Epigenetic regulation of glucosinolate biosynthesis and the contribution of regulatory variation to interspecific diversity in glucosinolate profiles

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Simon Mitreiter

aus Bonn

Köln, 2021

Berichterstatter:	PD Dr. Tamara Gigolashvili
	Prof. Dr. Ute Höcker
Prüfungsvorsitzende:	Prof. Dr. Karin Schnetz
Tag der mündlichen Prüfung:	20.04.2021

Abstract

Plants of the order Brassicales produce glucosinolates (GSLs) as a biochemical defense against herbivores and pathogens. If the plant is attacked, these secondary metabolites are activated into toxic compounds such as thiocyanates, isothiocyanates and nitriles. Both intact GSLs and their degradation products influence plant-environment interactions in numerous ways. The regulation of GSL biosynthesis is known to be controlled by a complex consisting of MYB and bHLH transcription factors. Under stress conditions, in particular upon increased concentrations of the phytohormone jasmonate (JA), GSL biosynthesis is upregulated. Such inducible defenses are subject to the priming response, in which their output becomes more robust in response to repeated stimuli. The priming response is generally considered to be mediated by epigenetic mechanisms, but to date, no specific epigenetic regulation of GSL synthesis has been described. In this thesis, I aimed at studying the role of EMLs, a family of histone reader proteins, in this process. Via metabolite and transcript analysis of *Arabidopsis thaliana* (Arabidopsis) *eml* mutants, I could show that EMLs likely participate in the JA-mediated induction of GSL synthesis and the accompanying epigenetic shift. Additionally, EMLs were found to interact with bHLHs via colocalization and overexpression studies. This suggests that EMLs directly participate in the MYB-bHLH complex to epigenetically regulate GSL production.

Furthermore, a great diversity of GSL structures exists among the plants of the Brassicaceae family, with every species producing their own distinct GSL profiles. While certain interspecific differences have been attributed to particular alleles of GSL biosynthetic enzymes in the past, a major role of regulatory genes has also been postulated in giving rise to varied GSL profiles. This hypothesis has not been specifically studied to date, however. In this thesis, *MYB* and *bHLH* genes in Arabidopsis, *Arabis alpina*, *Cardamine hirsuta* and *Capsella rubella* were analyzed in regard to their phylogenetic relationship and their functional conservation. The results suggest that these genes can mainly contribute to GSL diversity in a broad fashion, affecting entire branches of GSL biosynthesis. GSL profiles of root and shoot tissues of the three non-Arabidopsis species were determined by liquid chromatography-mass spectrometry (LC-MS), revealing that *C. rubella* almost completely lacks indolic GSLs. The presence of numerous non-functional *bHLH* splice variants in this species was demonstrated, with overexpression studies in *C. rubella* cell culture showing that, together with low *MYB* expression, this alternative splicing contributes to the low indolic GSL content. Taken together, this thesis highlights how Brassicaceae plants regulate GSL biosynthesis via complex and multi-layered mechanisms, both on an individual and species-wide level.

Zusammenfassung

Pflanzen der Ordnung Brassicales (Kreuzblütlerartige) produzieren Glucosinolate (GSLs) als biochemische Verteidigung gegen Herbivoren und Pathogene. Wird die Pflanze angegriffen, werden diese Sekundärmetaboliten zu toxischen Verbindungen wie Thiocyanaten, Isothiocyanaten und Nitrilen aktiviert. Sowohl intakte GSLs als auch deren Zerfallsprodukte beeinflussen die Wechselwirkungen zwischen Pflanze und Umwelt auf vielfältige Weise. Die Regulation der GSL-Biosynthese wird bekanntermaßen durch einen aus MYB- und bHLH-Transkriptionsfaktoren (TFs) bestehenden Komplex kontrolliert. Unter Stressbedingungen, insbesondere einer erhöhten Konzentration des Phytohormons Jasmonat (JA), wird die GSL-Biosynthese hochreguliert. Solche Formen der induzierbaren Verteidigung unterliegen der Primingantwort, wobei die Reaktion durch wiederholte Stimuli robuster wird. Es wird davon ausgegangen, dass die Primingantwort durch epigenetische Mechanismen vermittelt wird, es wurde jedoch bis dato noch keine spezifische epigenetische Regulation der GSL-Synthese beschrieben. In dieser Arbeit zielte ich darauf ab, die Rolle von EMLs, einer Familie von Histon-Leseproteinen, in diesem Prozess zu untersuchen. Durch die Analyse von Metaboliten und Transkripten in Arabidopsis-*eml*-Mutanten konnte ich zeigen, dass EMLs wahrscheinlich an der JA-vermittelten Induktion der GSL-Synthese und der sie begleitenden epigenetischen Veränderung beteiligt sind. Außerdem konnte ich mittels Kolokalisations- und Überexpressionsversuchen feststellen, dass EMLs mit bHLHs interagieren. Dies legt nahe, dass EMLs unmittelbar mit dem MYB-bHLH-Komplex wechselwirken, um die GSL-Produktion epigenetisch zu regulieren.

Des Weiteren existiert eine große Vielfalt an GSL-Strukturen in der Pflanzenfamilie der Brassicaceen (Kreuzblütler), wobei jede Art ihre eigenen, unterschiedlichen GSL-Profile produziert. Während bestimmte zwischenartliche Unterschiede bereits bestimmten Allelen von Enzymen der GSL-Biosynthese zugeschrieben werden konnten, wurde darüber hinaus eine bedeutende Rolle von regulatorischen Genen in der Erzeugung von diversen GSL-Profilen postuliert. Diese Hypothese wurde bislang jedoch nicht direkt untersucht. In dieser Arbeit wurden *MYB*- und *bHLH*-Gene in Arabidopsis, *Arabis alpina*, *Cardamine hirsuta* und *Capsella rubella* in Hinblick auf ihre phylogenetische Verwandtschaft und ihre konservierte Funktion analysiert. Die Ergebnisse legen nahe, dass diese Gene hauptsächlich in einer breit angelegten Weise, die ganze Zweige der GSL-Biosynthese betrifft, zur GSL-Vielfalt beitragen können. GSL-Profile von Wurzel- und Sprossgewebe der drei Nicht-Arabidopsis-Arten wurden durch Flüssigkeitschromatographie-Massenspektrometrie (LC-MS) bestimmt, wobei eine fast vollständige Abwesenheit von indolischen GSLs in *C. rubella* zutage trat. Das Vorhandensein von zahlreichen nicht-funktionellen *bHLH*-Spleißvarianten wurde nachgewiesen, wobei Überexpressionsversuche in einer Zellkultur von *C. rubella* zeigten, dass dieses alternative Spleißen zusammen mit niedriger Expression von *MYB*-Genen zum niedrigen Gehalt an indolischen GSLs beiträgt. Zusammengenommen stellt diese Arbeit heraus, wie Brassicaceen die GSL-Biosynthese sowohl auf individueller als auch auf Artebene durch komplexe und mehrschichtige Mechanismen regulieren.

Table of Contents

1 Introduction[1]

1.1 Glucosinolates and their biochemistry

Glucosinolates (GSLs) are a small, but prominent family of secondary metabolites in the plant order Brassicales. GSLs not only shape the interaction between plants and their biotic and abiotic environment, but are also of considerable human interest due to their agronomic and nutritional properties. GSLs can be useful in agriculture when providing resistance against pests and pathogens, either in live crop plants or in biofumigation (the incorporation of plant material into the soil). However, too high GSL concentrations can be harmful in animal feed (Halkier and Gershenzon, 2006; Morris *et al.*, 2020).

Moreover, GSLs and their degradation products provide the typical pungent flavors of vegetables and spices such as cabbage, mustard or horseradish (Bell *et al.*, 2018; Wieczorek *et al.*, 2018). As part of the human diet, GSLs have anticarcinogenic effects, inducing detoxification enzymes that break down dietary carcinogens and suppressing tumor growth by arresting the cell cycle and promoting apoptosis in cancer cells (Ishida *et al.*, 2014; Katz *et al.*, 2018).

Structurally, GSLs are comprised of a sulfonated oxime group that is linked to a thioglucose group, and an amino acid-derived side chain (**Figure 1**). All GSL molecules can be classified into three categories, according to the amino acid from which they derive: aliphatic (from alanine, valine, leucine, isoleucine or methionine), benzenic (from phenylalanine or tyrosine) or indolic (from tryptophan).

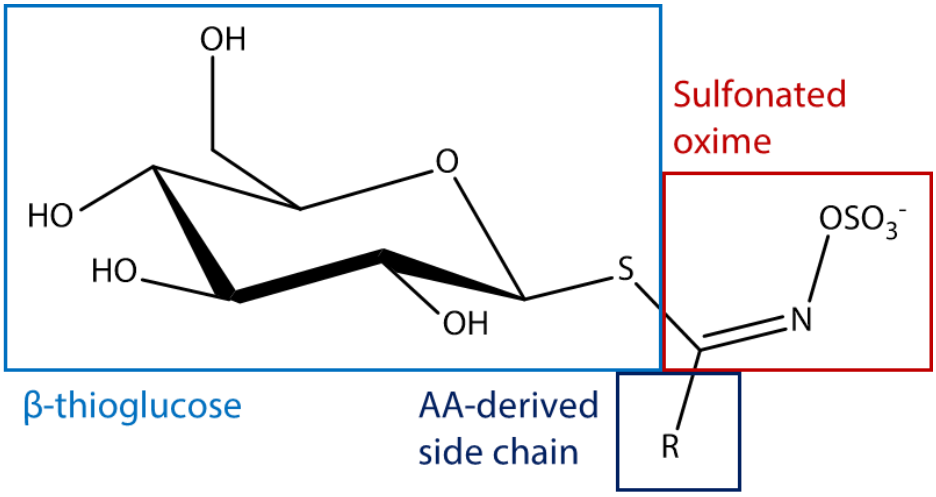


Figure 1: General structure of a GSL molecule.
The sulfonated oxime group is marked in red, the β-thioglucose group in light blue and the amino acid-derived side chain in dark blue. GSLs can be classified according to which amino acid the side chain is derived from.

[1]This introduction partially overlaps with a review in which I discuss the regulation of GSL biosynthesis. It was published prior to this thesis (Mitreiter and Gigolashvili, 2020).

Starting from the respective amino acid, the biosynthesis pathway proceeds through three phases (**Figure 2**). First, the carbon chain of non-tryptophan amino acids can be elongated in a reaction sequence of MAM, IPMI and IPMDH enzymes through which the molecules are cycled up to six or more times. Second, the GSL core structure is formed. The first step entails the conversion of amino acids into aldoximes and is catalyzed by the cytochrome P450 enzymes CYP79F1 and CYP79F2 for aliphatic amino acids, or respectively CYP79B2 and CYP79B3 for tryptophan. The aldoximes are then oxidized into nitrile oxides by CYP83A1 in the aliphatic and by CYP83B1 in the indolic pathway. This is followed by several further steps, some of which share the same enzymes between both pathways, until the final GSL structure is completed by a sulfation reaction (Sønderby *et al.*, 2010b).

Third, the GSL molecules can undergo a variety of secondary modifications, primarily at their side chains. Prominent examples include the hydroxylation and subsequent methoxylation of the main indolic GSL indole-3-methylglucosinolate (I3M) into 1- or 4-substituted derivates (4OH-I3M, 1MO-I3M, 4MO-I3M) by CYP81F and IGMT enzymes (Pfalz *et al.*, 2009; Pfalz *et al.*, 2011; Pfalz *et al.*, 2016) or the transformation of methylsulfinylalkyl into alkenyl and hydroxyalkyl GSLs by AOP2 and AOP3, respectively (Kliebenstein *et al.*, 2001; Neal *et al.*, 2010; Jensen *et al.*, 2015a). A number of other modifications including oxidation and benzoylation have been described. Not in all cases however, the responsible enzymes have been identified yet (Blažević *et al.*, 2020).

Notably, these pathways are only well-described for indolic and Met-derived GSLs. The non-methionine aliphatic and benzenic GSL synthesis pathways are thought to be ancestral and have been postulated from phylogenetic comparisons, with the non-Met pathway being predominantly co-opted by Met-GSLs in the Brassicaceae (Schranz *et al.*, 2011). The benzenic pathway has been proposed to partially overlap with the aliphatic one, but the precise mechanics remain to be elucidated (Wittstock and Halkier, 2000; Schranz *et al.*, 2011; Kittipol *et al.*, 2019).

Native GSLs are stable and possess little biological activity, having only been reported to act as feeding or oviposition cues for insects (Marazzi and Städler, 2004; Sun *et al.*, 2009). To fulfil most of their functions, they need to be activated by myrosinases, specific β-thioglucosidases that hydrolyze the glucose residue. Subsequently, the resulting aglycones can spontaneously rearrange into toxic isothiocyanates. However, the appropriate chemical conditions and the presence of specifier proteins promote the formation of other active compounds such as nitriles, thiocyanates, and epithionitriles (Wittstock and Burow, 2010; Wittstock *et al.*, 2016a).

The storage of GSLs and myrosinases is spatially distinct so that these two classes of compounds only come into contact following challenge to the plant. Specialized cell types can act as different storage locations: S-cells for the GSLs and myrosin cells for the "classical" myrosinases. GSL hydrolysis then occurs upon tissue disruption, for example by herbivory. Both components can also be differentially compartmentalized within

a single cell, where GSLs are localized to vacuoles and "atypical" myrosinases to endoplasmic reticulum bodies or the cytosol. Here, GSL hydrolysis occurs as part of the immune response (Wittstock and Burow, 2010; Wittstock *et al.*, 2016a; Sugiyama and Hirai, 2019).

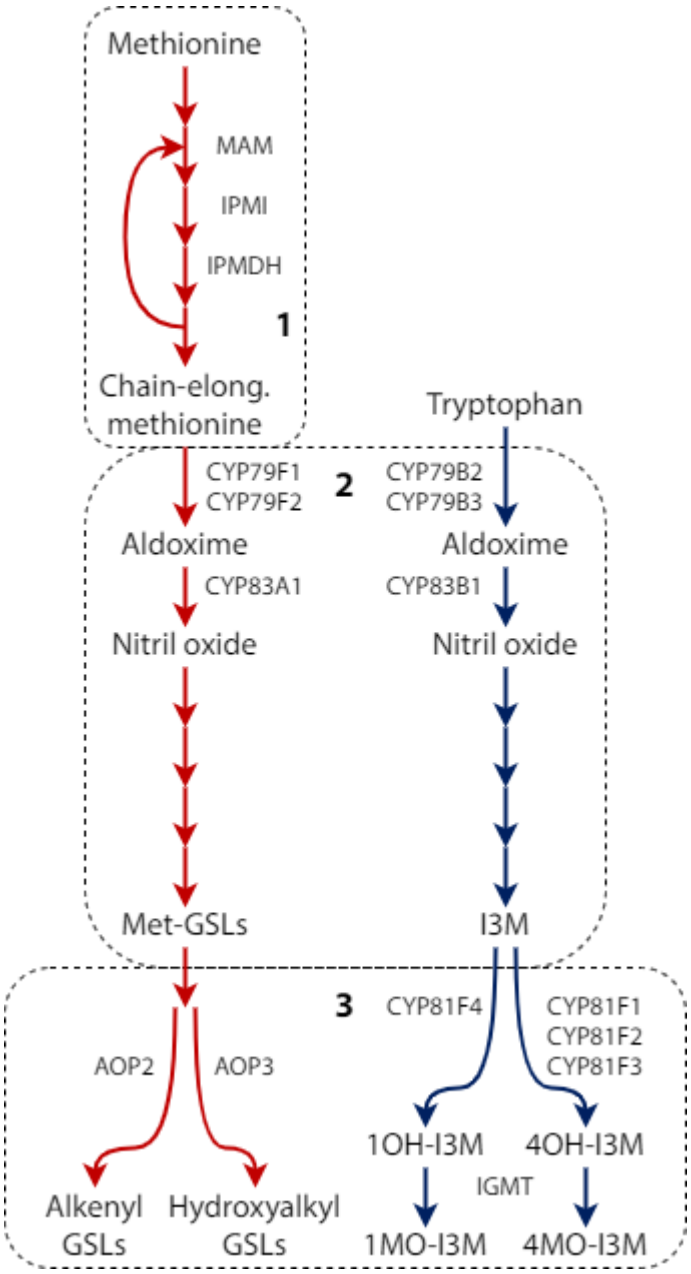


Figure 2: Simplified scheme for the biosynthesis of methionine- and tryptophan-derived GSLs.
Each arrow signifies one enzymatic reaction, with red arrows indicating the aliphatic methionine-derived pathway and blue arrows indicating the indolic tryptophan-derived pathway. Selected intermediates, enzymes or enzyme families are named; some of the unnamed enzymes are shared between the two pathways. Numbers indicate the three stages of GSL synthesis: 1, amino acid chain elongation; 2, core structure synthesis; 3, secondary side-chain modifications. Only selected secondary modifications are shown.

1.2 Regulation of glucosinolate biosynthesis by MYB and bHLH TFs

A tight regulation of GSL production is essential for Brassicales plants, primarily because of their role in shaping plant-environment interactions. GSL synthesis requires significant metabolic investments, such as increased photosynthetic requirements estimated at least 15% (Bekaert *et al.*, 2012) and a considerable proportion of the sulfur supply (Maruyama-Nakashita, 2017). However, the final cost-benefit calculation of which amounts and types of GSLs are produced is extremely complex and shaped by interactions with intra- and interspecific competitors, generalist and specialist herbivores, bacterial and fungal pathogens as well as mutualist pollinators (Mitreiter and Gigolashvili, 2020). In the wild, this ever-changing web of interconnected relationships leads to disruptive selection, i.e. the maintenance of diverse GSL profiles (Burow *et al.*, 2010; Züst *et al.*, 2012; Kerwin *et al.*, 2015).

The central regulatory mechanism of GSL production is the transcriptional control of biosynthesis gene expression by MYB and bHLH transcription factors (TFs). They constitute the key regulators that integrate diverse regulatory information via their mutual interactions and binding to gene promoters.

1.2.1 Subgroup 12 MYB factors

The most specific transcriptional regulators involved in GSL synthesis are MYB TFs from the R2R3 subfamily. This group of MYBs is characterized by two N-terminal MYB repeats, each containing three α-helices. The third helices of each of the two repeats coordinately mediate DNA binding to a range of specific sequences (Kelemen *et al.*, 2015). Additionally, many R2R3 MYBs have a disordered transcriptional activation domain (TAD) at their C-terminus (Feller *et al.*, 2011). The R2R3 MYBs are further divided into subgroups according to sequence similarity, which at least partly correspond with functional conservation (Kranz *et al.*, 1998; Stracke *et al.*, 2001; Dubos *et al.*, 2010). The six members of subgroup 12 positively regulate genes responsible for GSL biosynthesis and for connected biochemical reactions, such as amino acid biosynthesis and sulfur assimilation.

Three MYBs are central to either aliphatic or indolic GSL synthesis, respectively: The aliphatic clade consists of MYB28 (Gigolashvili *et al.*, 2007b; Hirai *et al.*, 2007; Sønderby *et al.*, 2007), MYB29 (Hirai *et al.*, 2007; Sønderby *et al.*, 2007; Gigolashvili *et al.*, 2008) and MYB76 (Sønderby *et al.*, 2007; Gigolashvili *et al.*, 2008), whereas the indolic clade consists of MYB34 (Celenza *et al.*, 2005), MYB51 (Gigolashvili *et al.*, 2007a) and MYB122 (Gigolashvili *et al.*, 2007a).

Indolic and aliphatic MYBs are partially redundant in their respective functions and can maintain GSL synthesis in the absence of the other clade members. MYB28 and MYB29 are key regulators of aliphatic GSL biosynthesis and MYB51 and MYB34 of indolic GSL biosynthesis, respectively. If both of these main factors are absent within one clade, the synthesis of aliphatic (Sønderby *et al.*, 2007; Beekwilder *et al.*, 2008; Sønderby *et al.*, 2010a) respectively indolic GSLs (Frerigmann and Gigolashvili, 2014) is almost completely abolished.

The aliphatic and indolic clades have been found to act antagonistically towards each other, exerting a type of reciprocal negative control to maintain homeostasis between aliphatic and indolic GSL production, either in the form of repression when one type of GSL overaccumulates, or as compensation when the accumulation of one type decreases (Gigolashvili *et al.*, 2008; Gigolashvili *et al.*, 2009; Sønderby *et al.*, 2010a).

1.2.2 Subgroup IIIe bHLH factors

Similar to the subgroup 12 MYB TFs, some members of the bHLH TF family are also essential for GSL biosynthesis. The bHLH proteins are characterized by their eponymous basic helix-loop-helix domain, which encompasses a stretch of 15 to 20 mostly basic amino acids, followed by two α-helices connected by a loop of variable length (Carretero-Paulet *et al.*, 2010). This highly conserved structure mediates homo- and heterodimerization, as well as DNA binding to E-box sequences (CANNTG), one of which is the G-box (CACGTG) (Fernández-Calvo *et al.*, 2011).

According to their phylogenetic relationships, bHLHs can be categorized into 25 to 32 different subgroups, depending on the exact criteria used (Heim *et al.*, 2003; Carretero-Paulet *et al.*, 2010; Pires and Dolan, 2010). Among these, subgroup IIIe contains the four closely related TFs bHLH04/MYC4, bHLH05/MYC3, bHLH06/MYC2 and bHLH28/MYC5. All of these interact in a complex with subgroup 12 MYBs and directly regulate GSL biosynthesis, albeit the most distantly related member bHLH28 only in a minor fashion (Schweizer *et al.*, 2013; Frerigmann *et al.*, 2014). The individual dimerization properties of each bHLH influence its respective function (Feller *et al.*, 2011). This is exemplified by bHLH06 which can – in contrast to other subgroup IIIe bHLHs that only homo- or heterodimerize – homotetramerize, thereby increasing its DNA binding affinity (Lian *et al.*, 2017).

In addition to the bHLH domain present in all bHLH TFs, the specific domain architecture of subgroup IIIe bHLHs is characterized by the N-terminal MYB-interacting region, also known as JAZ-interacting domain (JID) (Pattanaik *et al.*, 2008; Schweizer *et al.*, 2013). The JID facilitates interactions with MYB proteins via their MYC-interaction motif MIM, located outside of the MYB domain itself (Millard *et al.*, 2019). In the bHLHs, adjacent to the JID, resides a TAD that enables interaction with the Mediator protein MED25 for RNA polymerase II recruitment, thus conferring TF activity (Çevik *et al.*, 2012; Chen *et al.*, 2012). An ACT domain is located at the C-terminus, which has been shown in related bHLH factors from anthocyanin biosynthesis to mediate dimerization and has also been proposed to function in ligand binding (Feller *et al.*, 2006; Feller *et al.*, 2011).

The roles of subgroup IIIe bHLHs are not limited to GSL biosynthesis, but encompass numerous other functions that include developmental processes, and biotic and abiotic stress responses (Song *et al.*, 2017). The specificity in these distinct functions is mediated via interaction with other TFs: For example, while subgroup 12 MYBs are responsible for GSL synthesis, the role of subgroup IIIe bHLHs in stamen development is executed in interaction with subgroup 19 MYBs (Qi *et al.*, 2015).

1.2.3 The MYB-bHLH transcription complex

Subgroup 12 MYB and subgroup IIIe bHLH TFs together form a complex that is required for the direct transcriptional regulation and essential for the biosynthesis of all GSLs present in Arabidopsis. Consistent with the principle that MYBs are crucial for specific gene activation, whereas bHLHs integrate hierarchically higher-order signals and relay this information (Pireyre and Burow, 2015; Chezem and Clay, 2016; Brkljacic and Grotewold, 2017), all subgroup 12 MYBs can interact with all subgroup IIIe bHLHs, and the type of GSL synthesized (aliphatic or indolic) is solely dependent on which MYBs are present in the transcriptional complex.

A specific feature that has often been observed for the GSL-regulating MYB–bHLH complex is the uncoupling of *MYB* transcript levels and GSL metabolite levels, which highlights a more complex regulation (Sønderby *et al.*, 2007; Sønderby *et al.*, 2010a; Justen and Fritz, 2013; Frerigmann and Gigolashvili, 2014; Seo *et al.*, 2016; Seo *et al.*, 2017). The precise mechanisms at work here remain unclear and might involve concentration-dependent metabolic feedback or further transcriptional regulation, either independently from the MYB–bHLH complex or as part of it.

There is a limited number of TFs that have been described to regulate GSL biosynthesis by interfering with MYB–bHLH function. For example, the brassinosteroid-dependent TF BES1 binds to subgroup 12 MYBs to competitively attenuate bHLH function (Liao *et al.*, 2020), while SDI1 directly interacts with MYB28 and inhibits its transcriptional activity to limit GSL production under sulfur starvation (Aarabi *et al.*, 2016). Some proteins have also been reported to influence GSL accumulation independently of the MYB-bHLH complex, such as the IQ67 domain protein IQD1 (Levy *et al.*, 2005) and the DOF transcription factor OBP2 (Skirycz *et al.*, 2006). These regulators may directly induce GSL synthesis in response to wounding and herbivory.

1.2.4 Hormonal influences on the MYB-bHLH transcription complex

The function of the MYB-bHLH complex is strongly influenced by various phytohormones. GSL production is also responsive to many abiotic factors; this has been proposed to confer a precociously increased defense against potential attacks under harsh conditions, but also a direct alleviation of stress conditions mediated by signaling properties of GSL hydrolysis products (Del Carmen Martínez-Ballesta *et al.*, 2013; Salehin *et al.*, 2019). However, these abiotic influences seem to be generally mediated by the established hormonal and transcriptional factors.

The plant hormones reported to influence GSL levels include jasmonates (JA), salicylic acid (SA), ethylene (ET), abscisic acid (ABA) and brassinosteroids (BR). Out of these, BR seems to be the major negative regulator, while the others generally induce GSL accumulation. Their relationship can be described in a simplified model in which SA mediates the response to biotrophic pathogens, ET and JA to necrotrophic pathogens, and ABA and JA to herbivores. JA, SA, ET and ABA induce GSL production, whereby SA can

have a negative effect in high concentrations (Guo *et al.*, 2013b). The signaling mechanisms of all mentioned hormones influence each other in a highly complex network (Pieterse *et al.*, 2012; Verma *et al.*, 2016; Li *et al.*, 2019).

Of all hormones, the strongest and best-described regulators are the JAs, lipid-derived molecules that are essential for the integration of diverse input and output signals with respect to developmental processes and stress responses (Howe *et al.*, 2018). In contrast to the other hormones, they have the ability to profoundly increase GSL concentrations, up to a several-fold multiplication (Mikkelsen *et al.*, 2003; Guo *et al.*, 2013b; Frerigmann and Gigolashvili, 2014; Thiruvengadam *et al.*, 2016; Miao *et al.*, 2017). It has been well documented that the synthesis of indolic GSLs is more strongly induced by JA than that of aliphatic GSLs (van Dam *et al.*, 2009; Augustine and Bisht, 2015; Zang *et al.*, 2015).

