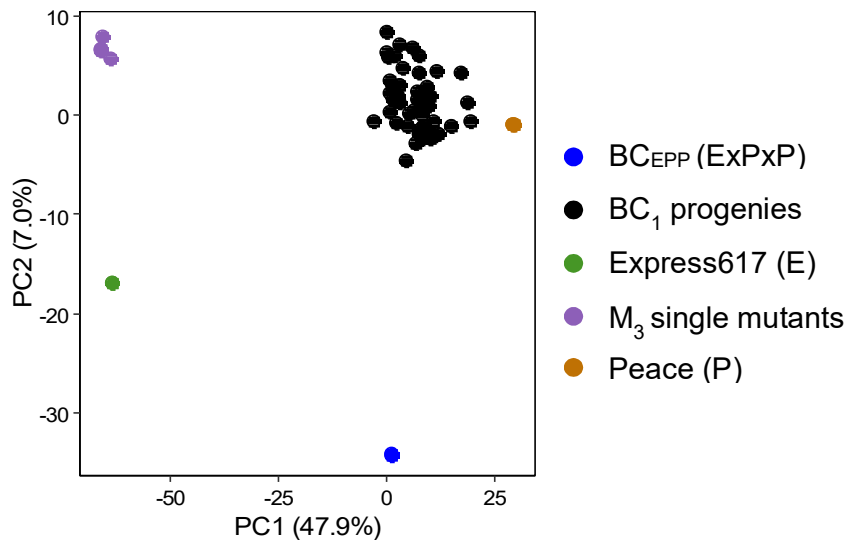
\*Premature stop codon, n.a: not applicable

Confirmed single  $M_3$  mutants were used to make hand crosses in two ways (Supplementary Figure 14, Supplementary Table 16). (i)  $M_3$  single mutants were first crossed with each other (referred to as ' $M_3 \times M_3$ '). The resulting  $F_1$  offspring were selfed to generate the two  $F_2$  populations 205872 and 205873. (ii) Heterozygous *BnGTR2* double mutants were backcrossed with the winter-type Express617 and the spring-type Peace to produce the  $BC_1F_2$  populations 210631 and 210710, respectively.

#### 4.3.4 Genotyping $BC_1$ plants using the *Brassica* 19K SNP array

I selected 51  $BC_1$  plants originating from the backcross of a *BnGTR2* double mutant (seed code 200431\_2) with the spring-type rapeseed Peace for genotyping with a *Brassica* 19K SNP array. Additionally, I genotyped the Express617 (E) and Peace (P) parents and the  $M_3$  single mutants (seed codes 190849\_9 and 190850\_12) corresponding to the selected  $BC_1$  plants. Moreover, a wildtype  $BC_1$  progeny of the two non-mutagenized parent genotypes named ' $BC_{EPP}$  (ExPxP)' was used as a control.

Principal component analyses revealed the population structure of the  $BC_1$  individuals with respect to the two parental genotypes (Figure 17). The principal components PC1 and PC2 explained 47.9% and 7.0% of the total genetic variation, respectively. As expected, the parents (Express617 and Peace) and the  $M_3$  single mutants were clustered distantly. Furthermore, the ' $BC_{EPP}$  (ExPxP)' control plants were observed to be more similar to the mutant  $BC_1$  progenies. Whereas the  $M_3$  single mutants were more similar to the Express617 wildtype. Due to the high mutation load in EMS mutants, a considerable genetic diversity was observable in the PC2, explaining 7.0% of the genetic variation. Due to their EMS mutation loads, the  $M_3$  single mutants and the mutant  $BC_1$  progenies were distant to the wildtype Express617 and the  $BC_{EPP}$  (ExPxP) controls, respectively.

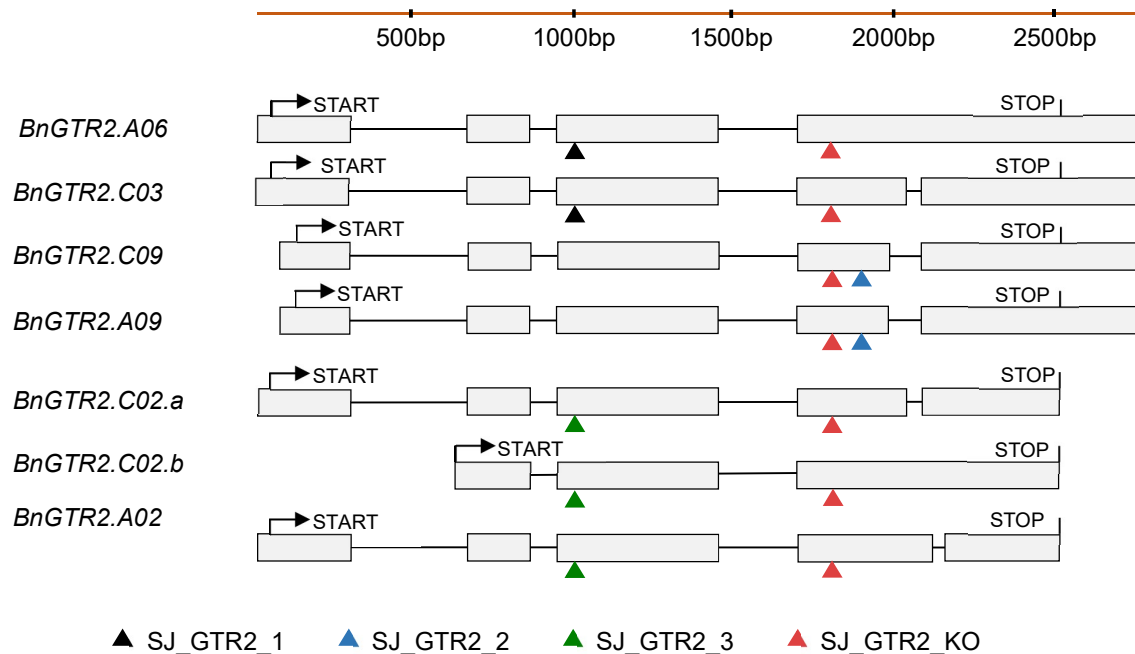


**Figure 17:** Principal component analysis showing relatedness of parents, mutant lines, and their offspring. Homozygous  $M_3$  plants carrying mutations in *BnGTR2* paralogs (genotypes  $E_1E_1F_eF_e$  and  $E_eE_eF_1F_1$ ) were crossed to produce  $F_1$  hybrids (genotype  $E_1E_pF_1F_p$ ).  $F_1$  hybrids were backcrossed with the spring oilseed rape Peace to produce the  $BC_1$  progenies. 51 *BnGTR2*  $BC_1$  plants were genotyped with a *Brassica* 19K SNP array. Non-mutagenized Peace (P) and Express617 (E) plants and a  $BC_1$  offspring of non-mutagenized Express617 and Peace named ' $BC_{EPP}$  (ExPxP)' were used as controls. PC1 and PC2 were calculated using TASSEL5 (Bradbury et al., 2007) and displayed using the 'ggplot2' package in R (Wickham, 2009).

I analyzed the share of parent-specific alleles in the genotyped individuals. The genomic share of the recurrent Peace parent in the genotyped BC<sub>1</sub> plants was between 69.4-85.2%. As expected, the Peace genome share in the 'BC<sub>EPP</sub> (ExPxP)' control was similar to the mutant BC<sub>1</sub> individuals, observed at 69.1%. Consequently, I selected a BC<sub>1</sub> double mutant of *BnGTR2* (seed code 210536\_5) possessing 82.4% Peace genome share that was selfed to generate the segregating BC<sub>1</sub>F<sub>2</sub> population (seed code 210710).

#### 4.3.5 Development of final CRISPR-Cas9 constructs for the targeted mutagenesis of *BnGTR2*

Since seven paralogs for *BnGTR2* were identified in rapeseed, significant phenotypic effects might not be achieved with the sole knock-out of the two highest expressed paralogs *BnGTR2.C03* and *BnGTR2.A06*. Therefore, I also considered the simultaneous knock-out of multiple *BnGTR2* paralogs using CRISPR-Cas9 induced targeted mutagenesis. As the first step, I searched for exonic 5'-NGG-3' motifs conserved across the seven paralogs to target the whole gene family and then within paralogs to target the three sub-families independently. Consequently, the 20 bp upstream sequences were chosen as potential target sites. I targeted the seven paralogs of *BnGTR2* in two ways (Figure 18). For the complete knock-out of *BnGTR2*, I designed a construct for a single target site conserved across all seven paralogs (SJ\_GTR2\_KO). Furthermore, three independent constructs were designed for target sites specific to the three sub-families of *BnGTR2*. I called these constructs SJ\_GTR2\_1, SJ\_GTR2\_2 and SJ\_GTR2\_3 for paralogs of sub-families 1, 2 and 3, respectively (Supplementary Table 17).



**Figure 18:** Locations of selected target oligonucleotides within the coding regions of *BnGTR2* across seven paralogs (marked by triangles). Targets SJ\_GTR2\_1, SJ\_GTR2\_2 and SJ\_GTR2\_3 were designed as sub-family specific targets for sub-families 1, 2 and 3, respectively. SJ\_GTR2\_KO was specific to all seven paralogs of *BnGTR2*. The 20 bp targets are located adjacent to the 5'-NGG-3' protospacer adjacent motif. START and STOP refer to the translation start and stop sites, respectively.



Using the Darmor-*bzh* reference genome, a BLAST analysis did not reveal any off-target effects for the three sub-family targets. Two possible off-targets were detected for the target sequence SJ\_GTR2\_KO conserved across the whole gene family. However, an off-target effect is not expected since one sequence had a disrupted PAM and the other possessed a SNP in the fourth nucleotide upstream of the PAM (Supplementary Figure 16).

The integration of the sgRNA-oligonucleotide complex into the pCas9-TPC vector was confirmed via standard PCR with primers flanking the insertion site (Supplementary Figure 17A) and validation by Sanger sequencing the colony PCR products (Supplementary Figure 17B). As a result, individual pCas9-TPC vectors possessing the sgRNA-oligonucleotide complexes for the four target constructs were developed. The plasmid vectors were transformed into competent *Agrobacterium tumefaciens*. Single colonies were observed after 1-2 days of growth on a selection medium with rifampicin, gentamycin, and spectinomycin (1mg/ml each). Colony PCR of positive colonies revealed the successful integration of the final pCas9-TPC vector into *A. tumefaciens* for each of the constructs (Supplementary Figure 18).

#### 4.4 Discussion

GSLs are secondary metabolites well known for their role in the defense mechanism against insect pests and bacterial and fungal pathogens (Fahey et al., 2001). However, due to their anti-nutritive properties, the reduction of seed GSL content is an important breeding goal for rapeseed. While reduced GSLs in the seeds increase the commercial value of the rapeseed meal, reduced levels in the vegetative parts can compromise plant defenses. Therefore, a favorable outcome would be a seed-specific decrease without disrupting the GSL content elsewhere in the plant. To achieve a seed-specific GSL reduction in rapeseed, I targeted the GSL transporter *GTR2* demonstrated with a seed-specific transport mechanism in Arabidopsis (Nour-Eldin et al., 2012). Moreover, in transcriptomics and genome-wide association studies, the *GTR2* gene was found to be strongly associated with the seed GSL content in rapeseed (Lu et al., 2014; Lu et al., 2019; S. Liu et al., 2020), making it a promising candidate for functional studies.

For functional analyses on *BnGTR2*, my work was structured into two phases. In the first phase, the random EMS-induced mutagenesis approach was used. I identified EMS-induced nonsense mutations in two significantly expressed *BnGTR2* paralogs using the EMS mutagenized rapeseed population. Furthermore, I combined single M<sub>3</sub> mutants of these two paralogs to generate *BnGTR2* double mutants. In the next phase, I used targeted mutagenesis using the CRISPR-Cas9 mediated genome editing approach to knock out multiple *BnGTR2* paralogs simultaneously. As a basis for future studies, this chapter described the production of material that would enable comprehensive functional analyses of the GSL transporter *BnGTR2* in rapeseed.

Using *in silico* analyses, I identified seven *BnGTR2* paralogs in rapeseed. Using the Darmor-*bzh* oilseed rape reference genome, all seven retrieved paralogs of the *GTR2* gene shared >86% polypeptide sequence similarity with the Arabidopsis ortholog. I based the selection of putative functional paralogs on conserved polypeptide sequences and exon-intron structures along with the presence of the 'ETFEK' conserved domain (Jørgensen, Olsen, et al., 2015). To identify potential candidates for functional analysis via EMS mutagenesis, I first analyzed the expression profiles of the seven paralogs in leaves and seeds covering the seed-setting and loading phases (15-45 DAP). Since pyramiding multiple functional mutations into one

individual is cumbersome, the objective was to identify the two most significantly expressed paralogs for mutant detection before the subsequent combination.

In general, for the significantly expressed *BnGTR2* paralogs, I observed a tenfold higher expression level in the leaves than in seeds. In line with the reports of Lu et al. (2019) on rapeseed, *BnGTR2.C03* and *BnGTR2.A06* showed higher expression levels in leaves. Similar to my observations, Lu et al. (2019) have noted that the expression of all paralogs was lower in seeds than the vegetative parts, flowers and the silique pericarp. In *B. juncea*, an increased expression of *GTR2* paralogs was reported in the silique walls (Nour-Eldin et al., 2017). These studies suggested that the seed-specific accumulation of GSLs cannot be directly correlated with a higher expression level of the *GTR2* gene in seeds. Although the average expression levels of the *BnGTR2* paralogs were significantly lower in whole seeds, expression levels within specific parts of the seeds might reveal a more significant role for the transporter. In early studies of Gijzen et al. (1989) using *in vitro* analyses, prematurely excised embryos showed GSL accumulation when grown in a media supplemented with 2-propenyl glucosinolate but were unable to do so in non-supplemented media. Therefore, an embryo-specific transport mechanism and an insignificant role for GSL biosynthesis in the seeds were implied. During seed loading, GSLs are expected to be translocated across several biological membranes after phloem unloading until storage in the embryo (Nour-Eldin & Halkier, 2009). In this regard, investigating the transporter activity in the intermediate membranes, especially the outer and inner integuments and between the endosperm and the embryo, could shed light on the GSL transport mechanism within the seeds.

In my study, the sub-family 2 paralogs (*BnGTR2.C09* and *BnGTR2.A09*) showed low expression levels in general. While *BnGTR2.A09* was weakly expressed in leaves, both were undetectable in the seeds. In comparison, Lu et al. (2019) found a strikingly higher expression of these two paralogs in the roots of rapeseed. Variable expression profiles of the *BnGTR2* paralogs and their tissue-specific roles make it difficult to fully understand the GSL transport mechanism. The question arises whether these paralogs are functionally redundant, especially since the coding regions are highly conserved. Interestingly, studies using *GTR2* functional mutants of *B. juncea* did not suggest a sub-functionalization arising from a substrate specificity for aliphatic, aromatic or indolic GSLs (Nambiar et al., 2021). In this regard, variable GSL composition across different tissues may arise from variable expression patterns of the paralogs. The regulation of GSL transport is not well-studied. Therefore, by addressing the role of an upstream regulatory mechanism, I propose an experiment to investigate the promoter regions of the seven *BnGTR2* paralogs. The presence of regulatory domains for the paralogs could help ascertain their sub-functionalization for tissue-specific activities.

I used the EMS-induced random mutagenesis for sub-family 1 paralogs *BnGTR2.C03* and *BnGTR2.A06* due to their significantly higher expression levels over other paralogs. High sequence conservation across the seven *BnGTR2* paralogs made paralog-specific primer design difficult. Therefore, I could not design paralog-specific primers flanking the ‘ETFEK’ conserved domain for *BnGTR2.C03* and *BnGTR2.A06*. Nevertheless, I detected EMS-induced premature nonsense mutations for both paralogs, conferring almost 50% truncation in the resulting proteins. However, the conserved ‘ETFEK’ domain was retained in both the truncated proteins.

Based on the findings of functional studies from *GTR2* orthologs in *B. rapa* and *B. juncea* (Nour-Eldin et al., 2017; Nambiar et al., 2021), I expected a similar seed-specific reduction in

GSL content from *BnGTR2* knock-out mutants in *B. napus*. Therefore, I combined EMS-induced functional mutants of the highly expressed sub-family 1 paralogs *BnGTR2.C03* and *BnGTR2.A06*. To generate plant material for future experiments, I developed segregating F<sub>2</sub> populations originating from M<sub>3</sub>xM<sub>3</sub> crosses and BC<sub>1</sub>F<sub>2</sub> populations originating from backcrosses with the winter-type Express617 and spring-type Peace rapeseed. Using this material, I propose phenotyping experiments by measuring the GSL content and composition in the leaves and seeds of progenies segregating for the mutant and wildtype alleles in the developed populations.

EMS-mutants harbor high doses of random background mutations (Jung & Till, 2021). Therefore, application in conventional breeding becomes a viable option only after backcrossing with a recurrent wildtype parent, preferably the EMS donor line or an elite variety. In this work, I generated backcrossed mutant populations expected to have a considerably reduced background mutation load. Backcrossing mutants with the spring-type Peace achieved two goals. Firstly, the mutation load was reduced in backcrossed progenies. Secondly, crosses with Peace resulted in the loss of vernalization requirement in the BC<sub>1</sub> generation. As a result, seeds could be harvested within 3-4 months after sowing. In comparison, plants of the winter-type Express617 required ~8 weeks of vernalization, extending the generation time between harvests to about six months. By using the speed-breeding strategy described by Karunarathna et al. (2021), the generation time was reduced by almost 50% and experiments could be performed in a time-effective manner.

Based on the observations from phenotyping experiments described in Chapter 3, an influence of the spring-type Peace genome can be expected in the *BnGTR2* mutants backcrossed with Peace. The generated plant material encompasses the same EMS-induced *BnGTR2* functional mutants in Peace and Express617 backcrossed populations. Therefore, this provides the opportunity to compare the impact of the background parental genomes across the two populations. The entire process of GSL accumulation is expected to be relatively faster in Peace than in Express617 because of the significantly shorter generation cycle of the spring-type Peace. As a proof of concept, using time-course experiments, a comparison of GSL accumulation in seeds and leaves with respect to the expression of *BnGTR2* paralogs can be made for the winter and spring ecotypes of rapeseed.

Using re-sequencing data from 588 diverse rapeseed accessions, the study of Lu et al. (2019) reported that sub-family 1 paralogs *BnGTR2.C03* and *BnGTR2.A06* were under selective sweeps associated with reduced seed GSL content. Although not identified in selective sweeps, the sub-family 2 paralog *BnGTR2.C09* and sub-family 3 paralogs *BnGTR2.A02* and *BnGTR2.C02.b* were located in major QTL regions associated with seed GSL content. This suggests that a further reduction in seed GSL content could be achieved by functionally characterizing these three paralogs in rapeseed.

It is expected that only multiple gene knock-outs are most likely to result in significant phenotypic effects in rapeseed (Braatz et al., 2018; Shah et al., 2018; Karunarathna et al., 2020). Since seven *BnGTR2* paralogs with high sequence conservation were detected, other gene copies can likely have a compensatory effect in the event of functional knock-out of single paralogs. Therefore, in the second phase of this work, I aimed to knock out all seven *BnGTR2* paralogs simultaneously using the CRISPR-Cas9 mediated targeted mutagenesis. Noteworthy, I used resynthesized rapeseed RS306 and the spring-type rapeseed Haydn for the *Agrobacterium*-mediated transformation of hypocotyls (results not shown). However, due to the small scale of the experiments and time restrictions, transformed explants could not be

produced since insufficient numbers of hypocotyls were used for transformation. Nevertheless, the CRISPR-Cas9 vector constructs developed by me in this work will provide an opportunity to study the functional role of the *BnGTR2* paralogs in the future.

## 4.5 Supplementary data

### 4.5.1 Supplementary figures

**Supplementary Figure 14:** Crossing schemes and pedigrees of *BnGTR2* mutant plants used in this study. Single mutants were crossed with each other to produce double mutants. A) M<sub>3</sub> single mutants were crossed to generate F<sub>1</sub> hybrids (genotype  $E_1E_eF_1F_e$ ). The F<sub>1</sub> hybrids were then selfed to generate F<sub>2</sub> populations 205872 and 205873. B) M<sub>3</sub> single mutants of *BnGTR2* paralogs were crossed to produce heterozygous double mutants (genotype  $E_1E_eF_1F_e$ ). The F<sub>1</sub> hybrids were backcrossed with the non-mutagenized spring oilseed rape Peace to generate backcross generation (BC<sub>1</sub>). BC<sub>1</sub> plants were selfed to generate the segregating BC<sub>1</sub>F<sub>2</sub> population (seed code 210710). Plants from the BC<sub>1</sub> generation were first genotyped for the presence of expected EMS-induced mutations and then genotyped with the *Brassica* 19K SNP array. Leaves and seeds from segregating F<sub>2</sub> plants were used for glucosinolate determination. Genotypes mentioned were selected in each generation for further experiments and follow the single letter allele codes as in Table 8. Alleles with ‘e’ and ‘p’ in subscript refer to non-mutagenized Express617 and Peace alleles, respectively.

**Supplementary Figure 15:** The binary vector system used for developing the final Cas9 vectors. A) pChimera vector with target protospacer and the chimeric sgRNA complex. The target oligonucleotide (named ‘Target’) is ligated next to the chimeric sgRNA, constituting the sgRNA-oligonucleotide complex. The chimeric sgRNA is under the influence of the *Arabidopsis thaliana* AtU6-26(P) promoter. B) The final pCas9-TPC vector ligated with the sgRNA-oligonucleotide complex (red box). The Cas9 expression cassette encoding the *Streptococcus pyogenes* Cas9 endonuclease is associated with the PcUbl4-2(P) promoter and the Pea3A(T) terminator. AmpR: ampicillin resistance gene, SpecR: spectinomycin gene, PPT cassette: phosphinothricin resistance gene. LB: left border, RB: right border. Vector sequences are annotated as per the nomenclature of Fauser et al. (2014).

**Supplementary Figure 16:** BLAST results obtained for four selected target sites A) SJ\_GTR2\_1, B) SJ\_GTR2\_2, C) SJ\_GTR2\_3 and D) SJ\_GTR2\_KO considered as target oligonucleotide sequences for the CRISPR-Cas9 mediated knock-out of *BnGTR2*. BLAST analysis was done using the Darmor-*bzh* reference genome on the CLC Genomics Workbench 7. The 20 bp target oligonucleotides and the 5′-NGG-3′ protospacer adjacent motifs (PAM) are annotated with blue and green arrows, respectively. Highlighted nucleotides represent SNPs. 20bp target oligonucleotides were selected as targets.

**Supplementary Figure 17:** Construction of final pCas9-TPC vectors ligated with the sgRNA-oligonucleotide complex. A) Confirmation of successful ligation of the sgRNA-oligonucleotide complex into the pCas9-TPC vector and B) Validation for using Sanger sequencing of colony PCR products. *E. coli* cells were transformed with the pChimera vector ligated with the designed target oligonucleotides (SJ\_GTR2\_1-3 and SJ\_GTR2\_KO). Colony PCR was done using primers pCas9\_F + pCas9\_R flanking the sgRNA-oligonucleotide complex on single *E. coli* colonies (numbered 1-12) growing in spectinomycin selection media. Arrowheads represent the expected fragment size (~1000 bp). +: non-ligated pCas9-TPC vector as positive control and (-): Negative control. Agarose gel electrophoresis conditions: 1% agarose, 100 V, 30 min. GeneRuler™ 1kb DNA ladder used for size reference.

**Supplementary Figure 18:** Detection of the successful transformation of final Cas9-TPC vectors into *A. tumefaciens*. Numbers 1-12 represent single colonies of *A. tumefaciens*

transformed with the corresponding target constructs. Colony PCR was performed with primers pCas9F + pCas9R. GE conditions: 1% agarose, 30min at 100 V. GeneRuler™ 1kb DNA ladder used for size reference. Arrowheads represent the expected fragment size (~1000 bp) +: positive control with empty pCas9-TPC vector, (-): Negative control.

## 4.5.2 Supplementary tables

**Supplementary Table 4:** Primers developed in this study for RT-qPCR analyses.

**Supplementary Table 5:** List of primers used for generating amplicons for mutant screening. One amplicon per homoeolog was used for mutant detection.

**Supplementary Table 7:** All EMS-induced mutations detected in this study for two *BnMYB28*, two *BnCYP79F1* and two *BnGTR2* paralogs. Mutation positions are relative to the translation start site.

**Supplementary Table 15:** Summary of mutation effects conferred by EMS-induced mutations in *BnGTR2* paralogs.

**Supplementary Table 16:** Summary of seed material generated for future studies using EMS-induced functional mutations in *BnGTR2*. Genotypes of parents used for hand crosses and the resulting offspring are shown.

**Supplementary Table 17:** Target oligonucleotide sequences selected for the CRISPR-Cas9 mediated knock-out of seven *BnGTR2* paralogs. Targets were designed based on sequence conservation within the three sub-families and across the entire gene family.

## 4.6 References

- Andersen, T. G., Nour-Eldin, H. H., Fuller, V. L., Olsen, C. E., Burow, M., & Halkier, B. A. (2013). Integration of biosynthesis and long-distance transport establish organ-specific glucosinolate profiles in vegetative *Arabidopsis*. *The Plant Cell*, 25(8), 3133-3145.
- Araki, R., Hasumi, A., Nishizawa, O. I., Sasaki, K., Kuwahara, A., Sawada, Y., Totoki, Y., Toyoda, A., Sakaki, Y., & Li, Y. (2013). Novel bioresources for studies of *Brassica oleracea*: identification of a kale MYB transcription factor responsible for glucosinolate production. *Plant Biotechnology Journal*, 11(8), 1017-1027.
- Augustine, R., Majee, M., Gershenzon, J., & Bisht, N. C. (2013). Four genes encoding MYB28, a major transcriptional regulator of the aliphatic glucosinolate pathway, are differentially expressed in the allopolyploid *Brassica juncea*. *Journal of Experimental Botany*, 64(16), 4907-4921. doi:10.1093/jxb/ert280
- Bisht, N. C., Gupta, V., Ramchiary, N., Sodhi, Y., Mukhopadhyay, A., Arumugam, N., Pental, D., & Pradhan, A. (2009). Fine mapping of loci involved with glucosinolate biosynthesis in oilseed mustard (*Brassica juncea*) using genomic information from allied species. *Theoretical and Applied Genetics*, 118(3), 413-421.
- Braatz, J., Harloff, H.-J., Emrani, N., Elisha, C., Heepe, L., Gorb, S. N., & Jung, C. (2018). The effect of *INDEHISCENT* point mutations on silique shatter resistance in oilseed rape (*Brassica napus*). *Theoretical and Applied Genetics*, 131(4), 959-971.

- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633-2635.
- Bu, X. Y., Wang, Y. Y., Chen, F. Y., Tang, B. B., Luo, C. Z., Wang, Y., Ge, X. P., & Yang, Y. H. (2018). An evaluation of replacing fishmeal with rapeseed meal in the diet of *Pseudobagrus ussuriensis*: growth, feed utilization, nonspecific immunity, and growth-related gene expression. *Journal of the World Aquaculture Society*, 49(6), 1068-1080.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., & Samans, B. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, 345(6199), 950-953.
- Chen, S., Glawischnig, E., Jørgensen, K., Naur, P., Jørgensen, B., Olsen, C. E., Hansen, C. H., Rasmussen, H., Pickett, J. A., & Halkier, B. A. (2003). *CYP79F1* and *CYP79F2* have distinct functions in the biosynthesis of aliphatic glucosinolates in *Arabidopsis*. *The Plant Journal*, 33(5), 923-937.
- Fahey, J. W., Zalcmann, A. T., & Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, 56(1), 5-51.
- Frerigmann, H., Glawischnig, E., & Gigolashvili, T. (2015). The role of MYB34, MYB51 and MYB122 in the regulation of camalexin biosynthesis in *Arabidopsis thaliana*. *Frontiers in Plant Science*, 6, 654.
- Friedt, W., Tu, J., & Fu, T. (2018). Academic and economic importance of *Brassica napus* rapeseed. In *The Brassica napus genome* (pp. 1-20): Springer.
- Gigolashvili, T., Yatusovich, R., Berger, B., Müller, C., & Flügge, U. I. (2007). The R2R3-MYB transcription factor *HAG1/MYB28* is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*, 51(2), 247-261.
- Gijzen, M., McGregor, I., & Séguin-Swartz, G. (1989). Glucosinolate uptake by developing rapeseed embryos. *Plant Physiology*, 89(1), 260-263.
- Halkier, B. A., & Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.*, 57, 303-333.
- Harloff, H.-J., Lemcke, S., Mittasch, J., Frolov, A., Wu, J. G., Dreyer, F., Leckband, G., & Jung, C. (2012). A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. *Theoretical and Applied Genetics*, 124(5), 957-969.
- Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., & Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nature Biotechnology*, 30(8), 798-802.
- Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., Araki, R., Sakurai, N., Suzuki, H., & Aoki, K. (2007). Omics-based identification of *Arabidopsis* *Myb* transcription factors regulating aliphatic glucosinolate biosynthesis. *Proceedings of the National Academy of Sciences*, 104(15), 6478-6483.

- Höfgen, R., & Willmitzer, L. (1988). Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Research*, 16(20), 9877.
- Ishida, M., Hara, M., Fukino, N., Kakizaki, T., & Morimitsu, Y. (2014). Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables. *Breeding Science*, 64(1), 48-59.
- Jørgensen, M. E., Olsen, C. E., Geiger, D., Mirza, O., Halkier, B. A., & Nour-Eldin, H. H. (2015). A Functional EXXEK Motif is Essential for Proton Coupling and Active Glucosinolate Transport by NPF2.11. *Plant & Cell Physiology*, 56(12), 2340-2350. doi:10.1093/pcp/pcv145
- Jung, C., & Till, B. (2021). Mutagenesis and genome editing in crop improvement: perspectives for the global regulatory landscape. *Trends in Plant Science*.
- Karunarathna, N. L., Patiranage, D. S. R., Harloff, H.-J., Sashidhar, N., & Jung, C. (2021). Genomic background selection to reduce the mutation load after random mutagenesis. *Scientific Reports*, 11(1), 19404. doi:10.1038/s41598-021-98934-5
- Karunarathna, N. L., Wang, H., Harloff, H. J., Jiang, L., & Jung, C. (2020). Elevating seed oil content in a polyploid crop by induced mutations in *SEED FATTY ACID REDUCER* genes. *Plant Biotechnology Journal*, 18(11), 2251-2266.
- Kim, Y. B., Li, X., Kim, S.-J., Kim, H. H., Lee, J., Kim, H., & Park, S. U. (2013). MYB transcription factors regulate glucosinolate biosynthesis in different organs of Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Molecules*, 18(7), 8682-8695.
- Kittipol, V., He, Z., Wang, L., Doheny-Adams, T., Langer, S., & Bancroft, I. (2019). Genetic architecture of glucosinolate variation in *Brassica napus*. *Journal of Plant Physiology*, 240, 152988. doi:<https://doi.org/10.1016/j.jplph.2019.06.001>
- Liu, S., Huang, H., Yi, X., Zhang, Y., Yang, Q., Zhang, C., Fan, C., & Zhou, Y. (2020). Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome-wide association study. *Plant Biotechnology Journal*, 18(6), 1472-1484.
- Lu, G., Harper, A. L., Trick, M., Morgan, C., Fraser, F., O'Neill, C., & Bancroft, I. (2014). Associative Transcriptomics Study Dissects the Genetic Architecture of Seed Glucosinolate Content in *Brassica napus*. *DNA Research*, 21(6), 613-625. doi:10.1093/dnares/dsu024
- Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., Zhang, C., Chen, Z., Xiao, Z., & Jian, H. (2019). Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nature Communications*, 10(1), 1154.
- Mitreiter, S., & Gigolashvili, T. (2021). Regulation of glucosinolate biosynthesis. *Journal of Experimental Botany*, 72(1), 70-91.
- Nambiar, D. M., Kumari, J., Augustine, R., Kumar, P., Bajpai, P. K., & Bisht, N. C. (2021). GTR1 and GTR2 transporters differentially regulate tissue-specific glucosinolate contents and defence responses in the oilseed crop *Brassica juncea*. *Plant, Cell & Environment*.



Nour-Eldin, H. H., Andersen, T. G., Burow, M., Madsen, S. R., Jørgensen, M. E., Olsen, C. E., Dreyer, I., Hedrich, R., Geiger, D., & Halkier, B. A. (2012). NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature*, 488(7412), 531.

Nour-Eldin, H. H., & Halkier, B. A. (2009). Piecing together the transport pathway of aliphatic glucosinolates. *Phytochemistry Reviews*, 8(1), 53-67.

Nour-Eldin, H. H., Madsen, S. R., Engelen, S., Jørgensen, M. E., Olsen, C. E., Andersen, J. S., Seynnaeve, D., Verhoye, T., Fulawka, R., Denolf, P., & Halkier, B. A. (2017). Reduction of antinutritional glucosinolates in *Brassica* oilseeds by mutation of genes encoding transporters. *Nature Biotechnology*, 35, 377. doi:10.1038/nbt.3823

<https://www.nature.com/articles/nbt.3823#supplementary-information>

Rask, L., Andréasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., & Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Molecular Biology*, 42(1), 93-114.

Reintanz, B., Lehnen, M., Reichelt, M., Gershenzon, J., Kowalczyk, M., Sandberg, G., Godde, M., Uhl, R., & Palme, K. (2001). bus, a bushy *Arabidopsis CYP79F1* knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. *The Plant Cell*, 13(2), 351-367.

Sánchez-Pujante, P., Sabater-Jara, A., Belchí-Navarro, S., Pedreño, M., & Almagro, L. (2018). Increased glucosinolate production in *Brassica oleracea* var. *italica* cell cultures due to coronatine activated genes involved in glucosinolate biosynthesis. *Journal of Agricultural and Food Chemistry*, 67(1), 102-111.

Seo, M. S., Jin, M., Chun, J.-H., Kim, S.-J., Park, B.-S., Shon, S.-H., & Kim, J. S. (2016). Functional analysis of three *BrMYB28* transcription factors controlling the biosynthesis of glucosinolates in *Brassica rapa*. *Plant Molecular Biology*, 90(4), 503-516.

Shah, S., Karunarathna, N. L., Jung, C., & Emrani, N. (2018). An *APETALA1* ortholog affects plant architecture and seed yield component in oilseed rape (*Brassica napus* L.). *BMC Plant Biology*, 18(1), 380. doi:10.1186/s12870-018-1606-9

Sønderby, I. E., Geu-Flores, F., & Halkier, B. A. (2010). Biosynthesis of glucosinolates—gene discovery and beyond. *Trends in Plant Science*, 15(5), 283-290.

Textor, S., De Kraker, J.-W., Hause, B., Gershenzon, J., & Tokuhiya, J. G. (2007). MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. *Plant Physiology*, 144(1), 60-71.

Till, B. J., Zerr, T., Comai, L., & Henikoff, S. (2006). A protocol for TILLING and Ecotilling in plants and animals. *Nature Protocols*, 1(5), 2465.

Tripathi, M., & Mishra, A. (2007). Glucosinolates in animal nutrition: A review. *Animal Feed Science and Technology*, 132(1-2), 1-27.

- Velasco, P., Soengas, P., Vilar, M., Cartea, M. E., & del Rio, M. (2008). Comparison of glucosinolate profiles in leaf and seed tissues of different *Brassica napus* crops. *Journal of the American Society for Horticultural Science*, 133(4), 551-558.
- Wei, D., Cui, Y., Mei, J., Qian, L., Lu, K., Wang, Z.-M., Li, J., Tang, Q., & Qian, W. (2019). Genome-wide identification of loci affecting seed glucosinolate contents in *Brassica napus* L. *Journal of Integrative Plant Biology*, 61(5), 611-623. doi:10.1111/jipb.12717
- Wickham, H. (2009). *Elegant graphics for data analysis* (Vol. 35).
- Wiesner, M., Schreiner, M., & Zrenner, R. (2014). Functional identification of genes responsible for the biosynthesis of 1-methoxy-indol-3-ylmethyl-glucosinolate in *Brassica rapa* ssp. *chinensis*. *BMC Plant Biology*, 14(1), 124. doi:10.1186/1471-2229-14-124
- Wittkop, B., Snowdon, R., & Friedt, W. (2009). Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. *Euphytica*, 170(1), 131-140.
- Yang, J., Wang, J., Li, Z., Li, X., He, Z., Zhang, L., Sha, T., Lyu, X., Chen, S., & Gu, Y. (2021). Genomic signatures of vegetable and oilseed allopolyploid *Brassica juncea* and genetic loci controlling the accumulation of glucosinolates. *Plant Biotechnology Journal*.
- Yin, L., Chen, H., Cao, B., Lei, J., & Chen, G. (2017). Molecular characterization of MYB28 involved in aliphatic glucosinolate biosynthesis in Chinese Kale (*Brassica oleracea* var. *alboglabra* Bailey). *Frontiers in Plant Science*, 8, 1083.
- Zuluaga, D. L., Graham, N. S., Klinder, A., van Ommen Kloeke, A. E., Marcotrigiano, A. R., Wagstaff, C., Verkerk, R., Sonnante, G., & Aarts, M. G. (2019). Overexpression of the MYB29 transcription factor affects aliphatic glucosinolate synthesis in *Brassica oleracea*. *Plant Molecular Biology*, 101(1), 65-79.