Besides the direct transcriptional control of *MYB* and *bHLH* genes, JA induction of GSL synthesis occurs via the MYC and ERF branches of the JA signaling pathway. While the ERF branch uses other components, the MYC branch directly targets MYC/bHLH TFs: In the absence of JA, JAZ repressors bind via their Jas domain to the JID and TAD domains of the bHLHs. Simultaneously, they recruit the co-repressors TPL and TPR, either directly or via the NINJA adapter. This not only prevents DNA binding, but also hinders the recruitment of MYBs and MED25 to form an active transcriptional complex (Zhang *et al.*, 2015; Zhang *et al.*, 2017; Howe *et al.*, 2018).

When plants are attacked, specific molecular patterns associated with pathogens, herbivores or mechanical damage are perceived by plant pattern recognition receptors. The perception of herbivory or necrotrophic pathogens induces JA synthesis; however, the downstream mechanisms that mediate this process are incompletely understood (Zhang *et al.*, 2017). The enzyme JAR1 conjugates jasmonic acid with isoleucine to produce the most bioactive jasmonate, JA-Ile (Fonseca *et al.*, 2009). The presence of JA-Ile then promotes the binding of the JAZ proteins to the E3 ubiquitin ligase complex SCF^{COI1} (a process directly promoted by MED25), resulting in their ubiquitination and degradation by the 26S proteasome (Chini *et al.*, 2007; Thines *et al.*, 2007; An *et al.*, 2017; Howe *et al.*, 2018). Thus, the equilibrium of bHLH binding is shifted away from interaction with the JAZ repressors towards the formation of functional MYB-bHLH complexes and the transcription of their target genes (Goossens *et al.*, 2017).

Whereas all subgroup IIIe bHLHs are regulated by JAZ proteins (Chini *et al.*, 2007; Fernández-Calvo *et al.*, 2011; Niu *et al.*, 2011; Qi *et al.*, 2015), this has not been demonstrated for subgroup 12 MYBs yet. bHLH06 is particularly important for GSL synthesis in response to JA, whereas bHLH04 and bHLH05 play a more important role under non-stress conditions. However, JA responses are constitutively activated in a mutant in which the JAZ binding ability of bHLH05 is impaired (Frerigmann *et al.*, 2014). The transcription of *bHLH04, bHLH05* and *bHLH06*, but not *bHLH28,* is induced by JA (Song *et al.*, 2017).

The responsiveness of individual *MYBs* to JA and other elicitors is highly variable (Mitreiter and Gigolashvili, 2020). A differential regulation in response to herbivore and pathogen attack is also reflected in the type of GSL produced: ABA and JA induce elevated levels of the modified indolic GSL 1MO-I3M, whereas SA and ET promote the production of 4MO-I3M. This specific regulation is thought to function independently from MYB proteins and remains to be elucidated (Wiesner *et al.*, 2013; Frerigmann and Gigolashvili, 2014; Zang *et al.*, 2015).

1.3 Epigenetic regulation of glucosinolate biosynthesis

Beyond the classical transcriptional control mediated by *cis*- and *trans*-elements, the epigenetic code adds another level to the regulatory machinery of plants. Epigenetic regulation allows for both the realization of developmental programs via cell differentiation (He *et al.*, 2011; Xiao and Wagner, 2015; Cheng *et al.*, 2019) and a flexible short-term response to stress (Boyko and Kovalchuk, 2008; Kim *et al.*, 2015; Pandey *et al.*, 2016; Alonso *et al.*, 2019). In regard to stress-induced epigenetic changes, there is a complex interplay between the maintenance of altered genes states, allowing for mitotic and meiotic inheritance of epialleles even in the absence of the original signal, and their regulatory plasticity in response to environmental stimuli (Berr *et al.*, 2012; Cavalli and Heard, 2019).

Epigenetic information is conveyed through the chromatin state and the resulting accessibility of the DNA for binding factors. In Arabidopsis, nine distinct chromatin states have been described (Sequeira-Mendes *et al.*, 2014). Overall, the chromatin state can be characterized on three interacting levels: chromatin remodeling, DNA methylation and histone modification (Boyko and Kovalchuk, 2008). The fundamental unit of chromatin is the nucleosome. It is comprised of 147 bp of DNA wrapped around the core histone octamer, which in turn contains two copies each of the histone proteins H2A, H2B, H3 and H4. Further proteins, such as the linker histone H1, organize the nucleosome into higher-order structures (Luger *et al.*, 1997; Ding and Wang, 2015).

Chromatin remodeling encompasses the incorporation of histone variants such as H2A.Z or H3.3 and changes in nucleosome positioning (Chodavarapu *et al.*, 2010; Zhu *et al.*, 2012), which in turn alter the chromatin landscape by influencing long-range intra- and inter-chromosomal contacts and thus the 3D chromatin structure (Brkljacic and Grotewold, 2017; Cavalli and Heard, 2019). DNA methylation functions via the reversible methylation of cytosine (or sometimes adenine) nucleotides. The methylation status is controlled by distinct processes of *de novo* methylation, maintenance regulation and active demethylation, whereby *de novo* methylation is induced by a siRNA-mediated pathway called RNA-directed DNA methylation (Matzke and Mosher, 2014; Kim *et al.*, 2015; Pandey *et al.*, 2016).

Histone modifications or marks can consist of numerous different posttranslational modifications at the core proteins, including acetylation, methylation and ubiquitination (Kouzarides, 2007; Ding and Wang, 2015). Histone acetylation is associated with transcriptional activation (Eberharter and Becker, 2002), while

the function of methylation and ubiquitination depends on the targeted residues (Alvarez-Venegas, 2010; Ding and Wang, 2015). In plants, H3K4me (i.e. methylation of lysine 4 in H3), H3K36me and H2Bub marks are generally considered as activating histone modifications, while H3K9me, H3K27me, H4K20me and H2AK121ub marks are considered as repressive (Alvarez-Venegas, 2010; Liu *et al.*, 2010; Bourbousse *et al.*, 2012; Zhou *et al.*, 2017; Cheng *et al.*, 2019).

These histone marks are then recognized by specialized proteins who can act either as histone readers, writers or erasers. Histone writers and erasers have the ability to add and remove histone modifications, while histone readers simply recognize certain histone marks and mediate the interaction with other protein components (Liu *et al.*, 2010; Cheng *et al.*, 2019). A prominent example are the two Polycomb repressive complexes PRC1 and PRC2, multiprotein complexes that contain reader and writer components. Both PRCs deposit repressive marks: PRC1 modifies histones with H2A121ub, PRC2 with H3K27me (Kim and Sung, 2014; Xiao and Wagner, 2015).

LHP1 is a Chromo domain H3K27me reader that has been hypothesized to play a role in the recruitment of PRC1 to its targets, although its precise function remains unclear (Turck *et al.*, 2007; Zhou *et al.*, 2017). Interestingly, the *lhp1* mutant has slightly reduced levels of indolic GSLs and strongly elevated levels of aliphatic GSLs (Ludwig-Müller *et al.*, 1999; Kim *et al.*, 2004; Roux *et al.*, 2014), while the induction of GSL synthesis by heat stress and JA is considerably reduced (Bennett *et al.*, 2005).

The RNA-binding heterogeneous nuclear ribonucleoprotein LIF2 interacts with LHP1 and is a putative suppressor of plant immunity that mediates the balance between growth and defense. In the Arabidopsis *lif2* mutant, the content of all GSLs was significantly reduced, even though the mutant exhibited a basal primed defense state and showed increased resistance against multiple pathogens. These findings might reflect a complex deregulation, which interferes with JA and SA synthesis and signaling and also affects the transcription of *MYB28, MYB34* and *bHLH06* (Roux *et al.*, 2014). Moreover, the chain elongation pathway of aliphatic GSL synthesis has been associated with distinct histone modification patterns, although no effect on transcription was detected (Xue *et al.*, 2015).

These observations are however not the only evidence for an epigenetic regulation of GSL biosynthesis: A well-known phenomenon that occurs upon abiotic stress, wounding or pathogen infection in plants is the priming of defense mechanisms, by which responses become more rapid and robust in response to smaller stimuli – a process that has been clearly demonstrated for JA- and SA-mediated resistance. (Kim and Felton, 2013; Conrath *et al.*, 2015; Mauch-Mani *et al.*, 2017). The priming memory is thought to be stored epigenetically (Lämke and Bäurle, 2017; Ramirez-Prado *et al.*, 2018; Roberts and López Sánchez, 2019; Koç *et al.*, 2020). Because GSLs are components of inducible defense, their synthesis is part of the priming response (Rasmann *et al.*, 2012; Bakhtiari *et al.*, 2018) and thus likely a target of epigenetic regulation upon stress-related induction (Rasmann *et al.*, 2012).

Indeed, there are reports of JA-mediated mechanisms which could form the basis for a general epigenetic priming response in this context: Interaction between components of JA signaling and epigenetic factors have been shown to influence the transcription of some JA target genes, with JAZ proteins recruiting the repressive histone deacetylase HDA6 (Zhu *et al.*, 2011; Kim *et al.*, 2017) and MED25 recruiting the activating histone acetyltransferase HAC1 (An *et al.*, 2017), whose activity is required in the priming of pattern-triggered immunity (Singh *et al.*, 2014).

However, despite this circumstantial evidence, no direct epigenetic regulation of GSL biosynthesis has been described to date. This is slightly better understood in anthocyanin biosynthesis, a process which is also controlled by MYB-bHLH complexes: In maize, this complex is formed by the MYB factor C1 and the bHLH factor R and interacts with the histone reader RIF1 (Hernandez *et al.*, 2007).

As the regulations of anthocyanins and GSLs are closely related, this suggests the Arabidopsis homologs of RIF1, the EML family, as potential candidates to directly exert epigenetic control over GSL production. These family contains the four members EML1–4, proteins that are characterized by the combination of ENT, Agenet and coiled-coil domains (Tsuchiya and Eulgem, 2011). While the ENT domain has been shown to mediate homodimerization in RIF1 (Hernandez *et al.*, 2007), the Agenet represents a characteristic histone reader domain (Maurer-Stroh *et al.*, 2003; Zhao *et al.*, 2018).

EML1 and EML3 were shown to recognize H3K36 methylation and acetylation marks, with the highest affinity for H3K36me (Coursey *et al.*, 2018). While this is also true for the Agenet domain of EML1 alone, it additionally binds H3K4me (Zhao *et al.*, 2018) – which was not found for the full-length EML proteins, potentially because of an additional influence of the ENT domain on binding behavior (Coursey *et al.*, 2018). Moreover, EMLs have been described to function in immunity against oomycetes (Tsuchiya and Eulgem, 2011) and viruses (Coursey *et al.*, 2018) and in developmental processes such as flowering transition (Tsuchiya and Eulgem, 2011) and embryo formation (Milutinovic *et al.*, 2019), but they have not been characterized so far in relation to metabolic processes such as GSL biosynthesis.

1.4 Glucosinolate diversity and its evolution

A high degree of structural diversity exists among GSLs, with over 130 different GSLs documented in nature (Agerbirk and Olsen, 2012; Blažević *et al.*, 2020) and about 40 in Arabidopsis alone (Sønderby *et al.*, 2010b). There is considerable variation in the composition of the GSLs produced among different plant species (Bennett *et al.*, 2004; Bell, 2019; Czerniawski *et al.*, 2021), among different accessions within a single species (Kliebenstein *et al.*, 2005; Kissen *et al.*, 2016) and often even among different tissues within one plant (Brown *et al.*, 2003; van Dam *et al.*, 2009; Zang *et al.*, 2015). Additional variation in the myrosinase system leads to an even greater diversity in GSL breakdown products (Wittstock *et al.*, 2016a; Wittstock *et al.*, 2016b).

Non-methionine aliphatic and benzenic GSLs are present in all families of the Brassicales, whereas the indolic pathway is thought to have originated from the At-β whole-genome duplication (WGD) event that is ancestral to only a subset of the Brassicales. The synthesis of chain-elongated and Met-derived aliphatic GSLs – which accounts for the vast majority of GSL diversity in Arabidopsis (Reichelt *et al.*, 2002) – likely arose as the product of the second, more recent At-α WGD that is specific to one Brassicales family, the Brassicaceae (Schranz *et al.*, 2011; van den Bergh *et al.*, 2016; Barco and Clay, 2019).

In studies among the Brassicaceae, some intraspecific and interspecific differences in GSL profiles have been attributed to particular alleles of certain GSL biosynthesis enzymes. For example, the presence or absence of the aliphatic alkenyl and hydroxyalkyl GSLs in different Arabidopsis accessions is determined by the *AOP2* and *AOP3* genes (Kliebenstein *et al.*, 2001), while the absence of the indolic GSL 1MO-I3M in some Camelineae species might result from the loss of *CYP81F4* (Czerniawski *et al.*, 2021). However, these findings do not provide an extensive explanation on how the observed variation in GSL accumulation arises.

Another contributing factor may be variation in regulatory genes, which could in turn result in expression changes of biosynthesis genes. Such an effect of the regulatory machinery was postulated in a phylogenetic study by Windsor *et al.* (2005), who compared the interspecific variation in GSL profiles across the Brassicaceae. Many of the detected polymorphisms appeared to have a quantitative rather than a qualitative effect, and thus are thought to have arisen from variation in regulation. Kerwin *et al.* (2015) conducted a field trial with Arabidopsis, where mutant plants were used to recreate natural diversity in aliphatic GSL production. The mutant alleles that were shown to have an impact on fitness included subgroup 12 MYB TFs, again highlighting the potential evolutionary importance of variation in regulatory genes to maintain diverse GSL profiles. However, there have been no studies to date that directly address the question to which extent such regulatory variation contributes to metabolite variation.

To explore this potential role of TFs in the origin of GSL diversity, the degree of their conservation needs to be considered on two levels: First, how similar are the individual factors in one species or accession, i.e. are their affinities towards target genes considerably different? This would mean that relative variations in TF expression could cause corresponding changes in GSL biosynthesis enzyme expression and ultimately metabolite production. Such effects could e.g. give rise to tissue-specific differences in GSL accumulation. Second, how conserved are the TFs between species? Might orthologous factors have different affinities towards certain GSL biosynthesis genes, thus creating distinct GSL profiles?

In both cases, bHLHs are unlikely to mediate very specific changes. As they are recruited to target genes via interaction with the more specific MYBs, bHLHs would be expected to cause rather broad quantitative changes in GSL accumulation – between species or between tissues, as *bHLHs* exhibit very distinct expression patterns down to particular cell layers (Figueroa and Browse, 2015; Gasperini *et al.*, 2015; Qi *et al.*, 2015; Goossens *et al.*, 2017).

MYBs are also known to be expressed relatively distinctly across tissues (Gigolashvili *et al.*, 2009; Frerigmann and Gigolashvili, 2014). However, GSL profiles obtained from *MYB* overexpression lines show remarkably similar patterns of accumulation, with only MYB76 and MYB122 displaying a somewhat weaker inductive effect than their clade partners (Gigolashvili *et al.*, 2007a; Sønderby *et al.*, 2007; Malitsky *et al.*, 2008; Li *et al.*, 2013). The only other difference documented in literature is a slightly stronger influence of MYB28 on the production of long-chain aliphatic GSLs (those that have gone through more than two cycles of chain elongation) than the other aliphatic MYBs, MYB29 and MYB76 (Sønderby *et al.*, 2007; Beekwilder *et al.*, 2008; Sønderby *et al.*, 2010a; Li *et al.*, 2013).

Regarding the interspecific conservation of subgroup 12 MYBs, multiple studies have been conducted in *Brassica* species and other crops. Generally, *MYB* genes seem to be highly conserved. One notable exception is *MYB76*, which in Arabidopsis originates from a tandem (i.e. single gene) duplication relative to the genomes of more basal Brassicaceae (Hofberger *et al.*, 2013) and is also absent in *Brassica* species and *Raphanus sativus* (Mitsui *et al.*, 2015; Seo and Kim, 2017). In *Brassica,* asymmetrical evolution of polyploid genomes can contribute considerably to GSL diversity. This was demonstrated for *MAM* and *AOP2* biosynthesis genes in *B. oleracea* and *B. rapa* (Liu *et al.*, 2014); for *MYB* paralogs, divergent expression patterns have been documented in *B. juncea* (Augustine *et al.*, 2013) and *B. rapa* (Seo *et al.*, 2016; Seo *et al.*, 2017). While overall, overexpression of different *MYB* paralogs in these studies seemed to result in rather similar GSL profiles, some correlations between specific genes and specific GSL species have been made (Seo *et al.*, 2017). Moreover, Seo *et al.* (2016) proposed *MYB28* as a negative regulator of *AOP2* in *B. rapa*, which would represent a functional divergence compared to Arabidopsis. Taken together, the interspecies functional conservation of regulatory TFs remains largely unexplored in the context of GSL biosynthesis and its diversity.

1.5 Aims of the thesis

Over the last fifteen years, the transcriptional regulation of GSL biosynthesis by MYB and bHLH TFs has been well-described in Arabidopsis, with an increasing amount of information from other Brassicaceae species. In this thesis, my aim is to extend this knowledge in two different directions: First, by studying the connection between the MYB-bHLH transcriptional complex and a putative epigenetic regulation of GSL biosynthesis, and second, by exploring the contribution of *MYB* and *bHLH* genes to interspecific variation in GSL accumulation.

(i) Do EML proteins play a role in stress-mediated, targeted epigenetic regulation of GSL biosynthesis? To address this question, I analyzed publicly available omics data in regard to epigenetic marks of GSL synthesis genes, as well as the transcription of *EMLs*, in stressed and unstressed conditions. Furthermore, I performed microscopic localization and colocalization assays of MYB, bHLH and EML fusion proteins. This was followed by measuring JA-mediated GSL induction in various *eml* mutants both under prolonged and

repeated short-term stress conditions to obtain information about the role of EMLs in this process, how it behaves in regard to priming and how these two topics are connected. Finally, I conducted overexpression experiments of *EML* genes, together with *MYB* and *bHLH* genes, in Arabidopsis cell culture to examine their individual effects and combined interplay.

(ii) Do *MYB* and *bHLH* genes play a role in shaping interspecific diversity of GSL profiles? To address this question, I generated detailed GSL profiles for root and shoot tissue of the three Brassicaceae species *Arabis alpina*, *Cardamine hirsuta* and *Capsella rubella* via liquid chromatography-mass spectrometry (LC-MS), then established cell cultures of *A. alpina* and *C. rubella* for further experiments. I then analyzed genomic data of *MYB* and *bHLH* genes for the three species in comparison with Arabidopsis. Subsequently, I examined the conservation of various subgroup 12 MYBs from different species by overexpression in Arabidopsis cell culture and quantification of the resulting GSL accumulation. Finally, I investigated the alternative splicing of *bHLHs* in *C. rubella* and overexpressed *MYB*, *bHLH* and other constructs in cell culture with the goal of complementing the lack of indolic GSL production in *C. rubella*.

2 Materials and Methods

2.1 Materials

2.1.1 Enzymes

Enzymes used in this study are listed in **Table 1**.

Table 1: Enzymes used in this study.

Enzyme	Source
DNAse I	Thermo Scientific (Vilnius, Lithuania)
DreamTaq DNA Polymerase	Thermo Scientific (Vilnius, Lithuania)
Gateway BP Clonase II Enzyme Mix	Invitrogen (Carlsbad, CA, USA)
Gateway LR Clonase II Enzyme Mix	Invitrogen (Carlsbad, CA, USA)
Phusion High-Fidelity DNA Polymerase	Thermo Scientific (Vilnius, Lithuania)
Power SYBR Green Master Mix	Applied Biosystems (Warrington, UK)
Proteinase K	Invitrogen (Carlsbad, CA, USA)
RNAse A	Thermo Scientific (Vilnius, Lithuania)
Sulfatase from *Helix pomatia*	Sigma-Aldrich (Steinheim, Germany)
SuperScript II Reverse Transcriptase	Invitrogen (Carlsbad, CA, USA)

2.1.2 Kits

Molecular biology kits used in this study are listed in **Table 2**.

Table 2: Kits used in this study.

Kit	Source
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel (Düren, Germany)
Quantum Prep Plasmid Miniprep Kit	Bio-Rad (Hercules, CA, USA)

2.1.3 Buffers and solutions

All buffers and solutions (**Table 3**) were prepared with double-distilled water.

Table 3: Buffers and solutions used in this study.

Buffer/Solution	Components	
TAE buffer	40 mM	Tris/HCl, pH 8.0
	20 mM	HAc
	1 mM	EDTA-Na_2
10x DNA loading dye	30% (v/v)	Glycerol
	0.25% (w/v)	Bromphenol blue
gDNA extraction buffer	200 mM	Tris/HCl, pH 7.5
	250 mM	NaCl
	25 mM	EDTA-Na_2
	0.5% (w/v)	SDS
Acetate buffer	20 mM	HAc
	pH 5.0 with NaOH	

Buffer/Solution	Components
GSL extraction solution 1	80% (v/v) MeOH 50 µM Benzylglucosinolate
GSL extraction solution 2	80% (v/v) MeOH
LC-MS GSL extraction solution 1	80% (v/v) MeOH 5 µM Sinigrin
LC-MS GSL extraction solution 2	50% (v/v) MeOH 5 µM Sinigrin

2.1.4 Media

Media used for plant and bacterial growth are listed in **Table 4**. All media were prepared with double-distilled water.

Table 4: Media used in this study.

Medium	Components
½ MS	2 g/l MS including vitamins[1] 10 g/l Sucrose 8 g/l Agar (optional) pH 5.8 with KOH
MS Kallus	4.3 g/l MS basal salt mixture[2] 0.5% (v/v) Gamborg's vitamin solution[2] 20 g/l Glucose 3 mg/l 2,4-D 1 mg/l Kinetin 8 g/l Agar pH 5.8 with KOH
AT	4.4 g/l MS basal salt mixture[2] 0.5% (v/v) Gamborg's vitamin solution[2] 30 g/l Glucose 1 mg/l 2,4-D pH 5.8 with KOH
LB	10 g/l Tryptone 5 g/l Yeast extract 10 g/l NaCl 15 g/l Agar (optional)
LB light	10 g/l Tryptone 5 g/l Yeast extract 5 g/l NaCl
YEB	5 g/l Peptone 5 g/l Beef extract 1 g/l Yeast extract 5 g/l Sucrose 0.5 g/l $MgSO_4 \cdot 7\ H_2O$ 15 g/l Agar (optional)

Medium	Components
Psi	20 g/l Tryptone 5 g/l Yeast extract 5 g/l $MgSO_4 \cdot 7\ H_2O$ pH 7.6 with KOH
TfbI	30 mM KAc 100 mM RbCl 10 mM $CaCl_2$ 50 mM $MnCl_2$ 15% (v/v) Glycerol pH 5.8 with HAc
TfbII	10 mM MOPS 10 mM RbCl 75 mM $CaCl_2$ 15% (v/v) Glycerol pH 5.8 with NaOH

[1] Duchefa (Haarlem, Netherlands)
[2] Sigma-Aldrich (Steinheim, Germany)

2.1.5 Antibiotics

Antibiotics were added to the appropriate growth media to allow for selection. The compounds and their final concentrations are listed in **Table 5**.

Table 5: Antibiotics used in this study.

Antibiotic	Concentration *(E. coli)*	Concentration *(A. tumefaciens)*
Ampicilin (Amp)	100 µg/ml	-
Carbencilin (Carb)	-	100 µg/ml
Chloramphenicol (Chl)	-	75 µg/ml
Gentamycin (Gent)	7 µg/ml	25 µg/ml
Hygromycin (Hyg)	50 µg/ml	50 µg/ml
Kanamycin (Kan)	50 µg/ml	50 µg/ml 12 µg/ml (only GV3101.pMP90RK)
Rifampicin (Rif)	-	60 µg/ml 20 µg/ml (only SV0)

2.1.6 Plant material

2.1.6.1 *Arabidopsis thaliana* plants

Arabidopsis wildtypes used in this study are listed in **Table 6**. Furthermore, two different sets of mutants were used: Set I was kindly provided by the Eulgem Lab (University of California, USA) and is listed in **Table 7**. Set II was kindly provided by the Brkljacic Lab (Ohio State University, USA) and is listed in **Table 8**.

Table 6: Arabidopsis wildtypes used in this study.

Accession	ID
Col-0	CS70000
Col-3	CS908
Ler-0	CS20
Col-0/Ler-0	CS69827

Table 7: Set I mutant lines.
These lines were first described by Tsuchiya and Eulgem (2011).

Genotype	ID	Background	Method of gene targeting	Targeted genes
eml1-2	SALK_114038	Col-0	T-DNA insertion	*AT3G12140*
eml1-3	SALK_077088	Col-0	T-DNA insertion	*AT3G12140*
eml2-1	SALK_116222	Col-0	T-DNA insertion	*AT5G06780*
eml3-1	SALK_106147	Col-0	T-DNA insertion	*AT5G13020*
eml1-2/ eml2-1	Cross of SALK_114038 and SALK_116222	Col-0	T-DNA insertion	*AT3G12140*, *AT5G06780*
eml-quad1	Cross of SALK_114038 and SALK_116222 with independent amiRNA insertion	Col-0	T-DNA insertion, amiRNA silencing	*AT3G12140*, *AT5G06780* (T-DNA); *AT5G13020*, *AT2G44440* (amiRNA)
eml-quad2	same as *eml-quad1*	Col-0	T-DNA insertion, amiRNA silencing	same as *eml-quad1*
eml-quad3	same as *eml-quad1*	Col-0	T-DNA insertion, amiRNA silencing	same as *eml-quad1*

Table 8: Set II mutant lines.
These lines were first described by Milutinovic et al. (2019).

Genotype	ID	Background	Method of gene targeting	Targeted genes
eml1-2	CS69823	Ler-0	Transposon insertion	*AT3G12140*
eml2-1	CS69824	Col-0	T-DNA insertion	*AT5G06780*
eml3-4	CS69825	Col-0	T-DNA insertion	*AT5G13020*
eml4-1	CS860536	Col-3	T-DNA insertion	*AT2G44440*
eml1-2/ eml3-4	CS69826	Col-0/Ler-0	Transposon insertion, T-DNA insertion	*AT3G12140*, *AT5G13020*

2.1.6.2 Other Brassicaceae plants

Non-Arabidopsis plant lines used in this study are listed in **Table 9**.

Table 9: Other Brassicaceae plant lines used in this study.

Species	Accession	Source
Arabis alpina	Pajares	Albani Lab (MPIPZ, Cologne, Germany), Bucher Lab (University of Cologne, Germany)
Cardamine hirsuta	Oxford	Tsiantis Lab (MPIPZ, Cologne, Germany)
Capsella rubella	Monte Gargano	Bednarek Lab (Polish Academy of Sciences, Poznan, Poland)

2.1.6.3 Cultured cells

Cell culture lines used and/or generated in this study are listed in **Table 10**.

Table 10: Cell culture lines used in this study.

Species	Genotype/ Accession	Targeted genes	Source
Arabidopsis thaliana	Ler-0	-	Gigolashvili Lab
Arabidopsis thaliana	*bhlh04/05/06* (in Col-0)	*AT4G17880*, *AT5G46760*, *AT1G32640* (T-DNA insertion)	Gigolashvili Lab[1]
Arabis alpina	Pajares	-	this study
Capsella rubella	Monte Gargano	-	this study

[1]The *bhlh04/05/06* mutant line was first described in Frerigmann *et al.* (2014).

2.1.7 Bacteria

Bacterial strains used in this study are listed in **Table 11**.

Table 11: Bacterial strains used in this study.

Species	Strain	Resistance	Reference
Agrobacterium tumefaciens	GV3101.19K	Rif, Kan	Voinnet *et al.* (1999); Voinnet *et al.* (2003)
Agrobacterium tumefaciens	GV3101.pMP90RK	Rif, Gent, Kan	Koncz and Schell (1986)
Agrobacterium tumefaciens	SV0 (LBA4404.pBBR1MCS*vir*GN54D)	Rif, Chl	van der Fits *et al.* (2000)
Escherichia coli	DH5α	-	Thermo Scientific (Vilnius, Lithuania)
Escherichia coli	TOP10	-	Invitrogen (Carlsbad, CA, USA)

2.1.8 Genes

The genes that were studied in this thesis and their orthologs in the analyzed Brassicaceae species are listed in **Table 12**.

Table 12: Genes studied in this thesis and their orthologs in Brassicaceae species.

Gene	*A. thaliana*	*C. rubella*	*C. hirsuta*	*A. alpina*
bHLH04	*AT4G17880*	*Carubv10004392*	*CARHR224600*	*Aa_G152240*
bHLH05	*AT5G46760*	*Carubv10028441*	*CARHR174740*	*Aa_G324640*
bHLH06	*AT1G32640*	*Carubv10008586*	*CARHR032070*	*Aa_G16560*
bHLH28	*AT5G46830*	*Carubv10028430*	*CARHR174620*	*Aa_G295550*
EML1	*AT3G12140*	*Carubv10014171*	*CARHR086710*	*Aa_G5240*
EML2	*AT5G06780*	*Carubv10001509*	*CARHR206400*	*Aa_G386400*
EML3	*AT5G13020*	*Carubv10002698*	*CARHR200440*	*Aa_G155210*
EML4	*AT2G44440*	*Carubv10023267*	*CARHR139600*	*Aa_G288460*
MYB28	*AT5G61420*	*Carubv10027861*	*CARHR275180*	*Aa_G606850*

Gene	*A. thaliana*	*C. rubella*	*C. hirsuta*	*A. alpina*
MYB29	*AT5G07690*	*Carubv10001328*	*CARHR205480*	*Aa_G76360*
MYB76	*AT5G07700*	-	*CARHR205450*	-
MYB34	*AT5G60890*	*Carubv10027524*	*CARHR274570*	*Aa_G693340*
MYB51	*AT1G18570*	*Carubv10009563*	*CARHR019190*	*Aa_G271510*
MYB122	*AT1G74080*	*Carubv10020723*	*CARHR067240*	-

2.1.9 Vectors

Vectors used in this study are listed in **Table 13**. Full lists of Gateway entry and expression clones used and/or generated in this study can be found in **Supplementary Table 1** and **Supplementary Table 2**, respectively.

Table 13: Vectors used in this study.

Vector	Description	Resistance	Reference
pAUBERGINE	Gateway destination vector, 35S promoter, C-mCherry tag	Amp/Carb	GenBank ID: FR695418
pDONR201	Gateway donor vector	Kan	Invitrogen (Carlsbad, CA, USA)
pDONR207	Gateway donor vector	Gent	Invitrogen (Carlsbad, CA, USA)
pENTR/D-TOPO	Gateway donor vector	Kan	Invitrogen (Carlsbad, CA, USA)
pGWB2	Gateway destination vector, 35S promoter	Kan, Hyg	Nakagawa *et al.* (2007)
pGWB3	Gateway destination vector, no promoter, C-GUS tag	Kan, Hyg	Nakagawa *et al.* (2007)
pGWB5	Gateway destination vector, 35S promoter, C-GFP tag	Kan, Hyg	Nakagawa *et al.* (2007)
pGWB6	Gateway destination vector, 35S promoter, N-GFP tag	Kan, Hyg	Nakagawa *et al.* (2007)

2.1.10 Software

Software used in this study is listed in **Table 14**.

Table 14: Software used in this study.

Program	Version	Purpose	Reference
ApE	2.0.53c	Sequence editing and alignment	Davis (2020)
BLAST	2.11.0	Sequence search	Altschul *et al.* (1990)
Chromeleon	7.2.2.6394	Data processing and analysis (HPLC)	Thermo Scientific
draw.io	14.1.8	Figure design	Seibert Media
Image Lab	5.0.18	Image processing (gel electrophoresis)	Bio-Rad
Leica Application Suite	2.8.1	Image processing (fluorescence microscope)	Leica
MEGA X	10.2.2	Phylogenetic analysis	Kumar *et al.* (2018)

Program	Version	Purpose	Reference
Microsoft Office	2019	Data and text management, figure design	Microsoft
Primer3	4.1.0	Primer design	Koressaar and Remm (2007)
R with RStudio	4.0.3 (R), 1.2.1335 (RStudio)	Statistical analysis, figure design	R Core Team (2020); RStudio Team (2020)
QuantStudio Design & Analysis	1.4.3	Data processing and analysis (qPCR)	Applied Biosystems
ZEN black edition	16.0.1.306	Image processing (confocal microscope)	Carl Zeiss

2.2 Methods

2.2.1 Plant and bacteria cultivation

2.2.1.1 Plant cultivation

Plants were grown on VM soil (Einheitserde; Sinntal-Altengronau, Germany) under common greenhouse conditions for propagation or nucleic acid extraction. Otherwise, plants were grown on ½ MS medium, either under short-day (8 h light at 22 °C, 16 h darkness at 18 °C; 60% relative humidity; ca. 160 µmol/m²s light intensity) or long-day conditions (16 h light at 22 °C, 8 h darkness at 18 °C; 60% relative humidity; ca. 50 µmol/m²s light intensity). In these cases, seeds were sterilized before sowing for 2 min in 70% EtOH and 1 min in 100% EtOH, then either placed into liquid medium or dispersed across solid plates in a suspension of 0.1% agarose. In all cases, the sown seeds were stratified at 4 °C in darkness for 3 days before transfer to the greenhouse or growth chamber to increase germination rates. Plants grown in liquid ½ MS medium were shaken at 120 rpm.

2.2.1.2 Cell culture cultivation

For the generation of new callus cultures, plants were grown in liquid ½ MS medium under long-day conditions. 3 weeks (ca. 1 month for *A. alpina*) after germination, the root tissue was finely chopped by scalpel and spread upon MS Kallus plates in flat clumps of ca. 1 cm in diameter. These callus cultures were grown in darkness with all other conditions kept constant and maintained by a monthly transfer to fresh MS Kallus plates. Suspension cell cultures were generated by homogenizing callus cultures of at least several months of age in AT medium, grown at 22 °C and 120 rpm shaking and maintained by weekly dilution with fresh AT medium.

2.2.1.3 Bacteria cultivation

Escherichia coli were grown on LB plates at 37 °C and Agrobacteria (*Agrobacterium tumefaciens*) on YEB plates at 28 °C, with appropriate selective antibiotics present. After 1 d (*E. coli*) resp. 2–3 days (Agrobacteria; 3 days for GV3103) of growth, the bacterial cultures were stored at 4 °C and maintained by transfer to fresh LB and YEB plates every 3 to 4 weeks. Overnight cultures of *E. coli* and Agrobacteria were generated by

inoculating 2–4 ml of liquid LB resp. YEB medium (including antibiotics), then grown overnight at 37 °C resp. 28 °C. For long-term storage, glycerol stocks were generated from 750 µl of overnight culture mixed with 750 µl of 60% glycerol in LB resp. YEB medium. The stocks were immediately frozen in LN_2 and stored at -80 °C.

2.2.2 Molecular biology methods

2.2.2.1 Agarose gel electrophoresis

In cases where loading buffer was not already present in the sample, DNA was mixed appropriately with loading dye. Electrophoresis was then typically performed in a gel composed of 1% agarose and 0.5 µg/ml ethidium bromide in TAE buffer, at a voltage of 120–140 V. After sufficient separation was achieved, the bands were visualized with UV light using a Gel Doc XR+ system (Bio-Rad; Munich, Germany). Images were recorded and processed with Image Lab.

2.2.2.2 Genotyping

To verify the genotypes of individual plants, ca. 100 mg of freshly harvested leaf tissue was frozen in LN_2. It was homogenized for 30 s in a Tissue Lyser (Qiagen; Hilden, Germany), then 300 µl of gDNA extraction buffer were added and the mixture was again homogenized for 30 s. The samples were centrifuged at 14,000 rpm for 5 min, after which the supernatant was transferred into a fresh tube where it was mixed with 300 µl isopropanol and subsequently incubated at RT for 5 min to precipitate the DNA. Afterwards, the sample was again centrifuged at 14,000 rpm for 5 min, with the supernatant being discarded. Finally, the gDNA pellet was dried and re-dissolved in 50 µl H_2O.

Two different kinds of genotyping were performed: First, for the set II mutant lines, the wild-type backgrounds were verified for two different chromosomes using simple sequence length polymorphism markers (SSLPs; **Table 15**). Second, the presence of T-DNA insertions was verified for all mutants via PCRs using the LB/LP/RP system, where left and right primer (LP and RP) flank the genomic insertion site, while the left border primer (LB) binds to the T-DNA sequence. This means that if the T-DNA insertion is present, the combination of LB and RP primers will yield a band; if the wild-type allele is present, the combination of LP and RP primers will yield a band – i.e. a heterozygous plant will yield bands in both cases. The presence of the transposon insertion in the *eml1-2* line CS69823 was verified in the same way, apart from the use of a right border (RB) instead of an LB primer. The presence of the amiRNA construct in the *eml-quad* lines was verified using primers specific to the BASTA resistance gene *bar* as a marker for the construct. The primers used for genotyping are listed in **Supplementary Table 3**.

Table 15: SSLPs used for determination of the background accession in this study.

SSLP	Chromosome	Length in Col [bp]	Length in Ler [bp]
CIW5	4	164	144
NGA139	5	174	132

2.2.2.3 RNA isolation

Ca. 100 mg of leaf tissue were frozen in LN_2 and homogenized in cooled racks of a Tissue Lyser for 1 min, before 1 ml Trizol (Ambion; Carlsbad, CA, USA) was added to extract the RNA in a stabilized manner. The samples were homogenized again for 30 s, then centrifuged at 14,000 rpm and 4 °C for 10 min. The supernatant was transferred into a new reaction tube, where 200 µl of chloroform were added to solve organic components. The mixture was shaken vigorously and incubated at RT for 3 min, then centrifuged at 14,000 rpm and 4 °C for 15 min to separate the organic and the aqueous phases. The latter, which contained the RNA, was subsequently transferred into a new reaction tube where 500 µl isopropanol were added to precipitate the RNA. The mixture was shaken vigorously and incubated at RT for 10 min, then centrifuged at 14,000 rpm and 4 °C for 10 min. The resulting RNA pellet was washed with 1 ml 75% EtOH, followed by centrifugation at 8,500 rpm and 4 °C for 5 min. Finally, it was dried and re-dissolved in 35 µl H_2O. The concentration of the isolated RNA was determined at a Nanodrop photometer (Peqlab; Erlangen, Germany).

2.2.2.4 cDNA synthesis

As a first step, 5–10 µg of RNA was digested with DNase I to remove any remaining DNA contamination (**Table 16**). After 15 min of incubation at 37 °C, DNase activity was terminated by the addition of 2.5 µl of 25 mM EDTA-Na_2. The RNA was then denatured by incubation at 65 °C for 10 min.

Table 16: DNase digestion reaction.

Component	**Volume**
First Strand Buffer	5 µl
H_2O	19 – x µl
DNase I	1 µl
RNA	x µl

Afterwards, 10 µl of the denatured RNA was used as template for reverse transcription by SuperScript II (**Table 17**). The reaction was incubated at RT for 8 min to allow for primer annealing before the reverse transcriptase was added. Then, the samples were incubated at 42 °C for 75 min until the enzyme was inactivated by incubation at 70 °C for 15 min. The quality of the resulting cDNA was verified by a PCR with primers for the housekeeping gene *ACTIN 2* (**Supplementary Table 4**).

Table 17: Reverse transcription reaction.

Component	Volume
First Strand Buffer	7.5 µl
H_2O	21 µl
Oligo(dT)$_{18}$ primer (100 µM)	2 µl
dNTPs (10 µM each)	2.5 µl
DTT (100 mM)	5 µl
RiboLock RNase Inhibitor[1]	1 µl
SuperScript II Reverse Transcriptase	1 µl
RNA	10 µl

[1]Thermo Scientific (Vilnius, Lithuania)

2.2.2.5 Generation of chemically competent *E. coli*

E. coli DH5α bacteria were grown first on an LB plate without antibiotics, then as an overnight culture in Psi medium. Next, 500 ml of Psi medium were inoculated with 1 ml of the overnight culture. This new culture was grown to an OD_{600} of 0.5–0.6, then cooled on ice for 15 min, as all subsequent steps were performed at 4 °C. The bacteria were then centrifuged at 2,500 rpm for 10 min, after which the resulting bacterial pellet was resuspended in 200 ml TfbI. The cells were rested on ice for 15 min, then pelleted as before, resuspended in 20 ml TfbII and rested again on ice for 15 min. Finally, the bacteria were divided into 100 µl aliquots and stored at -80 °C for further use.

2.2.2.6 Generation of chemically competent Agrobacteria

SV0 Agrobacteria were grown first on a YEB plate with the appropriate antibiotics, then as two 10 ml overnight cultures in LB light medium. These bacteria were used to inoculate three 400 ml cultures that were grown overnight to an OD_{600} of 1.0. At this point, the cells were cooled on ice for 15 min, as all subsequent steps were performed at 4 °C. The bacteria were centrifuged at 4,000 rpm for 15 min, after which the resulting bacterial pellet was resuspended in 200 ml H_2O. The cells were then centrifuged as before, resuspended in 100 ml H_2O, centrifuged again and resuspended in 10 ml 10% glycerol. Subsequently, they were centrifuged yet again and resuspended in a volume of 10% glycerol equal to the one of the pellet (ca. 8 ml). Finally, the bacteria were divided into 50 µl aliquots and stored at -80 °C for further use.

2.2.2.7 Cloning

All constructs cloned in this study were generated using the Gateway cloning system by Invitrogen (Katzen, 2007). This technique encompasses the generation of entry clones and expression clones. Entry clones contain a sequence of interest and allow for its amplification and flexible transfer to various destination vectors, creating different types of expression clones. All entry clones produced in this study were generated via the BP reaction in pDONR donor vectors; all expression clones were generated via the LR reaction in pGWB destination vectors. Clones in other vectors were obtained from external sources.

At the beginning of the cloning process, cDNA was used as a template to generate a specific PCR product consisting of the desired sequence fused with the attB1 (5' end) and attB2 (3' end) sites, thus enabling the BP reaction. This reaction was performed with Phusion Polymerase to ensure the correct sequence; a list of primers used for this purpose can be found in **Supplementary Table 5**.

The resulting PCR product was purified via agarose gel electrophoresis. The band of the correct size was then excised from the gel and the contained DNA extracted with the NucleoSpin Gel and PCR Clean-up Kit according to the manufacturer's instructions. Afterwards, the BP reaction was performed and transformed into *E. coli* DH5α or TOP10 cells following the manufacturer's instructions, apart from some modifications: The reaction was conducted overnight and scaled down to a total volume of 5 µl, of which the entirety was then used for the transformation.

After distinct *E. coli* colonies had grown on selective LB plates, 4 ml overnight cultures were inoculated with bacteria from defined colonies. These cultures were grown overnight, after which the amplified entry plasmids were isolated from them using the Quantum Prep Plasmid Miniprep Kit, following the manufacturer's instructions. Subsequently, the resulting concentration of the plasmid was determined at a Nanodrop photometer. All entry clones were sequenced via LightRun Sanger sequencing (Eurofins; Cologne, Germany); the obtained sequences were analyzed with ApE.

Entry clones for which the correct sequence was verified could then serve as component of the LR reaction. This step, together with the subsequent transformation of the resulting expression vector into *E. coli* DH5α or TOP10 cells, was performed following the manufacturer's instructions apart from the same modifications as for the BP reaction. The plasmid was amplified in *E. coli,* then isolated via minipreparation as before. The identity of the obtained plasmid was verified by PCR with construct-specific primers. Oligonucleotides used specifically for sequencing and sequence verification are listed in **Supplementary Table 2**.

2.2.2.8 Electroporation transformation of Agrobacteria

Agrobacterium aliquots were thawed on ice, after which 2 µl of expression plasmid were added to the cells, followed by gentle mixing. After incubation on ice for 2 min, then mixture was transferred into an electroporation cuvette and shocked in an electroporator 2510 (Eppendorf; Hamburg, Germany) at 2500 V. Immediately, 500 µl of YEB medium were added, then the cells were allowed to recover at 28 °C for 2 h before they were plated on selective YEB plates. After distinct Agrobacterium colonies had grown on these plates, the presence of the correct plasmid in defined colonies was verified via colony PCR, where the template consisted of a small amount of bacterial culture added directly into the PCR reaction.

2.2.2.9 Transformation of cultured plant cells

Cultured plant cells were transformed with constructs of interest using the previously generated Agrobacteria. Overnight cultures were centrifuged at 4,000 rpm and 4 °C for 20 min, after which the

resulting bacterial pellet was resuspended in AT medium, at a 10% volume of the previous culture. Cultivated plant cells were freshly diluted and transferred into six-well plates in 4 ml aliquots. Subsequently, they were transformed by the addition of 20 µl of Agrobacterium suspension per well and per construct of interest. All cells were transformed with the anti-silencing strain GV3101.19K to increase expression of transgenes and to serve as a negative control.

2.2.2.10 Quantitative real-time PCR

qPCR was employed to quantify the expression of target genes from cDNA samples. Specific qPCR primers (**Supplementary Table 6**) were designed with Primer3 to amplify 70–150 bp fragments unique to the target genes. The primer quality was verified by a preceding regular PCR and a melting curve analysis after the main qPCR reaction. These reactions (**Tables 18–19**) were performed with two technical replicates per sample in optical 96-well plates, using a QuantStudio 5 system (Applied Biosystems; Singapore). Threshold and C_t values were determined with QuantStudio Design & Analysis; relative transcript levels were calculated as ΔC_t values (Schmittgen and Livak, 2008). Here, *PP2A* was used as internal reference gene with constant expression to normalize C_t values across conditions.

Table 18: qPCR reaction.

Component	Volume
2x SYBR Green Mix	10 µl
H_2O	7 µl
Primer forward (10 µM)	0.5 µl
Primer reverse (10 µM)	0.5 µl
cDNA template (1:10)	2 µl

Table 19: qPCR program.

Step	Temperature	Duration	Cycles
Denaturation (initial)	95 °C	180 s	1
Denaturation	95 °C	15 s	40
Annealing/Elongation	60 °C	60 s[1]	
Dissociation (initial)	95 °C	15 s	1 (Melting curve)
Annealing	60 °C	60 s	
Dissociation	Gradient of 0.15 °C/s until 95 °C[2]		

[1]Fluorescence intensity was measured at the end of this stage
[2]Fluorescence intensity was measured every second

2.2.3 Biochemical and analytical methods

2.2.3.1 Mass spectrometry

A. alpina, *C. hirsuta* and *C. rubella* seeds were stratified and grown in liquid ½ MS medium under long-day conditions for 16 d (*C. hirsuta*, *C. rubella*) resp. 23 d (*A. alpina*). Shoot and root tissues were separated, then dried on paper and frozen in LN_2 as samples of ca. 500 mg.

To extract intact GSLs, the samples were ground with mortar and pestle while being kept frozen with LN_2. About 40 mg of each of the samples was transferred into new reaction tubes, where 1 ml of LC-MS GSL extraction solution 1 was added per 20 mg of sample weight. The mixture was homogenized by vortexing, heated at 80 °C for 15 min, then cooled on ice and centrifuged at 14,000 rpm and 4 °C for 10 min. The supernatant was transferred into a fresh reaction tube stored on ice, while the pellet was again extracted with LC-MS GSL extraction solution 2. Both extracts were unified and centrifuged as before, after which 2 ml of each sample was transferred into a new reaction tube. These samples were then used by the Biocenter MS platform (University of Cologne, Germany) for LC-MS analysis.

The liquid chromatography served to generate additional information in form of retention time data and to reduce sample complexity via separation over a C18-CSH column. The samples then directly underwent electrospray ionization in negative mode and were measured in a quadrupole-quadrupole time-of-flight (Qq-TOF) mass analyzer setup, with the TOF analyzer recording the m/z (mass/charge) ratio to maximize accuracy. GSLs were identified from the entirety of detected compounds by their retention time, their mass and a sulfur-specific isotope pattern. Individual GSL identities were assigned via comparison to standards, or, where unavailable, to mass values of GSLs compiled by Sun *et al.* (2016), followed by an additional manual evaluation according to Blažević *et al.* (2020).

2.2.3.2 MeJA treatment

To simulate stress conditions, plants were induced with the phytohormone methyl jasmonate (MeJA). This principle was used for two different experimental approaches: First, plants were grown on MeJA-containing medium for a prolonged time. Second, plants were repeatedly exposed to MeJA for short times. The first approach focused on the induction of GSL production by MeJA, while the second one also considered the priming of the GSL response by repeated stimuli.

For the prolonged stress approach, sterilized seeds were sown on ½ MS plates, stratified and then grown under short-day conditions to prevent early flowering. After 9–10 days, the seedlings were transferred to fresh ½ MS plates containing 50 µM MeJA. They were grown under these conditions for ca. 2 weeks before they were harvested and frozen in LN_2.

For the priming approach, two different experimental setups were used: In the first, Col-0, Col-0/Ler-0, *eml1-2/eml2-1* and *eml1-2/eml3-4* seeds were sown on ½ MS plates, stratified and grown as in the prolonged stress approach. 14 days after germination, the seedlings were transferred to liquid ½ MS medium, where they were placed into "vases" made from cut pipette tips to prevent submersion in the medium. Three such vases were placed into each well of a six-well plate.

After one day of acclimation, the medium was exchanged with fresh one, either in the same composition as before or additionally containing 17 µM MeJA; this time point was termed "T0". 24 h later, the medium was

again exchanged with standard ½ MS for all plants. This procedure was repeated after 7 d and 14 d, with these time points termed "T1" and "T2". For T1 and T2, plants were harvested and frozen in LN_2 24 h after the start of the MeJA induction.

In the second experimental setup, a more precise time course of the priming process was measured. Col-0/Ler-0 and *eml1-2/eml3-4* seeds were sown directly on 20 x 20 mm cutouts of polypropylene mesh (PP-105/16; Franz Eckert; Waldkirch, Germany) floating on liquid ½ MS medium, in a density of ten seeds per cutout. They were treated as before, until 18 d after germination (T0), the medium was replaced, now using ½ MS with and without 25 µM MeJA. Plant shoots were harvested and frozen in LN_2 6 h, 12 h and 24 h after the start of the MeJA induction. Subsequently, the plants were washed with H_2O and returned to pure ½ MS medium as before. This procedure was repeated once after 7 days (T1).

2.2.3.3 GSL quantification

GSL concentration was determined in samples from plants and cultured cells. In both cases, samples were first lyophilized overnight, then weighed out to normalize GSL concentration against the dry weight. Samples were homogenized in a Tissue Lyser for 2 min, before 1 ml of GSL extraction solution 1, which contains the internal standard benzyl glucosinolate (BGS), was added. Subsequently, the samples were centrifuged at 14,000 rpm for 10 min, after which the supernatant was transferred into a new reaction tube. The pellets were re-extracted with GSL extraction solution 2 and centrifuged as before, with both extracts then being unified. For cell culture samples, this extract was centrifuged and transferred to a new tube once more to remove any remaining cell debris.

The extract was applied onto columns that contained 1 ml of DEAE-Sephadex (Sigma-Aldrich; Steinheim, Germany) equilibrated in acetate buffer, adsorbing the negatively charged GSLs. The columns were then washed with 2 ml H_2O and 2 ml acetate buffer, after which 500 µl of sulfatase (2.8 mg/ml in actate buffer) was added and left to incubate overnight on the sealed columns. The sulfatase catalyzed the cleavage of the sulfate group from the GSLs, thus enabling the elution of the resulting desulfo-GSLs from the columns with 1.5 ml H_2O. The obtained eluate was dried overnight in a vacuum concentrator, after which the remaining pellet was re-dissolved in 300 µl H_2O. The samples were then centrifuged at 14,000 rpm for 2 min for a final purification step, then transferred into coned HPLC vials.

The desulfo-GSL concentration was measured via HPLC at an UltiMate3000 system (Dionex; Germering, Germany), using a Spherisorb C18-ODS2 column (4.6 x 250 mm, 5 µm particle size) (Waters; Milford, MA, USA) and a compound linear gradient of H_2O (solvent A) and acetonitrile (solvent B) (**Table 20**) at a flow rate of 1 ml/min and a column oven temperature of 25 °C. Desulfo-GSLs were quantified by UV absorption at a wavelength of 229 nm.

Table 20: HPLC gradient for GSL quantification.
The gradient proceeded linearly between the given time points.

Time point	Percentage of solvent B
0 min	5%
2 min	5%
8.5 min	28%
12 min	34%
14 min	43%
16 min	50%
20 min	50%
20.5 min	100%
25 min	100%
25.1 min	5%
30 min	5%

Peak detection and area integration were performed with Chromeleon 7. Peak identities were assigned and GSL type-specific response factors calculated according to previous measurements of commercial standards. The final normalized GSL concentration was calculated according to **Equation 1**.

Equation 1: Calculation of GSL concentration. "GSL" refers to the GSL in question, "BGS" to the internal standard.

$$c_{GSL}\left[\frac{\mu mol}{g}\right] = \frac{peak\ area_{GSL} \times response\ factor_{GSL} \times amount_{BGS}\ [\mu mol]}{peak\ area_{BGS} \times response\ factor_{BGS} \times sample\ dry\ weight\ [g]}$$

2.2.3.4 Qualitative GUS transactivation assay

The comparative ability of TFs to induce GSL production in cultured cells was qualitatively measured via GUS transactivation. Cultured cells were transfected with TFs of interest and the promoter-reporter construct pGWB3-proCYP79B3, which contains the reporter gene *GUS* under the control of the promoter of the indolic biosynthesis gene *CYP79B3*. The induction of this promoter by different TFs of interest could then be quantified by proxy of *GUS* transcription, which in turn was visualized by the turnover of the GUS substrate X-Gluc, resulting in a qualitative change of blue coloration. 20 µl of 10 mg/ml X-Gluc in DMF was added per ml of cell culture 3–7 days after transfection, depending on the density of the cultured cells. The samples were subsequently shaken and incubated at 37 °C for 2–12 h, depending on the intensity of the reaction.