## 5 Closing discussion

The potential of rapeseed meal (RSM) to serve as a sustainable protein source is compromised by anti-nutritive compounds like glucosinolates (GSLs), a prominent and probably the most detrimental class of anti-nutritive compounds in Brassicaceae. Therefore, breeding programs have focused on improving the nutritional quality of the seed meal via the systematic reduction in the seed GSL content. The narrow genetic diversity arising from a short domestication history and intensive selective sweeps for specific agronomic traits is a delimiting aspect in rapeseed breeding (Chalhoub et al., 2014). Therefore, breeders are constantly seeking novel strategies for increasing the genetic variation in rapeseed. In this respect, EMS-induced mutagenesis is an effective non-transgenic technique to introduce novel variations in crops, especially in the European Union, where GMOs are legally restricted.

The results of my study present a significant contribution towards functional genomics and molecular breeding in rapeseed. In the first stage of my work, I established a novel protocol for detecting EMS-induced mutations on a whole-genome scale in rapeseed using whole-genome sequencing data from a population of 1,988 M<sub>2</sub> EMS mutants (Chapter 2). I detected putative functional mutations arising from canonical C→T and G→A transitions and non-canonical transitions and transversions. Furthermore, I assessed the distribution of EMS mutations at the genome level, which has not been explored before in rapeseed. Consequently, a user-friendly web-accessible database was developed to enable convenient mutation screening in practically any genomic region of interest. In the next phase of my work (Chapter 3), I aimed to reduce the GSL content in rapeseed via the knock-out of genes involved in the GSL biosynthesis and transportation pathway. I studied paralogs of gene families *BnMYB28* and *BnCYP79F1* involved in the core structure formation of aliphatic GSLs and *BnGTR2* for its seed-specific GSL transport mechanism. I have demonstrated the functional effect by knocking out genes *BnMYB28* and *BnCYP79F1* using an EMS mutagenized winter rapeseed Express617 population. EMS-induced functional mutants of the two gene families were detected with significantly reduced aliphatic GSLs in the seeds. To set the foundation for future studies, I have generated plant material encompassing EMS mutants of *BnGTR2* that can be analyzed to understand further the molecular basis of GSL transport in rapeseed (Chapter 4). In an additional work package, I developed CRISPR-Cas9 vector constructs for the targeted simultaneous knock-out of multiple *BnGTR2* paralogs.

### 5.1 The TILLING by whole-genome sequencing platform is an unprecedented resource for functional genomics and breeding in rapeseed

The rapeseed EMS resource developed by Harloff et al. (2012) has been instrumental in identifying EMS mutants with several beneficial agronomic traits in the past (Guo et al., 2014; Emrani et al., 2015; Braatz et al., 2018; Shah et al., 2018; Sashidhar et al., 2019; Karunarathna et al., 2020). I used this resource as the basis of my first research area (Chapter 2) to address the urgent need for high-throughput, reliable and convenient screening methods for mutant discovery in EMS mutants. Conventionally, EMS-induced mutation detection assays have relied on laborious and time-consuming techniques often oriented around limited gene families. Therefore, it was clear that the full potential of this resource could only be realized if more efficient mutation screening protocols were established. A major improvement over previous studies was recently achieved when an amplicon sequencing approach was used to detect EMS mutants from 12 gene families encompassing 22 homoeologs involved in phytic acid synthesis (Sashidhar et al., 2019) in rapeseed. Thus, EMS-induced mutations could be efficiently identified in a far more sensitive and high-

throughput manner over previous screenings. The next step was to develop this resource further using a TILLING by whole-genome sequencing (TbyWGS) approach, the golden standard for detecting mutations in EMS mutants. Since a comprehensive platform representing functional EMS mutants for rapeseed was lacking to date, the TbyWGS approach was adopted in my study.

The current database contains ~78 million high confidence EMS-induced mutations detected from whole-genome sequencing data from 1,988 M<sub>2</sub> plants combined in 497 fourfold (4x) pools. Each pool represents four segregating M<sub>2</sub> plants originating from the corresponding M<sub>1</sub> parents. I reason that my TbyWGS resource has high reliability due to two reasons. Firstly, the original resource from Harloff et al. (2012) is a ‘tried and tested’ platform of EMS-mutagenized rapeseed plants. In former studies on this resource, sufficient mutation frequencies (1/12 - 1/72 kb) were observed, allowing the detection of functional mutations across multiple gene families. Secondly, for mutation detection on the whole-genome level, I used a high-quality reference genome assembly from an advanced inbred line (>F<sub>11</sub>) of the winter-type rapeseed ‘Express617’ (Lee et al., 2020). Since the mutant platform is also developed with the M<sub>2</sub> offspring of the EMS treated inbred line Express617 (F<sub>11</sub>), erroneous mutant discovery due to residual heterozygosity is expected to be extremely low (<0.0005%). Moreover, I controlled the quality of mutations based on their mapping qualities (MQ≥30) and allele depths (AD) based on the expected Mendelian segregation patterns of the four M<sub>2</sub> plants in each 4x pool. Moreover, the successful cross-validation of seven previously detected mutations of conventional screenings (Supplementary Table 2) and twelve newly detected mutations from the NGS dataset (Supplementary Table 3) strongly supports the reliability of mutations detected with the TbyWGS approach.

I focused on analyzing the mutation types, their effects and distribution on the whole-genome level. My first priority was to characterize the mutation effects arising from canonical C→T and G→A transitions that comprised 55% of all high confidence mutations. In contrast to previous studies that estimated EMS mutation frequencies based exclusively on the amplicons screened, I could calculate a mutation frequency on the genome-wide level. I calculated a frequency of one EMS mutation per 23.6 kb of the Express617 genome (effective size ~925 Mb). Interestingly, this value was identical to the estimations of Tang et al. (2020), who reported a maximum frequency of 1/23.64 kb from a dataset of 20 whole-genome sequenced EMS mutants. Furthermore, I identified possible mutation hotspots by estimating mutation frequencies for the 19 chromosomes of *B. napus*. I speculated a chromosomal bias for EMS mutagenesis due to the variable distribution of EMS mutations observed for the A (35/Mb) and C (41/Mb) sub-genomes of rapeseed. Moreover, mutation frequencies were generally lower in the telomeric regions for most chromosomes. In this regard, the physical conformation and chromatin structure of *B. napus* sub-genome chromosomes during interphase could influence the accessibility for EMS-induced mutagenesis. This could shed light on a possible ‘non-random’ nature of EMS mutagenesis.

I extended my analyses to the remaining 45% non-canonical transitions and transversions. Strikingly, the distribution of canonical and non-canonical mutation types showed a similar pattern for nearly all chromosomes with reduced mutation densities towards the chromosomal ends. The occurrence of non-C→T and G→A transitions is not a novel observation. Such mutations have been reported in varying frequencies and accounted for 12% - 31% of all detected mutations from EMS mutants of several crops like rice (Till et al., 2007), maize (Lu et al., 2018), barley (Caldwell et al., 2004), soybean (Lakhssassi et al., 2020) and sunflower (Fanelli et al., 2021). In these studies, the occurrence of such mutations was attributed to

residual heterozygosities in the EMS mutants. The findings of my study are against this argument. Since an advanced F<sub>11</sub> inbred line of Express617 was used for mutant production, the residual heterozygosity in the mutants is expected to be lower than 0.0005%. Therefore, the called and filtered variants are expected to be reliable EMS-induced mutations. Using DNA ethylation experiments, Segal (1984) has demonstrated that almost all nucleotide substitutions and even InDels can be incorporated due to depurination or depyrimidation after the EMS treatment. The next question is, if EMS can confer all nucleotide changes, why are C→T and G→A transitions more apparent in comparison to others? Deeper analyses on the biochemical effect of EMS on nucleotides could help answer this question. Additionally, an asymmetrical efficiency is speculated for the conventionally used CELI endonuclease enzyme for detection of EMS mutations. Since CELI has a high preference for cleaving heteroduplexes of C→T and G→A transitions, detection rates for other mutations are lower. As a proof of concept, conventional mutation screening with the ENDO1 mismatch-specific endonuclease (Triques et al., 2007) capable of efficiently cleaving all mismatches could be done to support this argument.

As the database does not represent the entirety of the original resource of 7,680 M<sub>2</sub> plants, a possible expansion to more sequenced M<sub>2</sub> plants can be considered in the future. Presently, the 1,988 M<sub>2</sub> plants in the resource were sufficient in yielding putative functional mutants on a genome-wide level. However, only subsequent screenings across multiple candidate gene families will reveal if further expansions are required. In the future, further additions to this resource are anticipated. The collection of high confidence non-canonical mutations and insertion-deletion (InDel) variants along with their mutation effects on the protein level can be seamlessly integrated to extend the present resource. For the long-term availability of the mutant seed material for users, I propose two strategies for seed propagation from the mutants. Firstly, polycrosses can be done within the M<sub>3</sub> plants originating from the same M<sub>2</sub> family sequenced as 4x pools. This will yield segregating progenies that retain the mutations detected in the respective M<sub>2</sub> pool. However, this is expected to increase the mutation loads causing adverse phenotypic effects and even lethality in the progenies, thereby reducing the seed output. The more preferred option would be to cross M<sub>3</sub> mutants with cytoplasmic male sterile (CMS) lines to maintain mutations in a heterozygous state. This will simultaneously enable the reduction of the mutation load in the progenies.

In conclusion, my study significantly contributes to reverse genetics, molecular breeding and functional genomics in rapeseed. This was achieved by developing an unprecedented platform giving access to EMS-induced mutations screenable at the genomic scale. Moreover, exploiting EMS mutations in the upstream and downstream regulatory regions has not been considered before. This study provides the novel prospect of utilizing ‘regulatory mutants’ rather than ‘functional mutants’ in research areas focusing on fine-tuning traits through gene regulation. Access to the data through the web-based resource combined with procurable seed material increases the value of this study.

## **5.2 Selecting candidate genes controlling glucosinolate metabolism, a complex quantitative trait in Brassicaceae**

My second research area (Chapter 3) focused on reducing glucosinolate content in the seeds of rapeseed. I started with a comprehensive literature survey of potential candidate genes controlling GSL biosynthesis, regulation and transport in *Brassica*. Information regarding GSL metabolism was mainly sourced from *Arabidopsis*, the close relative of rapeseed (Halkier & Gershenzon, 2006; Sønderby et al., 2010) since most aspects of the GSL metabolic pathway have already been investigated (Blažević et al., 2020; Mitreiter &

Gigolashvili, 2021). Since GSL metabolism is a diverse and complex quantitative trait, multiple gene families are expected to be associated with it. However, while translating this knowledge to rapeseed, the complexity of the pathway is magnified since multiple homoeologs for the same *Arabidopsis* ortholog can exist in rapeseed. Therefore, a strategic candidate gene selection for functional analyses was required.

The core structure formation of aliphatic GSLs and their transportation to the seeds were of particular interest in my work. Therefore, for their prominent role in the metabolism of aliphatic GSLs in *Brassica* species, I selected the genes *MYB28* and *CYP79F1* (Kittipol et al., 2019; S. Liu et al., 2020) as promising candidates in my study. In addition, the gene *GTR2* for its importance in the seed-specific transport of GSLs (Nour-Eldin et al., 2017) was targeted to achieve the seed-specific reduction in GSL content without compromising the profile elsewhere in the plant. I used the genomic DNA and amino acid sequences of the three gene families as BLAST queries on the ‘Darmor-*bzh*’ rapeseed reference. As expected, several hits were revealed for the candidate genes. By aligning amino acid sequences, I could identify genuine paralogs by assessing the sequence conservation within gene copies and with the *Arabidopsis* orthologs. Therefore, I removed paralogs arising from incorrect genome annotations and poor sequence similarities from further analyses. Consequently, I identified three, two and seven paralogs for the genes *BnMYB28*, *BnCYP79F1* and *BnGTR2*, respectively, in rapeseed.

Combining multiple functional mutants into a single individual by crossing is a laborious process. Therefore, for functional analyses, the expression studies in my work enabled the selection of the two most promising paralogs from each gene family. I analyzed expression levels of all selected paralogs in leaves and seeds in a time-course experiment encompassing 15-45 DAP. It is shown that GSL biosynthesis has no prominent role in the seeds (Nour-Eldin & Halkier, 2009) and occurs primarily in the vegetative parts (Toroser et al., 1995; Reintanz et al., 2001). Since GSL biosynthesis increases at the time of seed setting (Brown et al., 2003; Gigolashvili et al., 2007), I expected that GSL biosynthesis genes show higher expression levels in leaves during the early stages of seed setting (15-25 DAP) before subsequent accumulation in seeds. It has been demonstrated that the *MYB28* TF controls the expression of the *CYP79F1* gene in *Arabidopsis* (Hirai et al., 2007). This was evident since a subsequent increase in *BnCYP79F1* expression (between 25-35 DAP) was observed after the *BnMYB28* expression was at its highest at 15 DAP in the leaves. In selecting candidates for their role in GSL biosynthesis, I prioritized paralogs *BnMYB28.C09* and *BnMYB28.A03* over *BnMYB28.Cnn* due to their significantly higher expression in the leaves. Only two paralogs for *BnCYP79F1* were identified, so I was able to select both for functional studies. Out of the two paralogs, I assumed *BnCYP79F1.C05* to be a more promising candidate due to its higher leaf expression levels and strong association with the aliphatic GSL content in rapeseed, as shown in associative transcriptomic studies of Kittipol et al. (2019).

The functional analyses for the *BnGTR2* paralogs were initially considered exclusively with the CRISPR-Ca9 approach since multiple paralogs were identified in rapeseed. However, expression analyses revealed two significantly expressed paralogs *BnGTR2.C03* and *BnGTR2.A06*. Therefore, I was able to include the two paralogs for functional studies using EMS mutagenesis. Early studies of Gijzen et al. (1989) have shown that to establish the defense mechanism in the crucial generative phase, i.e., seed setting and loading, accumulation of GSLs starts as early as 14-16 DAP in seeds. Therefore, I hypothesized that the expression of the transporter gene is expected to be high at around the same time when the biosynthesis phase is at its peak. Fitting to this hypothesis, I observed that the expression

level of paralog *BnGTR2.C03* was higher at the 15 DAP stage in leaves, reducing to a constant level between 25-45 DAP similar to that of *BnMYB28*. In line with previous studies (Brown et al., 2003; Andersen et al., 2013), I observed that the transport process was concurrent with the biosynthesis process.

### 5.3 *BnMYB28* and *BnCYP79F1* loss of function mutants possess low aliphatic GSL content in the seeds

Using EMS mutants, I demonstrated the effects of genes *BnMYB28* and *BnCYP79F1* on the GSL content and composition in rapeseed. I expected a complete knock-out for *BnMYB28* since both single mutants of *BnMYB28.C09* and *BnMYB28.A03* possessed premature nonsense mutations truncating >95% of the resulting proteins, including the essential DNA-binding R2, R3 and the nuclear localization signal (NLS) domains (Dubos et al., 2010). For *BnCYP79F1*, double mutants encompassed a nonsense mutation for *BnCYP79F1.C05* and a missense mutation for *BnCYP79F1.A06*. Unfortunately, none of the detected mutations for *BnCYP79F1.A06* were premature nonsense mutations. This was because the screened amplicon had only 38% cDNA coverage and five codons with the potential of converting to induced nonsense mutations after the induced EMS mutation. Therefore I utilized a missense mutation (glutamic acid to lysine) that induced a minor modification in the protein folding due to a change in the surface charge of residues (results not shown).

My results clearly showed a significant reduction in total GSL content in the seeds of *BnMYB28* mutants originating from direct M<sub>3</sub>xM<sub>3</sub> crosses. A significant reduction in the aliphatic GSLs progoitrin (55.3%) and glucobrassicinapin (87%) were detected compared to the non-mutagenized Express617 wildtype controls. In contrast, no reduction in gluconapin was observed in *BnMYB28* double mutants. As gluconapin is suggested to be a direct precursor of progoitrin in the biosynthetic pathway, this could point at a possible downregulation of the gene regulating this conversion by *BnMYB28*. Similarly, epiprogoitrin, which is also synthesized from gluconapin, was reduced in the double mutants by 51% compared to the Express617 wildtype controls. It is probable that the hydroxylation of both stereoisomeric products is similarly regulated. For the *BnMYB28* mutants, I propose a proof of concept experiment to analyze the expression profiles of downstream targets of *MYB28* (based on knowledge from Arabidopsis) like *MAM3* involved in the chain-elongation and *CYP79F1*, *CYP79F2* and *CYP83A1* involved in the core structure formation of aliphatic GSLs (Hirai et al., 2007). The experiment can ascertain the role of *BnMYB28* as a key regulator in the transcriptional control of the aliphatic GSL biosynthesis process in rapeseed. Additionally, a possible control of *BnMYB28* over secondary modifications would be a novel finding.

Interestingly for *BnCYP79F1*, the double mutants did not display a major reduction in the total seed GSL content. However, a 32.4% and 30.4% reduction specifically in the levels of C4 GSLs progoitrin and gluconapin, respectively, was achieved. In Arabidopsis, CYP79F1 metabolizes both short and long-chained aliphatic GSLs (Reintanz et al., 2001; Chen et al., 2003). The C5 glucobrassicinapin content was not significantly altered in the *BnCYP79F1* double mutants, This could be due to the residual gene function of the paralog *BnCYP79F1.A06* as only a missense mutation conferring a minor change to the protein folding could be used in this study to generate the double mutants. I speculated the exclusive role of paralog *BnCYP79F1.C05* in the metabolism of the short-chained C4 GSLs. Therefore, experiments with a complete knock-out mutant of *BnCYP79F1* with nonsense mutations for both paralogs could explain a possible sub-functionalization for metabolizing C4 and C5

GSLs. Moreover, the resulting mutants would be expected to yield more significant phenotypic effects.

#### 5.4 Introgression of the recurrent Peace parent masked mutation effects in backcrossed mutants

By backcrossing mutants with the spring-type rapeseed Peace, I could simultaneously achieve two objectives. Firstly, the background mutation load due to the random mutagenesis by EMS was reduced in the BC<sub>1</sub> generation. Secondly, the introgression of the spring genotype into the winter-type Express617 genome abolished the vernalization requirement for flowering in the BC<sub>1</sub> generation. Therefore, as a speed breeding option, I could significantly reduce the time between generations.

While this approach was advantageous, it revealed an interesting shortcoming. In terms of seed GSL content, the spring and winter-type rapeseed are very diverse in content and composition (Hasan et al., 2008). Initially, I anticipated a further reduction in the GSL content in *BnMYB28* and *BnCYP79F1* backcrossed mutants compared to the non-mutagenized Peace wildtype parent. This was because the spring-type Peace already possessed a low seed GSL content (15.7 µmol/g DW). Contrary to expectations, significant reductions in the seed GSL content were not achieved in backcrossed mutants. A general decrease in the GSL content in all backcrossed mutants irrespective of mutations in the candidate genes pointed to the influence of the recurrent Peace genome on the GSL metabolism in the backcrossed progenies. The same mutations that conferred significant phenotypic effects in non-backcrossed mutants did not display any significant effects in the backcrossed mutants. In this regard, the underlying genetic mechanisms controlling the GSL metabolism in the Peace genome are assumed to neutralize the mutation effects.

Although I selected mutants with a higher Peace share in the BC<sub>1</sub> generation, every BC<sub>1</sub> individual still segregates for Express617 and Peace alleles in the background. Considering the distant and diverse spring and winter-type genomes, a variable effect on the GSL metabolism could stem from a differing share of the two genomes in the backcrossed progenies. This could result in plants with a higher variation in GSL content and composition. In this respect, my analyses on the BC<sub>EPP</sub> controls (progeny of non-mutagenized Peace backcrossed with Express617) were instrumental in addressing this speculation. The GSL content in these plants showed a much higher variation (12.8–30.6 µmol/g DW) in the GSL content across the biological replicates than the Peace controls (14.3–17.5 µmol/g DW).

It would be interesting to identify possible loci in the spring ecotype that possibly have a greater role in GSL metabolism over candidates selected in my study. In this regard, the other subgroup 12 R2R3-MYB TFs *HAG2/MYB76* or *HAG3/MYB29* (Gigolashvili et al., 2008) could be speculated for a more significant role in the regulation of aliphatic GSLs in spring ecotypes. Complementation by the two TFs could explain the less significant phenotypic effects in *BnMYB28* backcrossed mutants. Furthermore, the *MYC* genes from the family of basic helix-loop-helix (bHLH) transcription factors and their association with the *MYB* TFs add another dimension to GSL biosynthesis regulation in Brassicaceae (Mitreiter & Gigolashvili, 2021). Using yeast (*Saccharomyces cerevisiae*) two-hybrid screenings, it has been shown that the *MYC2/bHLH06*, *MYC3/bHLH05*, *MYC4/bHLH04* (Schweizer et al., 2013), and *MYC5/bHLH28* (Frerigmann et al., 2014) TFs can influence the DNA-binding properties of the subgroup 12 *MYB* TFs. An interaction between the two TF families can then collectively influence the downstream GSL biosynthesis process.

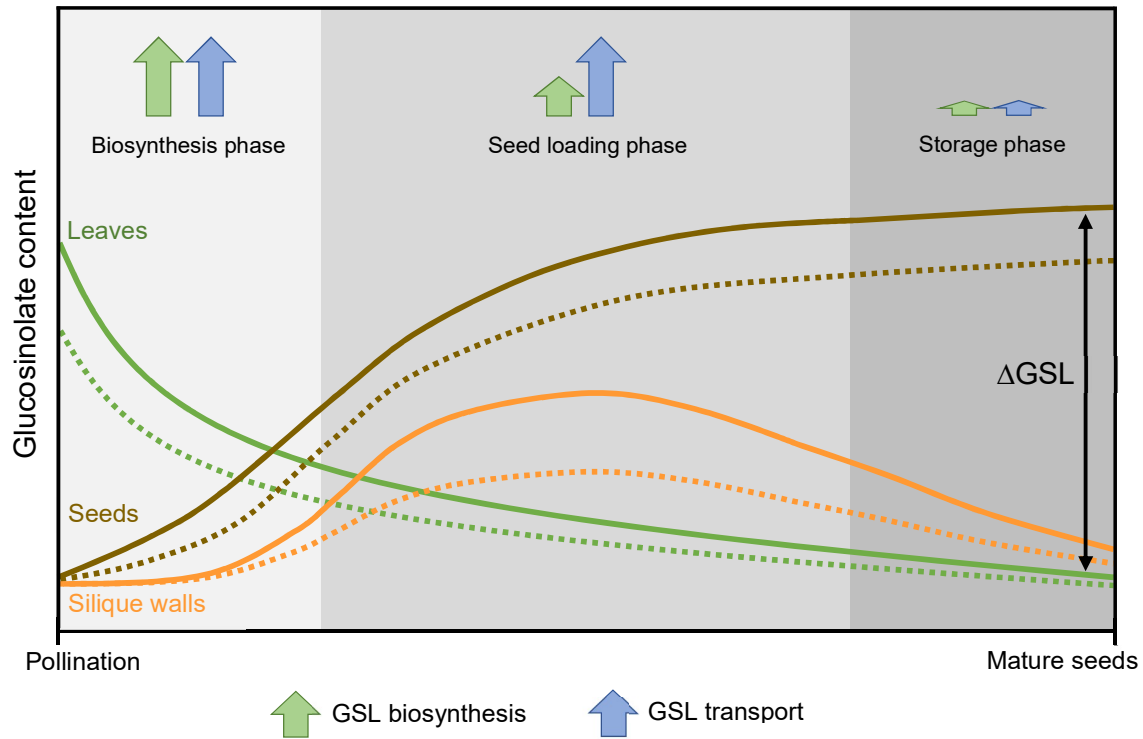


## 5.5 Controlling glucosinolate transport could dictate rapeseed breeding

An ambitious breeding goal for rapeseed has been the seed-specific GSL reduction without affecting other plant parts, especially the leaves. This is due to their prominent role in plant defense mechanisms against insect pests and even bacterial and fungal pathogens (Rask et al., 2000; Fahey et al., 2001). Since disrupting the GSL biosynthesis mechanism means GSL reduction in the whole plant, only a seed-specific reduction is desired. The studies of Nour-Eldin et al. (2017) in *B. rapa* and *B. juncea* and Nambiar et al. (2021) in *B. juncea* demonstrating the seed-specific GSL transport paves the future of breeding rapeseed for reduced seed GSL content.

In my last work package (Chapter 4), I have developed plant material encompassing EMS-induced loss of function mutants of the two most expressed *BnGTR2* paralogs, *BnGTR2.C03* and *BnGTR2.A06*. Double mutants originating from direct M<sub>3</sub>xM<sub>3</sub> crosses and BC<sub>1</sub> backcrosses with Peace and Express617 have been generated for further studies. I considered the simultaneous knock-out of multiple *BnGTR2* paralogs using the CRISPR-Cas9 mediated gene editing approach as an alternative to EMS random mutagenesis. For this, I developed vector constructs that target the entire gene family of seven paralogs or the three sub-families separately. Although plant transformation experiments failed due to the low number of hypocotyls transformed, the designed vector constructs are available for future experiments. The material generated in this work package can be utilized for the functional characterization of *BnGTR2* in rapeseed as a step towards understanding the source-sink relationship of GSLs.

I propose a model for the mobilization of GSLs from the leaves to the seeds through the silique walls across three growth phases, i.e., the biosynthesis, seed loading and storage phases (Figure 19). The GSL biosynthesis commences at an early growth stage in the vegetative parts (Toroser et al., 1995). The GSL content is expected to reach high levels in the leaves before seed loading, assuming endogenous degradation has not occurred. A redirection of GSLs to the inflorescence via transporters is expected as the plant establishes the defense mechanism in the generative organs. As the plant approaches maturity, the GSL biosynthesis and consequently the GSL content in the leaves are expected to diminish as accumulation into the seeds ensues during the loading phase. In the storage phase, both biosynthesis and transport activities are expected to be minimal after the mobilization of GSLs is completed in the seeds. The GSL content increases in the silique walls post pollination since they are capable of GSL biosynthesis (Nour-Eldin et al., 2012). Moreover, their role as the intermediary tissues between seeds and leaves suggests the translocation of the leaf GSLs to the silique walls before seed loading. The subsequent increase in the seed GSLs is concurrent with a decrease in the leaves and silique walls (Brown et al., 2003). In this regard, loss-of-function mutations in biosynthesis genes will reduce GSL content in the whole plant. However, a modified GSL transport mechanism using functional mutants of transporter genes can lower seed GSL content ( $\Delta$ GSL) without disrupting the GSL content elsewhere in the plant.



**Figure 19:** A proposed model for glucosinolate accumulation in different organs during three growth phases from pollination until maturity. The stages are divided into the biosynthesis, seed loading and storage phases. The lengths of green and blue arrows depict the putative levels of biosynthesis and transport activities, respectively. The GSL content in the leaves, silique walls and seeds is represented in green, orange, and brown curves, respectively. The dotted curves represent the putative GSL content in the respective tissues in mutants of biosynthesis genes.  $\Delta$ GSL is the difference in GSL content that can be achieved with modified transport activity.

GSL metabolism is a complex quantitative trait with several associated gene families (Chalhoub et al., 2014; Lu et al., 2019; Wei et al., 2019). Therefore, functional analyses for candidate genes, especially in the polyploid rapeseed genome, continue to be a challenge. Moreover, the vast diversity, heterogeneity of GSL types within and across *Brassica* species and tissue specificity make it difficult to modify the GSL metabolic process. As a result, there are still aspects of GSL biology that remain unknown, leaving open questions that could be addressed in future studies. Although biosynthesis and its regulation in source tissues are fairly known, the transportation process of GSLs to seeds and its regulations need more investigation. The *BnGTR2* paralogs are highly conserved. Therefore functional redundancies are expected. In *B. juncea*, studies have shown that the GSL transporters do not show substrate specificities for specific GSL types (Nambiar et al., 2021). Similarly, sub-functionalization within *BnGTR2* paralogs for the tissue-specific accumulation for specific GSL types could be rejected. However, it can be speculated that the tissue-specific GSL composition and distribution arises from a variable regulation of the *BnGTR2* paralogs. In the future, regulation of GSL transport could be an interesting approach for fine-tuning the distribution of GSLs in specific organs rather than modifying biosynthesis in the whole plant.

In rapeseed, the understanding of this complex trait has been significantly enhanced in recent years owing to advance sequencing-based studies (Nour-Eldin et al., 2017; Kittipol et al., 2019; Lu et al., 2019; S. Liu et al., 2020). Therefore, it is expected that candidate genes for functional studies can now be more easily selected on the basis of genomic and transcriptomic data in the future. In a future study, experiments will be done to compare the

expression profiles of major biosynthesis genes and GSL transporters with the accumulation of GSLs across the leaves, silique walls and seeds (Fangfei Niu, pers commun.). In this respect, understanding the role of the silique walls as an intermediary tissue between the source (leaves) and the sink (seeds) would be interesting.

To conclude, I established a TILLING by whole-genome sequencing resource for the detection of EMS mutations from the Express617 rapeseed EMS population. I demonstrated the reduced seed GSL content in *BnMYB28* and *BnCYP79F1* EMS mutants. Functional mutants of both genes possessed a significantly low progoitrin content in the seeds. Furthermore, I generated backcrossed mutants of both gene families using the winter-type Express617 and spring-type Peace. While mutant analyses in the spring-type Peace backcrossed mutants were suitable for proof of concept experiments, mutants in the winter-type rapeseed Express617 are better suited for rapeseed breeding in the Northern European climates. For future experiments, I developed vector constructs to start the CRISPR-Cas9 mediated targeted mutagenesis of the GSL transporter *BnGTR2*. Furthermore, I also generated plant material encompassing EMS mutants of the two highest expressed paralogs of *BnGTR2*. In this project, all experiments were done under greenhouse conditions. Therefore, a significant influence of the environment is expected on the GSL metabolism under field settings. Field data from different environmental conditions would be helpful in further validating the effects of the mutations.

## 6 Summary

Rapeseed (*Brassica napus* L.) is a major oil crop in the temperate and sub-tropical regions of the world. For utilization as animal feed, it is crucial to reduce the content of anti-nutritive compounds like glucosinolates (GSLs) in the rapeseed meal generated after oil extraction. The presence of especially the aliphatic GSLs has severe metabolic effects on poultry, fish and livestock. Since GSL metabolism is complex, functional analyses for underlying genes are still challenging for the polyploid rapeseed.

I developed a novel TILLING by whole-genome sequencing (TbyWGS) platform to detect EMS-induced mutations on the whole-genome level from a winter rapeseed EMS mutant population. The resource contains 78,150,284 high confidence EMS-induced C→T and G→A mutations originating from whole-genome sequencing data from 1,988 M<sub>2</sub> plants. On average, each plant possessed ~39,000 mutations with a frequency of 1/23.6 kb on the genome level. Roughly 82% of the mutations were located in 5 kb upstream or downstream (~28% each) of gene coding regions or within intergenic regions (26%). The remaining 18% were located within regions coding for genes. Of these, 0.4%, 7.0%, 4.1%, and 4.9% were predicted as nonsense, missense, synonymous and intronic mutations, respectively. Furthermore, the designed web-based resource incorporating whole-genome sequencing data enables the user-friendly detection of EMS mutations in any genomic region of interest.

For Arabidopsis orthologs *AtMYB28*, *AtCYP79F1* and *AtGTR2*, I identified three, two and seven paralogs in rapeseed, respectively. I analyzed the expression profiles of all paralogs in the winter-type rapeseed Express617 at 15, 25, 35 and 45 days after pollination in seeds and leaves and selected the two highest expressed paralogs from each gene family for functional analyses. I used a conventional gel-based approach for screening the EMS mutagenized winter rapeseed population and detected a total of 105 EMS-induced mutations across the six selected paralogs. For combining single functional mutants, I selected EMS-induced premature nonsense mutations for *BnMYB28.C09*, *BnMYB28.A03*, *BnCYP79F1.C05*, *BnGTR2.C03* and *BnGTR2.A06*, leading to a premature termination during protein synthesis. Since nonsense mutations could not be detected for *BnCYP79F1.A06*, I used a missense mutation conferring a glutamic acid to lysine change, causing a minor modification in the protein folding.

I combined single M<sub>3</sub> mutants of *BnMYB28* and *BnCYP79F1* to produce double mutants and performed backcrosses with the Express617 and Peace rapeseed genotypes to reduce the background mutation load. In direct M<sub>3</sub>xM<sub>3</sub> crosses, I developed two F<sub>2</sub> populations, 200527 and 200529, representing double mutants of *BnMYB28* and *BnCYP79F1*, respectively. In backcrossing experiments with Peace, I produced two BC<sub>1</sub>F<sub>2</sub> populations, 210465 and 210462, from double mutants of *BnMYB28* and *BnCYP79F1*, respectively. For GSL determination in the seeds and leaves, I analyzed the four segregating populations using two methods. An enzyme detection assay was used for GSL quantification. HPLC was used for a qualitative assessment of individual GSL types. I observed the most significant reduction in mutants originating from the M<sub>3</sub>xM<sub>3</sub> crosses.

In *BnMYB28* double mutants (seed code 200527), total GSL content in seeds (23.4 µmol/g DW) was reduced by 50.5% compared to the wildtype plants segregating within the same F<sub>2</sub> population (47.2 µmol/g DW) and 52.9% compared to the Express617 controls (49.7 µmol/g DW). For *BnCYP79F1* mutants (seed code 200529), although a significant reduction in total GSLs was not observed, the progoitrin levels were reduced significantly. Compared to the

Express617 control plants (36.3  $\mu\text{mol/g DW}$ ), progoitrin levels were decreased by 55.3% and 32.4% in double mutants of *BnMYB28* (16.20  $\mu\text{mol/g DW}$ ) and *BnCYP79F1* (24.5  $\mu\text{mol/g DW}$ ), respectively.

Using a *Brassica* 19K SNP array, I genotyped backcrossed mutants for a higher genome share of the recurrent Peace parent. Consequently, mutants of the BC<sub>1</sub> generation possessing 80-85% Peace genome share were selected. Due to higher Peace genome share, seed GSL contents were drastically reduced in the segregating F<sub>2</sub> (ranging from 12.5 to 23.1  $\mu\text{mol/g DW}$ ), but no significant reductions could be observed in the mutants compared to the controls in both populations 210465(*BnMYB28*) and 210462 (*BnCYP79F1*). The GSL content was generally lower in leaves than seeds, and was significantly reduced by 58.4% in backcrossed *BnMYB28* double mutants (0.32  $\mu\text{mol/g DW}$ ) compared to plants with wildtype alleles of the same population (0.77  $\mu\text{mol/g DW}$ ) and by 65.5% compared to the Peace wildtype (0.93  $\mu\text{mol/g DW}$ ).

For EMS mutagenesis of *BnGTR2*, single nonsense mutants of *BnGTR2.C03* and *BnGTR2.A06* were crossed to produce *BnGTR2* double mutants. Plants were self-pollinated to generate F<sub>2</sub> populations 205872 and 205873. A BC<sub>1</sub>F<sub>2</sub> population 210710 representing *BnGTR2* double mutants backcrossed with Peace was also produced. Segregating progenies can then be phenotyped for the GSL content and composition in future studies. For future experiments, I developed the final vector constructs for the CRISPR-Cas9 mediated targeted mutagenesis of seven *BnGTR2* homoeologs (*BnGTR2.C03*, *BnGTR2.A06*, *BnGTR2.C09*, *BnGTR2.A09*, *BnGTR2.A02*, *BnGTR2.C02.a* and *BnGTR2.C02.b*).

## 7 Zusammenfassung

Raps (*Brassica napus* L.) ist eine wichtige Ölpflanze in den gemäßigten und subtropischen Regionen der Welt. Für seine Verwendung als Tierfutter ist es entscheidend, den Gehalt an antinutritiven Verbindungen wie Glucosinolaten (GSL) im Rapsschrot, der nach der Ölgewinnung anfällt, zu reduzieren. Insbesondere aliphatische GSL können schwerwiegende Gesundheitsschäden bei Geflügel, Fischen und anderen Nutztieren verursachen. Da die Regulation des GSL-Stoffwechsels sehr komplex ist, sind funktionelle Analysen der beteiligten Gene in polyploidem Raps noch immer eine Herausforderung.