2.2.3.5 Fluorescence microscopy

The subcellular localization of N- and C-terminal GFP fusion proteins was studied for MYB, bHLH and EML fusion proteins by fluorescence microscopy of cultured Arabidopsis Ler-0 cells that were transfected with the respective constructs. Additionally, the colocalization of C-terminal fusion proteins GFP-EML2 and mCherry-bHLH05 was analyzed.

In all cases, microscopy was performed 5–7 d after transfection. 150 µl samples of cell culture were used, with an optional DAPI staining to visualize DNA, in which samples were stained with 15 µl of 250 µg/ml

DAPI for 1 h before microscopy. Images without DAPI were taken with an Eclipse E800 microscope in combination with a super-high pressure mercury lamp (Nikon; Amstelveen, Netherlands), while images with DAPI were taken with a LSM 700 confocal microscope (Carl Zeiss; Oberkochen, Germany), using the excitation and emission wavelengths shown in **Table 21**. Leica Application Suite or ZEN were used for image processing and analysis, respectively.

Table 21: Wavelengths used for fluorescence markers.

Fluorescence marker	Excitation	Emission
DAPI	405 nm	350–470 nm
GFP	488 nm	488–510 nm
mCherry	555 nm	555–584 nm

2.2.4 *In silico* methods

2.2.4.1 Statistical analyses

Statistical significance was calculated using analysis of variance (ANOVA). In experiments with one independent variable, one-way ANOVA was used in combination with Tukey's post-hoc test. In experiments with multiple independent variables, two-way ANOVA was used to additionally analyze the interaction between variable pairs. Quantitative data is always displayed together with the 95% confidence interval. All statistical calculations were performed in R with RStudio.

2.2.4.2 Phylogenetic analyses

Arabidopsis genomic and transcript data was obtained from TAIR (v10) (Huala *et al.*, 2001). Orthologs of Arabidopsis genes in other Brassicaceae species were detected using the following resources: Phytozome v12.1 (Goodstein *et al.*, 2011) for *C. rubella* (v1.0) (Slotte *et al.*, 2013), provided ortholog data from arabis-alpina.org for *A. alpina* (v5) (Willing *et al.*, 2015) and the BLAST function from chi.mpipz.mpg.de for *C. rubella* (v1.0) (Gan *et al.*, 2016). Phylogenetic trees were generated with the Maximum Likelihood Method from MUSCLE sequence alignments, using MEGA X.

The presence of *MYB76* orthologs in additional Brassicaceae species was verified using Phytozome for *Boechera stricta*, *Capsella grandiflora* and *Eutrema salsugineum*, the aforementioned ortholog data for *Arabidopsis lyrata*, *Arabis montbretiana*, *Leavenworthia alabamica*, *Schrenkiella parvula* and *Sisymbrium irio*, CamRegBase for *Camelina sativa* (Gomez-Cano *et al.*, 2020) and existing publications for *Aethionema arabicum* (Hofberger *et al.*, 2013), *Brassica* spp. (Seo and Kim, 2017) and *Raphanus sativus* (Mitsui et al., 2015).

2.2.4.3 Analysis of Plant Regulomics data

The Plant Regulomics database (Ran *et al.*, 2020) was used to gather transcriptomic data about genes of interest, as well as the binding of TFs and epigenetic factors. *EML1–4* were queried as single genes, while genes of aliphatic and indolic GSL biosynthesis were queried as lists to detect specific enrichments.

3 Results

3.1 Epigenetic regulation of GSL biosynthesis by EML histone readers

3.1.1 *EML1* transcripts are alternatively spliced

The Arabidopsis genome contains four *EML* genes, all of which encode for histone reader proteins comprising one ENT, Agenet and coiled-coil domain each, from N- to C-terminus (Tsuchiya and Eulgem, 2011). The Agenet domain conveys the histone reading activity (Maurer-Stroh *et al.*, 2003; Zhao *et al.*, 2018), while the ENT and coiled-coil domains might respectively contribute to EML function by enabling dimerization and protein-protein interaction. According to TAIR10, the genes *EML1*, *EML2* and *EML3* are annotated with three, two and three alternative splice variants, respectively (Huala *et al.*, 2001). In regard to *EML1*, the variants *EML1.1* and *EML1.2* are identical apart from differences in the UTR lengths. *EML1.3* additionally retains intron 9, which leads to an overall longer ORF, thus qualifying it as the representative gene model.

In addition to these three documented variants, two novel splice variants were isolated from Arabidopsis cDNA by cloning with primers specific for the 5′ and 3′ ends of the *EML1.3* CDS. While their ends are naturally identical to *EML1.3*, they retain additional introns each – the variants retaining intron 2 and 6 were termed *EML1.4* and *EML1.5*, respectively (**Figure 3**). Because these introns have respective lengths of 92 and 121 bp, their retention results in a frameshift and ultimately premature stop codons, thus drastically reducing ORF size. While none of the three domains described for EMLs are affected by the ORF differences in *EML1.1–1.3*, missense sequences begin in the ENT domain in *EML1.4* and towards the end of the Agenet domain in *EML1.5*, thus likely rendering any resulting proteins non-functional.

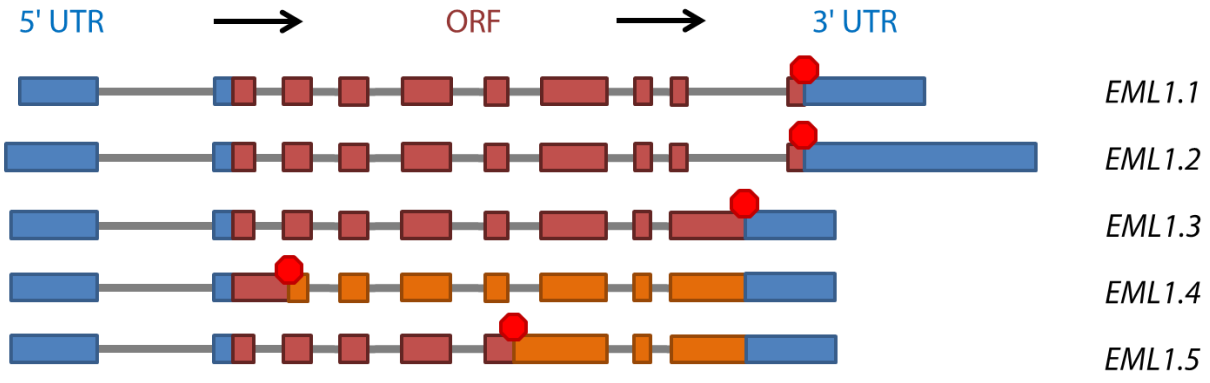


Figure 3: Newly identified *EML1* splice variants *EML1.4* and *EML1.5*.
Schematic representation of the splice variants of *EML1*, including the newly identified *EML1.4* and *EML1.5*. *EML1.3* is the representative gene model with the longest ORF and retains intron 9 relative to *EML1.1* and *EML1.2*. In contrast, *EML1.4* and *EML1.5* retain introns 2 and 6, respectively. In both cases, this leads to a frameshift and a premature stop codon. Thin gray lines indicate introns, while boxes indicate exons and red octagons signify stop codons. UTRs are shown in blue, the translated ORFs in dark red, and the former CDS after stop codons introduced by frameshift in orange.

3.1.2 Omics data implies a potential role of EMLs in epigenetic regulation of GSL biosynthesis

While alternative splicing may present an additional mechanism to regulate EML function, it remains unclear if this function itself extends to the regulation of GSL biosynthesis. To supplement the circumstantial evidence presented earlier, publicly available omics data that was generated under stressed and unstressed conditions was retrieved from Plant Regulomics (Ran *et al.*, 2020), then analyzed in regard to (i) the presence of histone marks at GSL biosynthesis genes and (ii) the expression of *EMLs*.

For the first approach, datasets of binding factor enrichment were generated for lists of the genes responsible for aliphatic and indolic GSL biosynthesis (**Supplementary Table 7**), then filtered for H2AK121ub, H3K27me and H3K4me histone marks (others such as H3K36me were not part of the datasets). For the second approach, transcriptomic data was analyzed individually for all *EML* genes.

H2AK121ub and H3K27me are both repressive histone marks, with H2AK121ub an LHP1-indpendent PRC1 product (Zhou *et al.*, 2017) and H3K27me a PRC2 product (Alvarez-Venegas, 2010; Cheng *et al.*, 2019). In contrast, H3K4me acts as an activating signal (Liu *et al.*, 2010; Cheng *et al.*, 2019). H2AK121ub and H3K27me are enriched in both aliphatic and indolic GSL genes under unstressed wildtype conditions, while H3K4me marks are enriched under drought stress (**Supplementary Tables 8–9**). The H2AK121ub enrichment is slightly more pronounced for the indolic genes, while the H3K27me enrichment seems to be present almost exclusively in aliphatic genes. The H3K4me enrichment under drought stress is more prominent for the aliphatic than for the indolic genes – in the latter, also some unstressed samples appear to be enriched, although at a tendentially lower rate than the stressed ones.

Transcriptomic data for the individual *EML* genes (**Supplementary Figures 1–4**) reveal a differential responsiveness to such osmotic stresses, with *EML1* expression induced by drought and NaCl stress, and both *EML1* and *EML2* expression reduced in seeds imbibed in water for 24 h. On the contrary, *EML3* and *EML4* expression is strongly reduced under NaCl stress (plus mannitol treatment for *EML4*). In regard to other conditions that are known to influence GSL synthesis, there seems to be a certain correlation with *EML* transcription. *EML1*, *EML2* and *EML3* are upregulated under treatment with the bacterial peptide flg22; *EML2* and *EML3* are induced by inoculation with the powdery mildew *Golovinomyces cichoracearum* and Cabbage leaf curl virus, respectively. The only exception seems to be inoculation with the bacterium *Pseudomonas syringae*, which only induces expression of *EML4*, but reduces expression of *EML1* and *EML3*. *EML2* is upregulated under JA/gibberellic acid and ET treatments, while high concentrations of SA downregulate *EML1* and *EML3*. The response to ABA appears to be highly variable in both positive and negative directions, however. Notably, *EML* expression is increased in *bri1*, *bzr1/bri1* and *jaz* mutants, all of which encode proteins that participate in repression of GSL biosynthesis, the first two as components of BR signaling (Guo *et al.*, 2013a).

Taken together, there appears to be at least a partial correlation between *EML* expression and the induction of GSL synthesis, although the individual properties of the *EML* genes vary considerably. The omics data suggests that this induction may be accompanied or even caused by an epigenetic shift in the histone marks of GSL biosynthesis genes, resulting in their repression in an unstressed state and their activation under stress conditions. This putative epigenetic regulation is underlined by the enrichment of LHP1 localization to the aliphatic GSL genes (**Supplementary Table 8**).

3.1.3 EMLs colocalize with bHLH proteins

If EMLs indeed take part in the regulation of GSL biosynthesis, they most likely have to interact with the central MYB-bHLH transcription factor complex that controls this process. To elucidate the connection between these three protein groups, localization and colocalization experiments were conducted.

3.1.3.1 MYB51 is functional as an N-terminal, but not as a C-terminal GFP fusion protein

As a basis for further experiments, the localization of GFP-fusion proteins was examined for C- and N-terminally tagged MYB51 and bHLH05 in Arabidopsis cell cultures. While bHLH05 localized to subnuclear speckles for both C- and N-terminal GFP tags, MYB51 was localized evenly throughout the entire nucleus in both cases (**Figure 4A**).

This prompted us to control the functionality of the MYB51 fusion proteins. To this end, MYB51 was overexpressed in Arabidopsis Ler-0 cell culture without any tag, with a C-terminal and with an N-terminal GFP fusion (i.e. in pGWB2, pGWB5 and pGWB6 vectors). Subsequently, the resulting overproduction of indolic GSLs was measured. Untagged and N-GFP-tagged MYB51 caused a strong induction of I3M accumulation, with concentrations in the range of 30 – 50 µmol/g DW, indicating that the N-terminal tag did not negatively affect MYB51 functionality – especially not in regard to the pivotal MYB domain, which resides proximally to the N-terminus. On the contrary, the I3M level in cells overexpressing C-GFP-tagged MYB51 was indistinguishable from the negative control (**Figure 4B**). This indicates that a C-terminal GFP fusion disturbs MYB function, possibly by interfering with the disordered transcriptional activation domain that is located at the C-terminus (Feller *et al.*, 2011). Consequently, only N-GFP-tagged MYB51 proteins were used for further experiments.

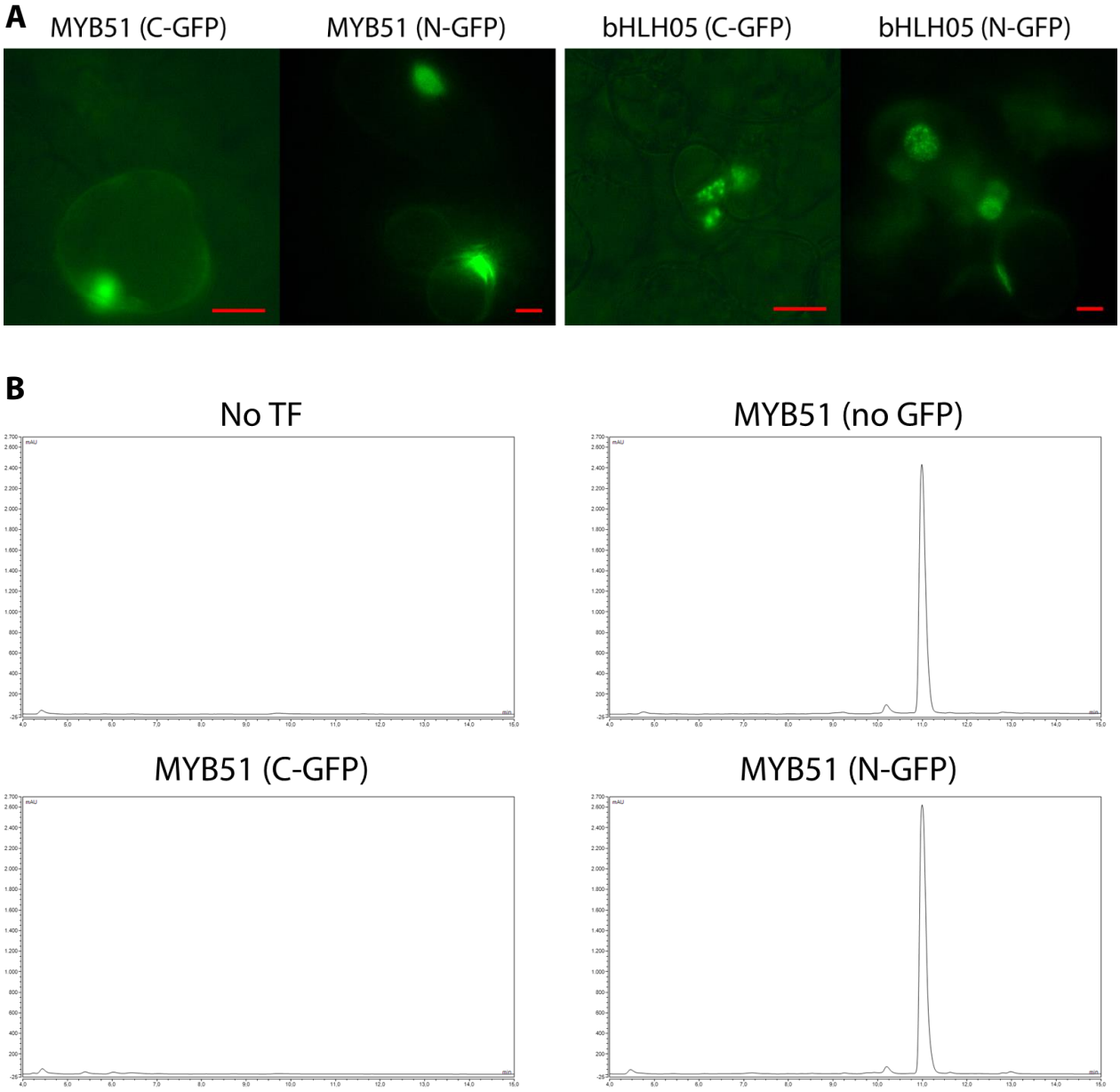


Figure 4: MYB51 is functional with an N-terminal, but not a C-terminal GFP tag.
(A) Localization of C- and N-terminally GFP-tagged MYB51 and bHLH05 proteins by fluorescence microscopy. bHLH05 localizes to subnuclear speckles, while MYB51 is localized evenly throughout the entire nucleus. Overexpression of the fusion proteins was performed using the pGWB5 (C-GFP) and pGWB6 (N-GFP) expression vectors. All images were taken five days after transfection, with the C-GFP images taken in Arabidopsis Ler-0 cell culture and the N-GFP images taken in Arabidopsis *bhlh04/05/06* (bHLH05) or *myb34/51/122* (MYB51) cell culture. The scale bars represent 10 µm. **(B)** Qualitative GSL profiles induced by overexpression of different MYB51 fusion proteins. The x-axis represents the retention time, the y-axis represents the absorbance and the major peak represents the main indolic GSL I3M. Untagged (in the pGWB2 vector) and N-GFP-fused (in pGWB6) MYB51 are able to induce I3M overproduction, while C-GFP-fused (in pGWB5) is indistinguishable from the negative control. GSL concentrations were determined in Arabidopsis Ler-0 cell culture, six days after transfection.

3.1.3.2 bHLH05 and EML2 colocalize in subnuclear structures

Subsequently, the same comparison was conducted for N- and C-terminally GFP-tagged EML constructs in Arabidopsis Ler-0 cell culture, using N-GFP fusions of EML1.3 and EML1.5 and C-GFP fusions of EML1.3 and EML2. The splice variant EML1.5 was included to determine whether its shortened sequence would affect the protein localization.

The N-terminally tagged EMLs localized homogenously throughout the nucleus, whereas the C-terminally tagged EMLs localized to subnuclear speckles, similar to the ones displayed by bHLH05 (**Figure 5A**). Consequently, a colocalization assay was performed using C-GFP-EML2 and C-mCherry-bHLH05 in Arabidopsis Ler-0 cell culture. Here, both proteins were colocalized in the speckles, which, as indicated by DAPI staining, represented regions of high DNA density (**Figure 5B**).

The logical next question that arises from the observed colocalization is if there is a direct protein-protein interaction between bHLHs and EMLs (and possibly MYBs and EMLs). A yeast two-hybrid assay between bHLH04/05/06 and EML1/2/3/4 has been performed by the Grotewold Lab (University of Michigan; unpublished data) and reveals a strong interaction between all combinations of the two protein groups. However, this finding remains to be complemented in the future in co-immunoprecipitation experiments.

3.1.4 JA-mediated induction of GSL synthesis is compromised in *eml1/eml3* double mutants

The previous findings imply an epigenetic regulation of GSL biosynthesis with a potential role of EMLs under stress conditions and demonstrate a direct interaction between EMLs and the bHLH TFs of the GSL transcription complex. This leaves open the question of a direct influence of EMLs on GSL accumulation. As the shift in epigenetic regulation seems to be caused by stress conditions, we focused on the function of EMLs in the JA-mediated induction of GSL biosynthesis – in which indolic GSLs are typically induced more strongly than aliphatic ones (Mitreiter and Gigolashvili, 2020) – to address this topic. To determine if the JA-mediated GSL induction is impaired in *eml* mutant plants, two different sets of mutants were employed, termed set I (**Table 7**) and set II (**Table 8**): After an initial genotyping, GSL induction was quantified, accompanied by measurements of biosynthetic gene expression by qPCR.

Set I encompasses the single mutants *eml1-2*, *eml1-3*, *eml2-1*, *eml3-1*, the double mutant *eml1-2/eml2-1* (*eml1/eml2* for short) and the three *quad* lines (*eml-quad1*, *eml-quad2* and *eml-quad3*) which consist of the *eml1/eml2* double mutant expressing an amiRNA construct that silences *EML3* and *EML4*. All of these lines exist in the Col-0 background (Tsuchiya and Eulgem, 2011). Set II encompasses the single mutants *eml2-1* (genotypically identical to set I), *eml3-4*, *eml4-1* and the double mutant *eml1-2/eml3-4* (*eml1/eml3* for short) in varying Col and Ler backgrounds (Milutinovic *et al.*, 2019).

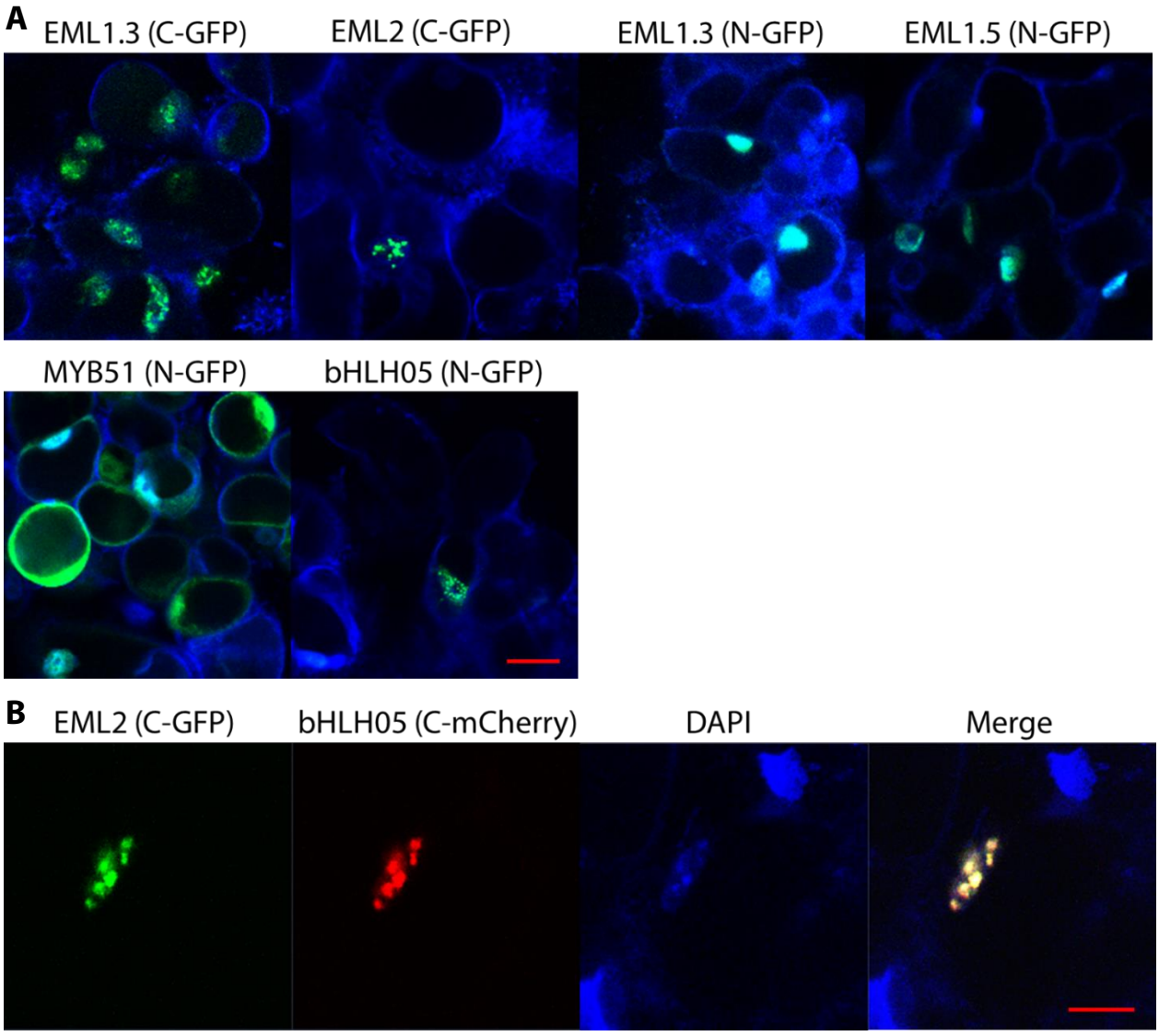


Figure 5: EML2 and bHLH05 colocalize in the same subnuclear structures.
(A) Localization of C- (in the pGWB5 expression vector) and N-terminally (in the pGWB6 expression vector) GFP-tagged EML proteins as well as N-terminally GFP-tagged MYB51 and bHLH05 proteins by fluorescence microscopy. C-GFP EML fusion proteins localize to more finely defined subnuclear structures than N-GFP fusion proteins. C-terminally tagged EMLs and bHLH05 localize to similar subnuclear speckles. DAPI staining (blue) was used to visualize DNA. Images were taken 6 to 7 days after transformation in Arabidopsis Ler-0 cell culture. The scale bar represents 20 µm. **(B)** Colocalization of C-terminally GFP-tagged EML2 (green) and C-terminally mCherry-tagged bHLH05 (red) fusion proteins by fluorescence microscopy. Overexpression of the fusion proteins was performed using the pGWB5 (C-GFP) and pAUBERGINE (C-mCherry) expression vectors. DAPI staining (blue) was used to visualize DNA. Because the colocalization of EML2 and bHLH05 coincides with DAPI, the proteins presumably localize to subnuclear structures with high DNA density. The images were taken six days after transfection in Arabidopsis Ler-0 cell culture. The scale bar represents 10 µm.

The T-DNA, transposon and amiRNA insertions as well as the respective wildtype backgrounds were verified for all mutants by genotyping. One notable finding was that *eml1/eml3*, which exists in a crossed Col-0/Ler-0 background, is homozygous for chromosomes 3 and 5 (which contain the *eml* mutations), but not for chromosome 4. Here, the line segregates into Col-0 and Ler-0 backgrounds, including the *AOP3* locus – a secondary GSL modification enzyme that is expressed in Ler-0, but not in Col-0 (Kliebenstein *et*

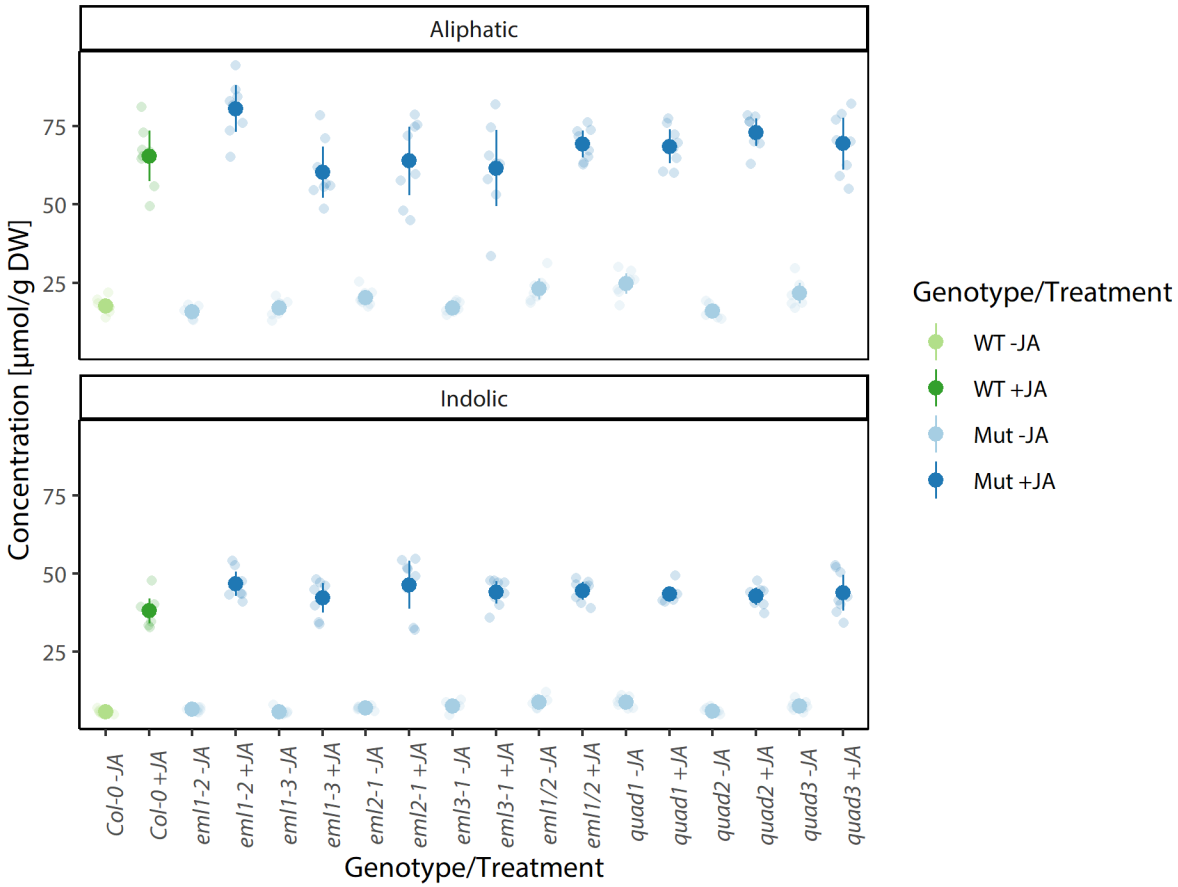
al., 2001). This difference results in a qualitatively changed profile of aliphatic GSLs between different *eml1/eml3* individuals. However, the ectopic expression of *AOP3* in Col-0 was shown to have no effect on the total amount of aliphatic GSL content (Jensen *et al.*, 2015b) and thus did not prevent further experiments. The expression of the amiRNA construct was not verified in this study.

As the first approach to study JA-mediated GSL induction in *eml* mutants, an experiment was conducted under prolonged stress conditions, where plants were grown on JA-containing medium and their GSL levels measured at a final harvest time point. To specifically examine the influence of *eml* mutations on the JA-dependent GSL induction, the data was analyzed with two-way ANOVA, considering the interaction between the genotype (wildtype vs. mutant) and the JA treatment as the main statistical factor. The other factors, i.e. the isolated effects of JA treatment and genotype, were less informative: JA treatment almost always had a highly significant effect on GSL levels, while the genotype by itself showed no notable differences to the significance levels of the JA-genotype interaction (**Supplementary Table 10**). This probably results from the observed pattern that GSL concentrations in *eml* mutant lines are only lowered relative to the wildtype under JA-stressed, but not under unstressed conditions. Thus, only the significance of the JA-genotype interaction was considered in the following paragraphs.

While set I and set II overlap in regard to the *eml2-1* line, they consist of otherwise different lines. Even though partially, the same genes are mutated, the T-DNA insertions are located at different sites and thus have different effects on EML function. Additionally, the amiRNA-mediated knockdowns of *EML3* and *EML4* in set II *quad* lines represent a fundamentally different mechanism of interfering with gene function than the T-DNA-mediated knockdowns. As a result, different observations were made between the two mutant sets: While a reduction of JA-mediated GSL induction could be detected in some lines belonging to set II, no such effect was observed in the set I lines.

In set I, JA-mediated GSL induction was not significantly impaired in any of the *eml* mutants – in fact, there was no notable decrease in either aliphatic or indolic GSL accumulation in any mutant (**Figure 6**). Interestingly, *eml1-2* and *eml2-1* exhibited slightly increased concentrations of indolic GSLs upon JA induction; *eml1-2* also of aliphatic GSLs. These effects contrast with the *eml1/eml2* double mutant as well as the three *quad* mutants, as all of these lines contain both of these mutations, but do not show any significant difference in comparison with the wildtype. This contradiction could theoretically be explained by epistatic effects between the mutant alleles or a potential metabolic feedback mechanism.

A



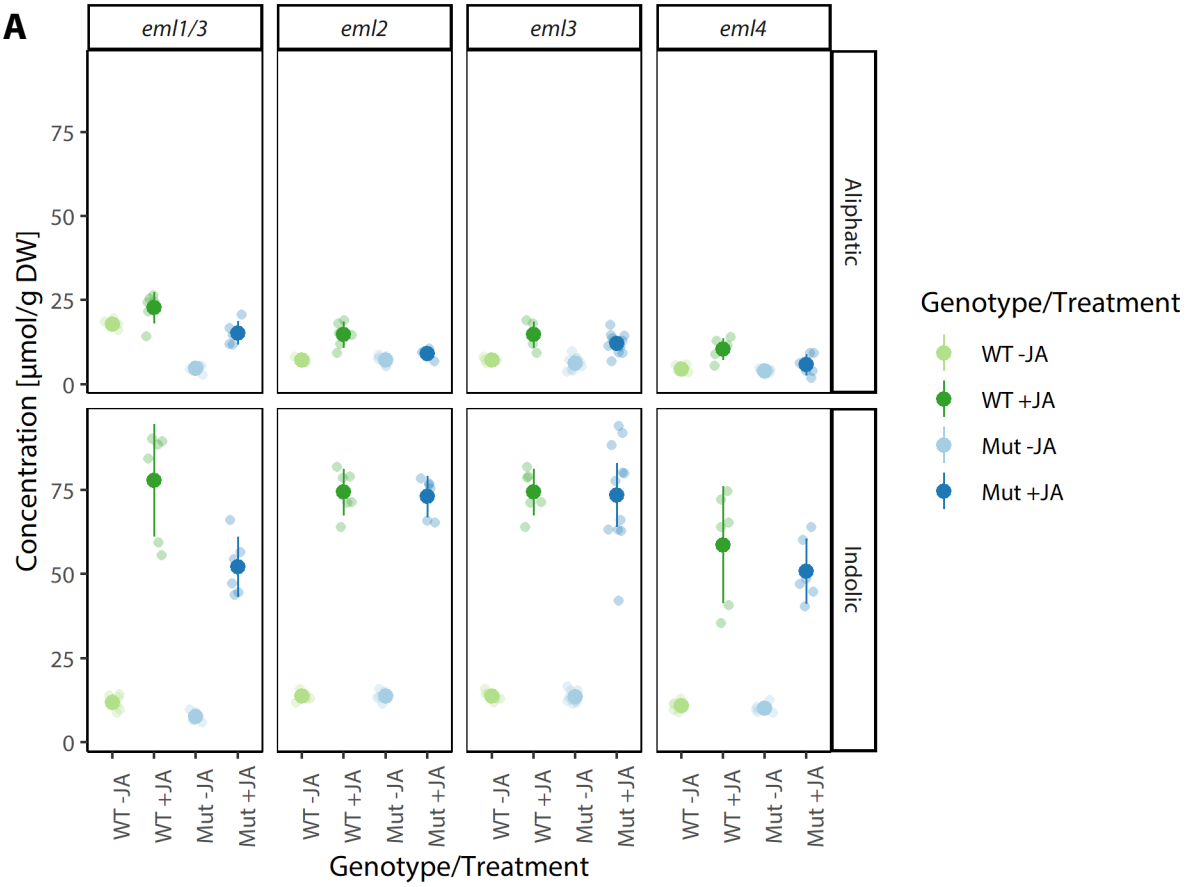
B

	eml1-2	***eml1-3***	***eml2-1***	***eml3-1***	***eml1/ eml2***	***quad1***	***quad2***	***quad3***
Aliphatic	** (0.0011)	- (0.38)	- (0.42)	- (0.52)	- (0.75)	- (0.43)	. (0.076)	- (0.99)
Indolic	** (0.0056)	- (0.15)	* (0.011)	- (0.15)	- (0.22)	- (0.41)	- (0.12)	- (0.15)

Figure 6: Set I *eml* mutants are not impaired in JA-mediated GSL induction.
(A) GSL quantification of set I *eml* mutants grown under JA presence. Error bars represent the 95% confidence interval. *eml1/eml2* is abbreviated to *eml1/2*. **(B)** Statistical significance of differences in GSL concentration between Col-0 wildtype and mutant plants, regarding the interaction between genotype (wildtype or mutant) and jasmonate presence. p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

With set II, four independent repetitions of the prolonged stress experiment were performed in total. However, the variation between data sets is too great to combine them for a statistical analysis. This might result from slight differences in cultivation conditions or technical reasons such as the quality of the HPLC columns used for analysis. One representative dataset is presented here, while the others can be found in **Supplementary Figures 5–7**. All repetitions displayed similar tendencies, however.

Among the set II *eml* mutants, the *eml1/eml3* double mutant displayed significantly reduced induction of both aliphatic and indolic GSLs by JA treatment (**Figure 7**). Among the single mutants, an analogous trend of reduced JA-mediated GSL induction can be observed, especially for *eml4-1* (*eml4* for short). The effect is however only statistically significant for aliphatic GSLs in *eml4* and *eml2-1* (*eml2* for short). Taken together, the prolonged stress experiments demonstrate an impaired GSL induction upon JA treatment in *eml* mutants, whereby this effect is highly dependent on the total reduction of *EML* expression levels.



B

	eml1/eml3	***eml2***	***eml3***	***eml4***
Aliphatic	* (0.026)	** (0.0041)	- (0.41)	* (0.049)
Indolic	** (0.0095)	- (0.73)	- (0.92)	- (0.39)

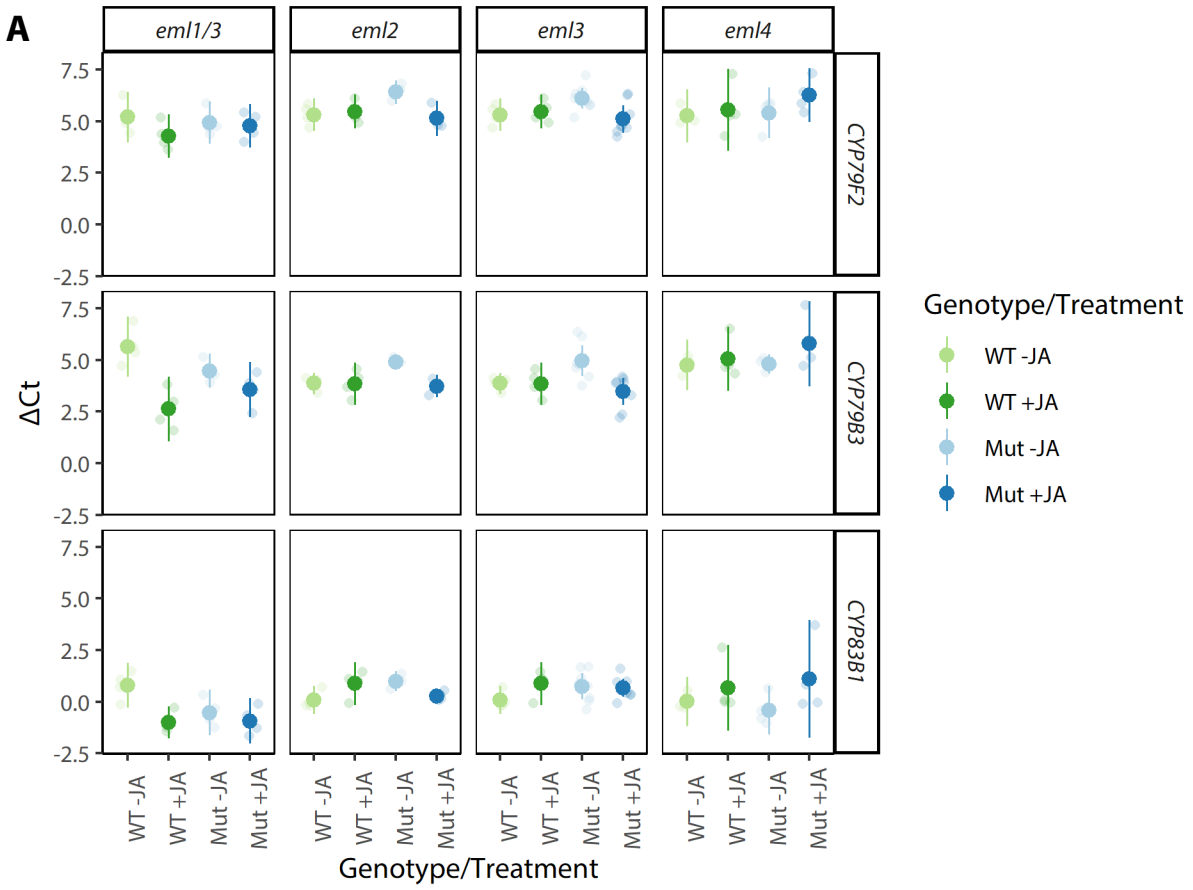
Figure 7: The set II *eml1/eml3* double mutant is impaired in JA-mediated GSL induction.
(A) GSL quantification of set II *eml* mutants grown under JA presence. Error bars represent the 95% confidence interval. *eml1/eml3* is abbreviated to *eml1/3*. **(B)** Statistical significance of differences in GSL concentration between mutant plants and their respective wildtypes, regarding the interaction between genotype (wildtype or mutant) and jasmonate presence. p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

If *eml* mutants produce less GSLs under JA-stressed conditions, GSL biosynthesis rate must be reduced, likely through a relative reduction in expression of GSL biosynthetic enzymes. To determine if the reduced metabolite levels observed in the second set of mutants correspond to transcript levels of GSL enzymes, a qPCR measurement was performed for three central genes: The core structure synthesis enzymes *CYP79F2* from the aliphatic and *CYP79B3* and *CYP83B1* from the indolic pathway. The samples used for this experiment were harvested simultaneously with those for the second prolonged stress experiment. Statistical significance was assessed in the same way as for the GSL concentrations, using ΔC_t values as the main measure.

Overall, the trends in *CYP* expression only partially corresponded with GSL concentrations (**Figure 8**). The *eml2* and *eml3* mutants mostly behaved significantly different from the wildtype, but this effect seemed to stem from a reduced *CYP* expression under unstressed, not under stressed conditions. Notably, these trends could not be confirmed in another repetition of this experiment. In contrast, *eml1/eml3*, the only mutant to exhibit a clear impairment in JA-mediated GSL induction, consistently displayed a reduced induction of *CYP* expression under JA conditions. This trend, however, is only significant for *CYP79B3*, which reflects the greater sensitivity of indolic GSLs to JA-mediated induction. Here and especially in the single mutants, a notable amount of uncoupling between *CYP* expression and GSL production occurs, pointing towards further regulatory mechanisms.

3.1.5 The compromised JA-mediated GSL induction affects the priming response in *eml1/eml3* mutants

The priming of defense responses is an adaptive strategy in plants, where upon the experience of stress defense mechanisms are set to an enhanced prime state. This means that in anticipation of further stresses, the defense responses become stronger and/or faster (Kim and Felton, 2013; Conrath *et al.*, 2015; Mauch-Mani *et al.*, 2017). The JA-mediated induction of GSLs is subject to priming, with some evidence for an epigenetic mechanism (Rasmann *et al.*, 2012; Bakhtiari *et al.*, 2018). The previous experiment demonstrated that this GSL induction is impaired in *eml* mutants – thus, I aimed to address the question how this effect relates to the priming response, i.e. if EMLs specifically play a priming-related role. To this end, two experiments were conducted in which plants were repeatedly exposed to JA for a short time of 24 h, after which the induced GSL levels were measured each time, with an additional time course in the second experiment.



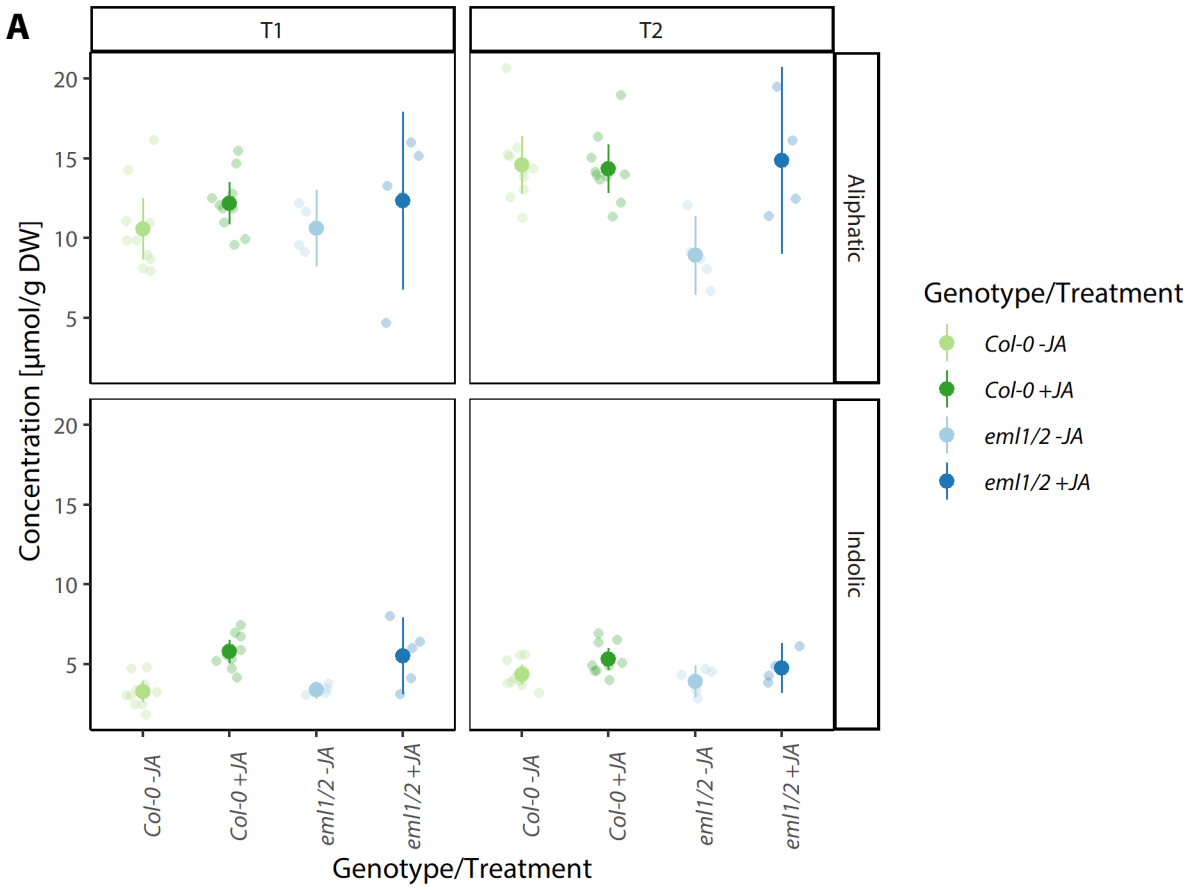
B

	eml1/eml3	***eml2***	***eml3***	***eml4***
CYP79F2	- (0.30)	* (0.011)	. (0.055)	- (0.56)
CYP79B3	* (0.026)	* (0.013)	* (0.036)	- (0.48)
CYP83B1	. (0.052)	** (0.0043)	- (0.12)	- (0.52)

Figure 8: Impairment of JA-mediated GSL induction in set II *eml* mutants is mostly not reflected on the transcript level. **(A)** Expression of GSL biosynthesis genes in set II *eml* mutants grown under JA presence. Error bars represent the 95% confidence interval. *eml1/eml3* is abbreviated to *eml1/3*. **(B)** Statistical significance of differences in GSL concentration between mutant plants and their respective wildtypes, regarding the interaction between genotype (wildtype or mutant) and jasmonate presence. p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

3.1.5.1 Induction of indolic GSLs is reduced by priming

In the first experiment, both *eml1/eml2* and *eml1/eml3* plants were compared to their respective wildtypes, Col-0 and Col-0/Ler-0. To interpret the data, three factors have to be taken into account in relation to the GSL content: the time point of measurement (TP; T1 or T2), the JA treatment (JA; with or without) and the genotype (GT; wildtype or mutant). Different questions can be answered by considering the interaction between different pairs of factors. Due to the differences between the mutants, it is useful to consider *eml1/eml2* and *eml1/eml3* separately, beginning with *eml1/eml2* (**Figure 9**).



B

	TP-JA	**JA-GT**
Aliphatic	- (0.91)	. (0.067)
Indolic	** (0.0083)	- (0.71)

Figure 9: The *eml1/eml2* double mutant is not impaired in JA-mediated priming of GSL induction.
(A) GSL quantification of the *eml1/eml2* mutant grown under repeated short JA stimuli. T1, 7 days after the initial induction; T2, 14 days. The induction of indolic GSLs is reduced at T2; Col-0 WT and *eml1/eml2* mutant plants do not behave differently. Error bars represent the 95% confidence interval. *eml1/eml2* is abbreviated to *eml1/2*. **(B)** Statistical significance of differences in GSL concentration, regarding the interaction between different factors: TP, time point (T1 or T2); JA, jasmonate presence; GT, genotype (wildtype or mutant). p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

The first question to consider is if plants produce different amounts of GSLs in response to JA compared between T1 and T2, i.e. if a priming response occurs in response to JA at all. For this, the TP-JA interaction was analyzed. It was indeed significant for indolic GSLs, although the direction of the priming effect was unexpected, with an overall reduced JA-mediated induction at the later time point. The aliphatic GSL content seemed to be more volatile and did not exhibit a priming effect.

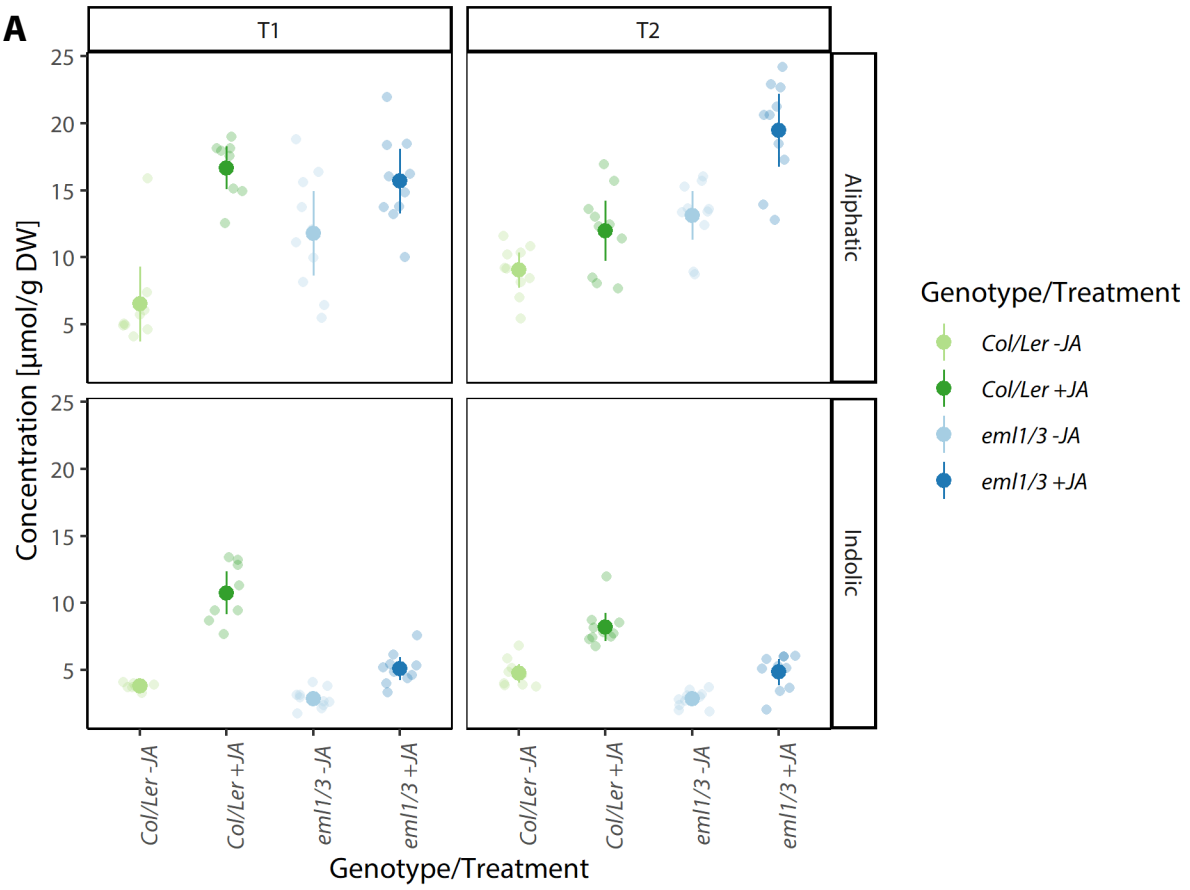
The second question to consider is if any difference exists in how the genotypes (wildtype and mutant) reacted to JA. This was analyzed via the JA-GT interaction, which also represents a control for consistent

effects between the priming and prolonged stress experiments. As *eml1/eml2* did not display any significant effect here, the mutations consistently had no influence on JA-mediated GSL induction, regardless of the experimental conditions. Thus, no further information can be drawn from experiments with this plant line.

The *eml1/eml3* mutants can then be analyzed in the same way as *eml1/eml2* (**Figure 10**). First, is there a priming response that alters the JA-dependent GSL induction from T1 to T2? Again, the induction was reduced for indolic GSLs over time; the overall TP-JA interaction was only mildly significant, however. Second, is there a difference between the amount of GSLs produced in response to JA between the wildtype and mutant? The JA-GT interaction was statistically significant for indolic, but not aliphatic GSLs. This means that although the experimental conditions were different, *eml1/eml3* plants exhibited the same tendency for impaired JA-mediated GSL induction as in the prolonged stress experiment.

This significant difference between *eml1/eml3* mutant and wildtype enables a more precise analysis than for *eml1/eml2*. Employing the TP-GT interaction, the observed effect of reduced JA-mediated GSL induction can be broken down specifically in regard to the genotype. For this, the samples with and without JA treatment were analyzed separately, so that the latter ones can act as a control. When considering these control samples, there was still a significant trend detectable for indolic GSLs, namely an increase over time, which could occur as a result of plant ageing.

In the JA-treated samples, other effects were dominant in this regard. In *eml1/eml3*, indolic GSL levels remained constant from T1 to T2, whereas in Col-0/Ler-0, the JA-mediated induction was notably attenuated for both indolic and aliphatic GSLs. Unexpectedly, the amount of JA-induced aliphatic GSLs in *eml1/eml3* significantly increaseed from T1 to T2. The underlying reason for this effect remains unclear, but might potentially be connected to epistatic effects between indolic and aliphatic GSL production.