Ich habe eine neuartige TILLING by Whole-Genome Sequencing (TbyWGS)-Plattform entwickelt, in der das komplette Genom von M<sub>2</sub> Familien einer Winterraps-EMS-Mutantenpopulation sequenziert wurde und damit EMS-induzierte Mutationen im gesamten Rapsgenom nachgewiesen werden können. Diese Ressource lieferte mit hoher Qualität 78.150.284 EMS-induzierte C→T- und G→A-Mutationen aus den Sequenzierungsdaten von 1.988 M<sub>2</sub>-Pflanzen. Im Durchschnitt besaß jede Pflanze circa 39.000 Mutationen mit einer Häufigkeit von 1/23,6 kb. Ungefähr 82% der Mutationen befanden sich entweder innerhalb von 5 kb stromaufwärts oder stromabwärts der kodierenden Genregionen (jeweils ~28%) oder in intergenischen Regionen (26%). Die restlichen 18% der Mutationen lagen innerhalb der kodierenden Regionen. Davon waren 0,4% Nonsense-, 7,0% Missense-, 4,1% synonyme Mutationen und 4,9% lagen in Introns. Eine aus den gesamten Sequenzierungsdaten entwickelte webbasierte Ressource ermöglicht eine benutzerfreundliche Identifikation von EMS-Mutationen in jeder beliebigen interessanten genomischen Region.

Für die Arabidopsis-Orthologe *AtMYB28*, *AtCYP79F1* und *AtGTR2* habe ich jeweils drei, zwei bzw. sieben Paraloge in Rapssamen identifiziert. Ich analysierte die Genexpression aller Paraloge im Winterraps Express617 in Samen und Blättern 15, 25, 35 und 45 Tage nach der Bestäubung und wählte die beiden am höchsten exprimierten Paraloge jeder Genfamilie für funktionelle Analysen aus. Ich habe eine konventionelle gelbasierte Methode zum Screening der EMS-mutagenisierten Winterraps-Population verwendet und insgesamt 105 EMS-induzierte Mutationen bei den sechs ausgewählten Paralogen identifiziert. Zur Kreuzung einzelner funktioneller Mutanten wählte ich EMS-induzierte Nonsense-Mutationen für die Gene *BnMYB28.C09*, *BnMYB28.A03*, *BnCYP79F1.C05*, *BnGTR2.C03* und *BnGTR2.A06* aus, die zu einem vorzeitigen Kettenabbruch bei der Proteinsynthese führen. Da für das Gen *BnCYP79F1.A06* keine Nonsense-Mutationen gefunden werden konnte, verwendete ich eine Missense-Mutation, die zu einem Austausch der Aminosäure Glutaminsäure gegen Lysin führt, und damit zu einer Änderung der Sekundärstruktur des Proteins.

Ich kombinierte einzelne M<sub>3</sub>-Mutanten von *BnMYB28* und *BnCYP79F1* durch Kreuzung zu Doppelmутanten und führte Rückkreuzungen mit den Raps-Genotypen Express617 und Peace durch, um den Anteil von Hintergrundmutationen zu reduzieren. In direkten M<sub>3</sub>xM<sub>3</sub>-Kreuzungen entwickelte ich zwei F<sub>2</sub>-Populationen, 200527 und 200529, die Doppelmутanten von *BnMYB28* bzw. *BnCYP79F1* enthielten. In Rückkreuzungsexperimenten mit Peace produzierte ich aus Doppelmутanten von *BnMYB28* bzw. *BnCYP79F1* zwei BC<sub>1</sub>F<sub>2</sub>-Populationen, 210465 und 210462. Zur GSL-Bestimmung in Samen und Blättern habe ich diese vier spaltende F<sub>2</sub>-Populationen mit zwei Methoden analysiert. Für die Bestimmung der Gesamt-GSL wurde ein Enzymassay durchgeführt. Zur Identifikation und Quantifizierung der einzelnen GLS wurde HPLC verwendet. Ich beobachtete die größte Reduktion bei Mutanten, die aus den M<sub>3</sub>xM<sub>3</sub>-Kreuzungen stammten.

Bei *BnMYB28*-Doppelmutanten ( $F_2$ -Population 200527) war der Gesamt-GSL-Gehalt in Samen gegenüber Pflanzen mit Wildtyp-Allelen der gleichen Population um 52,9% reduziert (d.h. von 47,2  $\mu\text{mol/g DW}$  auf 23,4  $\mu\text{mol/g DW}$ ) und gegenüber Express617-Kontrollen um 50,5% (49,7  $\mu\text{mol/g DW}$ ). Bei *BnCYP79F1*-Mutanten ( $F_2$ -Population 200529) wurde zwar keine signifikante Reduktion der Gesamt-GSL beobachtet, jedoch waren die Werte der prominentesten aliphatischen GLS-Verbindung Progoitrin signifikant reduziert. Im Vergleich zu den Express617-Kontrollpflanzen (36,3  $\mu\text{mol/g DW}$ ) waren die Progoitrinspiegel bei Doppelmutanten von *BnMYB28* bzw. *BnCYP79F1* um 55,3 % (16,20  $\mu\text{mol/g DW}$ ) bzw. 32,4 % (24,5  $\mu\text{mol/g DW}$ ) verringert.

Unter Verwendung eines *Brassica* 19K SNP-Arrays habe ich den Genomanteil des Rückkreuzungselters Peace in rückgekreuzten Mutanten der  $BC_1$ -Generation bestimmt und dann Mutanten mit 80-85% Peace-Genomanteil selektiert. In der spaltenden  $F_2$  waren die Samen-Glucosinolatgehalte aufgrund des höheren Peace-Genomanteils wesentlich reduziert (12,5 bis 23,1  $\mu\text{mol/g DW}$ ), jedoch wurden in den rückgekreuzten Mutanten von *BnMYB28* ( $F_2$ -Population 210465) und *BnCYP79F1* ( $F_2$ -Population 210462) keine signifikanten Reduktionen beobachtet. Der GSL-Gehalt in den Blättern lag weit unter den Samenwerten, war aber in rückgekreuzten *BnMYB28*-Doppelmutanten (0,32  $\mu\text{mol/g DW}$ ) signifikant um 65,0 % gegenüber Pflanzen mit Wildtyp-Allelen der gleichen Population reduziert (0,77  $\mu\text{mol/g DW}$ ) bzw. um 58,4 % im Vergleich zum Peace-Wildtyp (0,93  $\mu\text{mol/g DW}$ ).

Einzel-Nonsense-Mutanten von *BnGTR2.C03* und *BnGTR2.A06*, die in der EMS-Mutantenpopulation identifiziert wurden, wurden gekreuzt, um *BnGTR2*-Doppelmutanten zu erzeugen. Pflanzen wurden selbstbestäubt, um die  $F_2$ -Populationen 205872 und 205873 zu erzeugen. Eine mit Peace rückgekreuzte  $BC_1F_2$ -Population 210710, die *BnGTR2*-Doppelmutanten enthält, wurde ebenfalls erstellt. Die Nachkommen dieser spaltenden Populationen können dann in zukünftigen Studien hinsichtlich des GSL-Gehalts und der GLS-Zusammensetzung phänotypisiert werden. Für zukünftige Experimente mit CRISPR-Cas9 Genome Editing entwickelte ich transformationsfähige Vektoren mit Zielsequenzen für alle sieben *BnGTR2*-Paraloge (*BnGTR2.C03*, *BnGTR2.A06*, *BnGTR2.C09*, *BnGTR2.A09*, *BnGTR2.A02*, *BnGTR2.C02.a* und *BnGTR2.C02.b*).

8 Appendix

8.1 Supplementary tables

**Supplementary Table 1:** Summary of canonical transitions originating from the annotated and non-annotated regions of the Express617 genome. Regions of the Express617 reference assembly without chromosomal annotations for the A or C sub-genomes are denoted as non-annotated regions. Share of C→T and G→A transitions located within annotated and non-annotated regions was calculated as an average from 497 sequenced 4x pools.

	Length (bp)	Average SNPs /4x pool	%	Mutation frequency per Mb
chrA01- chrA10	295,690,213	41,413	26.5	35
chrC01- chrC09	468,939,566	76,819	49.1	41
Non-annotated regions	160,465,280	38,248	24.4	59
Total	925,095,059	156,480		

Chromosome lengths are based on the assembled Express617 reference genome (Lee et al. 2020).

**Supplementary Table 2:** Summary of validation experiments to confirm mutations in seven M<sub>2</sub> DNA pools harboring previously detected and characterized EMS mutations within four candidate gene families, *BnREF*, *BnSGT*, *BnSFAR4* and *BnSFAR1*. For each of the selected mutants, corresponding M<sub>2</sub> families were identified. Individual read alignments from each of the 4x pools (named with prefix ‘D’ or ‘IR’) representing the selected mutant M<sub>2</sub> families were visualized for regions harboring expected mutations.

Gene family	Gene name	Darmor- <i>bzh</i> gene ID <sup>[1]</sup>	Mutation position <sup>[2]</sup>	Mutation effect	Mutation type	Pool ID	Express617 gene ID <sup>[3]</sup>	Mutation position in Express617 reference <sup>[2]</sup>	References
<i>BnREF</i>	<i>BnREF1.C03</i>	<i>BnaC03g43810D</i>	G3329A	D392N	Missense	IR_154	<i>C03p045600</i>	C03_36975240	Emrani et al. (2015)
<i>BnSGT</i>	<i>BnSGT.C05</i>	<i>BnaC05g30670D</i>	G1054A	E352K	Missense	IR_229	<i>C05p028480</i>	C05_27555055	Harloff et al. (2012)
<i>BnSFAR4</i>	<i>BnSFAR4.A06a</i>		G497A	G166E	Missense	D533_1		A06_19593539	(Karunarathna et al., 2020)
	<i>BnSFAR4.A06a</i>	<i>BnaA06g18900D</i>	G764A	Splice site		D533_2	<i>A06p021720</i>	A06_19593806	
	<i>BnSFAR4.A06a</i>		C811T	P231L	Missense	IR_388		A06_19593853	



<i>BnSFAR4.A06a</i>			C924T	Q269*	Nonsense	D631	A06_19593966	
<i>BnSFAR1</i>	<i>BnSFAR1.Ann</i>	<i>Not annotated</i>	G411A	G103R	Missense	D436	<i>A03p056350</i>	A03_35121620

[1] Based on the Darmor-*bzh* reference genome (Genoscope).

[2] Relative to the translation START site.

[3] Based on the Express617 reference genome (Lee et al., 2020).

\*Premature stop codon mutation.

**Supplementary Table 3:** Summary of validation experiments to confirm the presence of mutations in 12 pools harboring EMS mutations within candidate gene families *BnMYB28*, *BnCYP79F1* and *BnGTR2*. 4x pools (named with prefix ‘D’) with nonsense mutations for the candidate genes were first identified. Genomic DNA was isolated separately from the leaf samples of the four individuals (#1-4) bulked in each pool. Standard PCR with paralog-specific primers was used to amplify regions encompassing the detected mutations. PCR fragments were Sanger sequenced to validate the presence of mutations.

Gene name	Express617 gene ID <sup>[1]</sup>	Pool ID	Mutation position <sup>[1]</sup>	% Read coverage of mutation in 4x pool (AD/DP) <sup>[2]</sup>	Flanking primers (5' → 3') <sup>[a]</sup>		M <sub>2</sub> genotypes <sup>[b]</sup>				Observed coverage of mutant alleles (%) <sup>[3]</sup>
					Forward primer	Reverse primer	#1	#2	#3	#4	
<i>BnMYB28</i>	<i>C09p006910</i>	D315	C 4710920 T	3/18 = 16.7	gagcttctctattctcatcctag	gaccgaccacctaagaccag	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>1</sub>A<sub>e</sub></i>	12.5
		D402	C 4710699 T	7/23 = 30.4			<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>2</sub>A<sub>e</sub></i>	12.5
		D599	C 4710860 T	5/21 = 23.8			<i>A<sub>3</sub>A<sub>e</sub></i>	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>3</sub>A<sub>e</sub></i>	25.0
	<i>A03p040980</i>	D602	C 21777647 T	3/23 = 13.0	gcattcttgggtgttttgaggg	gcgttggaactatcctcttc	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>4</sub>A<sub>e</sub></i>	<i>A<sub>e</sub>A<sub>e</sub></i>	<i>A<sub>e</sub>A<sub>e</sub></i>	12.5
<i>BnCYP79F1</i>	<i>A06p011120</i>	D465	G 6485927 A	3/12 = 25.0	cattgacgagagggtggagc	ccatcatgatcgctccgactttg	<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>1</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	12.5
		D522	C 6485583 T	5/29 = 17.2			<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>2</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	12.5
		D611	G 6485513 A	5/17 = 29.4			<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>3</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>3</sub>B<sub>e</sub></i>	25.0
		D628	G 6486481 A	3/24 = 12.5			<i>B<sub>4</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	12.5
	<i>C05p012900</i>	D40	G 8261289 A	9/16 = 56.3	cgcacttcgcagaccggcct	cggcctattccaggacgg	<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>5</sub>B<sub>e</sub></i>	<i>B<sub>e</sub>B<sub>e</sub></i>	<i>B<sub>5</sub>B<sub>e</sub></i>	25.0
<i>BnGTR2</i>	<i>A06p023060</i>	D215	C 21118518 T	6/24 = 25.0	cggtcgatacaagactctaactg	cttcttaaggttaggcacgagcac	<i>C<sub>e</sub>C<sub>e</sub></i>	<i>C<sub>1</sub>C<sub>e</sub></i>	<i>C<sub>e</sub>C<sub>e</sub></i>	<i>C<sub>1</sub>C<sub>e</sub></i>	25.0
		D386	G 21118742 A	5/27 = 18.5			<i>C<sub>2</sub>C<sub>e</sub></i>	<i>C<sub>e</sub>C<sub>e</sub></i>	<i>C<sub>e</sub>C<sub>e</sub></i>	<i>C<sub>e</sub>C<sub>e</sub></i>	12.5
		D569	C 21118768 T	7/22 = 31.8			<i>C<sub>3</sub>C<sub>e</sub></i>	<i>C<sub>3</sub>C<sub>e</sub></i>	<i>C<sub>e</sub>C<sub>e</sub></i>	<i>C<sub>e</sub>C<sub>e</sub></i>	25.0

*A<sub>e</sub>*, *B<sub>e</sub>* and *C<sub>e</sub>* are Express617 wild type alleles for the respective paralogs of *BnMYB28*, *BnCYP79F1* and *BnGTR2*, respectively. Mutant alleles (are represented by suffixes 1-5 (*A<sub>1</sub>-A<sub>4</sub>*, *B<sub>1</sub>-B<sub>5</sub>* and *C<sub>1</sub>-C<sub>3</sub>*).

[a] Primers used to amplify the detected mutants from the Express617 TILLING population via standard PCR.

[b] Alleles observed after Sanger sequencing of PCR products from the four segregating M<sub>2</sub> individuals constituting the 4x pools.

[1] Based on the Express617 reference genome (Lee et al., 2020).

[2] Share of reads possessing mutations (AD) out of all reads mapped at that position (DP).

[3] Share of mutant alleles in the 4 M<sub>2</sub> individuals observed after Sanger sequencing of PCR fragments.

**Supplementary Table 4:** Primers developed in this study for RT-qPCR analyses.

Gene name	Primer name	Primer sequence (5' → 3')	T <sub>m</sub> (°C)	Amplicon length (bp)
<i>BnMYB28.C09</i>	GSL_rt_2_6_F	acagttcacgcctcttactcc	60	144
	GSL_2_6_R	gcttcattgcttctcgtgga		
<i>BnMYB28.A03</i>	GSL_2_5_F	gcttcattgcttctcgtggt	64	194
	GSL_2_2_R	gcaaattctctggaggcgtgttgaca		
<i>BnMYB28.Cnn</i>	GSL_2_3_F	cgaacagggtattgatccca	58	264
	GSL_rt_2_3_R	gtgaccttagccgcaacttg		
<i>BnCYP79F1.C05</i>	GSL_3_1_F	cgccggaacacacgccatca	64	258
	GSL_rt_3_1_R	gcaaggaggtgtccgcttcaat		
<i>BnCYP79F1.A06</i>	GSL_rt_3_2_F	ggtatacaaaccagagcgacctc	64	222
	GSL_3_2_R	cctctagacttaacgggtccg		
<i>BnGTR2.A06</i>	GSL_1_10F2	ggcggccagatagcgtttctactgc	66	231
	GSL_rt_1_10R2	gatgggtccagctgacattg		
<i>BnGTR2.C03</i>	GSL_1_5F	gggtttctttagtggagcaggc	62	253
	GSL_1_5R	ccgcgaagaaaatcacacaagcc		
<i>BnGTR2.A09</i>	GSL_1_13F	cgatctactacttaacctggc	62	301
	GSL_rt_1_13R	gtcccaaacgcgtcggctatt		
<i>BnGTR2.C09</i>	GSL_1_16F	ggcagggtttctcgacaaggccgcgaa	68	246
	GSL_1_16R	gccgggatcatgaaacctcca		
<i>BnGTR2.C02.a</i>	GSL_rt_1_1F2	cgtgtacaggcccaagtgt	68	200
	GSL_1_1R2	caagatctgcgcgaacgtgaagggg		
<i>BnGTR2.C02.b</i>	GSL_1_20_F1	caagttggaagctgatggc	66	168

# Appendix

	GSL_rt_1_20_R	ggagggccttgaagactgga		
<i>BnGTR2.A02</i>	GSL_1_19F	gcctcaggtagtccatta	62	148
	GSL_1_21R	ggtcggtgtacttgagctta		
<i>BnACTIN2</i>	ACT1_F	tctggtgatggtgtgtctca	60	141
	ACT2_R	ggtaacatgtaccctctctcg		
<i>BnGAPDH</i>	GC1_F	ccgcttgcttcaacatcatt	60	160
	GC2_R	tctgaggtctttcgacgctg		

Primers were designed using the Darmor-*bzh* reference genome (Genoscope).

F= Forward primer, R= Reverse primer.

**Supplementary Table 5:** List of primers used for generating amplicons for mutant screening. One amplicon per paralog was used for mutant detection.

Standard PCR <sup>[1]</sup>				TILLING PCR <sup>[2]</sup>			
<i>B. napus</i> gene name	Darmor ID	Primer name	Primer sequence (5' → 3')	TILLING primer name	Primer sequence (5' → 3')	TILLING amplicon length (bp) <sup>[3]</sup>	cDNA coverage <sup>[4]</sup>
<i>BnMYB28.C09</i>	<i>BnaC09g05300D</i>	SJ_0082	acagttcacgcctcttactcc	SJ_0094	gagcttctctattctcatcctag	902	65%
		SJ_0083	ggcttgtgagtcacgggatcag	SJ_0095	gaccgaccacctaagaccag		
<i>BnMYB28.A03</i>	<i>BnaA03g40190D</i>	SJ_0084	catgaaacaccttgacgcta	SJ_0096	gcattcttgggtgttttgaggg	1,706	73%
		SJ_0085	ctcgggattactaacctgaaggc	SJ_0097	gcgttggaactatcctcttc		
<i>BnCYP79F1.C05</i>	<i>BnaC05g12520D</i>	SJ_0086	gacaatgatgatgagccttacc	SJ_0098	ccgacttcgcagaccggcct	1,516	59%
		SJ_0087	cctctagacttaacgggtcca	SJ_0099	cggcctattccaggacgg		
<i>BnCYP79F1.A06</i>	<i>BnaA06g11010D</i>	SJ_0088	gttgtttggaaggagacatatc	SJ_0100	cattgacgagagggtggagc	1,180	38%
		SJ_0089	cctctagacttaacgggtccg	SJ_0101	ccatcatgatcgtcccactttg		
<i>BnGTR2.C03</i>	<i>BnaC03g51560D</i>	SJ_0090	ctgttgaggaagaagtggg	SJ_0102	ggtcgatacaagactctaaccg	1058	49%
		SJ_0091	cttcttaaggtaggcacgagga	SJ_0103	gtcatcaaaaagacaacgtagg		
<i>BnGTR2.A06</i>	<i>BnaA06g22160D</i>	SJ_0092	ggtattatcaggaaatgagacgttt	SJ_0104	cggtcgatacaagactctaactg	1113	50%
		SJ_0093	gccatgacaagtatatgactaccg	SJ_0105	cttcttaaggtaggcacgagcac		

[1] Standard PCR was done using unlabeled primers.

[2] TILLING PCR was done using a combination of unlabeled and infrared (IR) labeled primers. Forward and reverse primers were labeled with DY681 and DY781 probes, respectively. The TILLING PCR was done as a nested PCR using amplicons from the standard PCR as the template.

[3] Length of the amplicon analyzed for detection of EMS-induced mutations.

[4] Share of the total cDNA encompassed in the analyzed amplicon for detection of EMS-induced mutations.

PCR conditions: 94°C for 2 min, 36 cycles of 94°C for 30s, 58-68°C for 30s-1 min 30s and 72°C for 1 min, followed by final elongation at 72°C for 5 min.

**Supplementary Table 6:** HPLC performance of commercial glucosinolate standards.

Compound	Side chain	Retention Time (min)	HPLC calibration (linear regression) <sup>[a]</sup>
Glucobrassicin	3-me-sulfinylpropyl	5.79	(area-3367)/162966
Progoitrin	(2 <i>R</i> )-2 hydroxy-3-butenyl	6.43	(area+1631)/164513
Epiprogoitrin	(2 <i>S</i> )-2 hydroxy-3-butenyl	6.60	(area+18347)/157248
Sinigrin	2-propenyl	6.66	(area-11372)/199282
Glucoraphanin	4-me-sulfinylbutyl	6.65	(area+19466)/174317
Gluconapin	3-butenyl	7.69	(area-5783)/175044
Glucobrassicinapin	4-pentenyl	8.77	(area-9690)/162867
Glucotropaeolin	benzyl	8.91	(area-9122)/198955
Glucobrassicin	3-Indolylmethyl	9.40	(area-11784)/552644
Gluconasturtiin	phenethyl	10.00	(area-5888)/164863

[a] Conversion from (area under the peak) to (nmol GSL).

**Supplementary Table 7:** All EMS-induced mutations detected in this study for two *BnMYB28*, two *BnCYP79F1* and two *BnGTR2* paralogs. Mutation positions are relative to the translation start site.

<i>B. napus</i> gene name	M2 plant name <sup>[1]</sup>	M2 mutant zygosity <sup>[2]</sup>	Mutation position on gDNA	cDNA change	Amino acid change	Mutation type
<i>BnMYB28.C09</i>	70_F4	Hom	G 12 C	G 12 C	K 4 N	Missense
	70_H4	Het	G 12 C	G 12 C	K 4 N	Missense
	49_F2	Het	G 26 A	G 26 A	G 9 E	Missense
	66_B11	Het	G 28 A	G 28 A	E 10 K	Missense
	71_B4	Hom	G 28 A	G 28 A	E 10 K	Missense
	72_E9	Het	G 31 A	G 31 A	G 11 R	Missense
	71_E9	Het	G 43 A	G 43 A	G 15 R	Missense
	71_H9	Hom	G 43 A	G 43 A	G 15 R	Missense
	8_E5	Het	G 43 A	G 43 A	G 43 R	Missense
	64_C10	Het	G 45 A	G 45 A	A 15 A	Silent
	3_H2	Het	G 51 A	G 51 A	W 17 *	Nonsense
	11_H7	Het	G 51 A	G 51 A	W 17 *	Nonsense
	54_F10	Het	C 78 T	C 78 T	I 26 I	Silent
	43_E3	Hom	G 97 A	G 97 A	G 33 R	Missense
	54_C6	Het	G 98 A	G 98 A	G 33 E	Missense
	12_A9	Hom	C 122 T	C 122 T	P 41 L	Missense
	54_B9	Het	G 133 A	G 133 A	G 45 R	Missense
	54_C6	Het	G 250 A	G 149 A	G 50 E	Missense
	7_C6	Het	C 289 T	C 188 T	P 63 L	Missense
	7_D6	Hom	C 289 T	C 188 T	P 63 L	Missense
	76_D11	Hom	G 303 A	G 202 A	G 68 S	Missense

	64_A8	Het	G 303 A	G 202 A	G 68 S	Missense
	68_H6	Hom	G 303 A	G 202 A	G 68 S	Missense
	68_A11	Het	G 323 A	G 222 A	E 74 E	Silent
	45_B8	Het	C 335 T	C 234 T	I 78 I	Silent
	11_A5	Het	G 329 A	G 329 A	Q 228 Q	Silent
<i>BnMYB28.A03</i>	54_F3	Hom	G 50 A	G 50 A	W 17 *	Nonsense
	48_C8	Hom	G 51 A	G 51 A	W 17 *	Nonsense
	48_D8	Hom	G 51 A	G 51 A	W 17 *	Nonsense
	54_E1	Het	G 61 A	G 61 A	E 21 K	Missense
	48_H3	Het	G 223 A	G 134 A	G 45 E	Missense
	48_D3	Hom	G 317 A	G 228 A	Q 76 Q	Silent
	50_E3	Het	C 706 T	C 421 T	P 141 S	Missense
	56_E12	Hom	G 737 A	G 452 A	S 151 N	Missense
	50_E3	Het	C 899 T	C 614 T	T 205 I	Missense
<i>BnCYP79F1.C05</i>	52_C8	Het	C 424 T	C 424 T	E 142 *	Nonsense
	64_C7	Hom	C 560 T	C 560 T	T 187 I	Missense
	64_F6	Het	C 606 T	C 606 T	T 202 T	Silent
	54_A6	Het	C 654 T	C 654 T	F 218 F	Silent
	53_B5	Hom	G 770 A	G 770 A	G 257 D	Missense
	63_E7	Hom	G 850 A	G 850 A	E 284 K	Missense
	65_E3	Hom	C 956 T	C 956 T	P 319 L	Missense
	65_A8	Het	C 1399 T	C 1107 T	D 369 D	Silent
	66_F6	Het	G 1401 A	G 1109 A	R 370 K	Missense
	53_C5	Het	C 1430 T	C 1138 T	L 380 L	Silent
	53_D5	Het	C 1430 T	C 1138 T	L 380 L	Silent
	56_A4	Het	C 1430 T	C 1138 T	L 380 L	Silent
	61_A10	Het	C 1521 T	C 1229 T	T 410 I	Missense
	61_C10	Het	C 1521 T	C 1229 T	T 410 I	Missense
<i>BnCYP79F1.A06</i>	61_A2	Het	G 938 A	G 938 A	G 313 E	Missense
	45_D7	Het	G 1379 A	G 1090 A	E 364 K	Missense
	41_A5	Het	G 1384 A	G 1095 A	V 365 V	Silent
	46_B6	Het	G 1384 A	G 1095 A	V 365 V	Silent
	53_B9	Het	G 1385 A	G 1096 A	V 366 M	Missense
	53_C9	Het	G 1385 A	G 1096 A	V 366 M	Missense
	46_F5	Het	G 1394 A	G 1105 A	D 368 N	Missense
	44_F12	Het	G 1399 A	G 1110 A	R 370 R	Silent
	44_H12	Hom	G 1399 A	G 1110 A	R 370 R	Silent

	44_G12	Het	G 1399 A	G 1110 A	R 370 R	Silent
	63_G10	Het	C 1413 T	C 1124 T	S 375 F	Missense
	69_F7	Hom	C 1472 T	C 1183 T	P 395 S	Missense
	74_A7	Het	C 1472 T	C 1183 T	P 395 S	Missense
	69_H7	Het	C 1472 T	C 1183 T	P 395 S	Missense
	43_F5	Het	C 1473 T	C 1184 T	P 395 L	Missense
	43_G5	Het	C 1473 T	C 1184 T	P 395 L	Missense
	48_A12	Het	C 1473 T	C 1184 T	P 395 L	Missense
	48_D12	Het	C 1473 T	C 1184 T	P 395 L	Missense
	76_A4	Het	C 1473 T	C 1184 T	P 395 L	Missense
	75_B9	Het	C 1489 T	C 1200 T	V 400 V	Silent
	42_B3	Het	G 1511 A	G 1222 A	D 408 N	Missense
	61_A2	Het	G 1511 A	G 1222 A	D 408 N	Missense
	51_C7	Hom	C 1519 T	C 1230 T	T 410 T	Silent
	55_B2	Hom	C 1520 T	C 1231 T	L 411 F	Missense
	63_E8	Hom	C 1538 T	C 1249 T	P 417 S	Missense
	44_G6	Het	G 1852 A	G 1289 A	G 430 D	Missense
	44_H12	Hom	G 1898 A	G 1335 A	E 445 E	Silent
	44_F12	Het	G 1898 A	G 1335 A	E 445 E	Silent
	44_G12	Het	G 1898 A	G 1335 A	E 445 E	Silent
<i>BnGTR2.C03</i>	70_F06	Hom	C 878 T	C 418 T	H 140 Y	Missense
	73_F02	Het	C 887 T	C 427 T	P 143 S	Missense
	52_E07	Hom	G 935 A	G 475 A	A 159 T	Missense
	52_F07	Hom	C 1322 T	C 862 T	L 288 F	Missense
	72_H06	Hom	C 1344 T	C 884 T	P 295 L	Missense
	50_A07	Hom	C 1354 T	C 894 T	Y 298 Y	Silent
	73_A09	Hom	C 1514 T	C 960 T	P 320 P	Silent
	49_F09	Hom	G 1533 A	G 979 A	D 327 N	Missense
	73_E07	Het	G 1541 A	G 987 A	K 329 K	Silent
	73_G07	Het	G 1541 A	G 987 A	K 329 K	Silent
	56_C10	Hom	G 1618 A	G 1064 A	W 355 *	Nonsense
	52_G8	Hom	C 1676 T	C 1122 T	F 374 F	Silent
	56_A05	Het	C 1488 T	C 934 T	L 496 F	Missense
	55_C09	Hom	G 1508 A	G 954 A	L503 L	Silent
<i>BnGTR2.A06</i>	14_F12	Hom	C 901 T	C 482 T	L 161 L	Silent
	3_F04	Hom	C 907 T	C 487 T	L163 L	Silent
	73_A03	Het	G 1007 A	G 587 A	R 196 Q	Missense

3_F04	Hom	C 1026 T	C 606 T	F 202 F	Silent
3_D08	Het	C 1062 T	C 642 T	I 214 I	Silent
3_F04	Hom	C 1063 T	C 643 T	L 215 L	Silent
5_F06	Hom	G 1140 A	G 720 A	M 240 I	Missense
16_D07	Hom	G 1153 A	G 733 A	V 245 M	Missense
49_D06	Het	C 1161 T	C 741 T	F 247 F	Silent
49_D03	Hom	G 1188 A	G 768 A	K 256 K	Silent
49_A03	Het	G 1188 A	G 768 A	K 256 K	Silent
7_B11	Het	C 1225 T	C 805 T	Q 267 *	Nonsense
16_A07	Hom	C 1618 T	C 1109 T	T 370 I	Missense

[1] Name of the single M<sub>2</sub> mutant as per the Express617 EMS mutant resource (Harloff et al., 2012).

[2] Zygosity of EMS-induced mutations observed in single M<sub>2</sub> individuals.

Silent mutations do not confer changes in polypeptide sequences.

\* Premature stop codon mutation.

**Supplementary Table 8:** Summary statistics of functional effects conferred by EMS-induced mutations in *BnMYB28* and *BnCYP79F1* paralogs. One amplicon per paralog was used for screening EMS-induced mutations.

Gene name	cDNA coverage	Number of M <sub>2</sub> pools screened [1]	Number of detected mutations				Mutation frequency (1/kb) [2]
			Nonsense mutations	Missense mutations	Silent mutations	Splice site mutations	
<i>BnMYB28.C09</i>	65%	13	2	19	5	0	1/37.8
<i>BnMYB28.A03</i>	73%	4	3	5	1	0	1/67.0
<i>BnCYP79F1.C05</i>	59%	4	1	7	6	0	1/38.0
<i>BnCYP79F1.A06</i>	38%	9	0	19	10	0	1/31.5
Total = 78							Average = 1/43.5 kb

[1] The number of eight-fold (8x) two-dimensional (2-D) M<sub>2</sub> pools used for screening mutations.

[2] Calculated as the number of mutations per M<sub>1</sub> plant based on the number of analyzed M<sub>2</sub> families (8x 2D pools) and amplicon lengths. Formula given by Harloff et al. (2012).



Appendix

**Supplementary Table 9:** Mutant genotyping in the M<sub>3</sub> generation to select crossing parents and combination of single mutations.

M <sub>3</sub> seed code	Gene name	M <sub>2</sub> genotype	Germination rate (%)	Mutation position on gDNA <sup>[1]</sup>	cDNA change	AA change <sup>[2]</sup>	Number of M <sub>3</sub> plants genotyped	No. M <sub>3</sub> genotypes observed		
								Homozygous WT	Heterozygous mutant	Homozygous mutants
190623	<i>BnMYB28.C09</i>	<i>A<sub>1</sub>A<sub>e</sub>B<sub>e</sub>B<sub>e</sub></i>	80	G 51 A	G 17 A	W 17 *	12	6	4	2
190624	<i>BnMYB28.C09</i>	<i>A<sub>1</sub>A<sub>e</sub>B<sub>e</sub>B<sub>e</sub></i>	67	G 51 A	G 17 A	W 17 *	10	1	8	1
190625	<i>BnMYB28.A03</i>	<i>A<sub>e</sub>A<sub>e</sub>B<sub>1</sub>B<sub>1</sub></i>	60	G 50 A	G 50 A	W 17 *	4	0	0	4
190628	<i>BnCYP79F1.C05</i>	<i>C<sub>1</sub>C<sub>e</sub>D<sub>e</sub>D<sub>e</sub></i>	60	C 424 T	C 424 T	E 142 *	9	3	2	4
190630	<i>BnCYP79F1.A06</i>	<i>C<sub>e</sub>C<sub>e</sub>D<sub>1</sub>D<sub>1</sub></i>	80	G 1379 A	G 1090 A	E 364 K	11	0	0	11
190849	<i>BnGTR2.C03</i>	<i>E<sub>1</sub>E<sub>1</sub>F<sub>e</sub>F<sub>e</sub></i>	93	G 1618 A	G 1064 A	W 355 *	9	0	0	9
190850	<i>BnGTR2.A06</i>	<i>E<sub>e</sub>E<sub>e</sub>F<sub>1</sub>F<sub>e</sub></i>	93	C 1225 T	C 805 T	Q 267 *	12	3	7	2

Plants were genotyped using Sanger sequencing of PCR fragments encompassing expected EMS-induced mutations.

All genotypes have been named using designated allele codes for *BnMYB28*, *BnCYP79F1* (Table 6) and *BnGTR2* (Table 8).

[1] Position of EMS-induced mutation on the gDNA from the translational start site.

[2] Position of amino acid change on the polypeptide chain.

\*Premature stop codon.

**Supplementary Table 10:** Genotypes of parents used for hand crosses and the offspring derived thereof made in this study to combine single EMS mutants of *BnMYB28* and *BnCYP79F1* and genotypes of selected crossing parents and their progenies used for phenotyping experiments.

Crossing type	Parental genotypes	Genotype of selected F <sub>1</sub> progeny	F <sub>2</sub> seed code	Genotypes used for phenotyping experiments <sup>[1]</sup>
M <sub>3</sub> x M <sub>3</sub>	<i>A<sub>e</sub>A<sub>e</sub>B<sub>1</sub>B<sub>1</sub></i> x <i>A<sub>1</sub>A<sub>1</sub>B<sub>e</sub>B<sub>e</sub></i>	<i>A<sub>1</sub>A<sub>e</sub>B<sub>1</sub>B<sub>e</sub></i>	200527 <sup>[a]</sup>	<i>A<sub>e</sub>A<sub>e</sub>B<sub>e</sub>B<sub>e</sub></i> , <i>A<sub>1</sub>A<sub>1</sub>B<sub>e</sub>B<sub>e</sub></i> , <i>A<sub>e</sub>A<sub>e</sub>B<sub>1</sub>B<sub>1</sub></i> , <i>A<sub>1</sub>A<sub>1</sub>B<sub>1</sub>B<sub>1</sub></i>
	<i>C<sub>e</sub>C<sub>e</sub>D<sub>1</sub>D<sub>1</sub></i> x <i>C<sub>1</sub>C<sub>1</sub>D<sub>e</sub>D<sub>e</sub></i>	<i>C<sub>1</sub>C<sub>e</sub>D<sub>1</sub>D<sub>e</sub></i>	200529 <sup>[a]</sup>	<i>C<sub>e</sub>C<sub>e</sub>D<sub>e</sub>D<sub>e</sub></i> , <i>C<sub>1</sub>C<sub>1</sub>D<sub>e</sub>D<sub>e</sub></i> , <i>C<sub>e</sub>C<sub>e</sub>D<sub>1</sub>D<sub>1</sub></i> , <i>C<sub>1</sub>C<sub>1</sub>D<sub>1</sub>D<sub>1</sub></i>
[(M <sub>3</sub> x M <sub>3</sub> ) x Peace] x Peace	[( <i>A<sub>1</sub>A<sub>e</sub>B<sub>1</sub>B<sub>e</sub></i> x <i>A<sub>p</sub>A<sub>p</sub>B<sub>p</sub>B<sub>p</sub></i> ) x <i>A<sub>p</sub>A<sub>p</sub>B<sub>p</sub>B<sub>p</sub></i> ]	<i>A<sub>1</sub>A<sub>p</sub>B<sub>1</sub>B<sub>p</sub></i>	210465 <sup>[b]</sup>	<i>A<sub>e</sub>A<sub>p</sub>B<sub>e</sub>B<sub>p</sub></i> , <i>A<sub>1</sub>A<sub>1</sub>B<sub>e</sub>B<sub>p</sub></i> , <i>A<sub>e</sub>A<sub>p</sub>B<sub>1</sub>B<sub>1</sub></i> , <i>A<sub>1</sub>A<sub>1</sub>B<sub>1</sub>B<sub>1</sub></i>
[(M <sub>3</sub> x Peace) x Peace] x [(M <sub>3</sub> x Peace) x Peace]	[( <i>C<sub>1</sub>C<sub>1</sub>D<sub>e</sub>D<sub>e</sub></i> x <i>C<sub>p</sub>C<sub>p</sub>D<sub>p</sub>D<sub>p</sub></i> ) x <i>C<sub>p</sub>C<sub>p</sub>D<sub>p</sub>D<sub>p</sub></i> ] x [( <i>C<sub>e</sub>C<sub>e</sub>D<sub>1</sub>D<sub>1</sub></i> x <i>C<sub>p</sub>C<sub>p</sub>D<sub>p</sub>D<sub>p</sub></i> ) x <i>C<sub>p</sub>C<sub>p</sub>D<sub>p</sub>D<sub>p</sub></i> ]	<i>C<sub>1</sub>C<sub>p</sub>D<sub>1</sub>D<sub>p</sub></i>	210462 <sup>[b]</sup>	<i>C<sub>e</sub>C<sub>p</sub>D<sub>e</sub>D<sub>p</sub></i> , <i>C<sub>1</sub>C<sub>1</sub>D<sub>e</sub>D<sub>p</sub></i> , <i>C<sub>e</sub>C<sub>p</sub>D<sub>1</sub>D<sub>1</sub></i> , <i>C<sub>1</sub>C<sub>1</sub>D<sub>1</sub>D<sub>1</sub></i>

All genotypes have been named as per designated allele codes (refer to Table 6). Non-mutated wildtype alleles from Express617 and Peace are represented by the ‘e’ and ‘p’ suffixes in subscript, respectively.

[1] For quantitative and qualitative glucosinolate determination.

[a] Segregating F<sub>2</sub> population originating from heterozygous *BnMYB28* and *BnCYP79F1* double mutants (originating from direct M<sub>3</sub>xM<sub>3</sub> crosses).

[b] Segregating BC<sub>1</sub>F<sub>2</sub> population originating from heterozygous *BnMYB28* and *BnCYP79F1* double mutants (originating from backcrosses with Peace).

**Supplementary Table 11:** Summary statistics of marker information and samples genotyped with the *Brassica* 19K SNP array. BC<sub>1</sub> individuals were first genotyped for the presence of expected EMS-induced mutations and then genotyped with the SNP array. Marker information was calculated as an average of all genotyped individuals.

Sample type	No. samples genotyped			
	No. BC <sub>1</sub> individuals	No. F <sub>1</sub> parents	No. wild type controls <sup>[1]</sup>	No. polymorphic SNP markers <sup>[2]</sup>
<i>BnMYB28</i> double mutants <sup>[a]</sup>	73	6	4	8,961
<i>BnCYP79F1</i> single mutants <sup>[b]</sup>	86	4	4	8,334

[a] SNP array genotyping with *BnMYB28* BC<sub>1</sub> double mutants of populations BC<sub>A</sub>-BC<sub>F</sub> (genotype *A<sub>1</sub>A<sub>p</sub>B<sub>1</sub>B<sub>p</sub>*).

[b] SNP array genotyping with *BnCYP79F1.C05* (genotype *C<sub>1</sub>C<sub>p</sub>D<sub>e</sub>D<sub>p</sub>*) and *BnCYP79F1.A06* (genotype *C<sub>e</sub>C<sub>p</sub>D<sub>1</sub>D<sub>p</sub>*) BC<sub>1</sub> single mutants of populations BC<sub>G</sub>-BC<sub>J</sub>.

[1] Wildtype progenitors Express617(E) and Peace(P) and the ‘BC<sub>EPP</sub> (ExpXP)’ plants were used as controls.

[2] Polymorphic across Express617 and Peace genotypes.

**Supplementary Table 12:** Summary statistics of the background Peace genome share for BC<sub>1</sub> double mutants (*BnMYB28*) and single mutants (*BnCYP79F1*) genotyped with the *Brassica* 19K SNP array.

Gene name <sup>[a]</sup>	BC <sub>1</sub> population <sup>[b]</sup>	Seed code of F <sub>1</sub> parents	BC <sub>1</sub> seed code	No. plants genotyped	Peace genome share (%)
<i>BnMYB28</i> ( <i>A<sub>1</sub>A<sub>p</sub>B<sub>1</sub>B<sub>p</sub></i> )	BC <sub>A</sub>	200465_14	206139	11	72.4 - 83.8
	BC <sub>B</sub>	200465_20	206140	10	70.6 - 83.3
	BC <sub>C</sub>	200465_32	206142	9	69.3 - 81.6
	BC <sub>D</sub>	200465_36	206143	3	70.1 - 75.7
	BC <sub>E</sub>	200465_41	206144	5	69.7 - 80.6
	BC <sub>F</sub>	200465_50	206145	35	70.2 - 85.5
<i>BnCYP79F1.C05</i> ( <i>C<sub>1</sub>C<sub>p</sub>D<sub>e/p</sub>D<sub>e/p</sub></i> )	BC <sub>G</sub>	191579_3	200503	22	63.0 - 76.3
	BC <sub>H</sub>	191579_4	200504	21	62.3 - 78.5
<i>BnCYP79F1.A06</i> ( <i>C<sub>e/p</sub>C<sub>e/p</sub>D<sub>1</sub>D<sub>p</sub></i> )	BC <sub>I</sub>	191581_3	200508	28	52.2 - 77.9
	BC <sub>J</sub>	191581_5	200509	15	63.0 - 73.6

[a] Selected genotypes of BC<sub>1</sub> individuals are mentioned in parenthesis (refer to Table 6 for the list of all allele codes).

[b] Independent BC<sub>1</sub> populations originating from 10 distinct F<sub>1</sub> parental backgrounds were genotyped; populations BC<sub>A</sub>-BC<sub>F</sub> originating from six independent *BnMYB28* F<sub>1</sub> double mutants and populations BC<sub>G</sub>-BC<sub>J</sub> representing single mutants of *BnCYP79F1* backcrossed with Peace.

**Supplementary Table 13:** Summary statistics of individual glucosinolates in seeds of *BnMYB28* EMS mutants and controls. Analyses were done using high-performance liquid chromatography.

Glucosinolate type <sup>[a]</sup>	Mutants backcrossed with Peace <sup>[1]</sup>								F <sub>2</sub> populations <sup>[2]</sup>					
	Population 210465				Controls				Population 200527				Controls	
	<i>BnMYB28</i> DM	SD	<i>BnMYB28</i> WT	SD	BC <sub>EPP</sub> (ExPxP)	SD	Peace (P)	SD	<i>BnMYB28</i> DM	SD	<i>BnMYB28</i> WT	SD	Express617 (E)	SD
Glucobrassicin	n.d		n.d		n.d		n.d		0.53	0.25	0.30	0.08	0.10	0.01
Progoitrin	5.80	3.09	6.23	2.74	10.14	4.822	7.37	1.12	16.20	6.52	41.22	13.05	36.32	5.09
Epiprogoitrin	0.16	0.02	0.16	0.03	0.15	0.012	n.d		0.57	0.16	1.37	0.41	1.18	0.27
Sinigrin + Glucoraphanin	0.11	0.07	0.07	0.03	0.57	0.115	0.17	0.06	1.12	0.86	0.27	0.21	0.64	0.08
Gluconapoleiferin	0.16	0.04	0.08	0.04	0.09	0.01	0.08	0.03	0.04	0.02	0.11	0.07	n.d	
Glucoallysin	0.67	0.12	0.79	0.30	1.62	0.717	1.14	0.17	1.03	0.41	2.82	0.72	2.94	0.05
Gluconapin	2.12	0.45	2.38	0.73	4.54	1.050	3.25	0.32	6.05	2.70	5.66	1.06	9.93	2.52
4-OH-glucobrassicin	2.94	0.25	3.28	0.22	2.26	0.43	2.53	0.19	2.56	0.61	2.31	0.16	2.23	0.32
Glucobrassicinapin	0.33	0.24	0.48	0.25	0.66	0.417	0.51	0.08	0.64	0.48	4.33	1.30	5.27	0.88
Glucotropaeolin	1.11	0.16	0.91	0.11	1.10	0.259	1.19	0.09	0.75	0.14	0.76	0.06	0.50	0.08
Glucobrassicin	0.58	0.15	0.55	0.14	0.57	0.381	0.51	0.12	0.26	0.12	0.13	0.03	0.20	0.36
Gluconasturtiin	1.20	0.31	1.04	0.27	1.88	0.498	1.60	0.13	1.08	0.24	1.07	0.13	0.73	0.16
4-methoxyglucobrassicin	0.03	0.02	0.03	0.01	0.06	0.023	0.05	0.01	0.05	0.02	0.07	0.01	0.04	0.02
Neoglucobrassicin	0.09	0.06	0.08	0.02	0.07	0.03	0.06	0.01	0.08	0.06	0.16	0.07	0.04	0.01
<b>Sum (μmol/g DW)</b>	<b>15.29</b>		<b>16.07</b>		<b>23.70</b>		<b>18.45</b>		<b>30.96</b>		<b>60.58</b>		<b>60.12</b>	

[1] Experiments with plants originating after backcrosses with the spring-type rapeseed Peace.

[2] Experiments with plants originating from direct M<sub>3</sub>xM<sub>3</sub> crosses without backcrossing.

[a] Estimated based on comparison of retention time and co-chromatography with commercial standards. GSL concentrations were calculated as μmol/g dry weight of tissue analyzed (five biological replicates per genotype).

*BnMYB28*\_DM: *BnMYB28* double mutant (genotype *A<sub>1</sub>A<sub>1</sub>B<sub>1</sub>B<sub>1</sub>*), *BnMYB28*\_WT: Plants with wildtype alleles for *BnMYB28* segregating within the same F<sub>2</sub> population, Express617 and Peace: non-mutagenized parents, BC<sub>EPP</sub> (ExPxP): BC<sub>1</sub> progeny of non-mutagenized Express617 and Peace.

n.d: Not detectable SD: Standard deviation

# Appendix

**Supplementary Table 14:** Summary statistics of individual glucosinolates in seeds of *BnCYP79F1* EMS mutants and controls. Analyses were done using high-performance liquid chromatography.

	Mutants backcrossed with Peace <sup>[1]</sup>								F <sub>2</sub> populations <sup>[2]</sup>					
	Population 210462				Controls				Population 200529				Controls	
Glucosinolate type <sup>[a]</sup>	<i>BnCYP79F1</i> DM	SD	<i>BnCYP79F1</i> WT	SD	BC <sub>EPP</sub> (ExPxP)	SD	Peace	SD	<i>BnCYP79F1</i> DM	SD	<i>BnCYP79F1</i> WT	SD	Express617	SD
Glucobrassicin	0.01	0.01	0.01	0.02	n.d		n.d		0.01	0.01	0.04	0.02	0.10	0.01
Progoitrin	14.26	0.92	11.52	6.93	10.14	4.820	7.37	1.12	24.55	1.09	37.53	8.41	36.32	5.09
Epiprogoitrin	0.16	0.01	0.14	0.07	0.14	0.01	n.d		0.78	0.07	1.12	0.17	1.18	0.27
Sinigrin + Glucoraphanin	0.35	0.05	0.70	0.25	0.57	0.12	0.17	0.06	0.45	0.07	0.42	0.03	0.64	0.08
Glucotropaeolin	0.06	0.04	0.19	0.03	0.09	0.01	0.08	0.03	0.08	0.02	0.05	0.02	n.d	
Glucobrassicin	2.07	0.12	2.17	0.80	1.62	0.72	1.14	0.17	3.28	0.15	2.47	0.55	2.94	0.05
Glucobrassicin	3.75	0.15	5.72	3.47	4.54	1.05	3.25	0.32	6.91	0.20	8.91	0.73	9.93	2.52
4-OH-glucobrassicin	2.36	0.13	2.56	0.35	2.26	0.43	2.53	0.19	4.40	0.19	3.81	0.68	2.23	0.32
Glucobrassicin	0.83	0.12	0.71	0.55	0.66	0.42	0.51	0.08	4.02	0.15	3.33	0.68	5.27	0.88
Glucotropaeolin	0.92	0.12	1.06	0.14	1.10	0.26	1.19	0.08	1.02	0.04	0.90	0.13	0.50	0.08
Glucobrassicin	0.37	0.03	0.64	0.36	0.57	0.38	0.51	0.12	0.31	0.05	0.19	0.06	0.20	0.36
Gluconasturtiin	1.41	0.16	1.60	0.24	1.88	0.5	1.60	0.13	1.62	0.07	1.56	0.13	0.73	0.16
4-methoxyglucobrassicin	0.05	0.02	0.08	0.02	0.06	0.02	0.04	0.01	0.09	0.00	0.08	0.01	0.04	0.02
Neoglucobrassicin	0.14	0.02	0.31	0.25	0.07	0.03	0.06	0.01	0.12	0.02	0.09	0.04	0.04	0.01
<b>Sum (μmol/g DW)</b>	<b>26.74</b>		<b>27.41</b>		<b>23.70</b>		<b>18.45</b>		<b>47.64</b>		<b>60.50</b>		<b>60.12</b>	

[1] Experiments with plants originating after backcrosses with the spring-type rapeseed Peace.

[2] Experiments with plants originating from direct M<sub>3</sub>xM<sub>3</sub> crosses without backcrossing.

[a] Estimated based on comparison of retention time and co-chromatography with commercial standards. GSL concentrations were calculated as μmol/g dry weight of tissue analyzed (five biological replicates per genotype).

*BnCYP79F1*\_DM: *BnCYP79F1* double mutant (genotype *C<sub>1</sub>C<sub>1</sub>D<sub>1</sub>D<sub>1</sub>*), *BnCYP79F1*\_WT: Plants with wildtype alleles for *BnCYP79F1* segregating within the same F<sub>2</sub> population, Express617 and Peace: non-mutagenized parents, BC<sub>EPP</sub> (ExPxP): BC<sub>1</sub> progeny of non-mutagenized Express617 and Peace.

n.d: Not detectable SD: Standard deviation

**Supplementary Table 15:** Summary of mutation effects conferred by EMS-induced mutations in *BnGTR2* paralogs.

Gene name	cDNA coverage	Number of detected mutations					Mutation frequency (1/kb) <sup>[2]</sup>
		Number of M <sub>2</sub> pools screened <sup>[1]</sup>	Nonsense mutations	Missense mutations	Silent mutations	Splice site mutations	
<i>BnGTR2.C03</i>	49%	4	1	7	6	0	1/25.7
<i>BnGTR2.A06</i>	50%	6	1	4	8	0	1/44.5
Total = 27							Average = 1/35.1 kb

[1] The number of eight-fold (8x) two-dimensional (2-D) M<sub>2</sub> pools used for screenings.

[2] Calculated as the number of mutations per M<sub>1</sub> plant based on the number of analyzed M<sub>2</sub> families (8x 2D pools) and amplicon lengths. Formula given by Harloff et al. (2012).

**Supplementary Table 16:** Summary of seed material generated for future studies using EMS-induced functional mutations in *BnGTR2*. Genotypes of parents used for hand crosses and the resulting offspring are shown.

Crossing type	Parental genotypes	Genotype of selected F <sub>1</sub> progeny	F <sub>2</sub> seed code <sup>[a]</sup>
M <sub>3</sub> x M <sub>3</sub>	<i>E<sub>e</sub>E<sub>e</sub>F<sub>1</sub>F<sub>1</sub></i> x <i>E<sub>1</sub>E<sub>1</sub>F<sub>e</sub>F<sub>e</sub></i>	<i>E<sub>1</sub>E<sub>e</sub>F<sub>1</sub>F<sub>e</sub></i>	205872 <sup>[1]</sup> 205873 <sup>[1]</sup>
[(M <sub>3</sub> x M <sub>3</sub> ) x Express617] x Express617	[( <i>E<sub>1</sub>E<sub>e</sub>F<sub>1</sub>F<sub>e</sub></i> x <i>E<sub>e</sub>E<sub>e</sub>F<sub>e</sub>F<sub>e</sub></i> ) x <i>E<sub>e</sub>E<sub>e</sub>F<sub>e</sub>F<sub>e</sub></i> ]	<i>E<sub>1</sub>E<sub>e</sub>F<sub>1</sub>F<sub>e</sub></i>	210631 <sup>[2]</sup>
[(M <sub>3</sub> x M <sub>3</sub> ) x Peace] x Peace	[( <i>E<sub>1</sub>E<sub>e</sub>F<sub>1</sub>F<sub>e</sub></i> x <i>E<sub>p</sub>E<sub>p</sub>F<sub>p</sub>F<sub>p</sub></i> ) x <i>E<sub>p</sub>E<sub>p</sub>F<sub>p</sub>F<sub>p</sub></i> ]	<i>E<sub>1</sub>E<sub>p</sub>F<sub>1</sub>F<sub>p</sub></i>	210710 <sup>[2]</sup>

[1] Segregating F<sub>2</sub> population originating from heterozygous *BnGTR2* double mutants (originating from direct M<sub>3</sub>xM<sub>3</sub> crosses).

[2] Segregating BC<sub>1</sub>F<sub>2</sub> populations originating from heterozygous *BnGTR2* double mutants (originating from backcrosses with Express617 and Peace).

[a] The F<sub>2</sub> populations will yield segregating progenies for quantitative and qualitative glucosinolate determination.

**Supplementary Table 17:** Target oligonucleotide sequences selected for the CRISPR-Cas9 mediated knock-out of seven *BnGTR2* paralogs. Targets were designed based on sequence conservation within the three sub-families and across the entire gene family.

Gene name	<i>B. napus</i> paralog targeted <sup>[1]</sup>	Chromosomal location	Construct name <sup>[2]</sup>	Sequences 5' → 3' <sup>[3]</sup>
<i>BnGTR2.A06</i>	<i>BnaA06g22160D</i>	A06	SJ_GTR2_1	F ATTGggaacagctctctcatgtat
<i>BnGTR2.C03</i>	<i>BnaC03g51560D</i>	C03		R AAACatacatgagagagctgttcc
<i>BnGTR2.A09</i>	<i>BnaA09g06190D</i>	A09	SJ_GTR2_2	F ATTGctctggaggtttcatgatcc
<i>BnGTR2.C09</i>	<i>BnaC09g05810D</i>	C09		R AAACggatcatgaaacctccagag
<i>BnGTR2.C02.a</i>	<i>BnaC02g42280D</i>	C02	SJ_GTR2_3	F ATTGggaacagccctcacgtgtac
<i>BnGTR2.C02.b</i>	<i>BnaC02g42260D</i>	C02		R AAACgtacacgtgagggtgttcc
<i>BnGTR2.A02</i>	<i>BnaA02g33530D</i>	A02		F ATTGttgtgagagtgttcctata
All seven paralogs			SJ_GTR2_KO	R AAACtataggaagcactctcaca

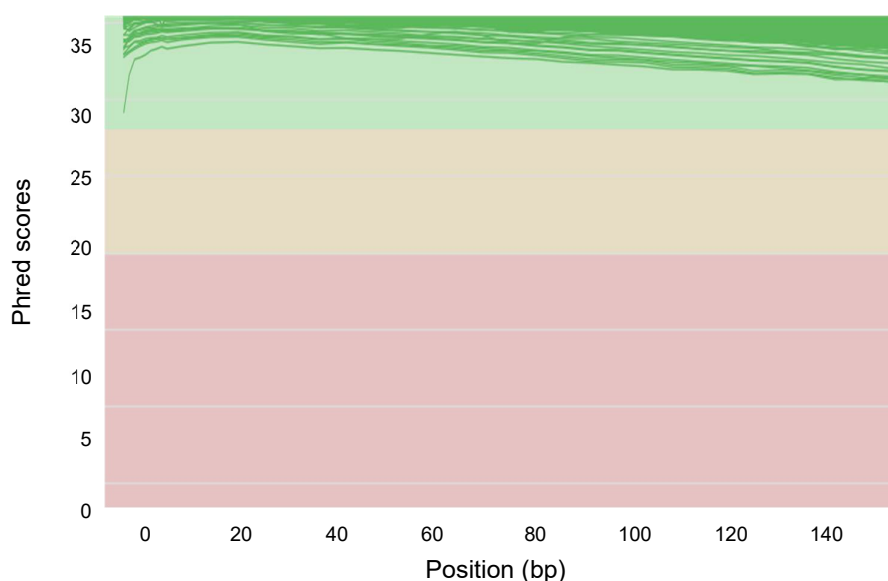
[1] As per the Darmor-*bzh* reference genome (Genoscope).

[2] Name of the final Cas9 vector harboring the sgRNA oligonucleotide complex and the Cas9 endonuclease.

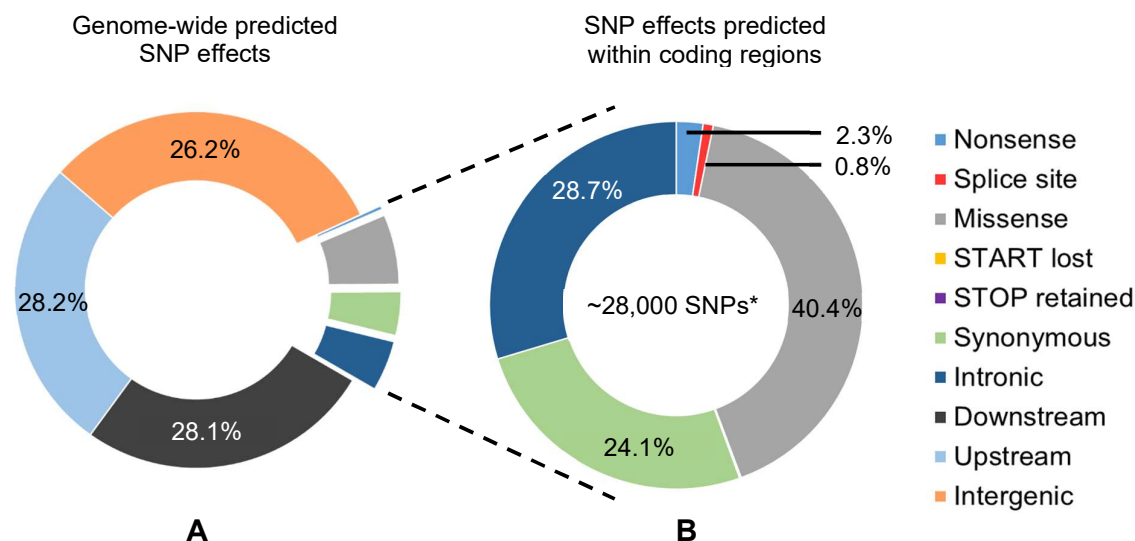
[3] Nucleotides in uppercase represent adapter sequences added to the protospacer for ligation into the pChimera vector.

F=Forward sequence, R=Reverse sequence.

## 8.2 Supplementary figures

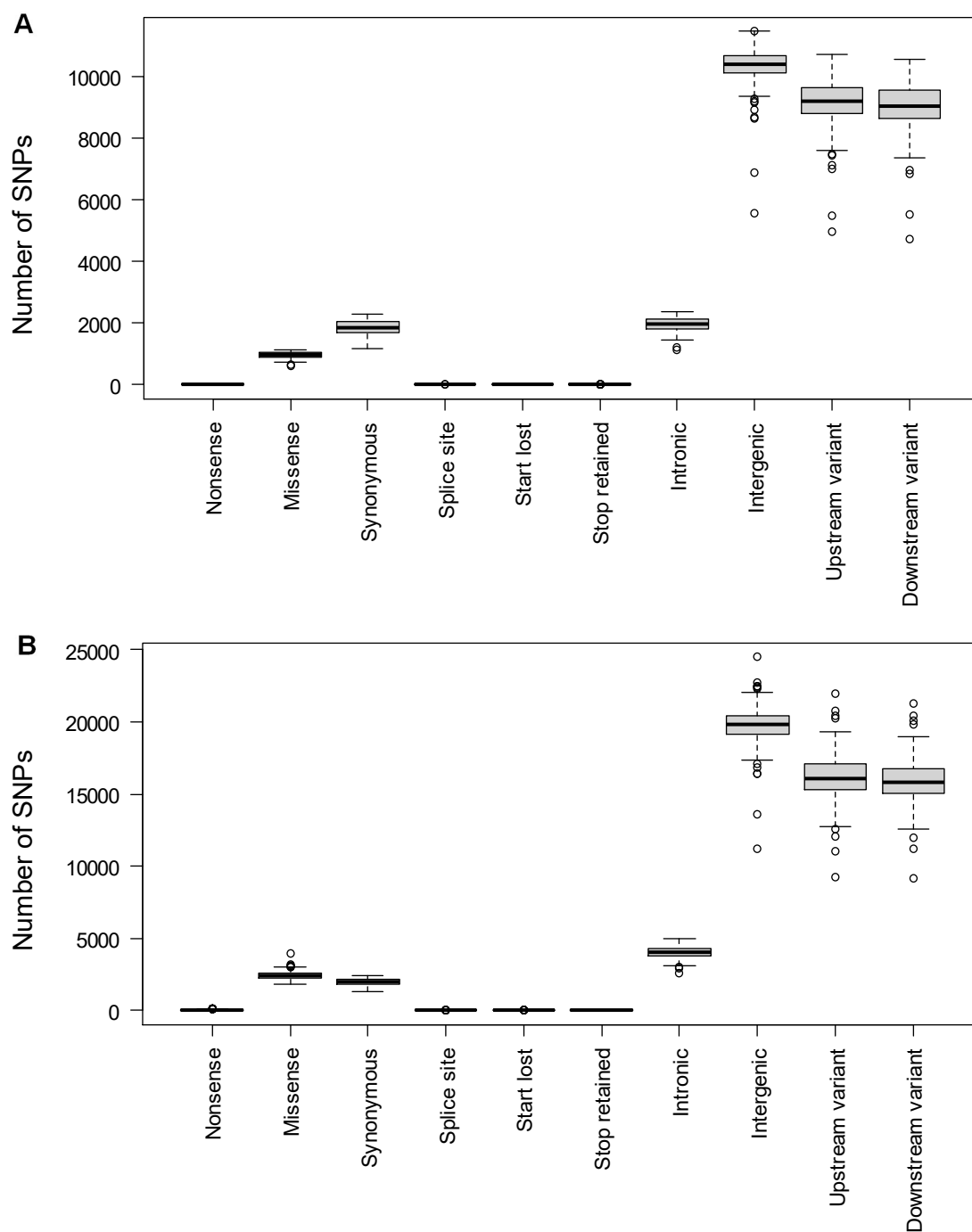


**Supplementary Figure 1:** Screenshot of the FastQC reports showing a graphical overview of quality checks for raw reads from all sequenced 4x pools. Mean quality scores represent Phred scores for individual bases called across the length 150 bp paired-end reads. Phred scores for all 4x pools are within the optimum levels of >30 (zone marked in green). Raw reads with Phred score = 20-26 are tolerated (yellow zones), Phred score  $\leq 20$  (red zones) are rejected. Analysis was done via FastQC v0.11.5 and MultiQC.

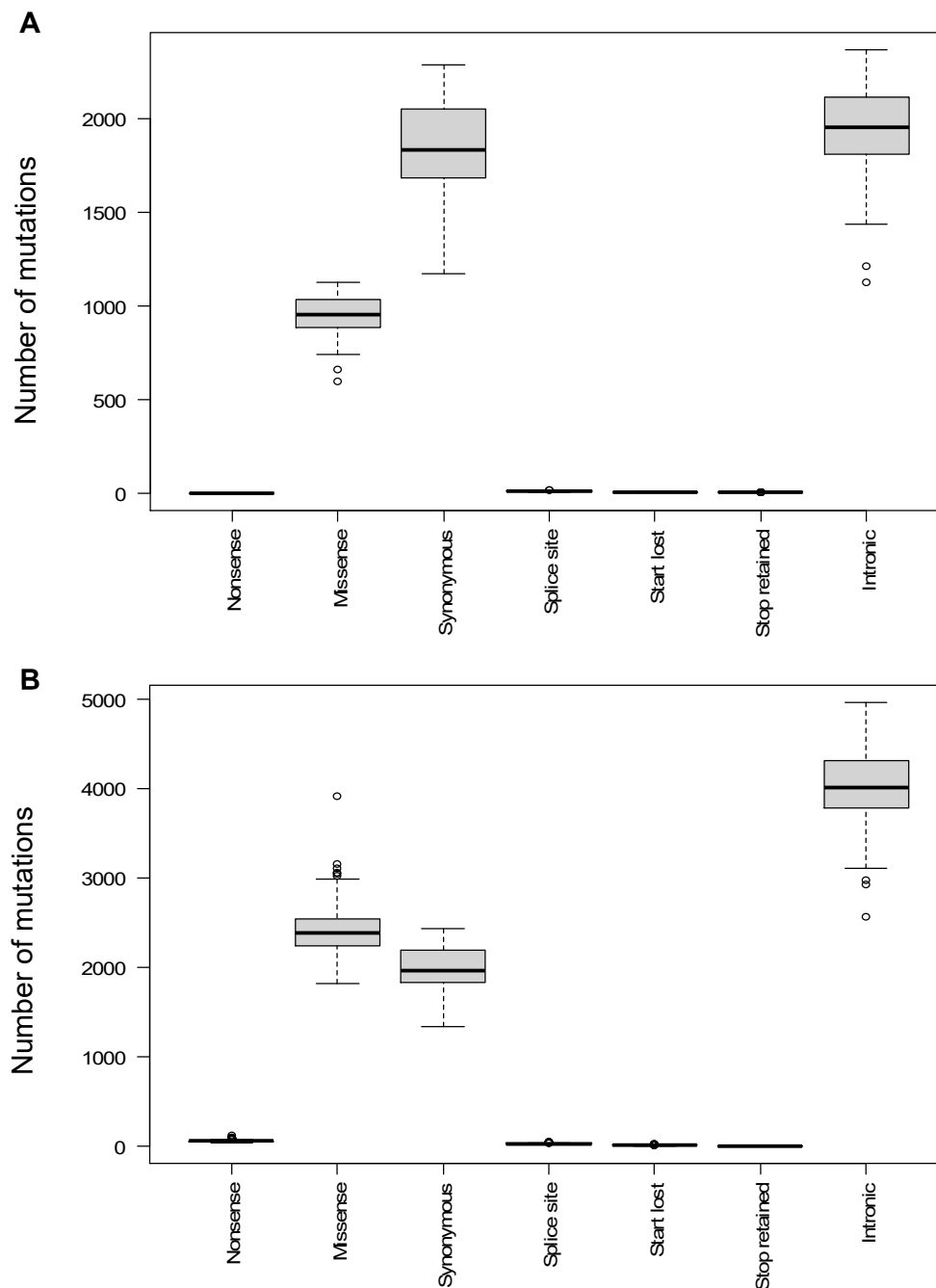


**Supplementary Figure 2:** Share of EMS-induced mutations predicted to have putative functional effects A) on a genome-wide scale and B) specifically within coding regions. SNP effects were first predicted on a genome-wide scale for all predicted gene models on the Express617 reference genome. The share of mutation effects within coding regions was calculated using SNPs that were present within and including the translation START and STOP sites. Splice site variants include the sum of splice acceptor and donor site mutations. Upstream and downstream variants encompass 5 kb genomic regions from the START and STOP sites, respectively. The Ensembl Variant Effect Predictor tool was used for SNP effect prediction using the Express617 reference genome (Lee et al., 2020). \*Reflects the mean SNP counts from 497 IRFFA 4x pools.