B

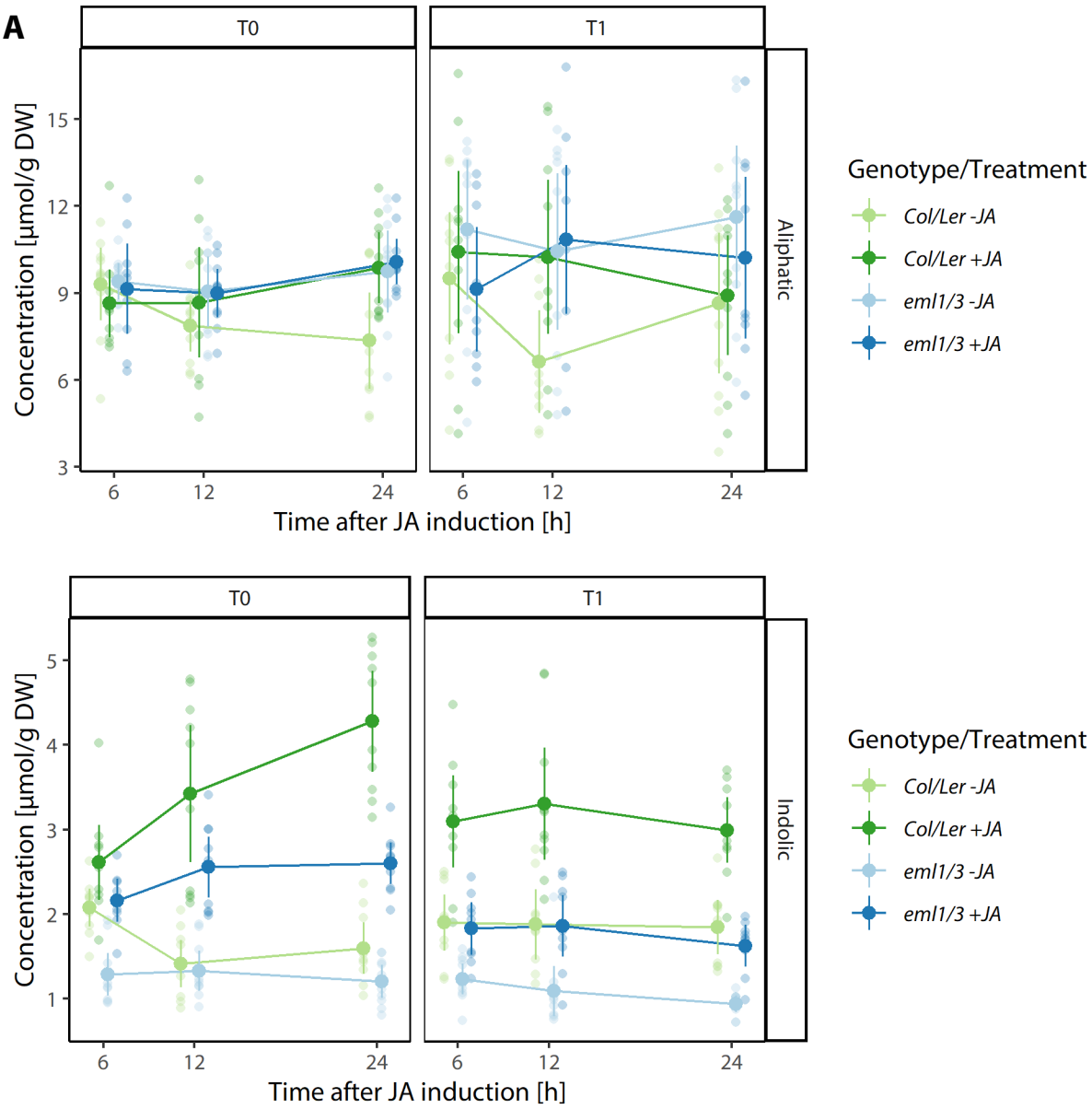
	TP-JA	JA-GT	TP-GT (for -JA)	TP-GT (for +JA)
Aliphatic	- (0.23)	- (0.46)	- (0.57)	*** (0.00019)
Indolic	. (0.075)	*** (6.3e-06)	* (0.036)	* (0.025)

Figure 10: The *eml1/eml3* double mutant is impaired in JA-mediated priming of indolic GSL induction.
(A) GSL quantification of the *eml1/eml3* mutant grown under repeated short JA stimuli. T1, 7 days after the initial induction; T2, 14 days. The induction of indolic GSLs is reduced at T2 in the Col-0/Ler-0 WT, while it remains low at both time points for the *eml1/eml3* mutant. Error bars represent the 95% confidence interval. Col-0/Ler-0 and *eml1/eml3* are abbreviated to Col/Ler and *eml1/3*, respectively. **(B)** Statistical significance of differences in GSL concentration, regarding the interaction between different factors: TP, time point (T1 or T2); JA, jasmonate presence (-JA, no jasmonate; +JA, 17 µM jasmonate); GT, genotype (wildtype or mutant). p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

3.1.5.2 Induction of indolic GSLs is accelerated by priming

The second priming experiment was designed with three key differences to the first one: First, only the genotypes *eml1/eml3* and Col-0/Ler-0 were used. Second, instead of the later time point T2, the initial induction T0 was observed, as the effects on GSL induction might be greater here. Third, an additional time course was introduced for both of the T0 and T1 time points, with measurements performed 6, 12 and 24 h after induction, enabling an observation of the temporal dimension of the priming response in addition to the quantitative one (**Figure 11**).

A



B

	TP-JA	TP-TAI (for -JA)	TP-TAI (for +JA)	JA-GT
Aliphatic	- (0.78)	- (0.54)	- (0.21)	* (0.016)
Indolic	* (0.019)	- (0.65)	* (0.018)	*** (0.00055)

	TP-GT (for -JA)	TP-GT (for +JA)	TAI-GT (for T0, -JA)	TAI-GT (for T0, +JA)	TAI-GT (for T1, -JA)	TAI-GT (for T1, +JA)
Aliphatic	- (0.11)	- (0.91)	- (0.12)	- (0.97)	- (0.60)	- (0.50)
Indolic	* (0.013)	- (0.18)	* (0.010)	* (0.024)	- (0.62)	- (0.90)

Figure 11: JA-mediated priming accelerates indolic GSL induction.
(A) GSL quantification of the *eml1/eml3* mutant grown under repeated short JA stimuli. T0, initial induction; T1, 7 days after the initial induction. The induction of indolic GSLs is reduced at T1 in regard to the concentration, but accelerated in regard to the speed. The overall capacity for GSL induction is severely compromised in the *eml1/eml3* mutant. Error bars represent the 95% confidence interval. Aliphatic and indolic GSL concentrations are mapped on different scales to better visualize the time course. Col-0/Ler-0 and *eml1/eml3* are abbreviated to Col/Ler and *eml1/3*, respectively. **(B)** Statistical significance of differences in GSL concentration, regarding the interaction between different factors: TP, time point (T0 or T1); JA, jasmonate presence (-JA, no jasmonate; +JA, 25 µM jasmonate); GT, genotype (wildtype or mutant); TAI, time after induction. p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

Again, the TP-JA interaction was used as a parameter for the presence of a priming effect between time points, i.e. does the JA-induced GSL concentration differ between T0 and T1? Despite the experimental differences, the results were very similar to the previous experiment: Strikingly, indolic GSLs were induced to a lesser degree at T1 than at T0.

Due to the additional variable of time after induction (TAI), this quantitative analysis of the priming response can be complemented by a temporal one. For this, the shape of the time course in relation to the time point was analyzed via the TP-TAI interaction, separating the samples between JA control and treatment. Unsurprisingly, there was no significant effect in the untreated control. Among the JA-treated plants, no significant effect existed for aliphatic GSLs either, although GSL concentrations over time seemed to be more volatile at T1. In contrast, there was a striking difference between the time courses of indolic GSLs at T0 and T1, with the JA-mediated induction becoming much faster in T1, so that the maximal GSL levels were already reached after 6 h, regardless of the genotypes.

With both a quantitative and a temporal priming effect present, it becomes interesting to examine the influence of the *eml* mutations on the priming response. Analysis of the JA-GT interaction reveals a significant difference in how *eml1/eml3* mutants reacted to JA treatment compared to the Col-0/Ler-0 wildtype, with a weaker effect for aliphatic and a stronger effect for indolic GSLs. This pattern is consistent with all previous observations.

To determine how the genotype relates to the quantitatively different GSL levels between T0 and T1 (i.e. the overall concentrations regardless of the time course shape), the TP-GT interaction was analyzed. Again, JA-treated and -untreated samples were considered separately. In the untreated control, indolic GSL levels again increased with the age of Col-0/Ler-0 plants, but this effect was absent in *eml1/eml3*. Aliphatic GSLs might exhibit an age-dependent increase for both genotypes, although this effect was obscured by diurnal variation and thus not statistically significant. In the JA-treated samples, aliphatic GSL concentrations showed a slight overall increase from T0 to T1, potentially either due to ageing or epistasis with indolic GSLs, although this interaction was not statistically significant. Due to the strongly differing time courses, there was no significant effect for the indolic GSLs either. However, the amount of GSLs after 24 h (i.e. the same time after induction as in the previous experiment) was clearly attenuated from T0 to T1.

To address the connection between *eml* mutations and the temporal aspect of the priming, i.e. the time course of GSL concentration, the data points were separated both according to JA treatment and to time point (T0 or T1), after which the TAI-GT interaction was analyzed. Here, no significant differences between genotypes were detected in regard to aliphatic GSLs. In regard to indolic GSLs however, statistically significant effects occured at T0: In the JA-untreated control, wildtype plants exhibit a notably increased GSL content after 6 h (which then vanished), while mutant plants did not. JA-treated plants displayed a much more notable difference in the time course of GSL induction: Here, indolic GSL levels continually increased from 6 to 24 h after induction in the wildtype, while they were not only drastically lower, but also already stagnated after 12 h in *eml1/eml3*. At T1, the time courses of GSL induction were highly similar between wildtype and mutant both in the presence and in the absence of JA. This may however be the result of the accelerated induction in JA-treated plants. If shorter time frames under 6 h had been measured, a difference between genotypes might have been present in a similar way as at T0.

In summary, a clear priming of indolic GSL induction was detected in all experiments, both on a quantitative and a temporal level. While the induction was accelerated over time as expected, it is highly interesting that the amount of induced GSLs decreased, and not increased, with repeated stimuli. In contrast, aliphatic GSLs were only mildly affected by JA induction and thus were not subject to a comparable priming response to indolic GSLs. In *eml1/eml3* plants, the JA-mediated induction of indolic GSLs was severely compromised, both under prolonged and repeated short-term stress conditions. As the quantitative induction decreased over the course of priming, so did the effect of the mutations, thus making it difficult to identify a priming-specific effect of *eml* mutations on GSL concentrations.

3.1.6 Overexpression of *EML* genes profoundly affects GSL production

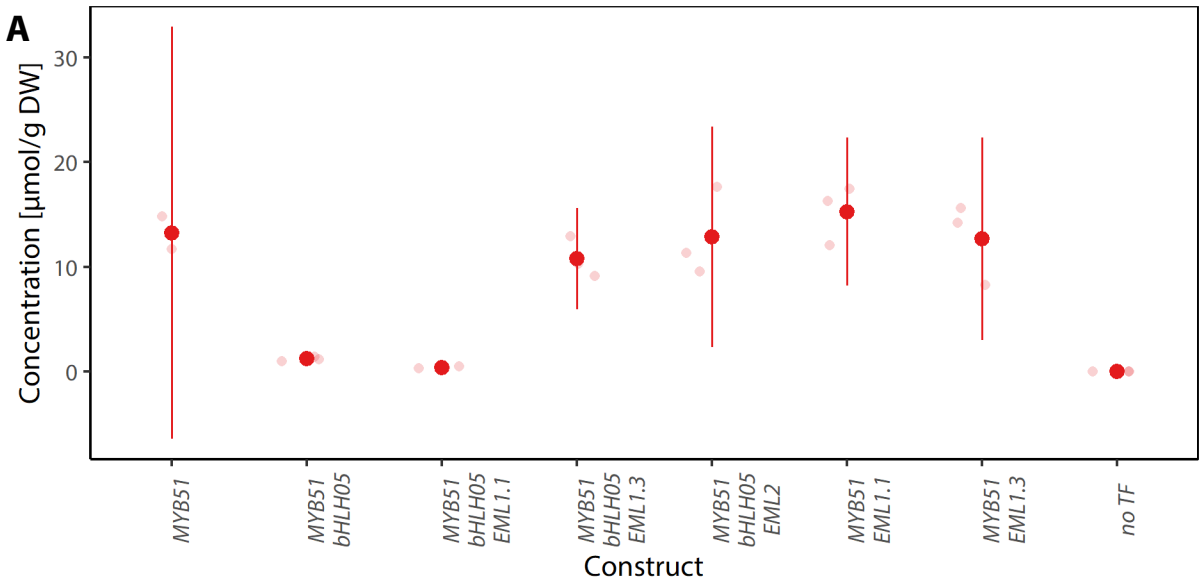
The analysis of *eml* mutants revealed a clear impairment of JA-mediated GSL induction. To further characterize EML function, I studied the effect of the overexpression (OX) of *EML* genes on GSL biosynthesis. To this end, *EML* constructs were overexpressed either alone or together with *MYB51* and *bHLH05,* i.e. other components of the transcription complex. The potentially non-functional splice variant *EML1.1* was also included to assess its functionality. These overexpressions were performed in Arabidopsis Ler-0 cell culture and followed by quantification of the resulting indolic GSL concentrations. Additionally, qualitative assessments of the transcriptional level, i.e. the expression of GSL biosynthesis genes, were made via GUS staining, using the promoter-reporter construct pGWB3-proCYP79B3.

While overexpression of *EMLs* alone had no influence on GSL levels (data not shown), a combination with other TFs led to remarkable differences (**Figure 12A**). However, interpretation of the data is complicated by the putative existence of a bHLH-mediated feedback mechanism that decreases GSL production at high metabolite levels (Bahlmann, 2020). In accordance with this mechanism, the OX of *MYB51* alone led to a significant increase in GSL levels, while the additional OX of *bHLH05* lowered them again, to similar levels

as found in cells expressing no TFs. The combined expression of *EML* genes together with *MYB51* did not significantly change GSL concentrations, although this picture changed when *EMLs* were co-expressed with both *MYB51* and *bHLH05.* The combination of *MYB51*, *bHLH05* and *EML1.3* or *EML2* raised the GSL concentration back to the level of the isolated *MYB51* OX. However, when *MYB51* and *bHLH05* were combined with *EML1.1*, the GSL concentration remained unchanged relative to *MYB51* and *bHLH05*.

This implies that while EMLs cannot induce GSL biosynthesis on their own, the full-length EMLs can interact with bHLH05 in a manner that influences GSL synthesis – while the splice variant EML1.1 cannot exert the same function. The intensity of the GUS staining closely reflected GSL concentrations, with *MYB51* OX resulting in a uniquely strong coloration (**Figure 12B**).

It is important to note that the induction of GSL production by the MYB-bHLH complex, especially in combination with the feedback mechanism, is a highly dynamic process that can drastically change over time. Therefore, future research should include overexpressing the constructs used in this experiment in the existing Arabidopsis *bhlh04/05/06* cell culture, where the lack of native bHLH factors would mitigate the influence of the feedback mechanism, potentially enabling a clearer view of EML interaction with the MYB-bHLH complex and its influence on GSL concentrations.



B

	MYB51	***MYB51 bHLH05***	***MYB51 bHLH05 EML1.1***	***MYB51 bHLH05 EML1.3***	***MYB51 bHLH05 EML2***	***MYB51 EML1.1***	***MYB51 EML1.3***	**no TF**
Significance	b	a	a	b	b	b	b	a

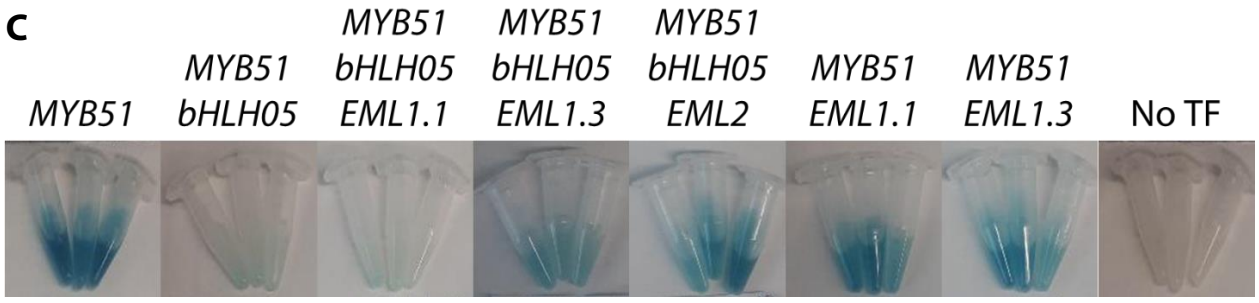


Figure 12: *EML* overexpression overcomes the bHLH-mediated feedback effect in indolic GSL synthesis.
(A) GSL quantification of Arabidopsis Ler-0 cell culture overexpressing *MYB, bHLH* and *EML* constructs, five days after transfection. Error bars represent the 95% confidence interval. **(B)** Statistical significance levels of differences in GSL concentration between the samples. p-values were calculated using one-way ANOVA and Tukey's post-hoc test. **(C)** Qualitative GUS staining with the promoter-reporter construct *proAtCYP79B3:GUS*.

3.2 Exploring the regulatory basis of GSL variation

There is a large variation in GSL profiles between Brassicales species (Bell, 2019; Blažević *et al.*, 2020). To obtain an exemplary overview over differences in GSL accumulation and to examine how those might be influenced by variation in the MYB- and bHLH-mediated transcriptional regulation of GSL biosynthesis, Arabidopsis was compared with three other Brassicaceae species: *Arabis alpina*, *Capsella rubella* and *Cardamine hirsuta*. These species were selected on the basis of varying degrees of relatedness to Arabidopsis (**Figure 13**) and the availability of genomic resources.

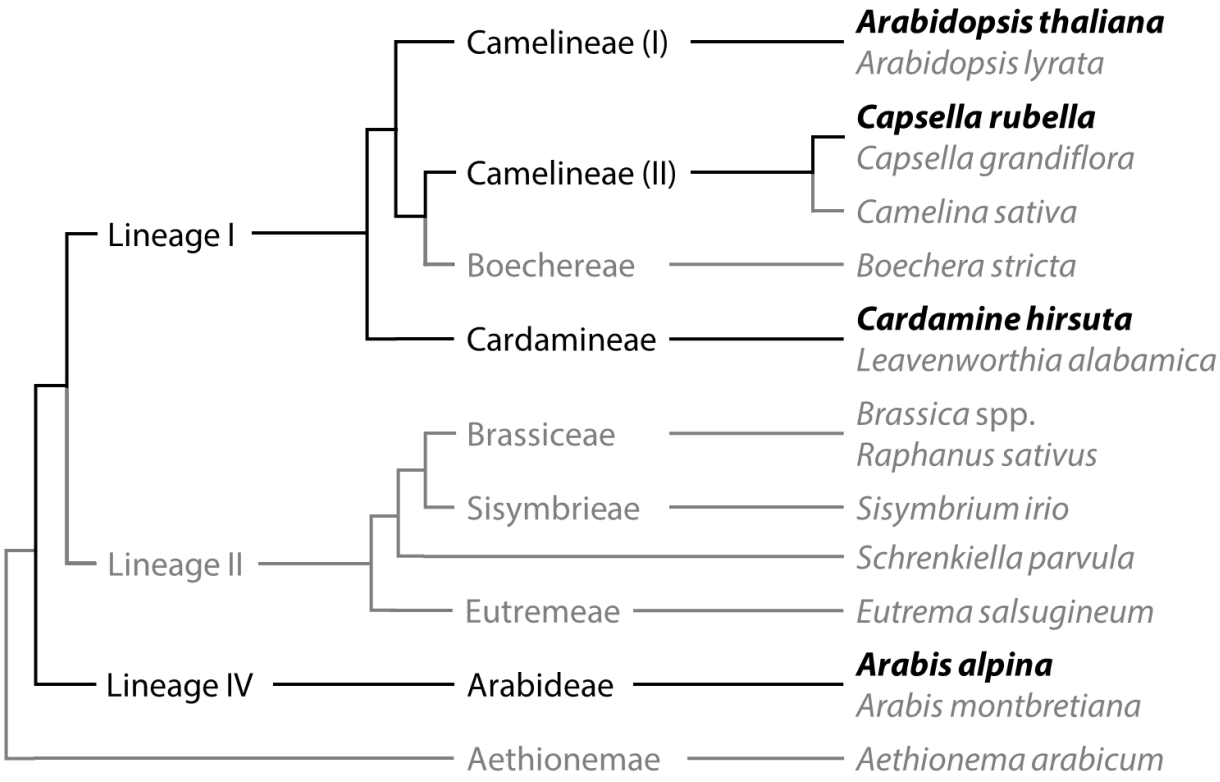


Figure 13: Phylogenetic relationship between the Brassicaceae species used in this study.
Simplified representation of Brassicaceae phylogeny after Nikolov *et al.* (2019). The first taxonomic level represents the lineage, the second level the tribe and the third the species. Species that were used experimentally are marked in black and bold, while species used only for phylogenetic comparisons are marked in gray. The Camelineae tribe does not represent a monophyletic group and is split into clades I and II.

3.2.1 Establishment of Brassicaceae cell cultures

Cell cultures provide a flexible system to analyze a high amount of different transformations without dependence on the plant life cycle, i.e. there is no time-intensive selection process and the cells always exist in the same developmental stage, facilitating analysis. Thus, I aimed at establishing root cell-derived callus and suspensions cell cultures for *A. alpina*, *C. rubella* and *C. hirsuta.*

C. rubella was cultivated under the standard conditions previously described for Arabidopsis cell culture, including 3 mg/l 2,4-D and 1 mg/l kinetin for the callus culture and 1 mg/ml 2,4-D for the suspension culture (**Table 4**). Here, the synthetic auxin 2,4-D supports cell dedifferentiation, while the cytokinin kinetin supports cell growth (Li *et al.*, 2012). Calluses had already grown robust and homogenous enough for transfer into suspension culture after three generations. The suspension culture appeared as a mixture of free cells suspended in the medium and lumps of sizes up to 2–3 mm.

In contrast, tissue cultures of *A. alpina* grew significantly more slowly and formed hardened calluses that exhibited oxidative browning (Jones and Saxena, 2013). The hardening was overcome by growing the first callus generations on modified MS Kallus medium with only 70% of the regular agar content, while growth could be improved by the addition of 2.5 g/l casein hydrolysate (Duchefa; Haarlem, Netherlands). The browning was managed by removing affected parts of the callus during each transfer. To accelerate the growth of the suspension culture, an increased 2,4-D concentration of 2 mg/ml was possible, but not required.

C. hirsuta calluses redifferentiated into root tissue under standard conditions of 3 mg/l (13.5 µM) 2,4-D. To prevent this development, 2,4-D concentration was increased, in accordance with a report by Maeda *et al.* (2008), which describes increased callus formation in *Cardamine yezoensis* for higher 2,4-D concentrations. Consequently, conditions of 3.5 mg/l (16 µM), 4.4 mg/l (20 µM) and 6 mg/l (27 µM) 2,4-D were tested. Here, redifferentiation was inhibited, but simultaneously, growth ceased so that further propagation of the calluses was impossible. Neither the substitution of 2,4-D by 5 mg/l (27 µM) NAA, another synthetic auxin, nor the addition of casein hydrolysate was able to prevent redifferentiation or to induce growth of undifferentiated calluses, respectively. Therefore, the establishment of a *C. hirsuta* cell culture was not possible.

3.2.2 Mass spectrometry reveals GSL diversity of Brassicaceae species

While the GSL profiles for different Arabidopsis tissues are already well-described (Brown *et al.*, 2003), this is only partially the case for the three other Brassicaceae species used in this study. Therefore, root and shoot tissue samples were analyzed for GSL content by LC-MS to establish a knowledge basis for further experiments (**Table 22**). Quantifying the different GSL profiles is the prerequisite for the identification of notable differences, which then in turn can be attributed to changes either in the biosynthetic or in the regulatory machinery.