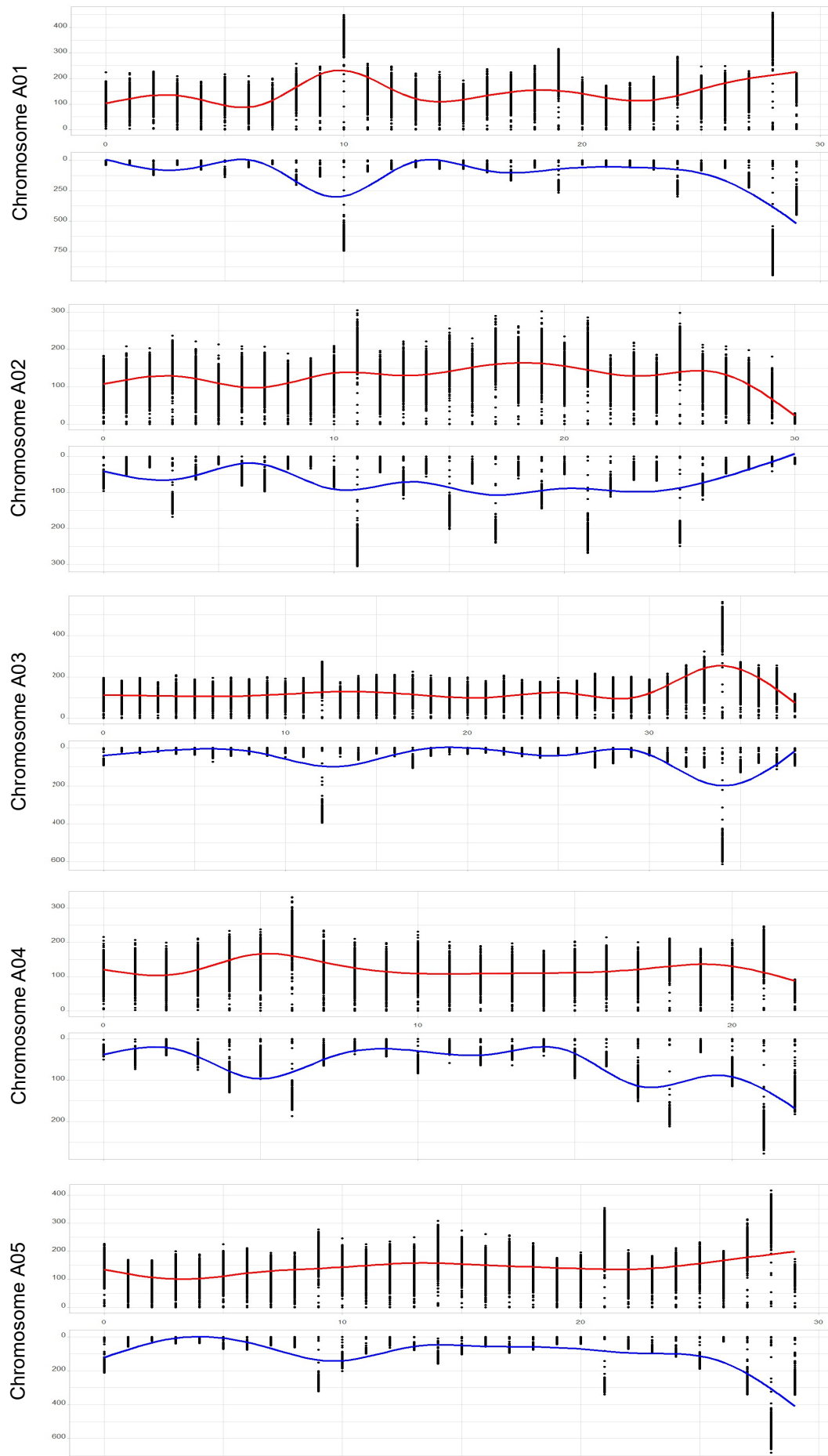


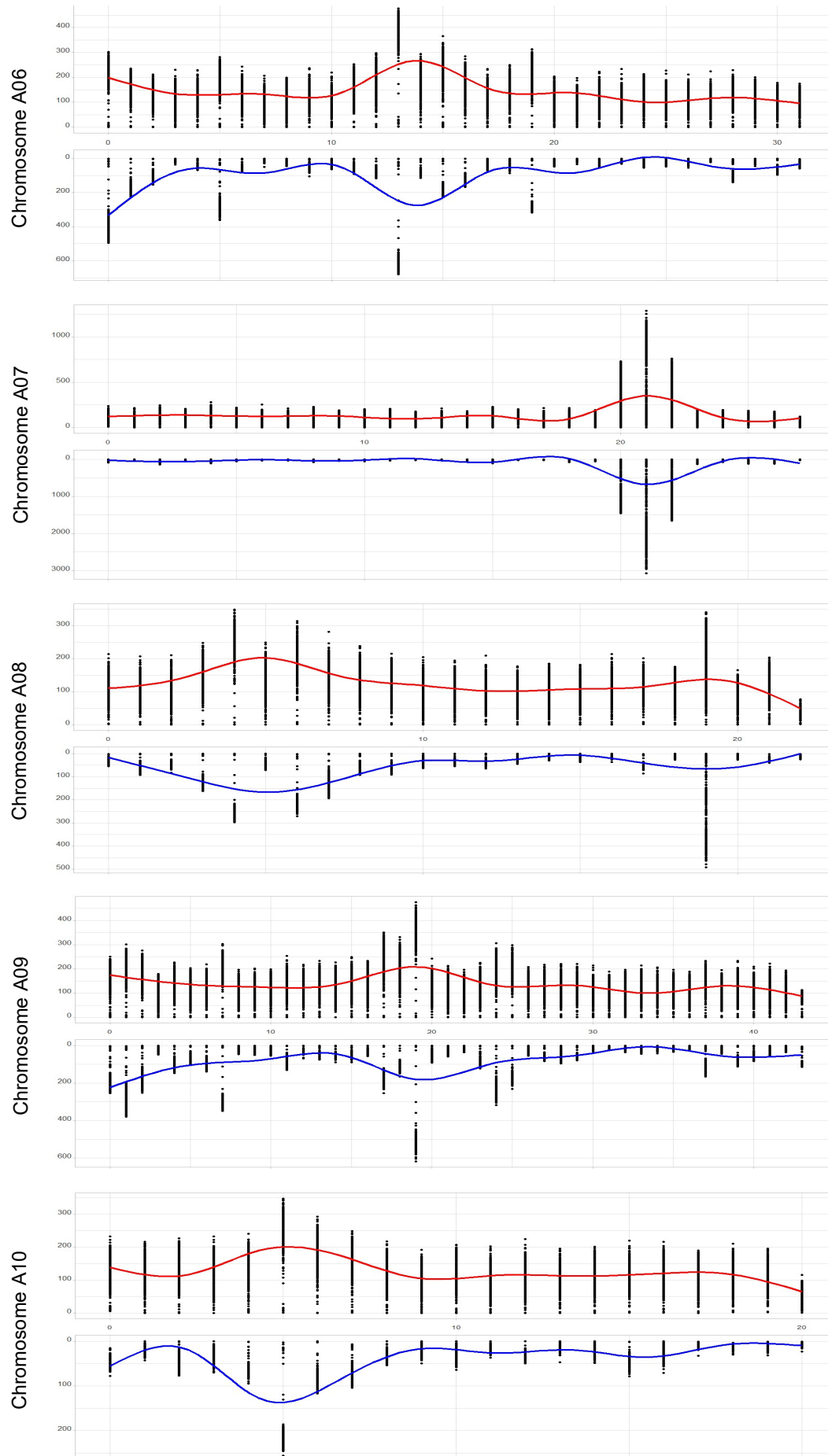


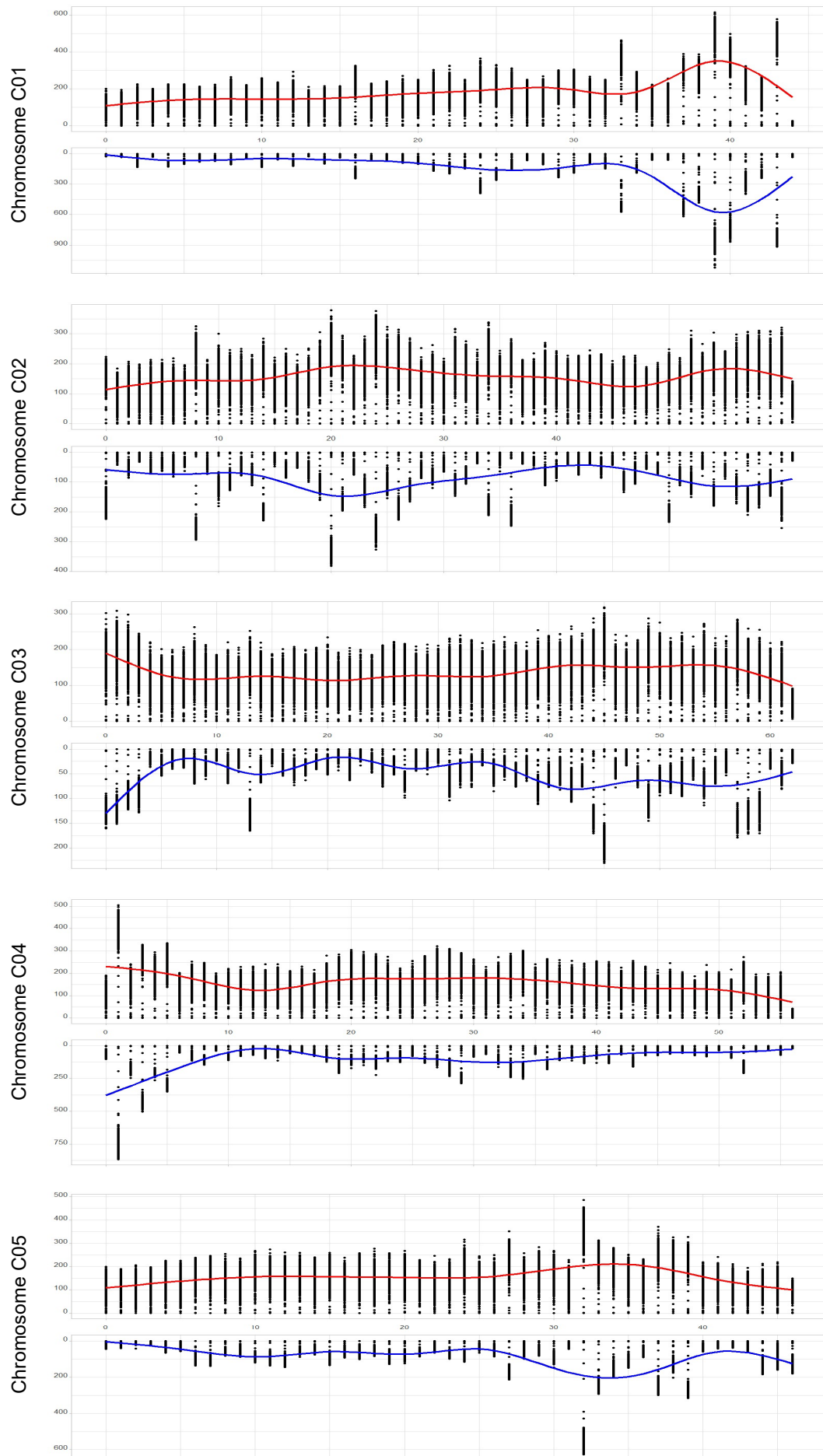
**Supplementary Figure 3:** Characterization of ‘non-canonical’ type A) Transitions (T→C and A→G) and B) Transversions (C→A, G→T, A→C, T→G, T→A, A→T, G→C and C→G) with predicted effects on a genome-wide scale. Boxplots show the distribution of SNP effects as a mean of 497 4x pools. Mutation effects are predicted using filtered SNPs with DP≥10, AD=12.5%-60% and MQ≥30. The Ensembl Variant Effect Predictor tool was used in offline mode. Mutation effects were predicted within the gene models of the Express617 genome. All gene IDs were extracted from the general feature format (GFF) of the Express617 reference genome. Splice site variants include acceptor and donor site mutations. Upstream and downstream variants are located 5 kb from the translation START and STOP sites, respectively. DP: Read depth, AD: Allele depth and MQ: Mapping quality.

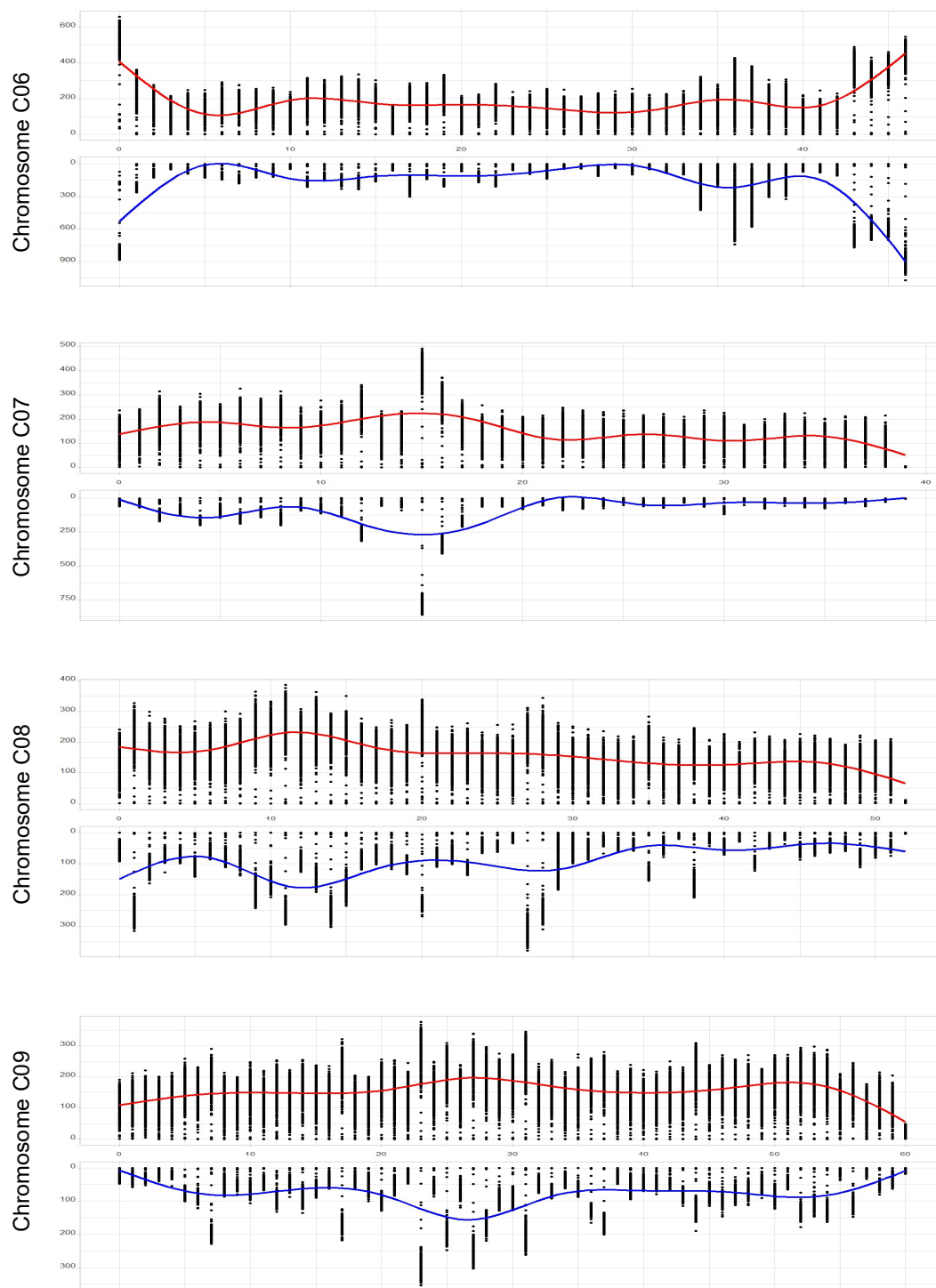


**Supplementary Figure 4:** Characterization of ‘non-canonical’ A) Transitions (T→C and A→G) and B) Transversions (C→A, G→T, A→C, T→G, T→A, A→T, G→C and C→G) with predicted effects exclusively within coding regions. Boxplots show the distribution of SNP effects as a mean of 497 4x pools. Mutation effects are predicted using filtered SNPs with DP ≥10, AD =12.5%-60% and MQ ≥30. The Ensembl Variant Effect Predictor release 99 was used in offline mode. Mutation effects were predicted within the gene models of the Express617 genome. All gene IDs were extracted from the general feature format (GFF) of the Express617 reference genome. Splice site variants include acceptor and donor site mutations. Upstream and downstream variants are located within 5 kb from the translation START and STOP sites, respectively. DP: Read depth, AD: Allele depth and MQ: Mapping quality.



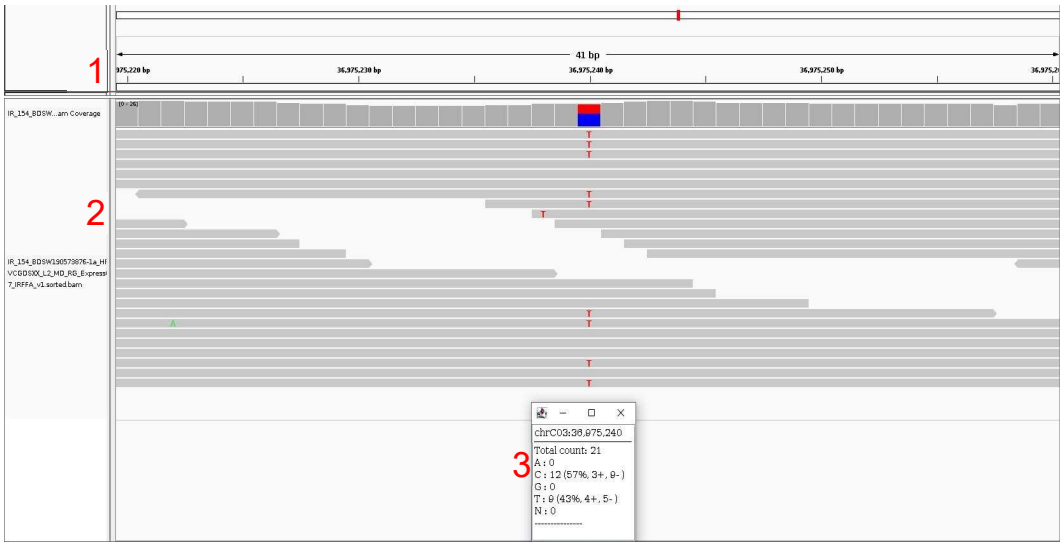






**Supplementary Figure 5:** Distribution and density of filtered mutations per 1 Mb windows across all sequenced pools for chromosomes A01-A10 and C01-C09. The red and blue lines represent the running average of the frequency distribution of filtered canonical (C→T and G→A transitions) and non-canonical type mutations (all other nucleotide substitutions), respectively across all pools. Black dots reflect the corresponding mutation densities per 1 Mb non-overlapping windows for each of the sequenced pools. The x-axis represents the length of chromosomes in Mb and each interval corresponds to a 1 Mb window. The y axis shows the number of SNPs located within each window. To smoothen the curves, the General Additive Model (GAM) in the ggplot2 package of R was used.

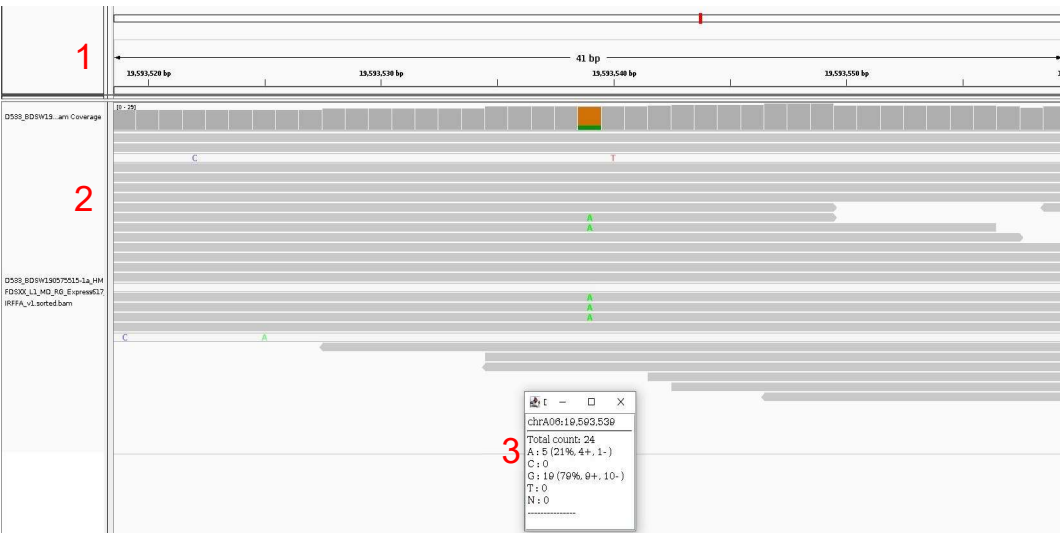
IR\_154



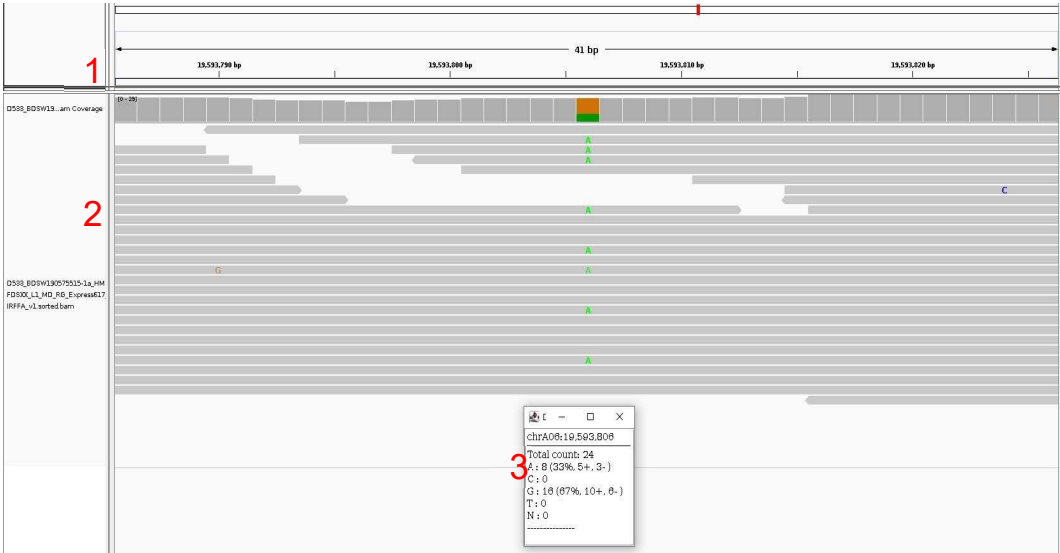
IR\_229



D533\_1



D533\_2



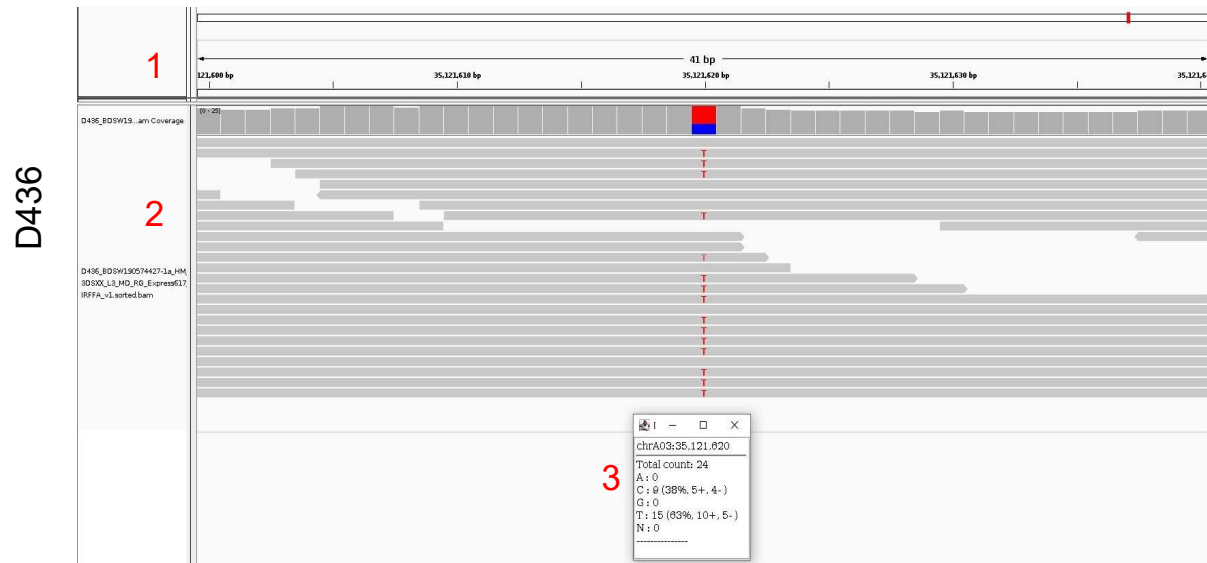
IR\_388



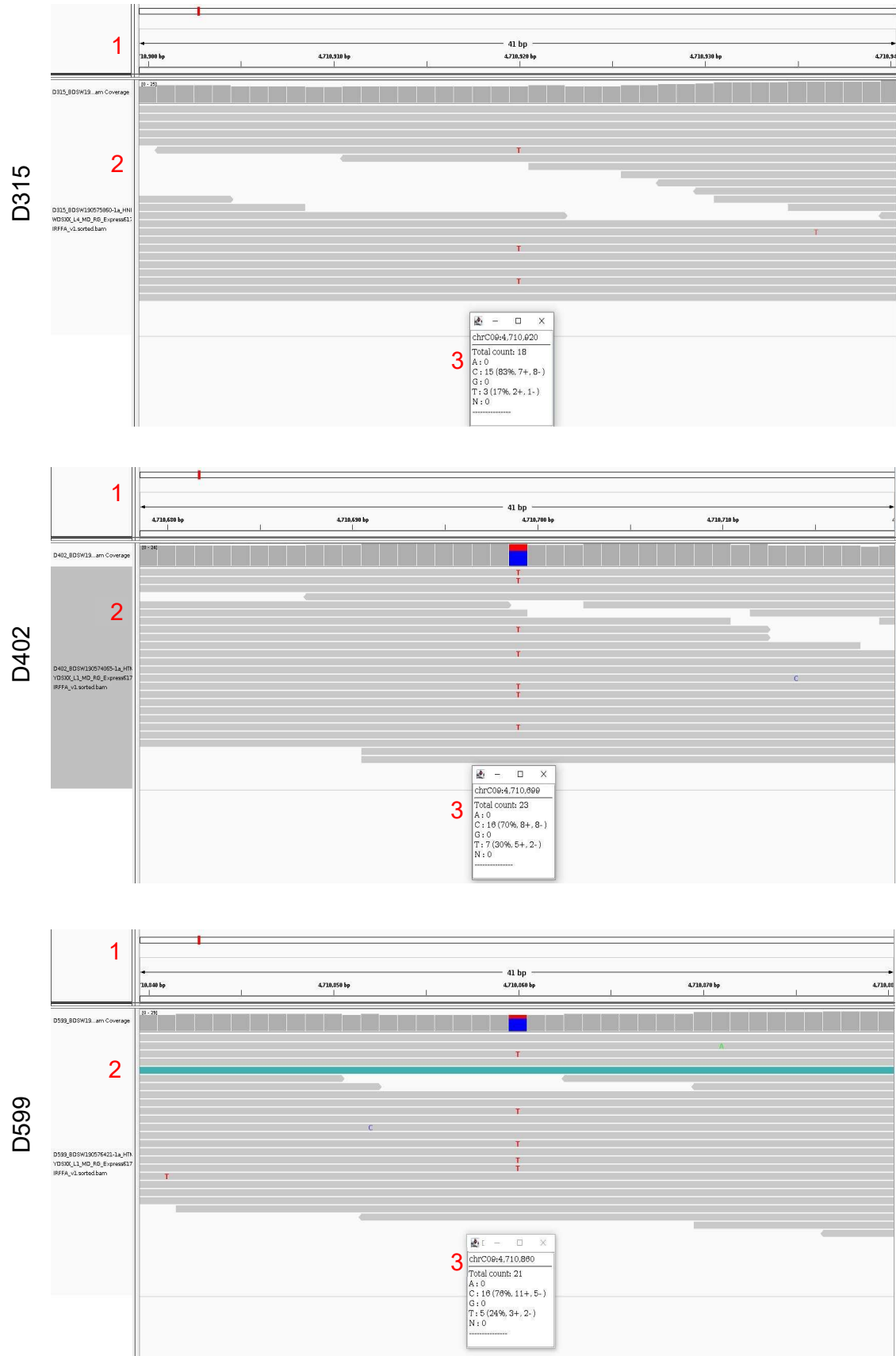
D631

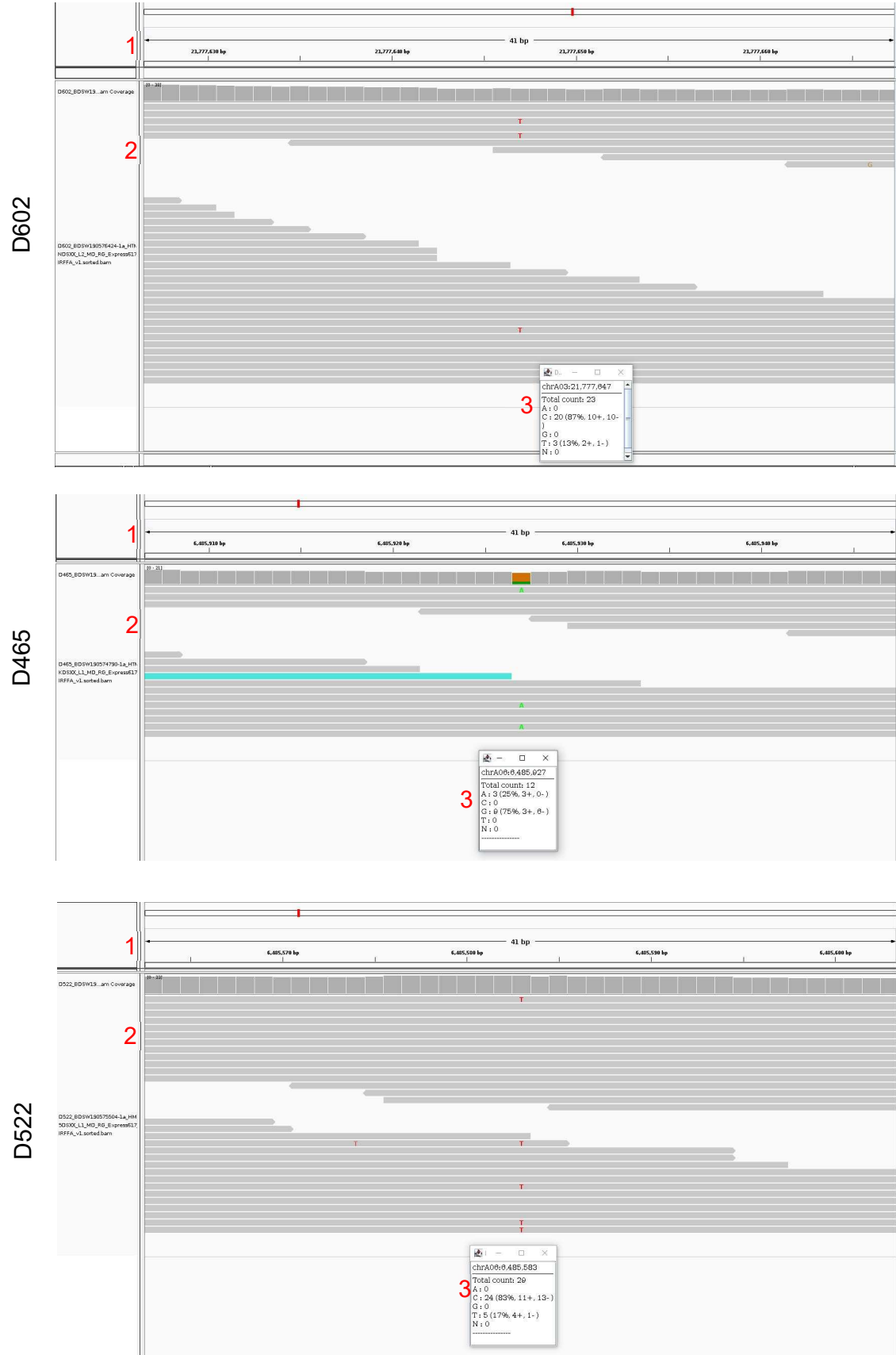


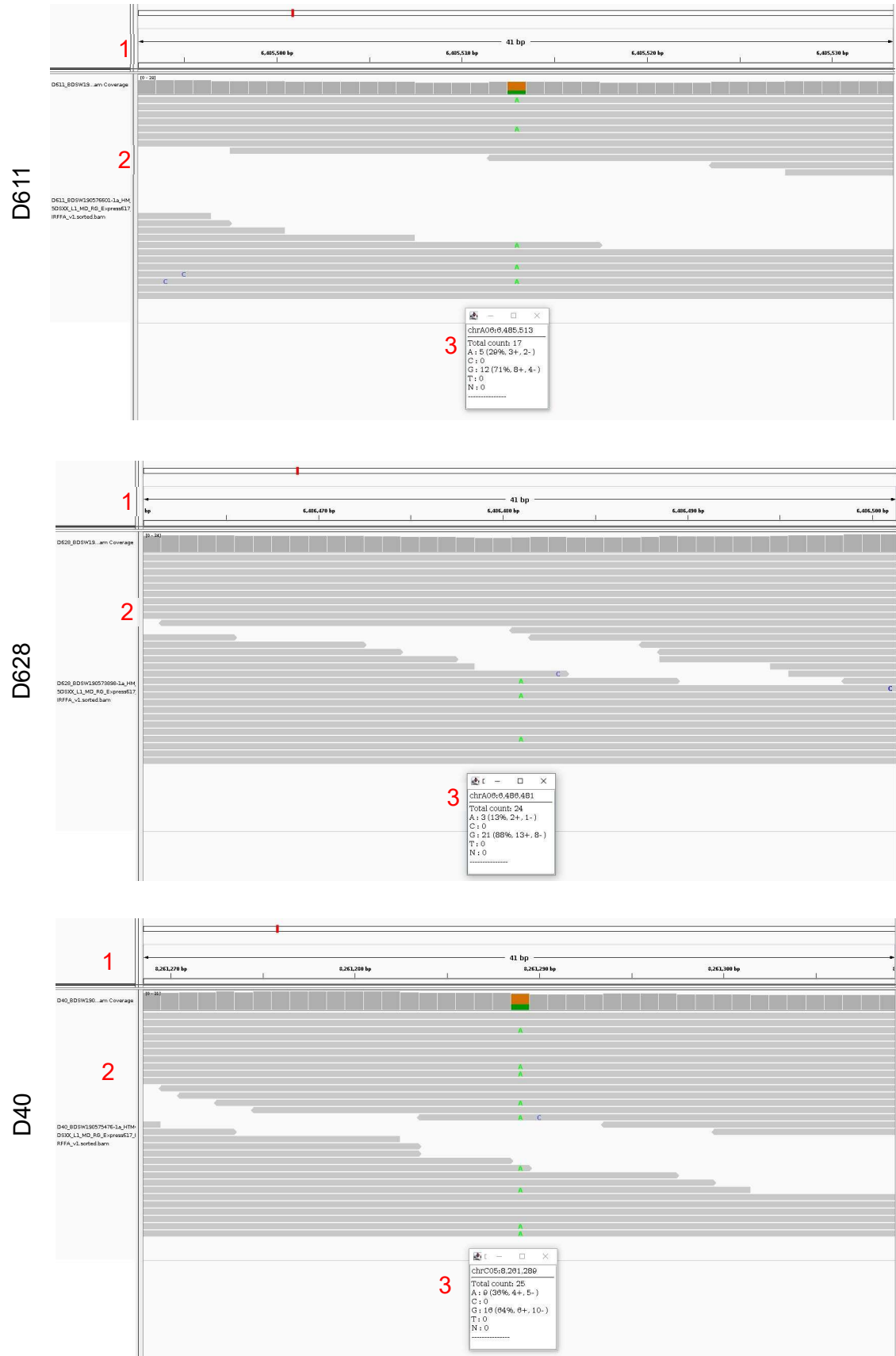


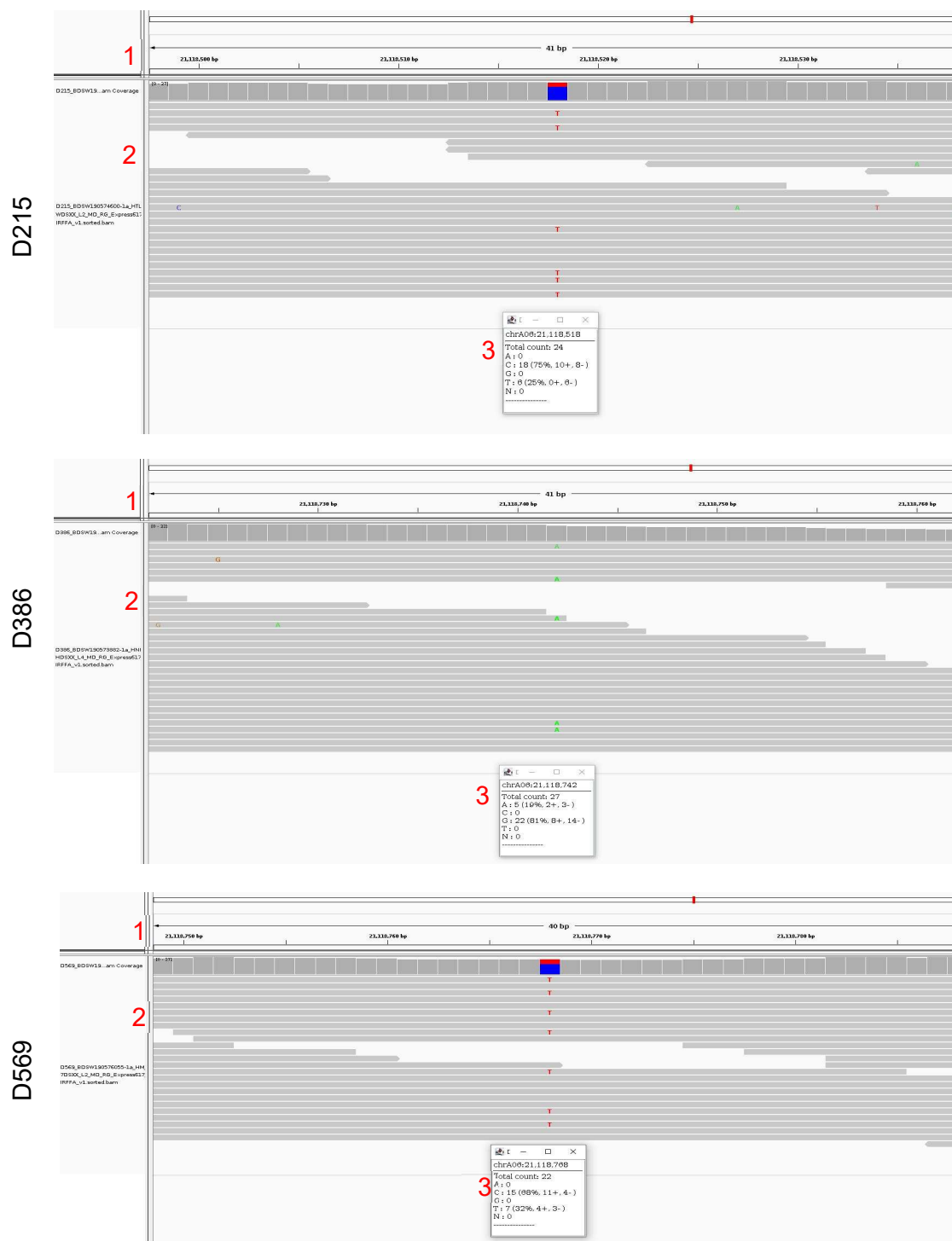


**Supplementary Figure 6:** Screenshot of the Integrative Genomics Viewer (IGV) showing read alignments from the seven selected 4x pools harboring previously detected EMS mutations. Pool IDs are named with the prefix ‘IR’ or ‘D’. (1) Tracks showing 41 bp regions with 20 bases up- and downstream of the detected mutations. (2) Coverage tracks showing read alignments with horizontal bars representing individual reads. SNPs are marked in different colors. (3) Inset windows representing the count details for each of the called nucleotides (A: Adenine, C: Cytosine, G: Guanine, T: Thymine or N: Unknown) and their share from all mapped reads at that position.