Table 22: GSL profiles of Brassicaceae root and shoot tissue.
Overview of the GSLs detected in LC-MS analysis of root and shoot tissue from *A. alpina* (Aa), *C. hirsuta* (Ch) and *C. rubella* (Cr). GSLs are sorted by type and mass/charge (m/z) ratio. Individual compounds are identified by a combination of m/z and retention time (RT) values. Possible identities for these molecules are provided with full name, number according to Blažević *et al.* (2020), type of GSL and the amino acid of origin. Concentrations are provided as an area ratio relative to the internal standard, which had a concentration of 0.5 µmol/g FW. GSLs in bold were verified by comparison to standards. A, aliphatic; BOA, benzoyloxyalkyl; HOA, hydroxyalkyl; MSAA, methylsulfanylalkyl; MSIA, methylsulfinylalkyl; MSOA, methylsulfonylalkyl; B, benzenic; I, indolic; n.d., not detected. Standard deviations for the concentration values are provided in **Supplementary Table 11**.

m/z ratio	RT (min)	Possible Identities	Number	Type	Derived from	Relative concentrations					
						Aa root	Aa shoot	Ch root	Ch shoot	Cr root	Cr shoot
360.0429	6	1-Methylethyl	56	A-Alkyl	Val	4,33E-02	8,40E-01	traces	traces	traces	1,15E-05
374.0585 a	6,8	1-Methylpropyl; 2-Methylpropyl; n-Butyl[1]	61, 62, 13[1]	A-Alkyl	Ile, Leu, ?	2,38E-01	1,15E+00	3,10E-04	traces	traces	3,18E-05
374.0585 b	7	1-Methylpropyl; 2-Methylpropyl; n-Butyl[1]	61, 62, 13[1]	A-Alkyl	Ile, Leu, ?	2,05E-02	7,60E-01	traces	traces	traces	1,57E-05
402.0898	9,05	n-Hexyl[1]; 3-Methypentyl[1]; 4-Methylpentyl[1]	20[1], 58[1], 59[1]	A-Alkyl	?, Ile, Leu	4,30E-03	3,74E-02	3,38E-03	1,69E-01	traces	n.d.
416.1050	10	n-Heptyl	–[3]	A-Alkyl	?	traces	1,26E-02	2,17E-03	6,82E-02	2,19E-03	n.d.
372.0429	6,4	But-3-enyl	12	A-Alkenyl	Met	3,36E-03	3,67E-01	1,28E-02	5,83E-01	traces	6,04E-04
386.0585	7,4	Pent-4-enyl; 3-Methylbut-3-enyl	101, 52[1]	A-Alkenyl	Met, Leu	traces	traces	1,91E-03	9,80E-02	n.d.	n.d.
494.0796	9,7	4-(Benzoyloxy)butyl; 1-(Benzoyloxymethyl)propyl[1]	5, 7[1]	A-BOA	Met, Ile	n.d.	n.d.	traces	traces	n.d.	n.d.
390.0534 a	5,19	1-(Hydroxymethyl)propyl; 2-Hydroxy-2-methylpropyl; 3-Hydroxybutyl[1]; 4-Hydroxybutyl[1]	30, 31, 25[1], 26[1]	A-HOA	Ile, Leu, Met, Met	n.d.	n.d.	n.d.	traces	n.d.	n.d.
390.0534 b	5,4	1-(Hydroxymethyl)propyl; 2-Hydroxy-2-methylpropyl; 3-Hydroxybutyl[1]; 4-Hydroxybutyl[1]	30, 31, 25[1], 26[1]	A-HOA	Ile, Leu, Met, Met	3,19E-01	2,96E-02	n.d.	n.d.	n.d.	n.d.
406.0306	6,79	3-(Methylsulfanyl)propyl[2]	95	A-MSAA	Met	traces	traces	traces	traces	traces	1,51E-04
420.0457	7,45	**4-(Methylsulfanyl)butyl**	84	A-MSAA	Met	1,43E-03	3,44E-03	traces	traces	traces	1,02E-05
434.0619	8,33	**5-(Methylsulfanyl)pentyl**	94	A-MSAA	Met	traces	traces	traces	traces	n.d.	n.d.
448.0775	9,2	6-(Methylsulfanyl)hexyl	88	A-MSAA	Met	traces	traces	n.d.	traces	n.d.	n.d.
462.0932	10,05	7-(Methylsulfanyl)heptyl	87	A-MSAA	Met	traces	traces	n.d.	n.d.	traces	n.d.
476.1088	10,82	8-(Methylsulfanyl)octyl	92	A-MSAA	Met	1,18E-02	9,02E-03	n.d.	n.d.	9,32E-03	n.d.
490.1245	11,6	9-(Methylsulfanyl)nonyl[1]	89[1]	A-MSAA	Met	1,24E+00	5,03E-01	traces	n.d.	4,05E-01	6,69E-04
504.1401	12,5	10-(Methylsulfanyl)decyl[1]	85[1]	A-MSAA	Met	2,45E-01	4,26E-02	traces	n.d.	4,13E-02	1,41E-04
422.0255	4,68	**3-(Methylsulfinyl)propyl**	73	A-MSIA	Met	traces	traces	traces	traces	1,33E-04	1,35E-04
436.0411	5,11	**4-(Methylsulfinyl)butyl**	64	A-MSIA	Met	traces	traces	4,18E-04	2,64E-03	2,33E-04	6,52E-05

450.0568	5,65	**5-(Methylsulfinyl)pentyl**	72	A-MSIA	Met	traces	traces	5,56E-04	traces	traces	n.d.
464.0724	6,26	6-(Methylsulfinyl)hexyl[4]	67	A-MSIA	Met	traces	traces	traces	traces	n.d.	n.d.
478.0881	6,92	7-(Methylsulfinyl)heptyl	66	A-MSIA	Met	traces	traces	traces	n.d.	traces	n.d.
492.1037	7,7	8-(Methylsulfinyl)octyl	69	A-MSIA	Met	4,04E-03	1,37E-02	traces	traces	6,44E-03	1,45E-04
506.1194 a	7,95	9-(Methylsulfinyl)nonyl[5]	68[5]	A-MSIA	Met	n.d.	n.d.	2,63E-03	traces	n.d.	n.d.
506.1194 b	8,45	9-(Methylsulfinyl)nonyl[5]	68[5]	A-MSIA	Met	2,15E-01	6,57E-01	n.d.	n.d.	5,16E-01	6,69E-03
520.1350	9,18	10-(Methylsulfinyl)decyl	65	A-MSIA	Met	3,23E-01	5,29E-01	traces	n.d.	7,60E-01	7,46E-03
522.1143 a	8,22	9-(Methylsulfonyl)nonyl[5]	79[5]	A-MSOA	Met	traces	traces	n.d.	n.d.	1,05E-03	n.d.
522.1143 b	8,55	9-(Methylsulfonyl)nonyl[5]	79[5]	A-MSOA	Met	traces	traces	traces	traces	2,26E-03	1,61E-05
522.1143 c	9	9-(Methylsulfonyl)nonyl[5]	79[5]	A-MSOA	Met	traces	1,69E-02	n.d.	n.d.	traces	4,54E-05
408.0429	7,4	**Benzyl**	11	B	Phe	traces	traces	1,67E-01	1,42E+00	traces	1,52E-04
422.0585	8,3	**2-Phenylethyl**	105	B	Phe	1,14E-02	1,75E-02	2,11E-02	7,04E-01	traces	9,94E-06
447.0538	7,87	**Indol-3-ylmethyl**	43	I	Trp	2,95E-03	3,00E-03	1,00E-01	1,25E-01	traces	n.d.
463.0487	7	**4-Hydroxyindol-3-ylmethyl**	28	I	Trp	1,86E-02	1,44E-02	1,39E-03	traces	traces	n.d.
477.0643 a	8,5	**4-Methoxyindol-3-ylmethyl**	48	I	Trp	8,73E-03	9,65E-02	1,49E-01	4,36E-01	traces	n.d.
477.0643 b	9,05	**1-Methoxyindol-3-ylmethyl**	47	I	Trp	n.d.	n.d.	1,46E-02	traces	n.d.	n.d.

Aliphatic (Alkyl)	3,06E-01	2,80E+00	5,86E-03	2,37E-01	2,19E-03	5,90E-05
Aliphatic (Met-derived)[6]	2,36E+00	2,17E+00	1,83E-02	6,84E-01	1,74E+00	1,61E-02
Benzenic	1,14E-02	1,75E-02	1,88E-01	2,12E+00	traces	1,62E-04
Indolic	3,03E-02	1,14E-01	2,65E-01	5,61E-01	traces	n.d.
Total	2,71E+00	5,10E+00	4,77E-01	3,61E+00	1,74E+00	1,64E-02

[1]This GSL structure was marked as "partially characterized" in Blažević *et al.* (2020), lacking characterization by either MS or NMR spectroscopy.
[2]Instead of 3-(Methylsulfanyl)propyl-GSL, this mass could also represent 4-Mercaptobutyl-GSL (133). This is unlikely however as other methylsulfanylalkyl-GSLs occur in sequence, with similar distribution patterns among the samples.
[3]No isomer for this mass was described in Blažević *et al.* (2020). Clarke (2010) suggests n-Heptyl-GSL for this mass.
[4]Instead of 6-(Methylsulfinylhexyl)-GSL, this mass could also represent 3-Hydroxy-6-(methylsulfanyl)hexyl-GSL[1] (36). This is unlikely however as other methylsulfinylalkyl-GSLs occur in sequence, with similar distribution patterns among the samples.
[5]Although there is only one GSL described for this mass in literature, multiple isomers were detected. They might represent unknown structures or iso-GSLs.
[6]This calculation encompasses all non-alkyl aliphatic GSLs, i.e. all potentially Met-derived structures.

In all analyzed samples, a rather broad range of GSL structures could be identified, with 32 different GSLs detected in *A. alpina* root and shoot tissue each, 28 and 26 in *C. hirsuta* roots and shoots, respectively, and 27 and 17 in *C. rubella* roots and shoots, respectively. As it is the case for Arabidopsis, Met-derived GSLs always contributed the most to total structural diversity. Many of the detected Met-GSLs are products of secondary modification reactions, namely oxidation, alkenylation, benzoylation and hydroxylation – with greatly differing concentrations of these compound classes between species. Except for the oxidation from methylsulfinyl to methylsulfonyl GSLs, the responsible enzymes for these reactions have been well characterized (Augustine and Bisht, 2016). Previous studies about this topic suggest that variation in function or expression of the responsible enzymes is mainly responsible for these interspecies differences (Kliebenstein *et al.*, 2001; Liu *et al.*, 2014; Czerniawski *et al.*, 2021). Thus, these variations in GSL profiles do not lend themselves well for further study on the influence of regulatory factors.

Another notable observation is that not all detected masses that were identified as GSLs correspond to a well-described GSL structure. One, presumably representing *n*-heptyl GSL, occurs in relevant amounts in all three species, although this molecule has not been conclusively documented *in vivo* so far. Furthermore, isomers of 9-(methylsulfinyl)nonyl and 9-(methylsulfonyl)nonyl GSL were detected, even though none have been described in literature yet. Therefore, these structures could either represent novel GSLs or examples for the proposed iso-GSLs, molecules with structural modifications in the GSL backbone (Blažević *et al.*, 2020).

The most striking difference in the entire analysis lies in the total GSL content, which is relatively consistent for most samples, but is moderately reduced in *C. hirsuta* roots and extremely lowered in *C. rubella* shoots – more than a hundred times lower than the root concentration and several hundred times lower than the shoot concentrations of other species. Furthermore, the GSLs in *C. rubella* shoots are almost exclusively Met-derived. Because alkyl and benzenic GSLs are also uncommon in the model species Arabidopsis, their precise biosynthesis pathway and regulation remain unclear to date (Mitreiter and Gigolashvili, 2020), so that a study of their respective contribution to this peculiar absence would have to remain mostly speculative. This leaves the almost complete lack of indolic GSLs in *C. rubella*, which is highly unusual and has been suggested to precede a complete loss of this metabolite family in the evolutionary lineage of this species (Czerniawski *et al.*, 2021). As the specific regulation of indolic GSLs by MYBs and bHLHs is well described, *C. rubella* provides an excellent environment to study their significance in giving rise to interspecies GSL diversity.

3.2.3 Regulatory *MYB* and *bHLH* genes are largely conserved across Brassicaceae

The detailed GSL profiles presented in the previous section show numerous differences between species, some of which are likely the result of variation in GSL biosynthesis genes, while others could potentially stem from variations in the MYB-bHLH regulatory machinery. The question if this is actually the case can be addressed by analyzing the conservation of these regulatory factors, both sequence-wise and *in*

vitro, the latter of which was addressed by OX in Arabidopsis Ler-0 cell culture. On this basis, it can then be assessed if differences in GSL accumulation may indeed be explained by regulatory variation.

3.2.3.1 *MYB* sequences are slightly less conserved than *bHLH* sequences

Genomic data reveals that subgroup 12 *MYB* and subgroup IIIe *bHLH* genes are almost completely conserved among the studied Brassicaceae. *A. alpina* harbors two exceptions, as it lacks copies of *MYB122* and *MYB76*. The absence of *MYB122* represents a secondary loss, since this gene is already present in *Aethionema arabicum* of the Aethionemeae tribe, the basal sister group to the core Brassicaceae (Hofberger *et al.*, 2013; **Figure 13**).

In contrast, *MYB76* is a product of a later tandem duplication that is ancestral only to the so-called lineage I of the Brassicaceae (Beilstein *et al.*, 2006; Nikolov *et al.*, 2019): *MYB76* neither occurs in the basal *A. arabicum*, the lineage IV members *Arabis montbretiana* and *A. alpina*, nor the lineage II members *Brassica* spp., *Eutrema salsugineum*, *Raphanus sativus*, *Schrenkiella parvula* and *Sisymbrium irio*. While the lineage I species *Arabidopsis lyrata*, *A. thaliana*, *Boechera stricta*, *C. hirsuta*, *Camelina sativa* and *Leavenworthia alabamica* all possess copies of *MYB76*, the gene is absent in the *Capsella* species *C. grandiflora* and *C. rubella* (Goodstein *et al.*, 2011; Hofberger *et al.*, 2013; Mitsui *et al.*, 2015; Willing *et al.*, 2015; Seo and Kim, 2017; Gomez-Cano *et al.*, 2020; **Figure 13**).

This deletion seems to represent one of multiple deregulations in GSL synthesis in the Camelineae clade II ancestry (Czerniawski *et al.*, 2021). As both *MBY76* and *MYB122* only play comparatively minor roles in their respective branches of GSL synthesis in Arabidopsis (Mitreiter and Gigolashvili, 2020), it is unlikely that their absence significantly alters GSL profiles in *A. alpina* and *C. rubella*.

The respective sequence similarity of all *MYB* and *bHLH* genes was then analyzed by generation of a MUSCLE alignment and a corresponding phylogenetic tree (**Figure 14**). Both gene groups exhibit a high degree of conservation, but there exist some notable differences: While the relatedness of the *bHLHs* tidily reflects the established phylogeny of the involved species, this is not entirely the case for the *MYBs*. All ortholog genes cluster together, but the apparent relatedness of the particular *MYB* genes themselves does not always mirror the Brassicaceae phylogeny. For instance, *CrMYB29* clusters with *ChMYB29* rather than with *AtMYB29*, whereas *AtMYB122* clusters with *ChMYB122* instead of *CrMYB122*. The most prominent example is *CrMYB34*, which is notably dissimilar from the three other orthologs.

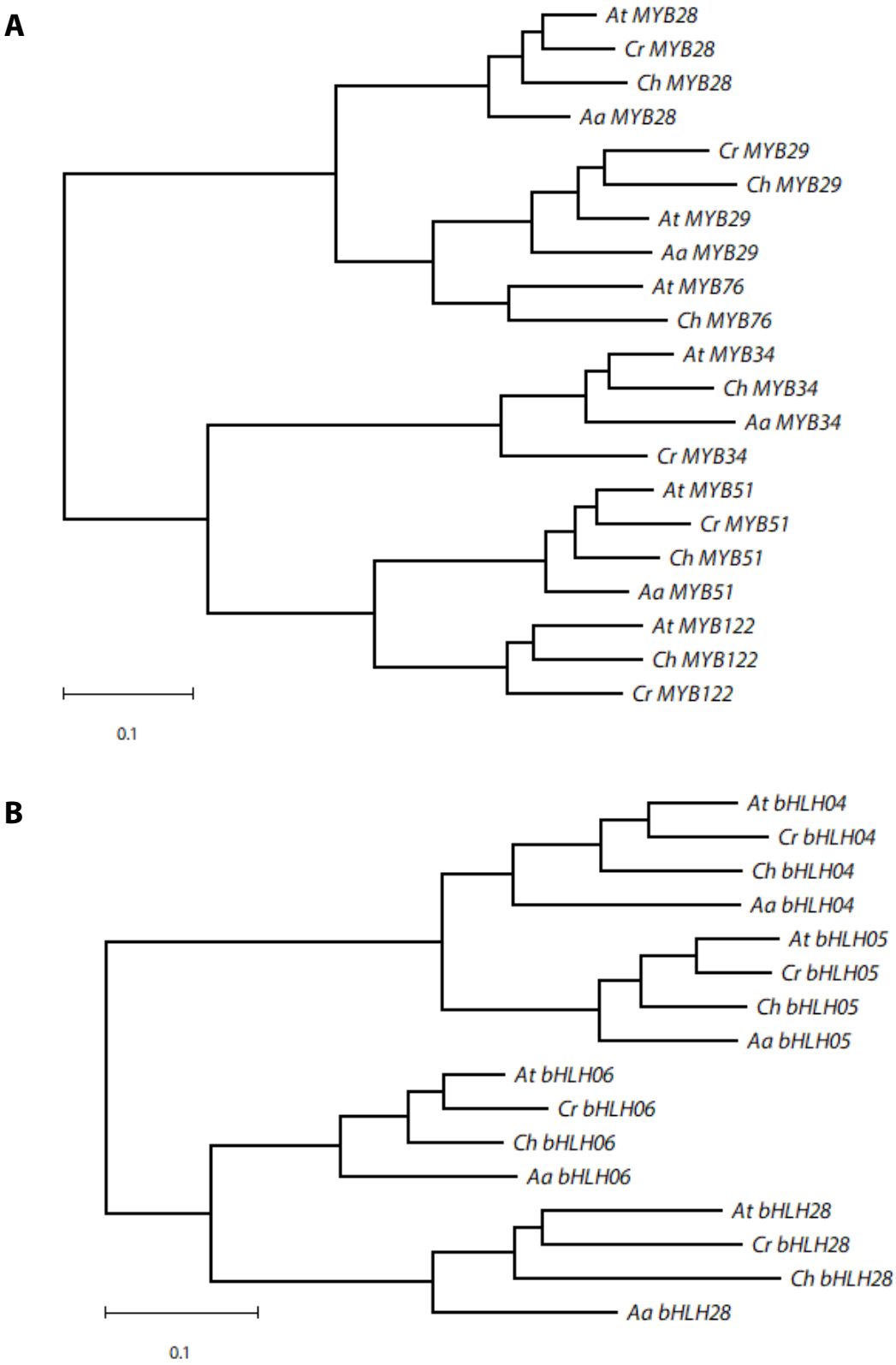


Figure 14: Evolutionary analysis of Brassicaceae *MYB* and *bHLH genes*.
The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model (Tamura and Nei, 1993). The trees with the highest log likelihoods are shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. **(A)** shows the relatedness of 21 *MYB* sequences with a total of 1396 positions in the final dataset, **(B)** shows the relatedness of 16 *bHLH* sequences with a total of 2219 positions in the final dataset.

3.2.3.2 MYB factors exhibit a high degree of functional conservation

Taken together, both subgroup 12 *MYB* and subgroup IIIe *bHLH* genes appear to be rather conserved in Brassicaceae, whereby some *CrMYB* genes may provide an exception as they do not fit into the phylogenetic tree as it would be expected. This, combined with the fact that MYBs provide specificity for the target genes in the MYB-bHLH complex, makes it especially interesting to experimentally test the influences of individual MYB TFs on GSL accumulation. To this end, I overexpressed various *MYBs* in cultured Arabidopsis Ler-0 cells (**Figure 15**). Using orthologous genes from different species enables the possibility to obtain information about (i) the functional similarity of different *MYBs* from one species and (ii) the functional similarity, i.e. conservation of orthologous *MYBs*, across different species.

Remarkably, aliphatic *MYBs* displayed a clear pattern of conservation where the *MYB28*, *MYB29* and *MYB76* orthologs each produced distinct GSL profiles that are highly similar among each other – in fact, the orthologous *MYBs* from different species behaved much more similarly to each other than to other *MYBs* from the same species. Most strikingly, OXs of *MYB28* induced a modest amount (less than 5 µmol/g DW) of the indolic GSL I3M, while this was not the case for the other aliphatic MYBs. Moreover, *MYB28* OXs led to a relatively high production of long-chain Met-GSLs (7–8C), while this effect was less pronounced for *MYB29* OXs and completely absent for *MYB76* OX. In contrast, OXs of indolic *MYBs* increased the production of main indolic GSL I3M, but did not significantly affect the accumulation of modified indolic or aliphatic GSLs. The profiles of *MYB34* and *MYB51* orthologs did not differ significantly from each other.

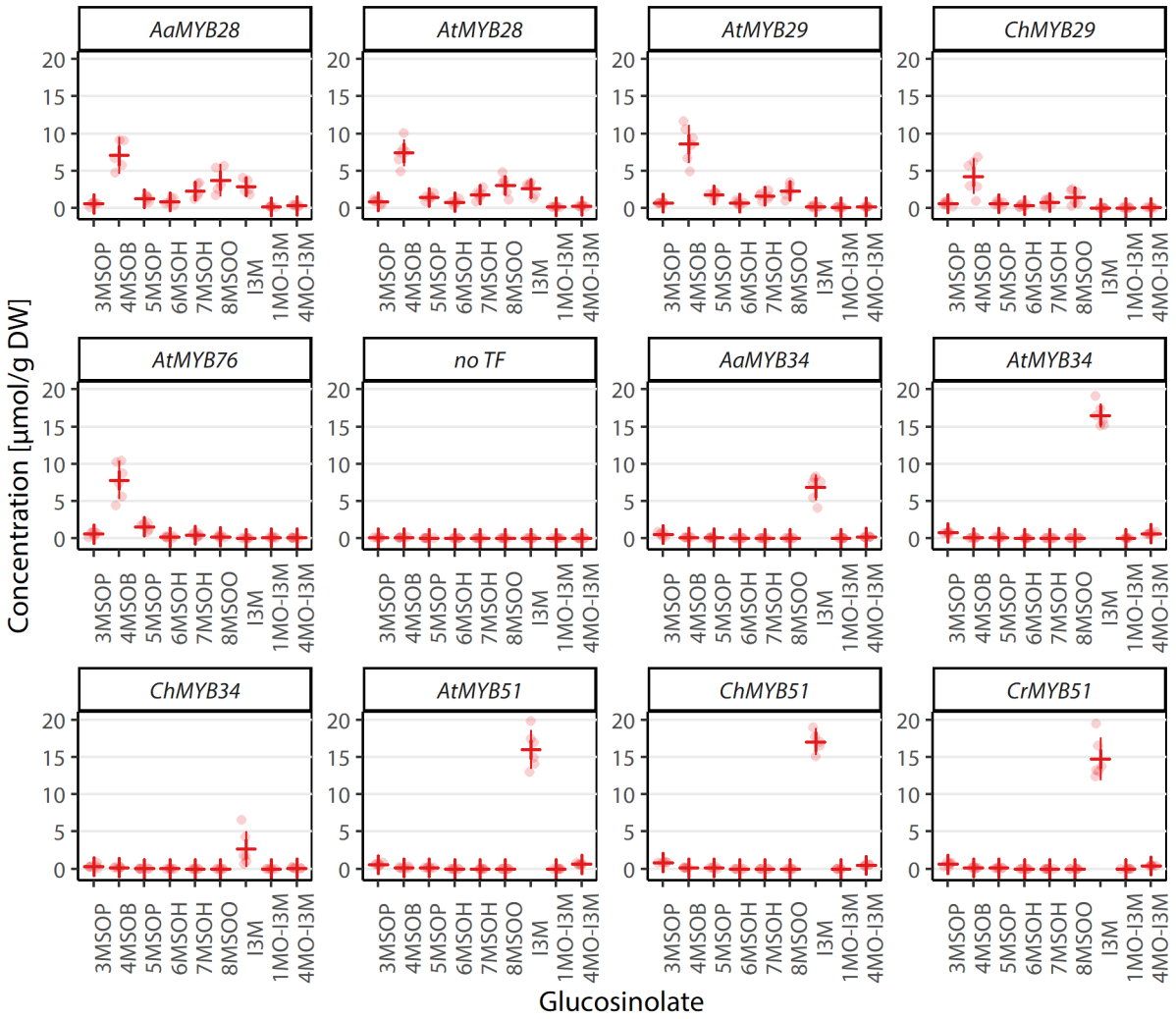


Figure 15: Orthologous MYB factors produce similar GSL profiles.
GSL quantification of Arabidopsis Ler-0 cell culture overexpressing *MYB* genes from different Brassicaceae species, seven days after transfection. Error bars represent the 95% confidence interval. GSL abbreviations: 3MSOP, 3-(methylsulfinyl)propyl: 4MSOB, 4-(methylsulfinyl)butyl: 5MSOP, 5-(methylsulfinyl)pentyl: 6MSOH, 6-(methylsulfinyl)hexyl: 7MSOH, 7-(methylsulfinyl)heptyl: 8MSOO, 8-(methylsulfinyl)octyl: I3M, indol-3-ylmethyl: 1MO-I3M, 1-methoxyindol-3-ylmethyl: 4MO-I3M, 4-methoxyindol-3-ylmethyl.

3.2.4 *C. rubella bHLHs* have numerous non-functional splicing variants

C. rubella produces extraordinarily low amounts of indolic GSLs, but in principle retains the synthetic capability to produce I3M, 4OH-I3M and 4MO-I3M. The previous experiment demonstrated that CrMYB51 is also functionally intact and does not differ from its Brassicaceae orthologs in regard to the GSL profile that it induces. These findings imply that both the core biosynthetic enzymes for indolic GSLs and their regulatory machinery are in principle operative, but the throughput of the pathway is drastically reduced. One hypothesis to explain this phenomenon is an overall downregulation of the pathway by changes in the MYB–bHLH regulatory machinery that decrease MYB and bHLH influence, but do not fully abolish their functionality.

One mechanism that could have such an impact is the presence of numerous shortened alternative splice variants of *CrbHLHs*, an observation made during the cloning of *C. rubella* genes and not yet reported for the other species used in this study. I investigated the functionality of these splice variants in the

induction of GSL biosynthesis via qualitative GUS staining and GSL quantification, with the goal of assessing if alternative splicing could potentially act as a regulatory mechanism that decreases indolic GSL production in *C. rubella*.

Multiple splice variants were isolated for all of the examined genes, *CrbHLH04*, *CrbHLH05*, and *CrbHLH06*. However, the full-length sequences could only be cloned for *CrbHLH04* and *CrbHLH06*, not for *CrbHLH05*. When comparing the missing parts of the spliced transcripts with available annotation about the domains of Arabidopsis and *C. rubella* bHLHs, it is notable that in most splice variants, at least one of the bHLH, TAD, JID and/or ACT domains is truncated (**Table 23**).

Table 23: *CrbHLH* genes are subject to alternative splicing.
List of isolated splice variants of *CrbHLH* genes with the missing base pairs relative to the full-length CDS. The location of the bHLH domains was taken from the UniProt entries, while the location of the TAD and JID domains was estimated from the UniProt entries of the Arabidopsis homologs. The location of the ACT domains was estimated from the report for AtbHLH06 by Feller *et al.* (2006). Variants that were used for further experiments are marked with an asterisk.

Gene		**Isolated splice variants**		**Truncated domains**
Name	**CDS length**	**Missing base pairs**	**Abbreviation**	
CrbHLH04	1845 bp	full length	*CrbHLH04 full**	none
		47 – 1575	*CrbHLH04 sh A**	JID, TAD, bHLH, ACT?
		193 – 1619	*CrbHLH04 sh B*	JID, TAD, bHLH, ACT?
CrbHLH05	1809 bp	67 – 1108	*CrbHLH05 sh A**	JID, TAD
		425 – 1539	*CrbHLH05 sh B**	TAD, bHLH, ACT?
		470 – 1033	*CrbHLH05 sh C*	TAD
		485 – 1557	*CrbHLH05 sh D*	TAD, bHLH, ACT?
		604 – 1275	*CrbHLH05 sh E*	none
CrbHLH06	1875 bp	full length	*CrbHLH06 full**	none
		46 – 1260	*CrbHLH06 sh A**	JID, TAD
		153 – 1275	*CrbHLH06 sh B*	JID, TAD

To determine the *in vitro* functionality of these shortened splice variants, several representative variants were selected and overexpressed in Arabidopsis *bhlh04/05/06* cell culture to prevent interference from native bHLHs. These OXs were conducted alone and in combination with *CrMYB51*, using *AtMYB51* and *AtbHLH05* as controls. Qualitative GUS staining with the promoter-reporter construct *proAtCYP79B3:GUS* was employed to evaluate the ability of the MYB-bHLH complex to activate gene transcription. Additionally, GSL quantification was employed to determine the resulting effect on the metabolite level.

GUS staining revealed a qualitative difference between the functional capabilities of the shortened and the full-length splice variants (**Figure 16**). Combined OXs of *MYB* and full-length *bHLH* constructs led to a stronger *GUS* expression than isolated *MYB* or *bHLH* OXs, while the combination of *MYB* and shortened *bHLHs* did not induce *GUS* expression beyond the level displayed by *MYB* OX alone. Interestingly, the effect of CrMYB51 on GUS expression seemed to be consistently stronger than for AtMYB51 – which might represent a functional difference or simply a result of a stronger silencing effect of Arabidopsis cells against Arabidopsis genes, compared to genes from other species.