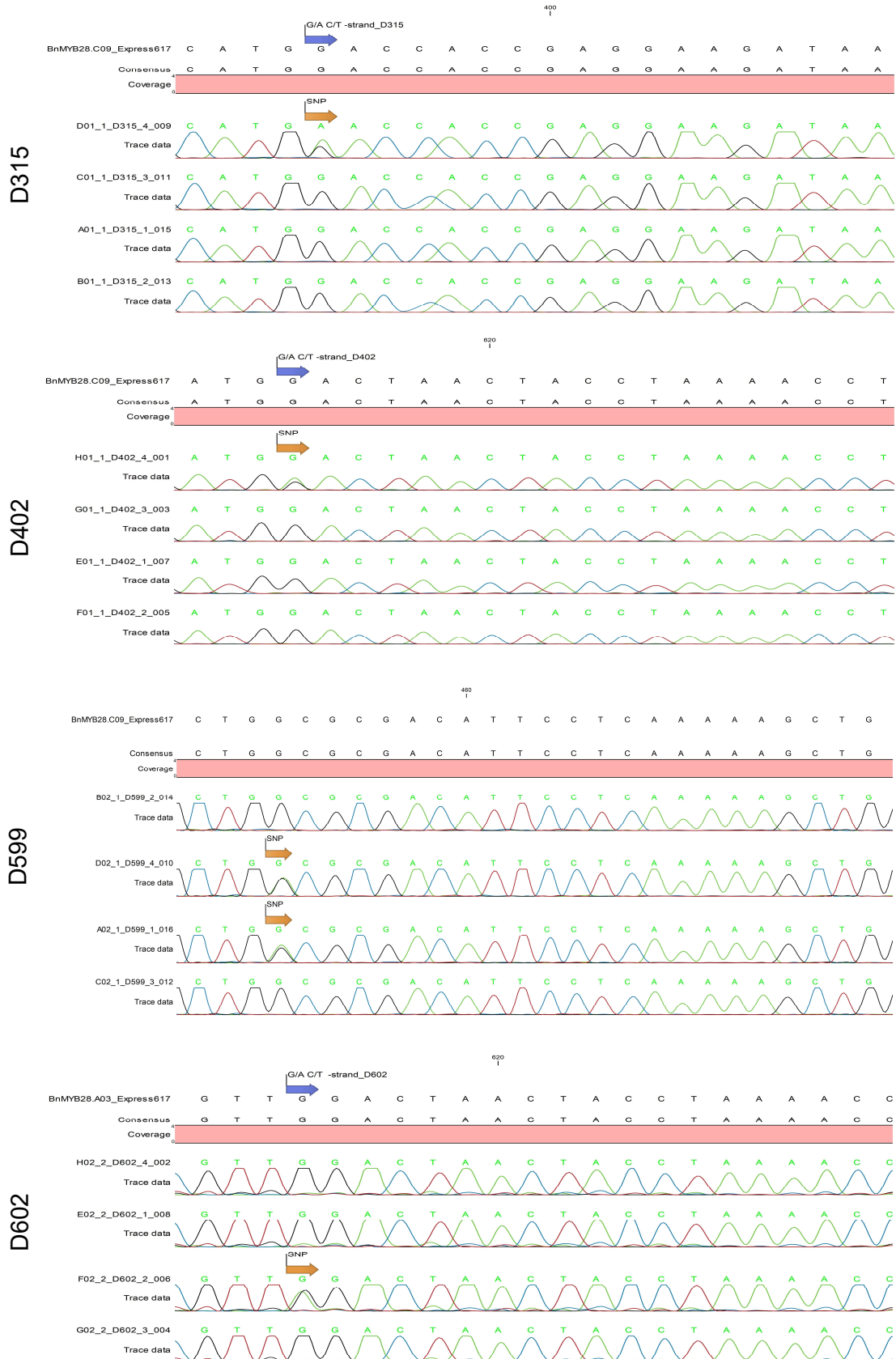


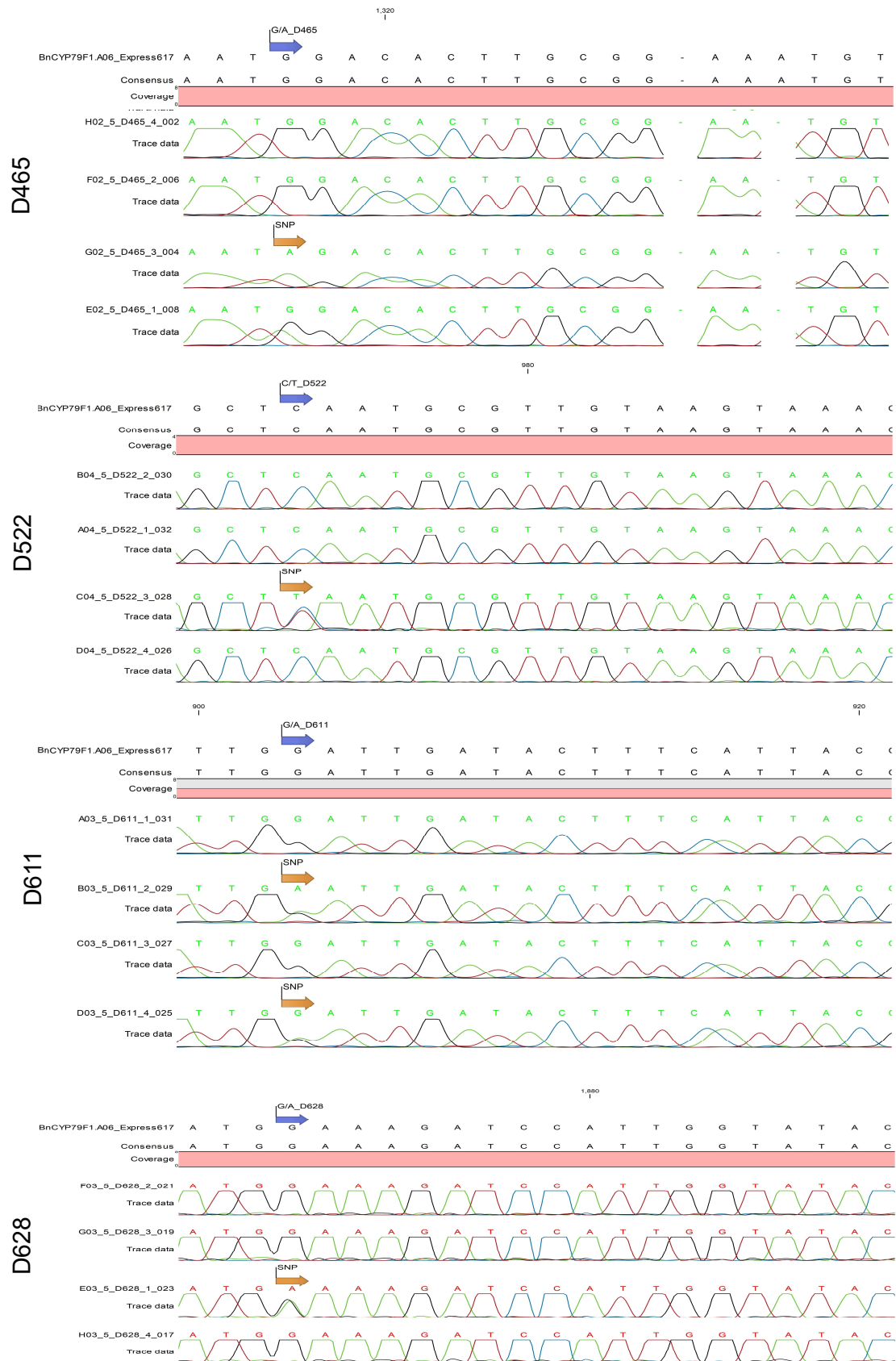


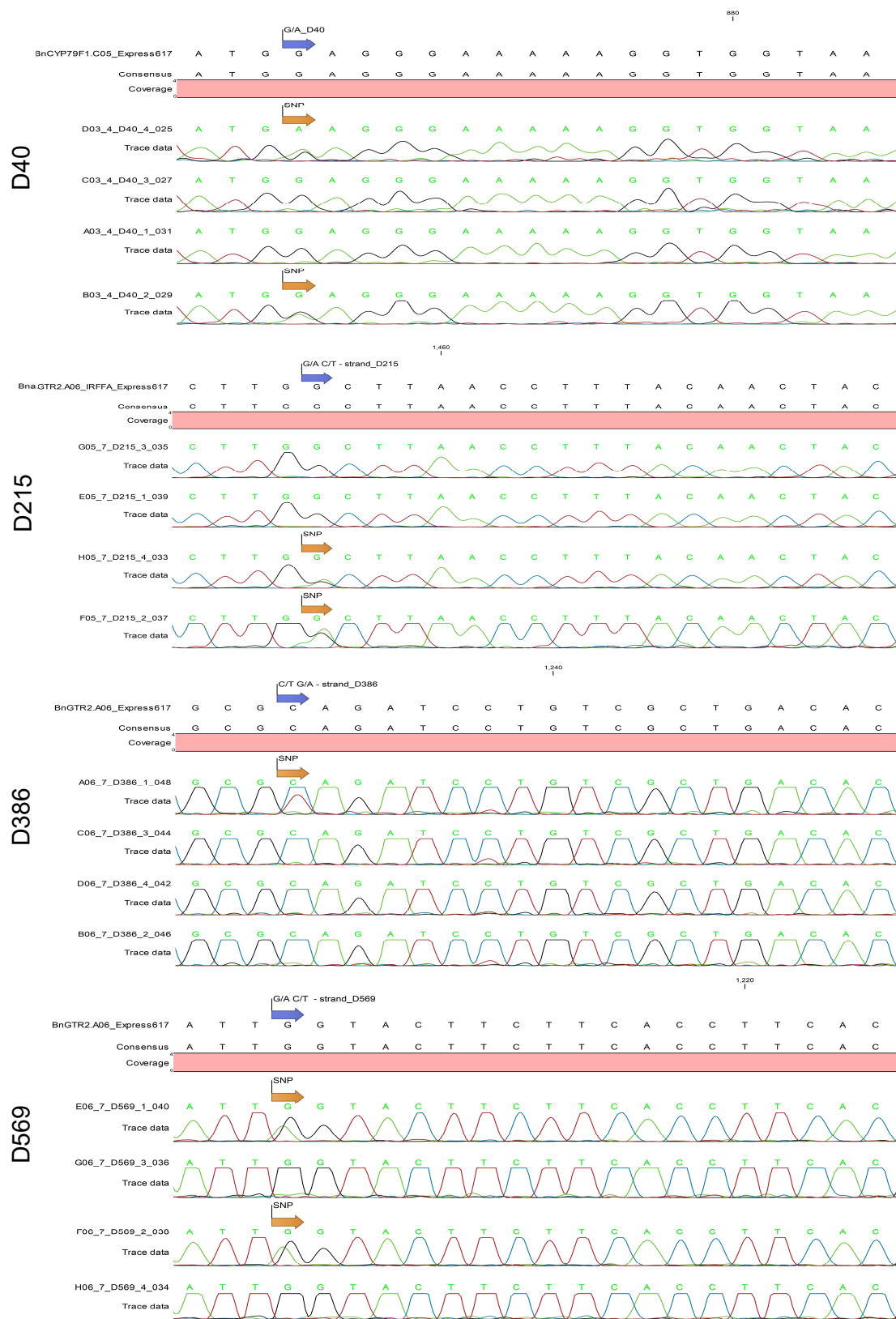




**Supplementary Figure 7:** Screenshot of the Integrative Genomics Viewer (IGV) showing read alignments from 12 fourfold pools harboring EMS mutations detected within candidate genes – *BnGTR2*, *BnMYB28* and *BnCYP79F1*. Pool IDs are named with the prefix ‘D’. (1) Tracks showing 41 bp regions with 20 bases up and downstream of the detected mutations. (2) Coverage tracks showing read alignments with horizontal bars representing individual reads. SNPs are marked in different colors. (3) Inset windows representing the count details for each of the called nucleotides (A: Adenine, C: Cytosine, G: Guanine, T: Thymine or N: Unknown) and their individual share from all mapped reads at that position.



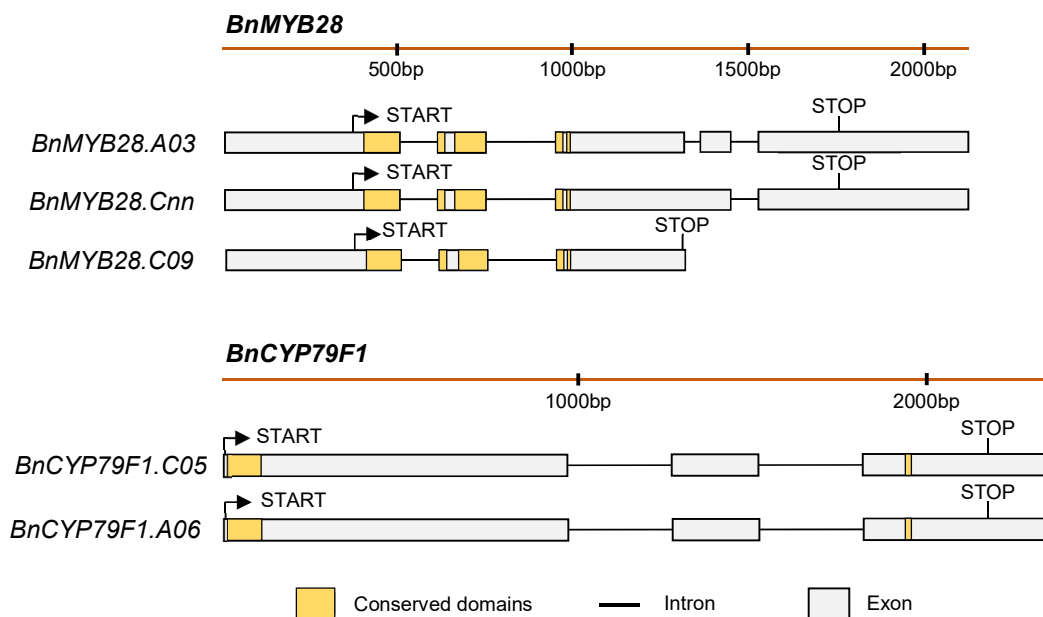




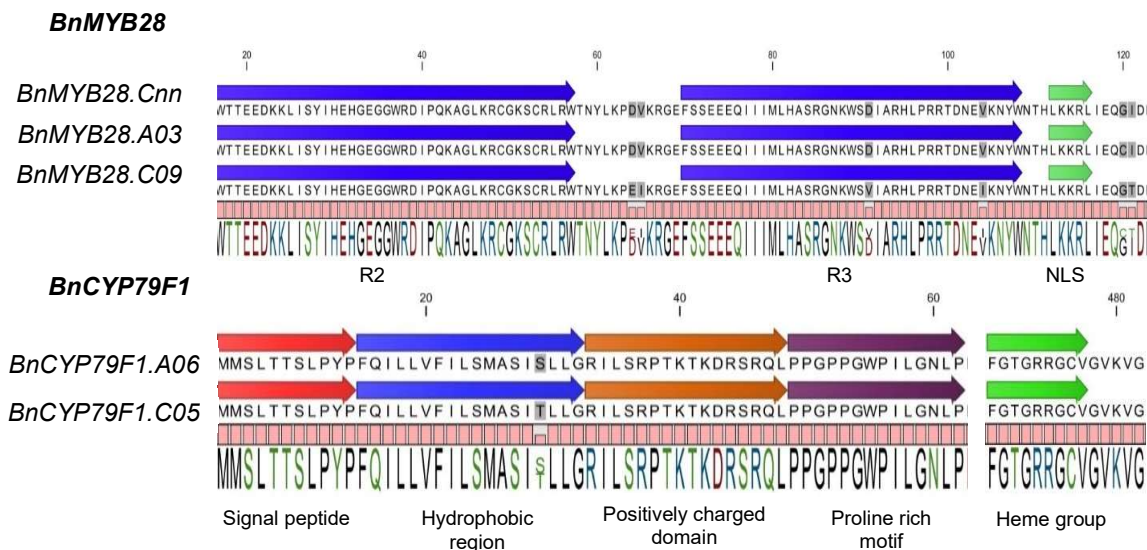
**Supplementary Figure 8:** Sanger sequencing results from experiments validating the presence of EMS-induced mutations detected in 12 fourfold pools (Supplementary Table 3). Using genomic DNA from the four representative individuals of the selected pools, PCR amplicons encompassing detected mutants were Sanger



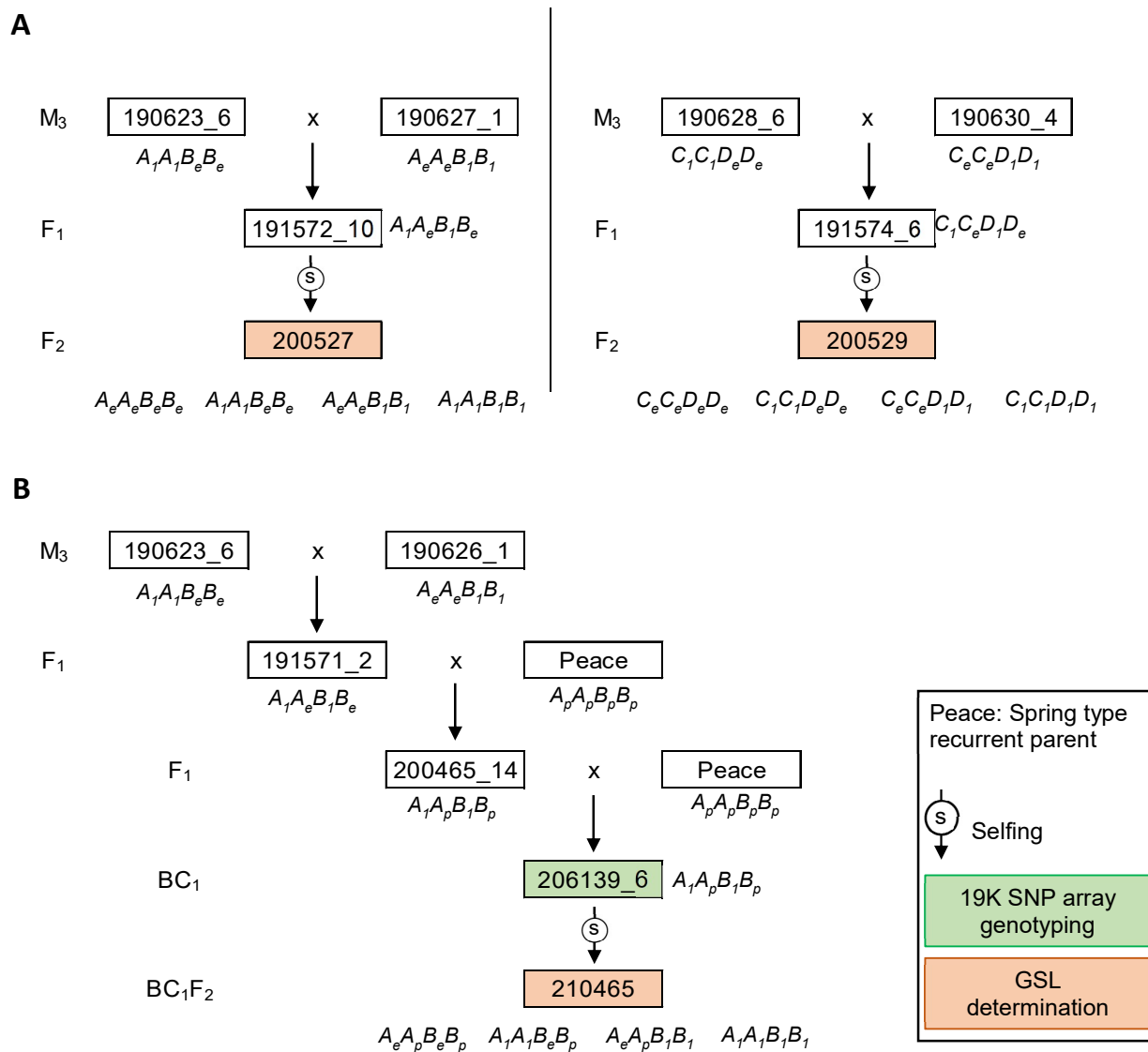
sequenced. Genomic sequences of the candidate genes were extracted from the Express617 reference genome. Location and type of detected mutations were annotated for the candidate genes (marked in blue arrows). Sanger sequencing reads were mapped to the corresponding candidate gene sequences to check for the presence of the expected mutations. Validated mutations from M<sub>2</sub> individuals are marked with orange arrows. Fourfold pools are named with the 'D' prefix. The four M<sub>2</sub> individuals corresponding to each of the pools have been named with the pool IDs followed by the '\_1', '\_2', '\_3' and '\_4' suffixes. Sequence analysis was done using the CLC Main Workbench 7.



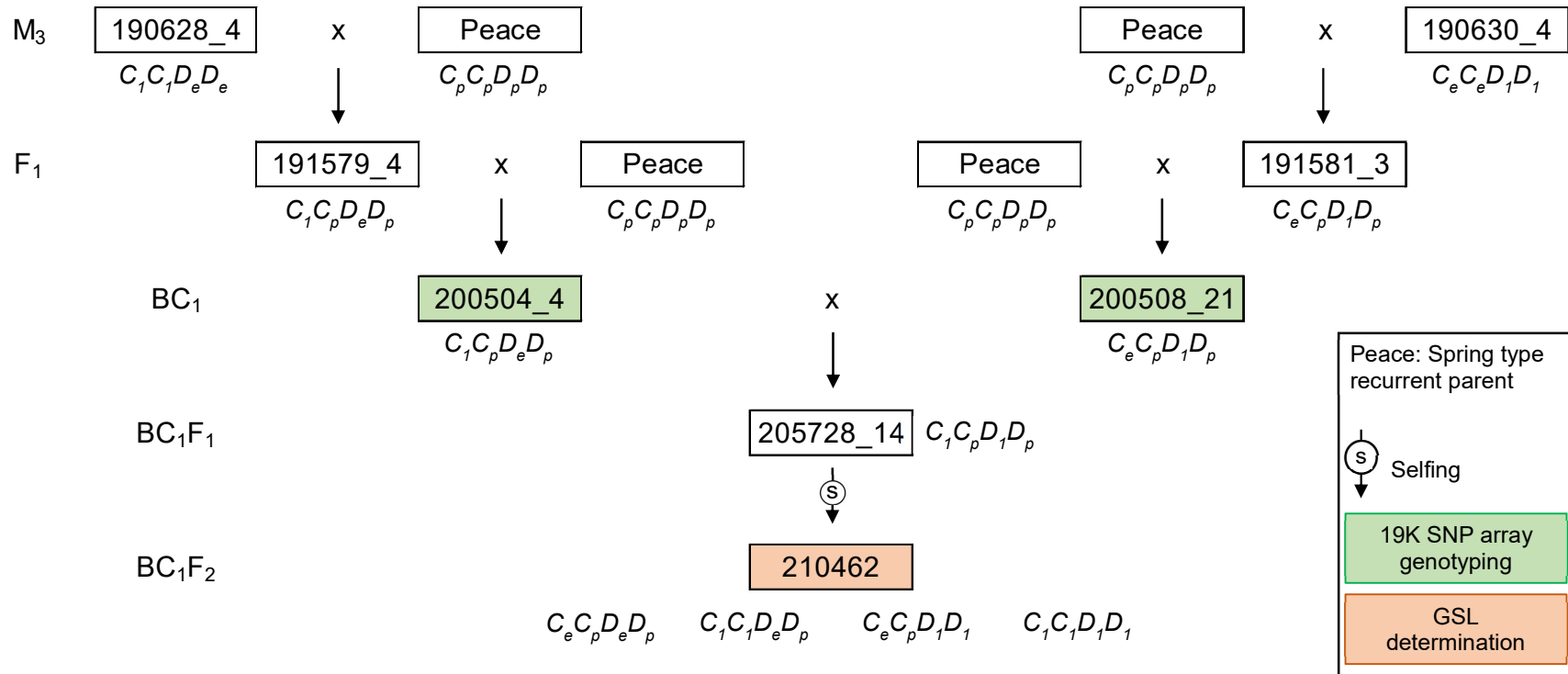
**Supplementary Figure 9:** Gene structures of three *BnMYB28* and two *BnCYP79F1* paralogs identified in rapeseed. START and STOP refer to the translation start and stop sites, respectively. Conserved domains characteristic of respective gene families are marked in yellow boxes. All gene models are based on the Darmor-*bzh* rapeseed reference genome. Paralog *BnMYB28.C09* was truncated but retained all conserved domains required for gene function. Information on the transcription start site for the *BnCYP79F1* paralogs is unavailable on the reference genome.



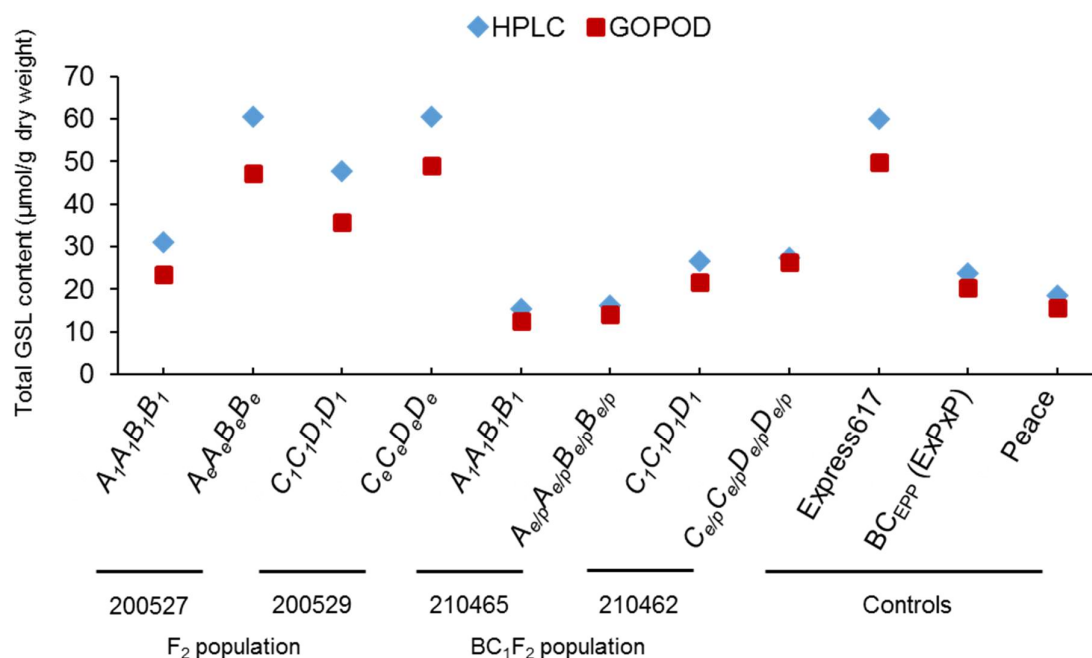
**Supplementary Figure 10:** Conserved domains observed in *MYB28* and *CYP79F1* paralogs of *B. napus*. The R2 and R3 DNA binding domains are highly conserved and characteristic of the MYB family transcription factors, followed by a nuclear localization signal (NLS) (Dubos et al., 2010). Five conserved domains have been reported from the family of cytochrome P450 enzymes. The Heme group is speculated to act as a catalytic domain vital for enzyme function (Reintanz et al., 2001).



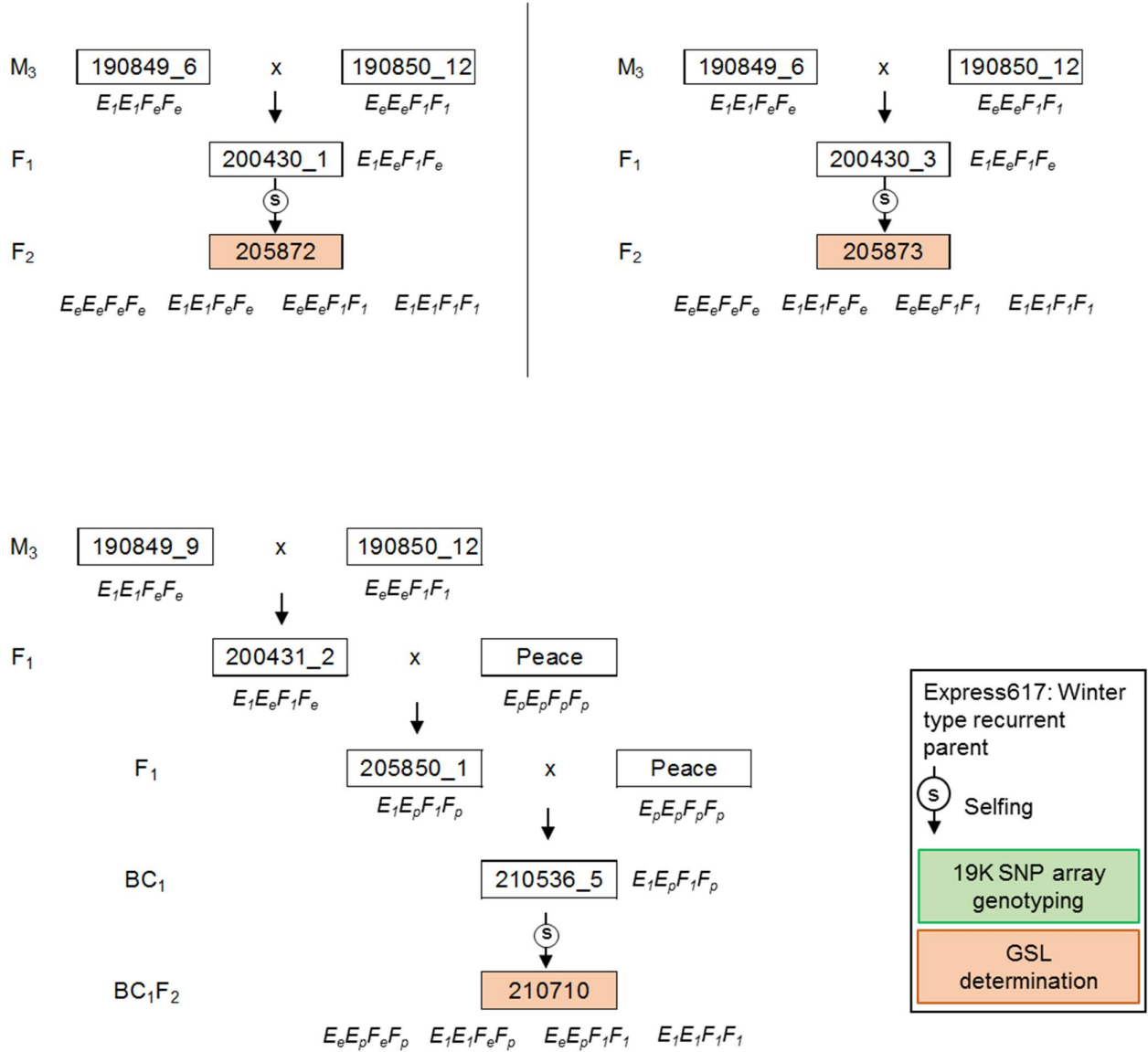
**Supplementary Figure 11:** Crossing schemes and pedigrees of *BnMYB28* and *BnCYP79F1* mutant plants used in this study. Single mutants of the same gene family were crossed with each other to produce double mutants. A) M<sub>3</sub> single mutants were crossed to generate F<sub>1</sub> hybrids (genotypes  $A_1A_eB_1B_e$  or  $C_1C_eD_1D_e$ ). The F<sub>1</sub> hybrids were then selfed to generate F<sub>2</sub> populations 200527 and 200529 representing *BnMYB28* and *BnCYP79F1* mutants, respectively. B) M<sub>3</sub> single mutants of *BnMYB28* paralogs were crossed to produce heterozygous double mutants of *BnMYB28* (genotype  $A_1A_eB_1B_e$ ). The F<sub>1</sub> hybrids were backcrossed with the non-mutagenized spring oilseed rape Peace to generate backcross generation (BC<sub>1</sub>). BC<sub>1</sub> plants were selfed to generate the segregating BC<sub>1</sub>F<sub>2</sub> population (seed code 210465). Plants from the BC<sub>1</sub> generation were first genotyped for the presence of expected EMS-induced mutations and then genotyped with the *Brassica* 19K SNP array. Leaves and seeds from segregating F<sub>2</sub> plants were used for glucosinolate determination. Genotypes mentioned were selected in each generation for further experiments and follow the single letter allele codes as in Table 6. Alleles with ‘e’ and ‘p’ in subscript refer to non-mutagenized Express617 and Peace alleles, respectively.



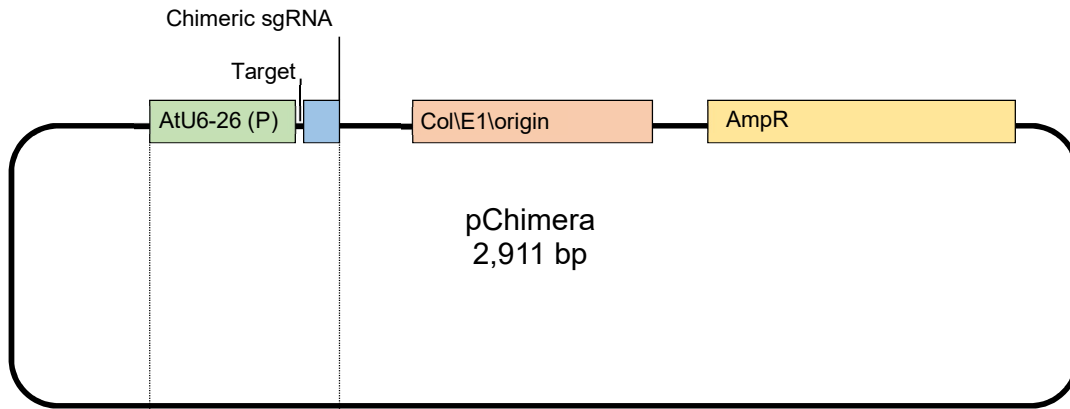
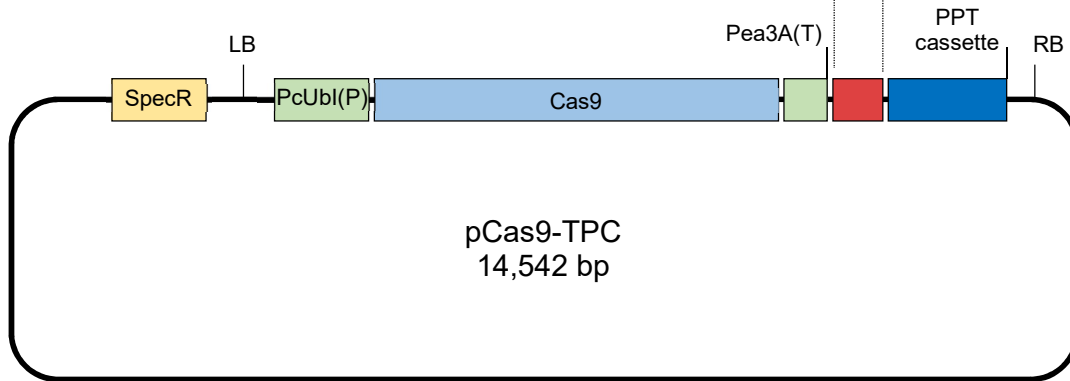
**Supplementary Figure 12:** Crossing schemes and pedigrees of backcrossed *BnCYP79F1* mutant plants used in this study. M<sub>3</sub> single homozygous mutants of *BnCYP79F1* paralogs (genotypes  $C_1C_1D_eD_e$  and  $C_eC_eD_1D_1$ ) were first backcrossed with Peace to generate corresponding BC<sub>1</sub> progenies. Plants from the BC<sub>1</sub> generation were first genotyped for the presence of expected EMS-induced mutations and then genotyped with the *Brassica* 19K SNP array. Single mutants of the BC<sub>1</sub> generation were then combined to produce the segregating BC<sub>1</sub>F<sub>2</sub> population 210462. The BC<sub>1</sub>F<sub>2</sub> population displayed a digenic segregation (1:15 for double mutants  $C_1C_1D_1D_1$ ). Genotypes mentioned were selected in each generation for further experiments and follow the single letter allele codes as in Table 6. Alleles with 'e' and 'p' in subscript refer to non-mutagenized from Express617 and Peace alleles, respectively.



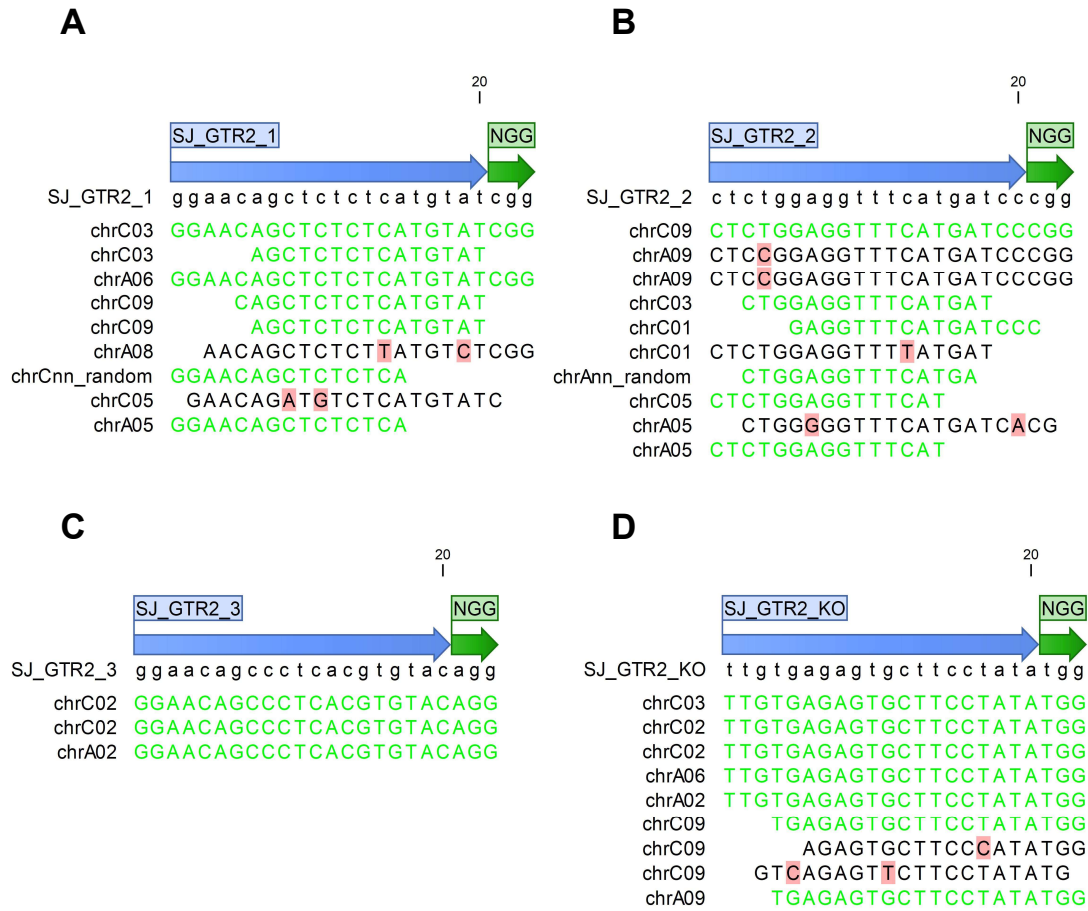
**Supplementary Figure 13:** Comparison of glucosinolate measurements by HPLC and the GOPOD enzymatic assay. Total GSL content determined by HPLC analyses is the sum of all major identified peaks identified from chromatograms (epiprogoitrin, glucobrassicinapin, glucoiberin, gluconapin, glucoraphanin, progoitrin, sinigrin, 4-hydroxyglucobrassicin, 4-methoxyglucobrassicin, glucobrassicin, gluconasturtiin, glucotropaeolin and neoglucobrassicin). Values were calculated after calibrating against GSL standards (Supplementary Table 6). Absolute GSL content was measured using the GOPOD enzymatic assay. *BnMYB28* and *BnCYP79F1* double mutants (genotypes  $A_1A_1B_1B_1$  and  $C_1C_1D_1D_1$ ) and wildtype plants from the same segregating population originating either from M<sub>3</sub>xM<sub>3</sub> crosses ( $A_eA_eB_eB_e$  and  $C_eC_eD_eD_e$ ) or from backcrosses with Peace ( $A_{e/p}A_{e/p}B_{e/p}B_{e/p}$  and  $C_{e/p}C_{e/p}D_{e/p}D_{e/p}$ ) were compared with non-mutagenized controls. For each of the analyzed genotypes, five biological replicates were analyzed.



**Supplementary Figure 14:** Crossing schemes and pedigrees of *BnGTR2* mutant plants used in this study. Single mutants were crossed with each other to produce double mutants. A) M<sub>3</sub> single mutants were crossed to generate F<sub>1</sub> hybrids (genotype  $E_1E_eF_1F_e$ ). The F<sub>1</sub> hybrids were then selfed to generate F<sub>2</sub> populations 205872 and 205873. B) M<sub>3</sub> single mutants of *BnGTR2* paralogs were crossed to produce heterozygous double mutants (genotype  $E_1E_eF_1F_e$ ). The F<sub>1</sub> hybrids were backcrossed with the non-mutagenized spring oilseed rape Peace to generate backcross generation (BC<sub>1</sub>). BC<sub>1</sub> plants were selfed to generate the segregating BC<sub>1</sub>F<sub>2</sub> population (seed code 210710). Plants from the BC<sub>1</sub> generation were first genotyped for the presence of expected EMS-induced mutations and then genotyped with the *Brassica* 19K SNP array. Leaves and seeds from segregating F<sub>2</sub> plants were used for glucosinolate determination. Genotypes mentioned were selected in each generation for further experiments and follow the single letter allele codes as in Table 8. Alleles with ‘e’ and ‘p’ in subscript refer to non-mutagenized Express617 and Peace alleles, respectively.

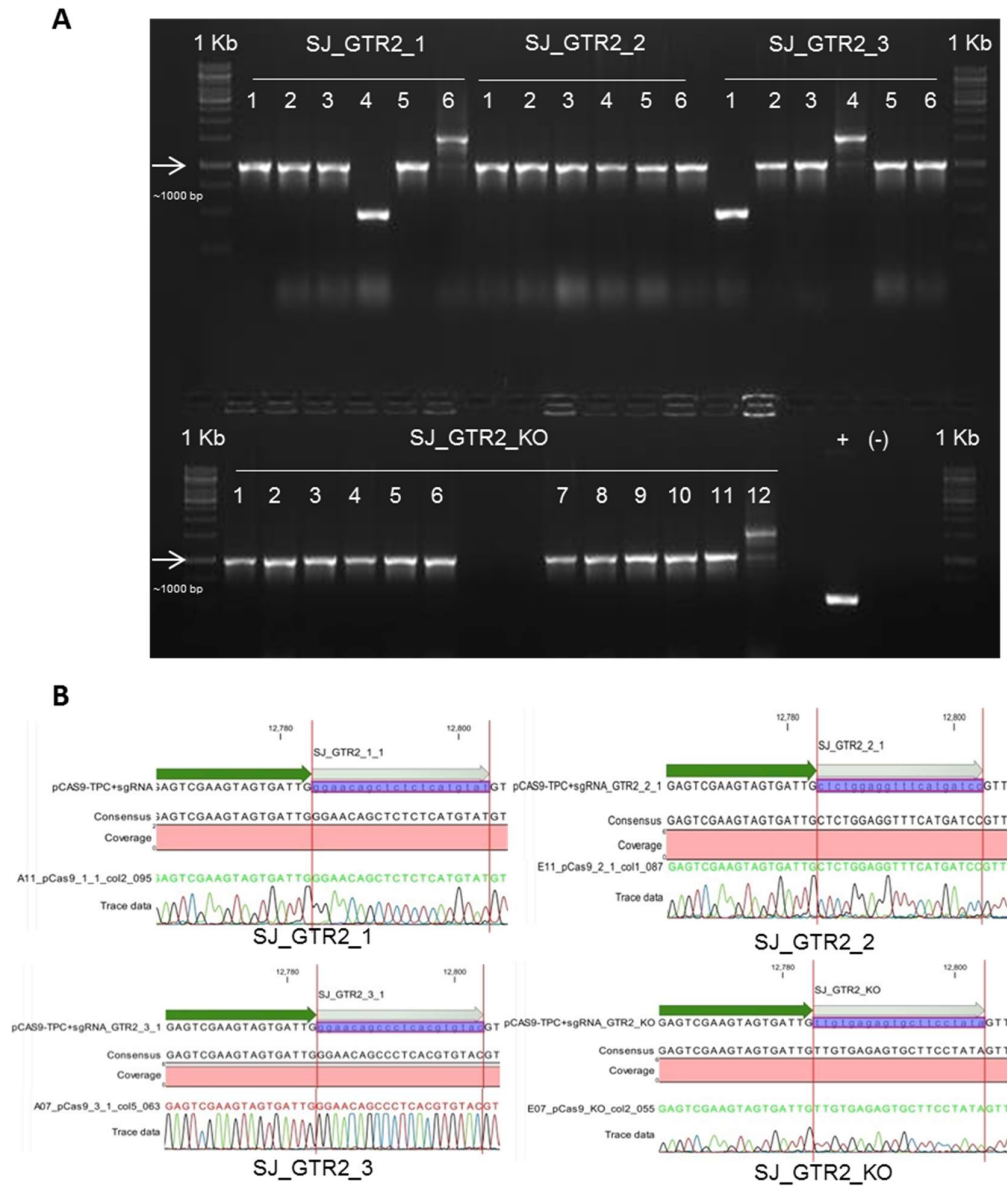
**A****B**

**Supplementary Figure 15:** The binary vector system used for developing the final Cas9 vectors. A) pChimera vector with target protospacer and the chimeric sgRNA complex. The target oligonucleotide (named 'Target') is ligated next to the chimeric sgRNA, constituting the sgRNA-oligonucleotide complex. The chimeric sgRNA is under the influence of the *Arabidopsis thaliana* AtU6-26(P) promoter. B) The final pCas9-TPC vector ligated with the sgRNA–oligonucleotide complex (red box). The Cas9 expression cassette encoding the *Streptococcus pyogenes* Cas9 endonuclease is associated with the PcUbl4-2(P) promoter and the Pea3A(T) terminator. AmpR: ampicillin resistance gene, SpecR: spectinomycin gene, PPT cassette: phosphinothricin resistance gene. LB: left border, RB: right border. Vector sequences are annotated as per the nomenclature of Fauser et al. (2014).



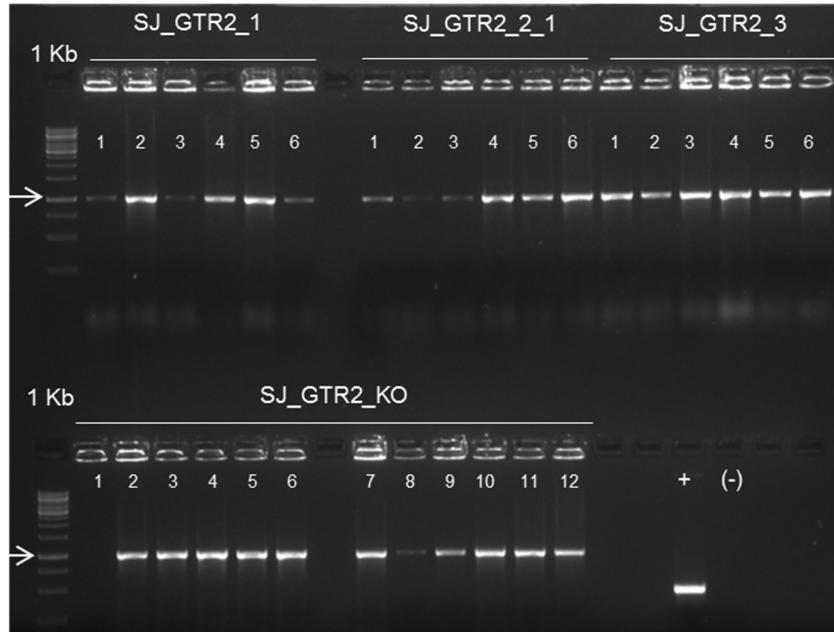
**Supplementary Figure 16:** BLAST results obtained for four selected target sites A) SJ\_GTR2\_1, B) SJ\_GTR2\_2, C) SJ\_GTR2\_3 and D) SJ\_GTR2\_KO considered as target oligonucleotide sequences for the CRISPR-Cas9 mediated knock-out of *BnGTR2*. BLAST analysis was done using the Darmor-*bzh* reference genome on the CLC Genomics Workbench 7. The 20 bp target oligonucleotides and the 5'-NGG-3' protospacer adjacent motifs (PAM) are annotated with blue and green arrows, respectively. Highlighted nucleotides represent SNPs. 20bp target oligonucleotides were selected as targets.





**Supplementary Figure 17:** Construction of final pCas9-TPC vectors ligated with the sgRNA-oligonucleotide complex.

A) Confirmation of successful ligation of the sgRNA-oligonucleotide complex into the pCas9-TPC vector and B) Validation for using Sanger sequencing of colony PCR products. *E. coli* cells were transformed with the pChimera vector ligated with the designed target oligonucleotides (SJ\_GTR2\_1-3 and SJ\_GTR2\_KO). Colony PCR was done using primers pCas9\_F + pCas9\_R flanking the sgRNA-oligonucleotide complex on single *E. coli* colonies (numbered 1-12) growing in spectinomycin selection media. Arrowheads represent the expected fragment size (~1000 bp). +: non-ligated pCas9-TPC vector as positive control and (-): Negative control. Agarose gel electrophoresis conditions: 1% agarose, 100 V, 30 min. GeneRuler™ 1kb DNA ladder used for size reference.



**Supplementary Figure 18:** Detection of the successful transformation of final Cas9-TPC vectors into *A. tumefaciens*. Numbers 1-12 represent single colonies of *A. tumefaciens* transformed with the corresponding target constructs. Colony PCR was performed with primers pCas9F + pCas9R. GE conditions: 1% agarose, 30min at 100 V. GeneRuler™ 1kb DNA ladder used for size reference. Arrowheads represent the expected fragment size (~1000 bp) +: positive control with empty pCas9-TPC vector, (-): Negative control.

## 9 References

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., & Tamiru, M. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology*, 30(2), 174-178.
- Andersen, T. G., & Halkier, B. A. (2014). Upon bolting the GTR1 and GTR2 transporters mediate transport of glucosinolates to the inflorescence rather than roots. *Plant Signaling & Behavior*, 9(1), e27740. doi:10.4161/psb.27740
- Andersen, T. G., Nour-Eldin, H. H., Fuller, V. L., Olsen, C. E., Burow, M., & Halkier, B. A. (2013). Integration of biosynthesis and long-distance transport establish organ-specific glucosinolate profiles in vegetative *Arabidopsis*. *The Plant Cell*, 25(8), 3133-3145.
- Andréasson, E., Jørgensen, L. B., Höglund, A.-S., Rask, L., & Meijer, J. (2001). Different Myrosinase and Idioblast Distribution in *Arabidopsis* and *Brassica napus*. *Plant Physiology*, 127(4), 1750-1763. doi:10.1104/pp.010334
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. In: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
- Araki, R., Hasumi, A., Nishizawa, O. I., Sasaki, K., Kuwahara, A., Sawada, Y., Totoki, Y., Toyoda, A., Sakaki, Y., & Li, Y. (2013). Novel bioresources for studies of *Brassica oleracea*: identification of a kale MYB transcription factor responsible for glucosinolate production. *Plant Biotechnology Journal*, 11(8), 1017-1027.
- Arora, L., & Narula, A. (2017). Gene editing and crop improvement using CRISPR-Cas9 system. *Frontiers in Plant Science*, 8, 1932.
- Augustine, R., & Bisht, N. C. (2015). Biotic elicitors and mechanical damage modulate glucosinolate accumulation by co-ordinated interplay of glucosinolate biosynthesis regulators in polyploid *Brassica juncea*. *Phytochemistry*, 117, 43-50.
- Augustine, R., Majee, M., Gershenzon, J., & Bisht, N. C. (2013). Four genes encoding MYB28, a major transcriptional regulator of the aliphatic glucosinolate pathway, are differentially expressed in the allopolyploid *Brassica juncea*. *Journal of Experimental Botany*, 64(16), 4907-4921. doi:10.1093/jxb/ert280
- Bai, M., Yuan, J., Kuang, H., Gong, P., Li, S., Zhang, Z., Liu, B., Sun, J., Yang, M., & Yang, L. (2020). Generation of a multiplex mutagenesis population via pooled CRISPR-Cas9 in soya bean. *Plant Biotechnology Journal*, 18(3), 721-731.
- Becker, H., Engqvist, G., & Karlsson, B. (1995). Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. *Theoretical and Applied Genetics*, 91(1), 62-67.
- Bibikova, M., Golic, M., Golic, K. G., & Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, 161(3), 1169-1175.
- Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B. J., Mendes, P., Roessner-Tunali, U., & Beale, M. H. (2004). Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*, 9(9), 418-425.

- Bischoff, K. L. (2021). Chapter 53 - Glucosinolates. In R. C. Gupta, R. Lall, & A. Srivastava (Eds.), *Nutraceuticals (Second Edition)* (pp. 903-909): Academic Press.
- Bisht, N. C., Gupta, V., Ramchiary, N., Sodhi, Y., Mukhopadhyay, A., Arumugam, N., Pental, D., & Pradhan, A. (2009). Fine mapping of loci involved with glucosinolate biosynthesis in oilseed mustard (*Brassica juncea*) using genomic information from allied species. *Theoretical and Applied Genetics*, 118(3), 413-421.
- Blažević, I., Montaut, S., Burčul, F., Olsen, C. E., Burow, M., Rollin, P., & Agerbirk, N. (2020). Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry*, 169, 112100.
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959), 1509-1512.
- Borgen, B. H., Thangstad, O. P., Ahuja, I., Rossiter, J. T., & Bones, A. M. (2010). Removing the mustard oil bomb from seeds: transgenic ablation of myrosin cells in oilseed rape (*Brassica napus*) produces *MINELESS* seeds. *Journal of Experimental Botany*, 61(6), 1683-1697. doi:10.1093/jxb/erq039
- Bourdon, D., & Aumaître, A. (1990). Low-glucosinolate rapeseeds and rapeseed meals: effect of technological treatments on chemical composition, digestible energy content and feeding value for growing pigs. *Animal Feed Science and Technology*, 30(3), 175-191. doi:[https://doi.org/10.1016/0377-8401\(90\)90014-Y](https://doi.org/10.1016/0377-8401(90)90014-Y)
- Braatz, J., Harloff, H.-J., Emrani, N., Elisha, C., Heepe, L., Gorb, S. N., & Jung, C. (2018). The effect of *INDEHISCENT* point mutations on silique shatter resistance in oilseed rape (*Brassica napus*). *Theoretical and Applied Genetics*, 131(4), 959-971.
- Braatz, J., Harloff, H.-J., Mascher, M., Stein, N., Himmelbach, A., & Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiology*, 174(2), 935-942.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633-2635.
- Brown, P. D., Tokuhisa, J. G., Reichelt, M., & Gershenzon, J. (2003). Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochemistry*, 62(3), 471-481.
- Bu, X. Y., Wang, Y. Y., Chen, F. Y., Tang, B. B., Luo, C. Z., Wang, Y., Ge, X. P., & Yang, Y. H. (2018). An evaluation of replacing fishmeal with rapeseed meal in the diet of *Pseudobagrus ussuriensis*: growth, feed utilization, nonspecific immunity, and growth-related gene expression. *Journal of the World Aquaculture Society*, 49(6), 1068-1080.
- Burow, M., & Halkier, B. A. (2017). How does a plant orchestrate defense in time and space? Using glucosinolates in *Arabidopsis* as case study. *Curr Opin Plant Biol*, 38, 142-147. doi:10.1016/j.pbi.2017.04.009
- Butruille, D., Guries, R., & Osborn, T. (1999). Increasing yield of spring oilseed rape hybrids through introgression of winter germplasm. *Crop Science*, 39(5), 1491-1496.

- Caldwell, D. G., McCallum, N., Shaw, P., Muehlbauer, G. J., Marshall, D. F., & Waugh, R. (2004). A structured mutant population for forward and reverse genetics in Barley (*Hordeum vulgare* L.). *The Plant Journal*, 40(1), 143-150.
- Carroll, D. (2011). Genome engineering with zinc-finger nucleases. *Genetics*, 188(4), 773-782.
- Carroll, D. (2021). A short, idiosyncratic history of genome editing. *Gene and Genome Editing*, 1, 100002. doi:<https://doi.org/10.1016/j.ggedit.2021.100002>
- Cartea, M. E., & Velasco, P. (2008). Glucosinolates in Brassica foods: bioavailability in food and significance for human health. *Phytochemistry Reviews*, 7(2), 213-229.
- Celenza, J. L., Quiel, J. A., Smolen, G. A., Merrih, H., Silvestro, A. R., Normanly, J., & Bender, J. (2005). The *Arabidopsis ATR1 Myb* transcription factor controls indolic glucosinolate homeostasis. *Plant Physiology*, 137(1), 253-262.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., & Samans, B. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, 345(6199), 950-953.
- Chao, H., Li, T., Luo, C., Huang, H., Ruan, Y., Li, X., Niu, Y., Fan, Y., Sun, W., & Zhang, K. (2020). BrassicaEDB: a gene expression database for *Brassica* crops. *International Journal of Molecular Sciences*, 21(16), 5831.
- Chen, K., & Gao, C. (2014). Targeted genome modification technologies and their applications in crop improvements. *Plant Cell Reports*, 33(4), 575-583. doi:10.1007/s00299-013-1539-6
- Chen, S., Glawischnig, E., Jørgensen, K., Naur, P., Jørgensen, B., Olsen, C. E., Hansen, C. H., Rasmussen, H., Pickett, J. A., & Halkier, B. A. (2003). *CYP79F1* and *CYP79F2* have distinct functions in the biosynthesis of aliphatic glucosinolates in *Arabidopsis*. *The Plant Journal*, 33(5), 923-937.
- Chen, S., Petersen, B. L., Olsen, C. E., Schulz, A., & Halkier, B. A. (2001). Long-distance phloem transport of glucosinolates in *Arabidopsis*. *Plant Physiology*, 127(1), 194-201.
- Chen, W., Wang, Y., Xu, L., Dong, J., Zhu, X., Ying, J., Wang, Q., Fan, L., Li, C., & Liu, L. (2019). Methyl jasmonate, salicylic acid and abscisic acid enhance the accumulation of glucosinolates and sulforaphane in radish (*Raphanus sativus* L.) taproot. *Scientia Horticulturae*, 250, 159-167.
- Chen, X., Tong, C., Zhang, X., Song, A., Hu, M., Dong, W., Chen, F., Wang, Y., Tu, J., & Liu, S. (2021). A high-quality *Brassica napus* genome reveals expansion of transposable elements, subgenome evolution and disease resistance. *Plant Biotechnology Journal*, 19(3), 615-630.
- Cheng, F., Liu, S., Wu, J., Fang, L., Sun, S., Liu, B., Li, P., Hua, W., & Wang, X. (2011). BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biology*, 11(1), 136. doi:10.1186/1471-2229-11-136
- Diepenbrock, W. (2000). Yield analysis of winter oilseed rape (*Brassica napus* L.): a review. *Field Crops Research*, 67(1), 35-49. doi:[https://doi.org/10.1016/S0378-4290\(00\)00082-4](https://doi.org/10.1016/S0378-4290(00)00082-4)
- Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213).

- Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). *MYB* transcription factors in *Arabidopsis*. *Trends in Plant Science*, 15(10), 573-581.
- Emrani, N., Harloff, H.-J., Gudi, O., Kopisch-Obuch, F., & Jung, C. (2015). Reduction in sinapine content in rapeseed (*Brassica napus* L.) by induced mutations in sinapine biosynthesis genes. *Molecular Breeding*, 35(1), 37. doi:10.1007/s11032-015-0236-2
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048.
- Fahey, J. W., Zalcmann, A. T., & Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, 56(1), 5-51.
- Fanelli, V., Ngo, K. J., Thompson, V. L., Silva, B. R., Tsai, H., Sabetta, W., Montemurro, C., Comai, L., & Harmer, S. L. (2021). A TILLING by sequencing approach to identify induced mutations in sunflower genes. *Scientific Reports*, 11(1), 9885. doi:10.1038/s41598-021-89237-w
- Fang, D. D., Naoumkina, M., Thyssen, G. N., Bechere, E., Li, P., & Florane, C. B. (2020). An EMS-induced mutation in a tetratricopeptide repeat-like superfamily protein gene (*Ghir\_A12G008870*) on chromosome A12 is responsible for the *li<sub>y</sub>* short fiber phenotype in cotton. *Theoretical and Applied Genetics*, 133(1), 271-282. doi:10.1007/s00122-019-03456-4
- Fausser, F., Schiml, S., & Puchta, H. (2014). Both CRISPR/C as-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *The Plant Journal*, 79(2), 348-359.
- Feng, J., Long, Y., Shi, L., Shi, J., Barker, G., & Meng, J. (2012). Characterization of metabolite quantitative trait loci and metabolic networks that control glucosinolate concentration in the seeds and leaves of *Brassica napus*. *New Phytologist*, 193(1), 96-108.
- Fenwick, G. R., & Curtis, R. F. (1980). Rapeseed meal and its use in poultry diets. A review. *Animal Feed Science and Technology*, 5(4), 255-298. doi:[https://doi.org/10.1016/0377-8401\(80\)90016-4](https://doi.org/10.1016/0377-8401(80)90016-4)
- Fenwick, R. G., Heaney, R. K., & Mullin, W. J. (1983). Glucosinolates and their breakdown products in food plants. *CRC Crit. Rev. Food Sci. Nutr.*, 18, 123-201.
- Ferreira, M. E., Satagopan, J., Yandell, B. S., Williams, P. H., & Osborn, T. C. (1995). Mapping loci controlling vernalization requirement and flowering time in *Brassica napus*. *Theoretical and Applied Genetics*, 90(5), 727-732. doi:10.1007/BF00222140
- Fiebig, H., & Arens, M. (1992). Glucosinolate (HPLC-Methode) -Gemeinschaftsarbeiten der DGF, 128. Mitteilung: Deutsche Einheitsmethoden zur Untersuchung von Fetten, Fettprodukten, Tensiden und verwandten Stoffen, 98. Mitt.: Analyse von Fettrohstoffen XII\*. 12. *Fett Wissenschaft Technologie*, 94, 199-203.
- Frerigmann, H., Berger, B., & Gigolashvili, T. (2014). bHLH05 is an interaction partner of MYB51 and a novel regulator of glucosinolate biosynthesis in *Arabidopsis*. *Plant Physiology*, 166(1), 349-369.
- Frerigmann, H., Glawischnig, E., & Gigolashvili, T. (2015). The role of MYB34, MYB51 and MYB122 in the regulation of camalexin biosynthesis in *Arabidopsis thaliana*. *Frontiers in Plant Science*, 6, 654.

- Friedt, W., & Snowdon, R. (2009). Oilseed rape. In *Oil crops* (pp. 91-126): Springer.
- Friedt, W., Tu, J., & Fu, T. (2018). Academic and economic importance of *Brassica napus* rapeseed. In *The Brassica napus genome* (pp. 1-20): Springer.
- Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, 32(3), 279-284.
- Fukushima, A., Kanaya, S., & Nishida, K. (2014). Integrated network analysis and effective tools in plant systems biology. *Frontiers in Plant Science*, 5(598). doi:10.3389/fpls.2014.00598
- Fussell, C. P., & Moens, P. (1987). The Rabl orientation: a prelude to synapsis. *Meiosis*, 275-299.
- Garcia, V., Bres, C., Just, D., Fernandez, L., Tai, F. W. J., Mauxion, J.-P., Le Paslier, M.-C., Bérard, A., Brunel, D., Aoki, K., Alseekh, S., Fernie, A. R., Fraser, P. D., & Rothan, C. (2016). Rapid identification of causal mutations in tomato EMS populations via mapping-by-sequencing. *Nature Protocols*, 11(12), 2401-2418. doi:10.1038/nprot.2016.143
- Gigolashvili, T., Berger, B., & Flügge, U.-I. (2009). Specific and coordinated control of indolic and aliphatic glucosinolate biosynthesis by R2R3-MYB transcription factors in *Arabidopsis thaliana*. *Phytochemistry Reviews*, 8(1), 3-13.
- Gigolashvili, T., Engqvist, M., Yatusевич, R., Müller, C., & Flügge, U. I. (2008). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytologist*, 177(3), 627-642. doi:10.1111/j.1469-8137.2007.02295.x
- Gigolashvili, T., Yatusевич, R., Berger, B., Müller, C., & Flügge, U. I. (2007). The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*, 51(2), 247-261.
- Gijzen, M., McGregor, I., & Séguin-Swartz, G. (1989). Glucosinolate uptake by developing rapeseed embryos. *Plant Physiology*, 89(1), 260-263.
- Gilchrist, E. J., Sidebottom, C. H., Koh, C. S., MacInnes, T., Sharpe, A. G., & Haughn, G. W. (2013). A mutant *Brassica napus* (Canola) population for the identification of new genetic diversity via TILLING and next generation sequencing. *PLoS One*, 8(12), e84303.
- Girke, A., Schierholt, A., & Becker, H. C. (2012). Extending the rapeseed gene pool with resynthesized *Brassica napus* L. I: Genetic diversity. *Genetic Resources and Crop Evolution*, 59(7), 1441-1447. doi:10.1007/s10722-011-9772-8
- Gulden, R. H., Warwick, S. I., & Thomas, A. G. (2008). The biology of Canadian weeds. 137. *Brassica napus* L. and *B. rapa* L. *Canadian Journal of Plant Science*, 88(5), 951-996.
- Guo, R., Qian, H., Shen, W., Liu, L., Zhang, M., Cai, C., Zhao, Y., Qiao, J., & Wang, Q. (2013). BZR1 and BES1 participate in regulation of glucosinolate biosynthesis by brassinosteroids in *Arabidopsis*. *Journal of Experimental Botany*, 64(8), 2401-2412. doi:10.1093/jxb/ert094
- Guo, Y., Harloff, H.-J., Jung, C., & Molina, C. (2014). Mutations in single *FT*- and *TFL1*-paralogs of rapeseed (*Brassica napus* L.) and their impact on flowering time and yield components. *Frontiers in Plant Science*, 5, 282.



- Halkier, B. A., & Du, L. (1997). The biosynthesis of glucosinolates. *Trends in Plant Science*, 2(11), 425-431.
- Halkier, B. A., & Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.*, 57, 303-333.
- Hansen, J. Ø., Øverland, M., Skrede, A., Anderson, D. M., & Collins, S. A. (2020). A meta-analysis of the effects of dietary canola / double low rapeseed meal on growth performance of weanling and growing-finishing pigs. *Animal Feed Science and Technology*, 259, 114302. doi:<https://doi.org/10.1016/j.anifeedsci.2019.114302>
- Harloff, H.-J., Lemcke, S., Mittasch, J., Frolov, A., Wu, J. G., Dreyer, F., Leckband, G., & Jung, C. (2012). A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. *Theoretical and Applied Genetics*, 124(5), 957-969.
- Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., & Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nature Biotechnology*, 30(8), 798-802.
- Hasan, M., Friedt, W., Pons-Kühnemann, J., Freitag, N., Link, K., & Snowdon, R. (2008). Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. *napus*). *Theoretical and Applied Genetics*, 116(8), 1035-1049.
- Hatzig, S., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.-H., Leckband, G., Abbadi, A., & Snowdon, R. J. (2018). Hidden Effects of Seed Quality Breeding on Germination in Oilseed Rape (*Brassica napus* L.). *Frontiers in Plant Science*, 9, 419-419. doi:10.3389/fpls.2018.00419
- Hemm, M. R., Ruegger, M. O., & Chapple, C. (2003). The *Arabidopsis* *ref2* Mutant Is Defective in the Gene Encoding CYP83A1 and Shows Both Phenylpropanoid and Glucosinolate Phenotypes. *The Plant Cell*, 15(1), 179-194. doi:10.1105/tpc.006544
- Henikoff, S., Till, B. J., & Comai, L. (2004). TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiology*, 135(2), 630-636.
- Henry, I. M., Nagalakshmi, U., Lieberman, M. C., Ngo, K. J., Krasileva, K. V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., & Tai, T. H. (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *The Plant Cell*, 26(4), 1382-1397.
- Himelblau, E., Gilchrist, E. J., Buono, K., Bizzell, C., Mentzer, L., Vogelzang, R., Osborn, T., Amasino, R. M., Parkin, I. A., & Haughn, G. W. (2009). Forward and reverse genetics of rapid-cycling *Brassica oleracea*. *Theoretical and Applied Genetics*, 118(5), 953-961.
- Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., Araki, R., Sakurai, N., Suzuki, H., & Aoki, K. (2007). Omics-based identification of *Arabidopsis* *Myb* transcription factors regulating aliphatic glucosinolate biosynthesis. *Proceedings of the National Academy of Sciences*, 104(15), 6478-6483.
- Höfgen, R., & Willmitzer, L. (1988). Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Research*, 16(20), 9877.
- Hopkins, R. J., Dam, N. M. v., & Loon, J. J. A. v. (2009). Role of Glucosinolates in Insect-Plant Relationships and Multitrophic Interactions. *Annual Review of Entomology*, 54(1), 57-83. doi:10.1146/annurev.ento.54.110807.090623



- Howell, P., Sharpe, A., & Lydiate, D. (2003). Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*Brassica napus*). *Genome*, 46(3), 454-460.
- Hu, S., Yu, C., Zhao, H., Sun, G., Zhao, S., Vyvadilova, M., & Kucera, V. (2007). Genetic diversity of *Brassica napus* L. Germplasm from China and Europe assessed by some agronomically important characters. *Euphytica*, 154(1), 9-16. doi:10.1007/s10681-006-9263-8
- Huang, T.-K., & Puchta, H. (2021). Novel CRISPR/Cas applications in plants: from prime editing to chromosome engineering. *Transgenic Research*, 1-21.
- Ishida, M., Hara, M., Fukino, N., Kakizaki, T., & Morimitsu, Y. (2014). Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables. *Breeding Science*, 64(1), 48-59.
- James, D. C., & Rossiter, J. T. (1991). Development and characteristics of myrosinase in *Brassica napus* during early seedling growth. *Physiologia Plantarum*, 82(2), 163-170.
- Jamil, I. N., Remali, J., Azizan, K. A., Nor Muhammad, N. A., Arita, M., Goh, H.-H., & Aizat, W. M. (2020). Systematic Multi-Omics Integration (MOI) Approach in Plant Systems Biology. *Frontiers in Plant Science*, 11, 944-944. doi:10.3389/fpls.2020.00944
- Jeschke, V., Kearney, E. E., Schramm, K., Kunert, G., Shekhov, A., Gershenzon, J., & Vassão, D. G. (2017). How Glucosinolates Affect Generalist Lepidopteran Larvae: Growth, Development and Glucosinolate Metabolism. *Frontiers in Plant Science*, 8(1995). doi:10.3389/fpls.2017.01995
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816-821.
- Jørgensen, M. E., Nour-Eldin, H. H., & Halkier, B. A. (2015). Transport of defense compounds from source to sink: lessons learned from glucosinolates. *Trends in Plant Science*, 20(8), 508-514. doi:<https://doi.org/10.1016/j.tplants.2015.04.006>
- Jørgensen, M. E., Olsen, C. E., Geiger, D., Mirza, O., Halkier, B. A., & Nour-Eldin, H. H. (2015). A Functional EXXEK Motif is Essential for Proton Coupling and Active Glucosinolate Transport by NPF2.11. *Plant & Cell Physiology*, 56(12), 2340-2350. doi:10.1093/pcp/pcv145
- Jung, C., & Till, B. (2021). Mutagenesis and genome editing in crop improvement: perspectives for the global regulatory landscape. *Trends in Plant Science*.
- Kaiser, F., Harloff, H. J., Tressel, R. P., Kock, T., & Schulz, C. (2021). Effects of highly purified rapeseed protein isolate as fishmeal alternative on nutrient digestibility and growth performance in diets fed to rainbow trout (*Oncorhynchus mykiss*). *Aquaculture Nutrition*.
- Kannan, B., Jung, J. H., Moxley, G. W., Lee, S. M., & Altpeter, F. (2018). TALEN-mediated targeted mutagenesis of more than 100 COMT copies/alleles in highly polyploid sugarcane improves saccharification efficiency without compromising biomass yield. *Plant Biotechnology Journal*, 16(4), 856-866.
- Karunarathna, N. L., Patiranage, D. S. R., Harloff, H.-J., Sashidhar, N., & Jung, C. (2021). Genomic background selection to reduce the mutation load after random mutagenesis. *Scientific Reports*, 11(1), 19404. doi:10.1038/s41598-021-98934-5

- Karunaratna, N. L., Wang, H., Harloff, H. J., Jiang, L., & Jung, C. (2020). Elevating seed oil content in a polyploid crop by induced mutations in *SEED FATTY ACID REDUCER* genes. *Plant Biotechnology Journal*, 18(11), 2251-2266.
- Kaul, S., Sharma, T., & K. Dhar, M. (2016). “Omics” Tools for Better Understanding the Plant–Endophyte Interactions. *Frontiers in Plant Science*, 7(955). doi:10.3389/fpls.2016.00955
- Kazama, T., Okuno, M., Watari, Y., Yanase, S., Koizuka, C., Tsuruta, Y., Sugaya, H., Toyoda, A., Itoh, T., & Tsutsumi, N. (2019). Curing cytoplasmic male sterility via TALEN-mediated mitochondrial genome editing. *Nature Plants*, 5(7), 722-730.
- Keck, A.-S., & Finley, J. W. (2004). Cruciferous vegetables: cancer protective mechanisms of glucosinolate hydrolysis products and selenium. *Integrative Cancer Therapies*, 3(1), 5-12.
- Kelly, P., Bones, A., & Rossiter, J. (1998). Sub-cellular immunolocalization of the glucosinolate sinigrin in seedlings of *Brassica juncea*. *Planta*, 206(3), 370-377.
- Kiddle, G. A., Doughty, K. J., & Wallsgrove, R. M. (1994). Salicylic acid-induced accumulation of glucosinolates in oilseed rape (*Brassica napus* L.) leaves. *Journal of Experimental Botany*, 45(9), 1343-1346.
- Kim, Y.-G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, 93(3), 1156-1160.
- Kim, Y. B., Li, X., Kim, S.-J., Kim, H. H., Lee, J., Kim, H., & Park, S. U. (2013). MYB transcription factors regulate glucosinolate biosynthesis in different organs of Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Molecules*, 18(7), 8682-8695.
- Kissen, R., Rossiter, J. T., & Bones, A. M. (2009). The ‘mustard oil bomb’: not so easy to assemble?! Localization, expression and distribution of the components of the myrosinase enzyme system. *Phytochemistry Reviews*, 8(1), 69-86.
- Kittipol, V., He, Z., Wang, L., Doheny-Adams, T., Langer, S., & Bancroft, I. (2019). Genetic architecture of glucosinolate variation in *Brassica napus*. *Journal of Plant Physiology*, 240, 152988. doi:<https://doi.org/10.1016/j.jplph.2019.06.001>
- Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., & Mitchell-Olds, T. (2001). Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiology*, 126(2), 811-825.
- Kondra, Z., & Stefansson, B. (1970). Inheritance of the major glucosinolates of rapeseed (*Brassica napus*) meal. *Canadian Journal of Plant Science*, 50(6), 643-647.
- Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R. H., Wang, X., & Borrill, P. (2017). Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences*, 114(6), E913-E921.
- Kroymann, J., Textor, S., Tokuhisa, J. G., Falk, K. L., Bartram, S., Gershenzon, J., & Mitchell-Olds, T. (2001). A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiology*, 127(3), 1077-1088.

- Lakhssassi, N., Lopes-Caitar, V. S., Knizia, D., Cullen, M. A., Badad, O., El Baze, A., Zhou, Z., Embaby, M. G., Meksem, J., & Lakhssassi, A. (2021). TILLING-by-Sequencing<sup>+</sup> Reveals the Role of Novel Fatty Acid Desaturases (GmFAD2-2s) in Increasing Soybean Seed Oleic Acid Content. *Cells*, 10(5), 1245.
- Lakhssassi, N., Zhou, Z., Liu, S., Piya, S., Cullen, M. A., El Baze, A., Knizia, D., Patil, G. B., Badad, O., & Embaby, M. G. (2020). Soybean TILLING-by-Sequencing<sup>+</sup> reveals the role of novel *GmSACPD* members in the unsaturated fatty acid biosynthesis while maintaining healthy nodules. *Journal of Experimental Botany*.
- Lee, H., Chawla, H. S., Obermeier, C., Dreyer, F., Abbadi, A., & Snowdon, R. (2020). Chromosome-Scale Assembly of Winter Oilseed Rape *Brassica napus*. *Frontiers in Plant Science*, 11(496). doi:10.3389/fpls.2020.00496
- Lee, Y.-H., Park, W., Kim, K.-S., Jang, Y.-S., Lee, J.-E., Cha, Y.-L., Moon, Y.-H., Song, Y.-S., & Lee, K. (2018). EMS-induced mutation of an endoplasmic reticulum oleate desaturase gene (*FAD2-2*) results in elevated oleic acid content in rapeseed (*Brassica napus* L.). *Euphytica*, 214(2), 1-12.
- Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., Qiu, J.-L., Chen, Y.-H., Li, J., & Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nature Biotechnology*, 38(7), 875-882. doi:10.1038/s41587-019-0393-7
- Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., Harper, A. L., Trick, M., Liu, S., & Gao, G. (2014). Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Research*, 21(4), 355-367.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Li, T., Liu, B., Spalding, M. H., Weeks, D. P., & Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nature Biotechnology*, 30(5), 390-392.
- Liersch, A., Bocianowski, J., & Bartkowiak-Broda, I. (2013). Fatty acid and glucosinolate level in seeds of different types of winter oilseed rape cultivars (*Brassica napus* L.). *Communications in Biometry and Crop Science*, 8(2), 39-47.
- Liu, S., Huang, H., Yi, X., Zhang, Y., Yang, Q., Zhang, C., Fan, C., & Zhou, Y. (2020). Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome-wide association study. *Plant Biotechnology Journal*, 18(6), 1472-1484.
- Liu, Y., Zhou, X., Yan, M., Wang, P., Wang, H., Xin, Q., Yang, L., Hong, D., & Yang, G. (2020). Fine mapping and candidate gene analysis of a seed glucosinolate content QTL, *qGSL-C2*, in rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics*, 133(2), 479-490. doi:10.1007/s00122-019-03479-x
- Liu, Z., Liang, J., Zheng, S., Zhang, J., Wu, J., Cheng, F., Yang, W., & Wang, X. (2017). Enriching Glucoraphanin in *Brassica rapa* Through Replacement of *BrAOP2.2/BrAOP2.3* with Non-functional Genes. *Frontiers in Plant Science*, 8(1329). doi:10.3389/fpls.2017.01329

- Long, L., Guo, D.-D., Gao, W., Yang, W.-W., Hou, L.-P., Ma, X.-N., Miao, Y.-C., Botella, J. R., & Song, C.-P. (2018). Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. *Plant Methods*, 14(1), 1-9.
- Lu, G., Harper, A. L., Trick, M., Morgan, C., Fraser, F., O'Neill, C., & Bancroft, I. (2014). Associative transcriptomics study dissects the genetic architecture of seed glucosinolate content in *Brassica napus*. *DNA research*, 21(6), 613-625.
- Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., Zhang, C., Chen, Z., Xiao, Z., & Jian, H. (2019). Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nature Communications*, 10(1), 1154.
- Lu, X., Liu, J., Ren, W., Yang, Q., Chai, Z., Chen, R., Wang, L., Zhao, J., Lang, Z., & Wang, H. (2018). Gene-indexed mutations in maize. *Molecular Plant*, 11(3), 496-504.
- Lu, Z., Cui, J., Wang, L., Teng, N., Zhang, S., Lam, H.-M., Zhu, Y., Xiao, S., Ke, W., Lin, J., Xu, C., & Jin, B. (2021). Genome-wide DNA mutations in *Arabidopsis* plants after multigenerational exposure to high temperatures. *Genome Biology*, 22(1), 160. doi:10.1186/s13059-021-02381-4
- Luo, J., Li, S., Xu, J., Yan, L., Ma, Y., & Xia, L. (2021). Pyramiding favorable alleles in an elite wheat variety in one generation by CRISPR-Cas9-mediated multiplex gene editing. *Molecular Plant*, 14(6), 847-850.
- Luo, M., Li, H., Chakraborty, S., Morbitzer, R., Rinaldo, A., Upadhyaya, N., Bhatt, D., Louis, S., Richardson, T., & Lahaye, T. (2019). Efficient TALEN-mediated gene editing in wheat. *Plant Biotechnology Journal*, 17(11), 2026.
- Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B., Yang, Z., Li, H., & Lin, Y. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular Plant*, 8(8), 1274-1284.
- Madloo, P., Lema, M., Francisco, M., & Soengas, P. (2019). Role of Major Glucosinolates in the Defense of Kale Against *Sclerotinia sclerotiorum* and *Xanthomonas campestris* pv. *campestris*. *Phytopathology*, 109(7), 1246-1256.
- Magrath, R., Herron, C., Giamoustaris, A., & Mithen, R. (1993). The inheritance of aliphatic glucosinolates in *Brassica napus*. *Plant Breeding*, 111(1), 55-72.
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, 339(6121), 823-826.
- Malka, S. K., & Cheng, Y. (2017). Possible interactions between the biosynthetic pathways of indole glucosinolate and auxin. *Frontiers in Plant Science*, 8, 2131.
- Martinez-Perez, E., Shaw, P., & Moore, G. (2001). The *Ph1* locus is needed to ensure specific somatic and meiotic centromere association. *Nature*, 411(6834), 204-207.
- Matthäus, B., & Brühl, L. (2003). Quality of cold-pressed edible rapeseed oil in Germany. *Food/Nahrung*, 47(6), 413-419.
- Mawson, R., Heaney, R., Piskula, M., & Kozłowska, H. (1993). Rapeseed meal-glucosinolates and their antinutritional effects. Part 1. Rapeseed production and chemistry of glucosinolates. *Die Nahrung*, 37(2), 131-140.

- Mithen, R. (1992). Leaf glucosinolate profiles and their relationship to pest and disease resistance in oilseed rape. In *Breeding for Disease Resistance* (pp. 71-83): Springer.
- Mitreiter, S., & Gigolashvili, T. (2021). Regulation of glucosinolate biosynthesis. *Journal of Experimental Botany*, 72(1), 70-91.
- Mochida, K., & Shinozaki, K. (2011). Advances in Omics and Bioinformatics Tools for Systems Analyses of Plant Functions. *Plant and Cell Physiology*, 52(12), 2017-2038. doi:10.1093/pcp/pcr153
- Moradpour, M., & Abdulah, S. N. A. (2020). CRISPR/dCas9 platforms in plants: strategies and applications beyond genome editing. *Plant Biotechnology Journal*, 18(1), 32-44.
- Mosa, K. A., Ismail, A., & Helmy, M. (2017). Omics and System Biology Approaches in Plant Stress Research. In *Plant Stress Tolerance: An Integrated Omics Approach* (pp. 21-34). Cham: Springer International Publishing.
- Mugford, S. G., Yoshimoto, N., Reichelt, M., Wirtz, M., Hill, L., Mugford, S. T., Nakazato, Y., Noji, M., Takahashi, H., Kramell, R., Gigolashvili, T., Flügge, U.-I., Wasternack, C., Gershenzon, J., Hell, R. d., Saito, K., & Kopriva, S. (2009). Disruption of Adenosine-5'-Phosphosulfate Kinase in *Arabidopsis* Reduces Levels of Sulfated Secondary Metabolites. *The Plant Cell*, 21(3), 910-927. doi:10.1105/tpc.109.065581
- Nambiar, D. M., Kumari, J., Augustine, R., Kumar, P., Bajpai, P. K., & Bisht, N. C. (2021). GTR1 and GTR2 transporters differentially regulate tissue-specific glucosinolate contents and defence responses in the oilseed crop *Brassica juncea*. *Plant, Cell & Environment*.
- Neequaye, M., Stavnstrup, S., Harwood, W., Lawrenson, T., Hundleby, P., Irwin, J., Troncoso-Rey, P., Saha, S., Traka, M. H., & Mithen, R. (2021). CRISPR-Cas9-mediated gene editing of *MYB28* genes impair glucoraphanin accumulation of *Brassica oleracea* in the field. *The CRISPR Journal*, 4(3), 416-426.
- Neugart, S., Hanschen, F., & Schreiner, M. (2020). 15B Glucosinolates in *Brassica*. *The Physiology of Vegetable Crops*, 389.
- Nie, S., Wang, B., Ding, H., Lin, H., Zhang, L., Li, Q., Wang, Y., Zhang, B., Liang, A., & Zheng, Q. (2021). Genome assembly of the Chinese maize elite inbred line RP125 and its EMS mutant collection provide new resources for maize genetics research and crop improvement. *The Plant Journal*.
- Nour-Eldin, H. H., Andersen, T. G., Burow, M., Madsen, S. R., Jørgensen, M. E., Olsen, C. E., Dreyer, I., Hedrich, R., Geiger, D., & Halkier, B. A. (2012). NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature*, 488(7412), 531.
- Nour-Eldin, H. H., & Halkier, B. A. (2009). Piecing together the transport pathway of aliphatic glucosinolates. *Phytochemistry Reviews*, 8(1), 53-67.
- Nour-Eldin, H. H., Madsen, S. R., Engelen, S., Jørgensen, M. E., Olsen, C. E., Andersen, J. S., Seynnaeve, D., Verhoye, T., Fulawka, R., Denolf, P., & Halkier, B. A. (2017). Reduction of antinutritional glucosinolates in *Brassica* oilseeds by mutation of genes encoding transporters. *Nature Biotechnology*, 35, 377. doi:10.1038/nbt.3823

<https://www.nature.com/articles/nbt.3823#supplementary-information>

- Oksman-Caldentey, K.-M., & Saito, K. (2005). Integrating genomics and metabolomics for engineering plant metabolic pathways. *Current Opinion in Biotechnology*, 16(2), 174-179. doi:<https://doi.org/10.1016/j.copbio.2005.02.007>
- Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., & Yeung, A. T. (1998). Mutation detection using a novel plant endonuclease. *Nucleic Acids Research*, 26(20), 4597-4602.
- Padilla, G., Cartea, M. E., Velasco, P., de Haro, A., & Ordás, A. (2007). Variation of glucosinolates in vegetable crops of *Brassica rapa*. *Phytochemistry*, 68(4), 536-545.
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., & Roazen, D. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178.
- Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *Journal of Experimental Botany*, 56(409), 1-14.
- Qu, C.-M., Li, S.-M., Duan, X.-J., Fan, J.-H., Jia, L.-D., Zhao, H.-Y., Lu, K., Li, J.-N., Xu, X.-F., & Wang, R. (2015). Identification of Candidate Genes for Seed Glucosinolate Content Using Association Mapping in *Brassica napus* L. *Genes*, 6(4), 1215-1229.
- Quijada, P. A., Udall, J. A., Lambert, B., & Osborn, T. C. (2006). Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.): 1. Identification of genomic regions from winter germplasm. *Theoretical and Applied Genetics*, 113(3), 549-561.
- Rahman, H. (2013). Review: Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. *Canadian Journal of Plant Science*, 93(3), 363-373. doi:10.4141/cjps2012-074
- Rahman, M. (2001). Production of yellow-seeded *Brassica napus* through interspecific crosses. *Plant Breeding*, 120(6), 463-472.
- Rask, L., Andréasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., & Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Molecular Biology*, 42(1), 93-114.
- Reintanz, B., Lehnen, M., Reichelt, M., Gershenzon, J., Kowalczyk, M., Sandberg, G., Godde, M., Uhl, R., & Palme, K. (2001). bus, a bushy *Arabidopsis* CYP79F1 knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. *The Plant Cell*, 13(2), 351-367.
- Renaud, J.-B., Boix, C., Charpentier, M., De Cian, A., Cochenne, J., Duvernois-Berthet, E., Perrouault, L., Tesson, L., Edouard, J., & Thinard, R. (2016). Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 nucleases. *Cell Reports*, 14(9), 2263-2272.
- Rousseau-Gueutin, M., Belser, C., Da Silva, C., Richard, G., Istace, B., Cruaud, C., Falentin, C., Boideau, F., Boutte, J., & Delourme, R. (2020). Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience*, 9(12), gaa137.
- Russo, M., Yan, F., Stier, A., Klasen, L., & Honermeier, B. (2021). Erucic acid concentration of rapeseed (*Brassica napus* L.) oils on the German food retail market. *Food Science & Nutrition*.



- Saghai-Maroofof, M. A., Soliman, K. M., Jorgensen, R. A., & Allard, R. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences*, 81(24), 8014-8018.
- Saikia, B., Singh, S., Debbarma, J., Velmurugan, N., Dekaboruah, H., Arunkumar, K. P., & Chikkaputtaiah, C. (2020). Multigene CRISPR/Cas9 genome editing of hybrid proline rich proteins (HyPRPs) for sustainable multi-stress tolerance in crops: the review of a promising approach. *Physiology and Molecular Biology of Plants*, 26(5), 857-869.
- Saito, K., & Matsuda, F. (2010). Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. *Annual Review of Plant Biology*, 61(1), 463-489.  
doi:10.1146/annurev.arplant.043008.092035
- Sánchez-Pujante, P., Sabater-Jara, A., Belchí-Navarro, S., Pedreño, M., & Almagro, L. (2018). Increased glucosinolate production in *Brassica oleracea* var. *italica* cell cultures due to coronatine activated genes involved in glucosinolate biosynthesis. *Journal of Agricultural and Food Chemistry*, 67(1), 102-111.
- Sashidhar, N., Harloff, H. J., & Jung, C. (2019). Identification of phytic acid mutants in oilseed rape (*Brassica napus*) by large scale screening of mutant populations through amplicon sequencing. *New Phytologist*.
- Sashidhar, N., Harloff, H. J., Potgieter, L., & Jung, C. (2020). Gene editing of three *BnITPK* genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnology Journal*, 18(11), 2241-2250.
- Schauer, N., & Fernie, A. R. (2006). Plant metabolomics: towards biological function and mechanism. *Trends in Plant Science*, 11(10), 508-516.  
doi:<https://doi.org/10.1016/j.tplants.2006.08.007>
- Schreiner, M., Krumbein, A., Knorr, D., & Smetanska, I. (2011). Enhanced Glucosinolates in Root Exudates of *Brassica rapa* ssp. *rapa* Mediated by Salicylic Acid and Methyl Jasmonate. *Journal of Agricultural and Food Chemistry*, 59(4), 1400-1405. doi:10.1021/jf103585s
- Schuster, J., Knill, T., Reichelt, M., Gershenzon, J., & Binder, S. (2006). Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. *The Plant Cell*, 18(10), 2664-2679.
- Schweizer, F., Fernández-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G., Lewsey, M. G., Ecker, J. R., Solano, R., & Reymond, P. (2013). *Arabidopsis* basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *The Plant Cell*, 25(8), 3117-3132.
- Sega, G. A. (1984). A review of the genetic effects of ethyl methanesulfonate. *Mutation Research/Reviews in Genetic Toxicology*, 134(2-3), 113-142.
- Seo, M. S., Jin, M., Chun, J.-H., Kim, S.-J., Park, B.-S., Shon, S.-H., & Kim, J. S. (2016). Functional analysis of three *BrMYB28* transcription factors controlling the biosynthesis of glucosinolates in *Brassica rapa*. *Plant Molecular Biology*, 90(4), 503-516.
- Seo, M. S., & Kim, J. S. (2017). Understanding of MYB Transcription Factors Involved in Glucosinolate Biosynthesis in Brassicaceae. *Molecules*, 22(9), 1549.

- Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J. D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitzel, T., Katagiri, F., Glazebrook, J., Law, M., & Goff, S. A. (2002). A High-Throughput *Arabidopsis* Reverse Genetics System. *The Plant Cell*, 14(12), 2985-2994. doi:10.1105/tpc.004630
- Seyis, F., Friedt, W., & Lühs, W. (2006). Yield of *Brassica napus* L. hybrids developed using resynthesized rapeseed material sown at different locations. *Field Crops Research*, 96(1), 176-180.
- Shah, S., Karunarathna, N. L., Jung, C., & Emrani, N. (2018). An *APETALA1* ortholog affects plant architecture and seed yield component in oilseed rape (*Brassica napus* L.). *BMC Plant Biology*, 18(1), 380. doi:10.1186/s12870-018-1606-9
- Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., Zhang, K., Liu, J., Voytas, D. F., & Zheng, X. (2013). Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Molecular Plant*, 6(4), 1365.
- Shelake, R. M., Pramanik, D., & Kim, J.-Y. (2019). Evolution of plant mutagenesis tools: a shifting paradigm from random to targeted genome editing. *Plant Biotechnology Reports*, 13(5), 423-445. doi:10.1007/s11816-019-00562-z
- Singh, N., Jayaswal, P. K., Panda, K., Mandal, P., Kumar, V., Singh, B., Mishra, S., Singh, Y., Singh, R., & Rai, V. (2015). Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. *Scientific Reports*, 5(1), 1-9.
- Smith, J., Bibikova, M., Whitby, F. G., Reddy, A., Chandrasegaran, S., & Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Research*, 28(17), 3361-3369.
- Sønderby, I. E., Geu-Flores, F., & Halkier, B. A. (2010). Biosynthesis of glucosinolates—gene discovery and beyond. *Trends in Plant Science*, 15(5), 283-290.
- Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., & Zhou, R. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants*, 6(1), 34-45.
- Stephenson, P., Baker, D., Girin, T., Perez, A., Amoah, S., King, G. J., & Østergaard, L. (2010). A rich TILLING resource for studying gene function in *Brassica rapa*. *BMC Plant Biology*, 10(1), 62.
- Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, 4(5), 447-456.
- Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., & Huala, E. (2007). The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research*, 36(suppl\_1), D1009-D1014. doi:10.1093/nar/gkm965
- Tadege, M., Sheldon, C. C., Helliwell, C. A., Stoutjesdijk, P., Dennis, E. S., & Peacock, W. J. (2001). Control of flowering time by *FLC* orthologues in *Brassica napus*. *The Plant Journal*, 28(5), 545-553.



- Takáč, T., Křenek, P., Komis, G., Vadovič, P., Ovečka, M., Ohnoutková, L., Pechan, T., Kašpárek, P., Tichá, T., & Basheer, J. (2021). TALEN-based *HvMPK3* knock-out attenuates proteome and root hair phenotypic responses to flg22 in barley. *Frontiers in Plant Science*, 12.
- Takeda, S., & Matsuoka, M. (2008). Genetic approaches to crop improvement: responding to environmental and population changes. *Nature Reviews Genetics*, 9(6), 444-457.
- Tanaka, A., Shikazono, N., & Hase, Y. (2010). Studies on biological effects of ion beams on lethality, molecular nature of mutation, mutation rate, and spectrum of mutation phenotype for mutation breeding in higher plants. *Journal of radiation research*, 51(3), 223-233.
- Tang, S., Liu, D. X., Lu, S., Yu, L., Li, Y., Lin, S., Li, L., Du, Z., Liu, X., & Li, X. (2020). Development and screening of EMS mutants with altered seed oil content or fatty acid composition in *Brassica napus*. *The Plant Journal*.
- Textor, S., De Kraker, J.-W., Hause, B., Gershenzon, J., & Tokuhsa, J. G. (2007). MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. *Plant Physiology*, 144(1), 60-71.
- Thangstad, O., Iversen, T.-H., Slupphaug, G., & Bones, A. (1990). Immunocytochemical localization of myrosinase in *Brassica napus* L. *Planta*, 180(2), 245-248.
- Thomson, M. J., Singh, N., Dwiyanti, M. S., Wang, D. R., Wright, M. H., Perez, F. A., DeClerck, G., Chin, J. H., Malitic-Layaoen, G. A., Juanillas, V. M., Dilla-Ermita, C. J., Mauleon, R., Kretschmar, T., & McCouch, S. R. (2017). Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. *Rice*, 10(1), 40. doi:10.1186/s12284-017-0181-2
- Till, B. J., Cooper, J., Tai, T. H., Colowit, P., Greene, E. A., Henikoff, S., & Comai, L. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biology*, 7(1), 1-12.
- Till, B. J., Zerr, T., Comai, L., & Henikoff, S. (2006). A protocol for TILLING and Ecotilling in plants and animals. *Nature Protocols*, 1(5), 2465.
- Toroser, D., Wood, C., Griffiths, H., & Thomas, D. (1995). Glucosinolate biosynthesis in oilseed rape (*Brassica napus* L.): studies with 35SO<sub>2</sub>- 4 and glucosinolate precursors using oilseed rape pods and seeds. *Journal of Experimental Botany*, 46(7), 787-794.
- Tripathi, M., & Mishra, A. (2007). Glucosinolates in animal nutrition: A review. *Animal Feed Science and Technology*, 132(1-2), 1-27.
- Tripathi, M., Mishra, A., Mondal, D., Misra, A., Prasad, R., & Jakhmola, R. (2008). Caecal fermentation characteristics, blood composition and growth of rabbits on substitution of soya-bean meal by unconventional high-glucosinolate mustard (*Brassica juncea*) meal as protein supplement. *animal*, 2(2), 207-215.
- Triques, K., Sturbois, B., Gallais, S., Dalmais, M., Chauvin, S., Clepet, C., Aubourg, S., Rameau, C., Caboche, M., & Bendahmane, A. (2007). Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea. *The Plant Journal*, 51(6), 1116-1125.
- Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K. J., Lieberman, M., Fass, J., Uauy, C., & Tran, R. K. (2011). Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiology*, 156(3), 1257-1268.

- Ulukapi, K., & Nasircilar, A. G. (2015). *Developments of gamma ray application on mutation breeding studies in recent years*. Paper presented at the International conference on advances in agricultural, biological & environmental sciences (AABES-2015), London, United Kingdom.
- Vageeshbabu, H. S., & Chopra, V. L. (1997). Genetic and Biotechnological Approaches for Reducing Glucosinolates from Rapeseed-Mustard Meal. *Journal of Plant Biochemistry and Biotechnology*, 6(2), 53-62. doi:10.1007/BF03263011
- Velasco, P., Soengas, P., Vilar, M., Cartea, M. E., & del Rio, M. (2008). Comparison of glucosinolate profiles in leaf and seed tissues of different *Brassica napus* crops. *Journal of the American Society for Horticultural Science*, 133(4), 551-558.
- Verkerk, R., Schreiner, M., Krumbein, A., Ciska, E., Holst, B., Rowland, I., De Schrijver, R., Hansen, M., Gerhäuser, C., & Mithen, R. (2009). Glucosinolates in Brassica vegetables: the influence of the food supply chain on intake, bioavailability and human health. *Molecular nutrition & food research*, 53(S2), S219-S219.
- Vermorel, M., Heaney, R. K., & Fenwick, G. R. (1986). Nutritive value of rapeseed meal: effects of individual glucosinolates. *Journal of the Science of Food and Agriculture*, 37(12), 1197-1202.
- Wang, B., Wu, Z., Li, Z., Zhang, Q., Hu, J., Xiao, Y., Cai, D., Wu, J., King, G. J., & Li, H. (2018). Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. *Plant Biotechnology Journal*, 16(7), 1336-1348.
- Wang, N., Li, F., Chen, B., Xu, K., Yan, G., Qiao, J., Li, J., Gao, G., Bancroft, I., & Meng, J. (2014). Genome-wide investigation of genetic changes during modern breeding of *Brassica napus*. *Theoretical and Applied Genetics*, 127(8), 1817-1829.
- Wang, N., Wang, Y., Tian, F., King, G. J., Zhang, C., Long, Y., Shi, L., & Meng, J. (2008). A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of *FAEI* point mutations by TILLING. *New Phytologist*, 180(4), 751-765.
- Wang, Y., Zong, Y., & Gao, C. (2017). Targeted mutagenesis in hexaploid bread wheat using the TALEN and CRISPR/Cas systems. In *Wheat Biotechnology* (pp. 169-185): Springer.
- Wei, D., Cui, Y., Mei, J., Qian, L., Lu, K., Wang, Z.-M., Li, J., Tang, Q., & Qian, W. (2019). Genome-wide identification of loci affecting seed glucosinolate contents in *Brassica napus* L. *Journal of Integrative Plant Biology*, 61(5), 611-623. doi:10.1111/jipb.12717
- Wells, R., Trick, M., Soumpourou, E., Clissold, L., Morgan, C., Werner, P., Gibbard, C., Clarke, M., Jennaway, R., & Bancroft, I. (2014). The control of seed oil polyunsaturate content in the polyploid crop species *Brassica napus*. *Molecular Breeding*, 33(2), 349-362.
- Wickham, H. (2009). *Elegant graphics for data analysis* (Vol. 35).
- Wiesner, M., Schreiner, M., & Zrenner, R. (2014). Functional identification of genes responsible for the biosynthesis of 1-methoxy-indol-3-ylmethyl-glucosinolate in *Brassica rapa* ssp. *chinensis*. *BMC Plant Biology*, 14(1), 124. doi:10.1186/1471-2229-14-124
- Wight, P., Scougall, R., Shannon, D., Wells, J., & Mawson, R. (1987). Role of glucosinolates in the causation of liver haemorrhages in laying hens fed water-extracted or heat-treated rapeseed cakes. *Research in veterinary science*, 43(3), 313-319.

- Winde, I., & Wittstock, U. (2011). Insect herbivore counteradaptations to the plant glucosinolate–myrosinase system. *Phytochemistry*, 72(13), 1566-1575.
- Wittkop, B., Snowden, R., & Friedt, W. (2009). Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. *Euphytica*, 170(1), 131-140.
- Wittstock, U., & Burow, M. (2010). Glucosinolate breakdown in *Arabidopsis*: mechanism, regulation and biological significance. *The Arabidopsis book/American Society of Plant Biologists*, 8.
- Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., & Wang, H. (2019). Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Molecular Plant*, 12(1), 30-43.
- Xu, L., Najeeb, U., Tang, G., Gu, H., Zhang, G., He, Y., & Zhou, W. (2007). Haploid and doubled haploid technology. *Advances in botanical research*, 45, 181-216.
- Yan, T., Yao, Y., Wu, D., & Jiang, L. (2021). BnaGVD: A Genomic Variation Database of Rapeseed (*Brassica napus*). *Plant and Cell Physiology*, 62(2), 378-383. doi:10.1093/pcp/pcaa169
- Yan, X., & Chen, S. (2007). Regulation of plant glucosinolate metabolism. *Planta*, 226(6), 1343-1352.
- Yang, B., Wen, X., Kodali, N. S., Oleykowski, C. A., Miller, C. G., Kulinski, J., Besack, D., Yeung, J. A., Kowalski, D., & Yeung, A. T. (2000). Purification, cloning, and characterization of the CEL I nuclease. *Biochemistry*, 39(13), 3533-3541.
- Yang, J., Wang, J., Li, Z., Li, X., He, Z., Zhang, L., Sha, T., Lyu, X., Chen, S., & Gu, Y. (2021). Genomic signatures of vegetable and oilseed allopolyploid *Brassica juncea* and genetic loci controlling the accumulation of glucosinolates. *Plant Biotechnology Journal*.
- Yasumoto, S., Matsuzaki, M., Hirokane, H., & Okada, K. (2010). Glucosinolate content in rapeseed in relation to suppression of subsequent crop. *Plant Production Science*, 13(2), 150-155.
- Yasumoto, S., Umemoto, N., Lee, H. J., Nakayasu, M., Sawai, S., Sakuma, T., Yamamoto, T., Mizutani, M., Saito, K., & Muranaka, T. (2019). Efficient genome engineering using Platinum TALEN in potato. *Plant Biotechnology*, 19.0805 a.
- Yin, L., Chen, H., Cao, B., Lei, J., & Chen, G. (2017). Molecular characterization of MYB28 involved in aliphatic glucosinolate biosynthesis in Chinese Kale (*Brassica oleracea* var. *alboglabra* Bailey). *Frontiers in Plant Science*, 8, 1083.
- Zerr, T., & Henikoff, S. (2005). Automated band mapping in electrophoretic gel images using background information. *Nucleic Acids Research*, 33(9), 2806-2812.
- Zheng, M., Zhang, L., Tang, M., Liu, J., Liu, H., Yang, H., Fan, S., Terzaghi, W., Wang, H., & Hua, W. (2020). Knockout of two *BnaMAX1* homologs by CRISPR/Cas9-targeted mutagenesis improves plant architecture and increases yield in rapeseed (*Brassica napus* L.). *Plant Biotechnology Journal*, 18(3), 644-654.

## References

- Zou, J., Zhu, J., Huang, S., Tian, E., Xiao, Y., Fu, D., Tu, J., Fu, T., & Meng, J. (2010). Broadening the avenue of intersubgenomic heterosis in oilseed *Brassica*. *Theoretical and Applied Genetics*, 120(2), 283-290.
- Zuluaga, D. L., Graham, N. S., Klinder, A., van Ommen Kloeke, A. E., Marcotrigiano, A. R., Wagstaff, C., Verkerk, R., Sonnante, G., & Aarts, M. G. (2019). Overexpression of the MYB29 transcription factor affects aliphatic glucosinolate synthesis in *Brassica oleracea*. *Plant Molecular Biology*, 101(1), 65-79.

## 10 Supplementary data on CD/DVD

The following supplemental data are available on a DVD and can be distributed upon request (contact: Prof. Dr. Christian Jung, [c.jung@plantbreeding.uni-kiel.de](mailto:c.jung@plantbreeding.uni-kiel.de)).

File name	Content	Format
TbyWGS dataset	Summary statistics of all SNP counts and mutation effects from whole-genome sequenced 4x pools (Chapter 2)	.xlsx
Gene sequences	Annotated gene sequences of <i>BnMYB28</i> , <i>BnCYP79F1</i> and <i>BnGTR2</i>	.clc
PAGE images	Polyacrylamide gel images for all TILLING mutants (Chapter 3 and Chapter 4)	.zip and .tif
Seed codes	Seed codes of the produced plant material	.xlsx
Brassica 19K SNP array	Raw data from <i>Brassica</i> 19K SNP array for: (1) Batch_1: <i>BnCYP79F1</i> BC <sub>1</sub> mutants, (2) Batch_2: <i>BnMYB28</i> BC <sub>1</sub> mutants and (3) Batch_3: <i>BnGTR2</i> BC <sub>1</sub> mutants	.xlsx
GSL determination	GOPOD and HPLC measurements from all analyzed <i>BnMYB28</i> and <i>BnCYP79F1</i> EMS mutants	.xlsx

## 11 Curriculum Vitae and Publications

### 11.1 Curriculum vitae

#### Personal Data

---

Name:	Srijan Jhingan
Date of birth:	June 1 <sup>st</sup> , 1994
Place of birth:	Delhi, India
Nationality:	Indian

#### Education

---

April 2018 - present	<b>Ph.D. student at the Plant Breeding Institute, Christian-Albrechts-University of Kiel, Germany</b>  Doctoral thesis: Functional analysis and mutagenesis of glucosinolate synthesis genes for breeding oilseed rape ( <i>Brassica napus</i> ) with lower glucosinolate content  Supervisors: Dr. Hans-Joachim Harloff and Prof. Dr. Christian Jung
October 2015 - March 2018	<b>Master of Science in AgriGenomics, Plant Breeding Institute, Christian-Albrechts-University of Kiel, Germany</b>  Masters thesis: QTL mapping of the bolting locus <i>B6</i> responsible for the biennial growth habit in leaf beet ( <i>Beta vulgaris</i> )  Supervisors: Dr. Nadine Dally, Prof. Dr. Christian Jung
July 2012 - June 2015	<b>Bachelor of Science in Botany (Honors), Sri Guru Tegh Bahadur Khalsa College, University of Delhi, India</b>

---

### 11.2 Oral and poster presentations

**Jhingan S.**, Harloff H-J., and Jung C. (2019). Breeding oilseed rape (*B. napus*) with lower glucosinolate content through functional analysis and mutagenesis. GPZ Conference Genome Research 2019, 28-29 March 2019, Hohenheim, Germany. Poster presentation

**Jhingan S.**, Harloff H-J., and Jung C. (2019). Breeding oilseed rape (*B. napus*) with lower glucosinolate content through functional analysis and mutagenesis. 15th International Rapeseed Congress, 16-19 June, 2019, Berlin, Germany. Poster presentation

**Jhingan S.**, Harloff H-J., and Jung C. (2019). Breeding oilseed rape (*B. napus*) with lower glucosinolate content through functional analysis and mutagenesis. Kiel Plant Center–Summer Symposium 2019, 22-23 July, 2019, Kiel, Germany. Poster presentation

**Jhingan S.**, Harloff H-J., Dreyer F., Abbadi A., and Jung C. (2020). Breeding oilseed rape (*B. napus*) with lower glucosinolate content through functional analysis and mutagenesis.

International Symposium of the Society for Plant Breeding (GPZ), 11-13 February, 2020, Tulln, Austria. Poster presentation

**Jhingan S.**, Harloff H-J., Dreyer F., Abbadi A., and Jung C. (2021). Breeding oilseed rape (*B. napus*) with lower glucosinolate content through functional analysis and mutagenesis. PLANT 2030 Status Seminar 2021, 10-11 March 2021. Online oral and poster presentation

## **12 Publications and declaration of own contribution**

Chapters 2, 3 and 4 include the work of doctoral student Srijan Jhingan alone.



## 13 Acknowledgements

First and foremost, I would like to thank Prof. Dr. Christian Jung for providing me with the opportunity and a platform to work and develop as a student and as a scientist in the Plant Breeding Institute. I am thankful for your critical suggestions and discussions that have directed the course of this project. I appreciate all the support and resources provided by you for the successful execution of this study and admire the high standard of scientific research set by you. This has always inspired me to push harder and excel at work and beyond.

I extend my deepest regards and gratitude to Dr. Hans-Joachim Harloff for his unconditional support and guidance during these past years of my life. You have been a source of inspiration and I am ever grateful for your invaluable support and guidance during this project. I thank you for your patience that was probably tested while indulging in countless discussions and solving problems I brought to you. I believe I have developed as a scientist, imbibing strong scientific morals and ethics under your supervision. As my supervisor, mentor and friend, I thank you for your support during all the highs and lows.

Special thanks to all the colleagues of the Plant Breeding Institute for having provided me with a very intellectually stimulating and nurturing work environment. You have helped me grow not just as a scientist but also as a human being. I have learned so much from each one of you. My heart goes out to Amar, Avneesh, Dilan, Nirosha, Sarah, Tahmina and Yixin for being my family in Germany. I am thankful to Moni and Sigi for being a crucial part of my support system. I think I am lucky to have enjoyed countless valuable moments that I have had the privilege to share and cherish with you. All of you have made Kiel feel like home away from “home”.

Above all, I would like to extend special thanks to Ishani for being the wind beneath my sails. You have kept me motivated, determined and grounded during the thick and thin of not only my Ph.D. journey but also in life. Words cannot explain the positivity and joy that you have brought into my life. Thank you for being what you are to me and being by my side when the going got tough.

Last but not least, I express my deepest gratitude to my parents for their boundless and unconditional love and support. I thank them for have always motivated me to pursue my goals and ambitions. They have always been and will continue to be the strongest pillars in my life.

Finally, I would like to thank the Bundesministerium für Bildung und Forschung (BMBF) for funding this work within the framework of the project IRFFA: Improved Rapeseed as Fish Feed in Aquaculture (grant number 031B0357B). I thank the Norddeutsche Pflanzenzucht Innovation (NPZi) GmbH for coordinating the IRFFA project and providing me with the mutant seed material crucial for the execution of the project. Additionally, I am grateful for the Sanger sequencing service provided by the Institute of Clinical Molecular Biology Kiel (IKMB). I also acknowledge the assistance provided by Jens Hermann and Prof. Dr. Wolfgang Bilger from the Department of Ecophysiology of Plants, CAU-Kiel for HPLC analytics.