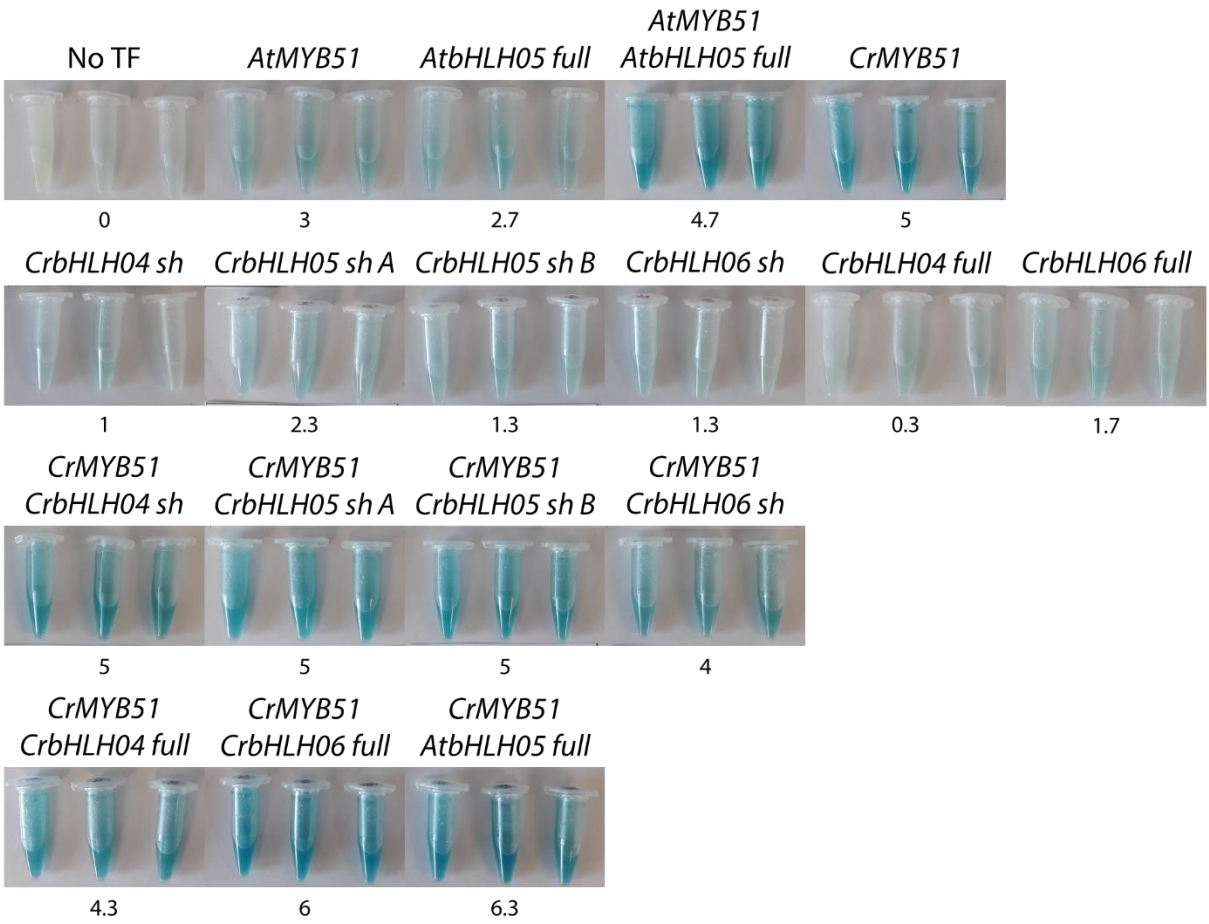
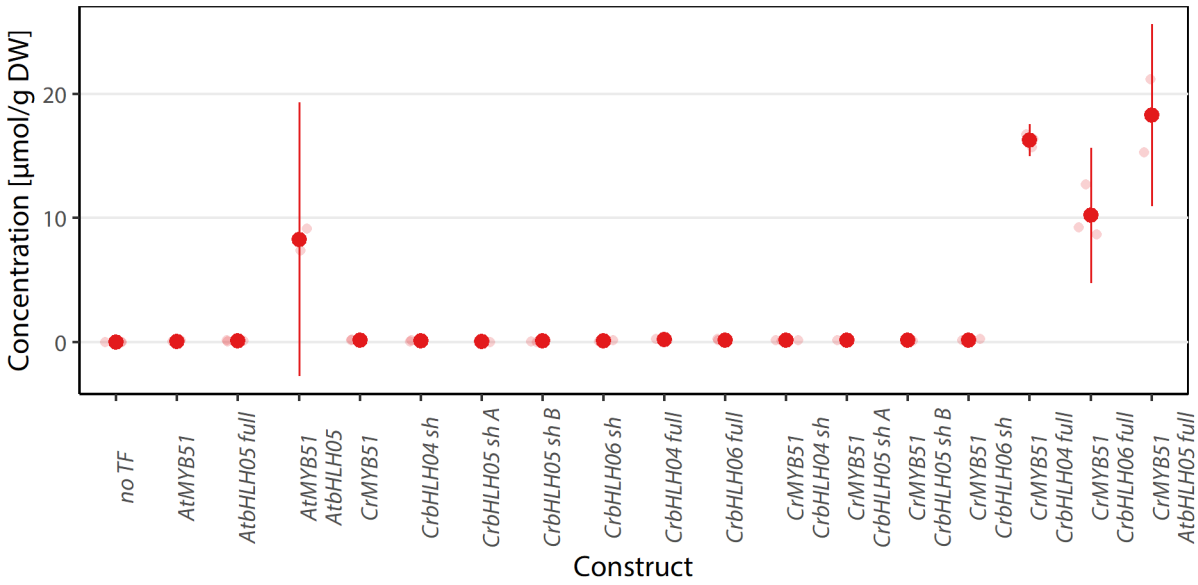


Figure 16: Shortened *CrbHLH* splice variants do not induce transcription of *CYP79B3*.
Qualitative GUS staining of Arabidopsis *bhlh04/05/06* cell culture overexpressing *bHLH* and *MYB* constructs as well as the *proAtCYP79B3:GUS* promoter-reporter construct, six days after transfection. The numbers indicate a relative quantification, with the scale reaching from 0 (no coloration) to 7 (maximum coloration). Values were calculated as the average of three biological replicates.

When GSL content was quantified in samples from the same experiment, the same tendencies as for the GUS staining could be observed, although even more clearly pronounced (**Figure 17**): Only the combined OX of *MYB* and full-length *bHLH* constructs led to an overproduction of indolic GSLs. *MYBs* alone had no notable effect on GSL accumulation due to the lack of native *bHLHs* in the cell culture, a gap that the shortened splice variants were unable to bridge.

A



B

	no TF	***AtMYB51***	***AtbHLH05 full***	***AtMYB51 AtbHLH05 full***	***CrMYB51***
Significance	a	a	a	b	a

	CrbHLH04 sh	***CrbHLH05 sh A***	***CrbHLH05 sh B***	***CrbHLH06 sh***	***CrbHLH04 full***	***CrbHLH06 full***
Significance	a	a	a	a	a	a

	CrMYB51 CrbHLH04 sh	***CrMYB51 CrbHLH05 sh A***	***CrMYB51 CrbHLH05 sh B***	***CrMYB51 CrbHLH06 sh***
Significance	a	a	a	a

	CrMYB51 CrbHLH04 full	***CrMYB51 CrbHLH06 full***	***CrMYB51 AtbHLH05***
Significance	c	b	c

Figure 17: Shortened *CrbHLH* splice variants do not induce GSL production.
(A) GSL quantification of Arabidopsis *bhlh04/05/06* cell culture overexpressing *bHLH* and *MYB* constructs, seven days after transfection. Error bars represent the 95% confidence interval. **(B)** Statistical significance levels of differences in GSL concentration between the samples. p-values were calculated using one-way ANOVA and Tukey's post-hoc test.

In conclusion, the newly identified alternative splicing of *bHLH* transcripts in *C. rubella* does produce non-functional transcripts, likely due to the truncation of domains vital for protein function. This mechanism has the potential to fulfil a regulatory function that decreases the overall GSL production – in this context, it is notable that neither the treatment of *C. rubella* cells with 10 µM JA in this study (data not shown) nor the inoculation of *C. rubella* plants with the fungus *Blumeria graminis* f. sp. *hordei* (Bednarek *et al.*, 2011) led to any induction of indolic GSL synthesis.

Additionally, indolic *MYB* expression is significantly lowered in all tissues of *C. rubella*, with *CrMYB34* expression being almost completely undetectable (P. Bednarek, unpublished data). I could confirm this observation by cloning attempts that were unsuccessful due to absence of *CrMYB34* transcripts from *C. rubella* leaf cDNA. If the alternative splicing of *bHLHs* fulfils regulatory functions, it could hypothetically act together with this decreased expression of indolic *MYB* genes, potentially dependent

on environmental or developmental conditions. However, the *in vivo* relevance of these mechanisms for GSL production in *C. rubella* remains to be elucidated in the future.

3.2.5 GSL accumulation in *C. rubella* cells can be partially complemented by *MYB* and *bHLH* overexpression

In *C. rubella*, both the biosynthetic pathway of indolic GSLs and its regulatory MYB and bHLH factors are in principle functional, as it was demonstrated in the previous experiments. Simultaneously, alternative splicing of *CrbHLHs* and low expression of *CrMYBs* were shown to potentially play a role in causing low indolic GSL production. To test this hypothesis, I tried to complement this lack of indolic GSLs in *C. rubella* cells by overexpression of *MYB* and *bHLH* genes, both from *C. rubella* and from other species.

OX of aliphatic *MYBs* revealed that *AaMYB28* and *AtMYB28*, but not *MYB29* or *MYB76* constructs, were able to induce a moderate accumulation of I3M at a concentration of about 1–2 µmol/g DW (**Figure 18A**). While there are quantitative differences to the OX of *MYBs* in Ler-0 cell culture, the induced profiles are remarkably similar in both species.

In another experiment, OX of indolic *MYB34* and *MYB51* factors induced small amounts of I3M production in the magnitude of 10–100 nmol/g DW. Additionally, the effect of *AtMYBs* on I3M accumulation was amplified up to 2.5-fold in combination with *AtbHLHs* (**Figure 18B**). Qualitative GUS staining with the promoter-reporter construct *proAtCYP79B3:GUS* indicated a limited transcriptional activation of this indolic GSL biosynthesis gene by indolic MYBs. This effect was considerably lower for *MYB28* orthologs and completely absent for *MYB29* and *MYB76* orthologs (**Supplementary Figure 8**).

Taken together, the overexpression of intact *MYB* and *bHLH* genes can increase indolic GSL levels in *C. rubella* cell culture. These levels do not reach the concentrations found in other species, however. To verify the role of *MYBs*, *bHLHs* and potential further genes in indolic GSL synthesis in *C. rubella*, further research is needed.

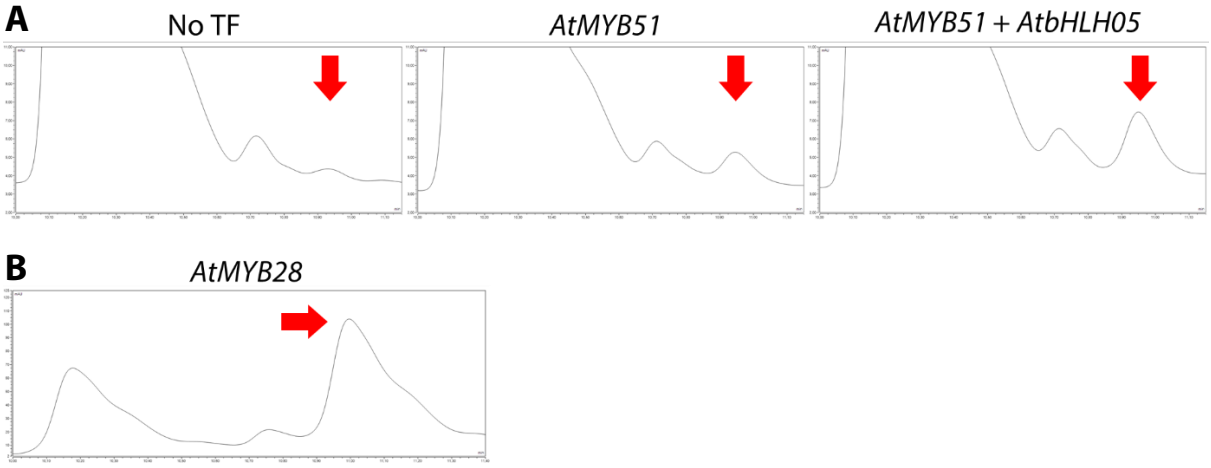


Figure 18: Overexpression of aliphatic and indolic MYBs can induce I3M production in *C. rubella* cells.
Qualitative GSL profiles induced in *C. rubella* cell culture by overexpression of different *MYB* and *bHLH* genes. The x-axis represents the retention time and the y-axis represents the absorbance. The peaks marked by a red arrow represent the main indolic GSL I3M, whereas the major peak to the left represents the reference BGS. **(A)** and **(B)** are not to scale, with the BGS peak representing an amount of 50 nmol in both cases.

4 Discussion

4.1 Epigenetic regulation of GSL biosynthesis by EML histone readers

The induction of GSL synthesis is a well-described part of the stress response in Brassicaceae plants and, as many other processes involved in it, is subject to priming (Rasmann *et al.*, 2012; Bakhtiari *et al.*, 2018). This means that GSL induction becomes more sensitive under repeated stress stimuli, an adaptation that is likely mediated via epigenetic mechanisms (Lämke and Bäurle, 2017; Ramirez-Prado *et al.*, 2018). Indeed, the knockout of *LHP1*, a PRC1 subunit, leads to the deregulation of GSL accumulation, as does the knockout of its interaction partner *LIF2* (Ludwig-Müller *et al.*, 1999; Kim *et al.*, 2004; Roux *et al.*, 2014). Nonetheless, no epigenetic regulation specific to GSL biosynthesis has been described to date. Thus, my aim in this study was to identify a potential mechanism for such a regulation. Through colocalization, mutant analysis and overexpression experiments I found that the EML histone reader family likely participates in the epigenetic control of GSL biosynthesis in interaction with the MYB-bHLH transcription complex.

4.1.1 EMLs likely take part in epigenetic regulation of GSL biosynthesis

The EML family were selected as candidates due to their homology to the maize chromatin-interacting factor RIF1, as both EMLs and RIF1 are characterized by the combination of ENT and Agenet domains. RIF1 takes part in the regulation of anthocyanin biosynthesis (Hernandez *et al.*, 2007), a process intimately connected to GSL biosynthesis – not only because both are part of secondary metabolism, but also because both are transcriptionally regulated by MYB-bHLH complexes from closely related subgroups. In Arabidopsis, anthocyanin production is controlled by subgroup 6 MYBs and subgroup IIIf bHLHs (Chen and Wang, 2019), both of which also interact with JAZ proteins to mediate a JA-dependent response (Qi *et al.*, 2011). Furthermore, the JA-mediated induction of both GSL and anthocyanin biosynthesis is reduced in the *lhp1* mutant (Bennett *et al.*, 2005). Therefore, we postulated that EMLs could fulfil a similar role in GSL synthesis as RIF1 in anthocyanin synthesis.

As part of this study, I gathered additional evidence for this by analysis of omics data from the Plant Regulomics database (Ran *et al.*, 2020). This revealed an enrichment of the repressive histone marks H3K27me (a PRC1 product) for aliphatic and H2AK121ub (a PRC2 product) for both aliphatic and indolic GSL biosynthesis genes under unstressed growth conditions. Under drought stress, the activating mark H3K4me becomes enriched, in particular for aliphatic genes (**Supplementary Tables 8–9**). This seems to reflect the observation that GSLs are often induced under drought stress (Del Carmen Martínez-Ballesta *et al.*, 2013; Variyar *et al.*, 2014), with aliphatic GSLs playing an especially important role in the mediation of stomatal closure (Salehin *et al.*, 2019). Histone marks of genes involved in the chain elongation of aliphatic GSLs have also been studied by Xue *et al.* (2015) – here, a particular pattern of H3K4me modifications was identified, although it seemingly did not influence the level of transcription. Additionally, analysis of Plant Regulomics data suggested that EML expression correlates

to some extent with conditions that induce GSL synthesis, such as immune response and various hormonal signals (**Supplementary Figures 1–4**).

Collectively, it seemed likely that EMLs function in stress response. For further experiments, the role of EMLs was considered in regard to the JA response, since JA represents the strongest positive hormonal influence on GSL synthesis. In fact, the JA response has been previously associated with other epigenetic mechanisms: JAZ proteins recruit the repressive histone deacetylase HDA6 to suppress transcription of target genes under uninduced conditions (Zhu *et al.*, 2011; Kim *et al.*, 2017), while MED25 recruits the activating histone acetyltransferase HAC1 in interaction with the MYB-bHLH complex (An *et al.*, 2017).

The JA-mediated GSL induction was measured in *eml* mutants by growing them in the presence of JA for a prolonged period of time. Out of the tested 12 mutant lines, an impairment in this response was detected in the lines *eml2-1*, *eml4-1* and *eml1/eml3* of set II, with the strongest effect in the double mutant *eml1/eml3* (**Figure 7**). As the function of EMLs is dosage-dependent (Milutinovic *et al.*, 2019), it is likely that the reduction in induced GSLs is highly dependent on the total reduction of *EML* expression levels. Notably, none of the T-DNA insertions and amiRNA constructs contained in set I (except for *eml3-1*) reduces the *EML* transcript levels significantly below 70% or 50% of the wildtype, respectively (Milutinovic *et al.*, 2019). Combined with the possibility that the expression of the amiRNA constructs could be reduced after several generations of seed propagation (which was not controlled for in this study due to time limitations), the overall impairment of EML function may not be strong enough in these mutants to obtain a noticeable effect on GSL induction (**Figure 6**). This is consistent with the fact that in the set II lines, the T-DNA insertions reduced *EML* expression to 30% of the wildtype and lower in all lines, rendering them completely undetectable in *eml1-2* and *eml3-4*. It is probable that this is the reason that an impairment of JA-induced GSL synthesis was only detectable in set II mutants, especially in *eml1/eml3*, which combines two strongly affected alleles.

This measurement of the metabolite level was complemented by a qPCR measurement of *CYP* biosynthesis genes (from both the aliphatic and the indolic pathways) in the set II mutants (**Figure 8**). Here, only the indolic genes *CYP79B3* and *CYP83B1* in *eml1/eml3* showed a trend that corresponds to the metabolite level. Neither the aliphatic gene *CYP79F2* in *eml1/eml3* nor any genes in the other lines exhibited a reduction in JA-mediated induction. These trends correspond generally well with the metabolite level, where indolic GSLs in *eml1/eml3* were most strongly affected. However, there also was a notable amount of uncoupling – for example, in *eml1/eml3,* a constitutive *CYP* upregulation in the absence of JA was observed. This disconnect between the transcript and the metabolite level is a well-described phenomenon in GSL biosynthesis (Mitreiter and Gigolashvili, 2020) that might be caused by an asynchronicity of the two levels or the interaction with further regulatory mechanisms, for example post-transcriptional or post-translational modifications of the biosynthetic enzymes. Another likely option is that changes in myrosinase activity, for example through lower expression rates, might lead to a decreased breakdown rate of GSLs. The nature of the mechanisms at work here needs to be identified through further research.

Building on the previous experiments in *eml* mutants, I aimed to specifically elucidate the connection between EML function and priming. Few previous studies exist on how GSL induction is affected by priming responses. Rasmann *et al.* (2012) examined transgenerational inheritance of a primed GSL production induced by herbivory, mechanical damage or JA application. Enhanced resistance against herbivores was found to be inherited over two generations, with 1MO-I3M concentration increased at least in the first progeny generation. In the parental generation, mutations affecting either JA perception or RNA-directed DNA methylation, showed decreased GSL concentrations both in induced and uninduced states, suggesting a role of these processes in priming.

Bakhtiari *et al.* (2018) conducted an experiment in which roots of *C. hirsuta* plants were induced with JA, after which the plants were subjected to aboveground herbivory. Indeed, JA-treated plants exhibited a priming response in the form of increased resistance against herbivores. However, the time course of GSL concentrations did not match expectations, with GSL levels remaining high in JA-induced plants 4 d after the initial treatment and similar GSL levels for both control and induced plants after an additional week of herbivory.

In the light of the aforementioned studies, a distinct approach was chosen in this study: I employed a repetition of the same stimulus, i.e. JA application, and recorded the same output, i.e. GSL induction. At a timeframe of 24 h, which should cover the greatest changes in GSL content (Zang *et al.*, 2015), GSL concentrations were measured on a quantitative level in both experimental setups and on a temporal level in the second setup.

Here, overall the same observation was made as in the prolonged stress experiment: GSL induction is significantly compromised in *eml1/eml3*, but not in *eml1/eml2* (**Figures 9–11**). However, no clear connection could be drawn between the priming response and EML function. The reason for this is that while a priming effect was clearly present for indolic GSLs, it did not take place according to expectations. While indolic GSL induction was clearly accelerated by priming, concentration-wise it consistently decreased over time. This effect could not be separated from the influence of the *eml* mutations.

In this context, it is important to note that the general definition of priming just requires the defense responses to become more robust over time, not necessarily both faster and stronger at the same time (Martinez-Medina *et al.*, 2016; Mauch-Mani *et al.*, 2017). Since indolic GSLs respond strongly to JA-mediated priming, while aliphatic GSLs do not, the composition of the GSL profile directly changes as a result. In fact, it is widely reported that aliphatic GSLs are less sensitive to JA induction than indolic GSLs and that both have differential effects on herbivores and pathogens (Bartlet *et al.*, 1994; Malitsky *et al.*, 2008; Müller *et al.*, 2010; Zhurov *et al.*, 2014).

Thus, the altered GSL profile caused by the priming response might in principle provide beneficial effects for plant fitness. Finally, it should be considered that the experimental setups used in this study did not directly harm the plant, so that no damage- or herbivore-associated patterns were generated.

The activation of this signaling mechanism might also have an influence on the nature of the priming response.

Another surprising observation was made in the second priming experiment: In multiple measurements of JA-untreated control plants, both aliphatic and indolic GSL content notably increases after 6 h, then decreases again afterwards (**Figure 11**). This might represent either a reaction to the medium exchange or a genuine diurnal pattern, as it has been previously reported for GSL synthesis (Rosa *et al.*, 1994; Huseby *et al.*, 2013; Burow and Halkier, 2017).

Additionally, a rather consistent trend observed in wildtype plants was a rise of GSL concentrations over time. This is in line with reports that document increased GSL contents in newly generated tissue with the ageing of plants (Brown *et al.*, 2003). Intriguingly, this trend was not apparent in *eml1/eml3* plants (**Figures 10–11**), suggesting the possibility that the mutation of *EML* genes also compromises the age-dependent regulation of GSL biosynthesis.

In summary of the priming experiments, the specific role of EMLs in this process remains uncharacterized. If the decrease in indolic GSL concentrations proves indeed to be physiological, a finer timescale is required to identify the influence of *eml* mutations on the acceleration of GSL induction with priming in future experiments. This context could also provide a prime opportunity to analyze the relative contributions of DNA methylation (Aller *et al.*, 2018) and histone modification on the regulation of GSL synthesis.

Moreover, it would be interesting to study the influence of *eml* mutations on the age-dependent accumulation of GSLs, as this could provide a hypothetical further role of EMLs in developmental processes. Finally, the entire GSL-myrosinase system should be taken into account in the future, as GSL hydrolysis is also dependent on plant age (Wentzell and Kliebenstein, 2008) and resulting isothiocyanate products themselves can function as a defense-priming compound (Schillheim *et al.*, 2018), implying the existence of even more complex regulatory networks.

4.1.2 EMLs likely function together with the MYB-bHLH complex

While the previously discussed experiments suggest a role of EMLs in the stress-, in particular JA-mediated GSL induction, localization and colocalization assays provided information about their mode of action. Both EML1 and EML2 localized to subnuclear speckles of high DNA density in living cells, with at least EML2 explicitly colocalizing with bHLH05 (**Figure 5**). Combined with the positive results of the yeast two-hybrid assays, that included all combinations of EMLs and subgroup IIIe bHLHs (except bHLH28), this strongly suggests an interaction between EML and bHLH proteins, and thus a participation of EMLs in the MYB-bHLH transcription complex.

This is supported by the GSL measurements of *EML* overexpressions. Here, *EML1 or EML2* alone have no influence, but combined with *MYB51* and *bHLH05*, the full-length proteins seemingly overcome the bHLH-mediated feedback that reduces GSL levels. These trends are fully reflected in the accompanying

GUS staining, although the *MYB51* OX alone seems to lead to a uniquely high transcription of target genes (**Figure 12**). The precise mechanism of EML action cannot be deduced from these experiments, but can be theorized to either function by titrating out bHLH05 molecules so that they cannot mediate the feedback mechanism anymore, or by direct participation in the MYB-bHLH complex in such a way that the feedback mechanism is disabled, for example by preventing bHLH05 dimerization.

This raises the question of what the molecular mechanism of EMLs looks like. EML1 and EML3 have been conclusively described as H3K36me histone readers, with potentially some affinity towards H3K4me (Coursey *et al.*, 2018; Zhao *et al.*, 2018). This contrasts with RIF1, which has been associated with H3K9 and H3K14 acetylation, although this is just an indirect connection (Hernandez *et al.*, 2007). These findings however only provide an incomplete overview about the histone reader properties of ENT-Agenet proteins, as multiple maize proteins related to RIF1 (Milutinovic *et al.*, 2019) as well as EML2 and EML4 remain uncharacterized. Based on the existing literature, they seem to exhibit an affinity for activating histone marks.

AIP1 is an Agenet domain protein that does not possess an ENT domain. It promotes flowering transition and has been described as a histone reader with the ability to recognize unmodified histones, but neither H3K9ac nor H3K14ac marks (Brasil *et al.*, 2015). Interestingly enough, AIP1 interacts with LHP1 – the same PRC1 component whose knockout leads to deregulation of GSL accumulation and induction. This makes it plausible that while EMLs read activating histone modifications, they could simultaneously influence PRC1 function via LHP1. How exactly this contributes to the activation of GSL synthesis under stress conditions has to remain unexplored for now, especially as the research on recruitment of the PRCs and their relationship to each other is still ongoing (Zhou *et al.*, 2017).

The role of the Agenet (histone reading) and ENT domains (dimerization) in EML function have been described relatively clearly in the literature. The importance of the ENT domain is underlined by the apparently disturbed localization of N-terminally GFP-tagged EML1 observed in this study (**Figure 5A**). In contrast, the function of the C-terminal coiled-coil domain in EMLs has not been studied so far. As it represents a typical protein-protein interaction domain (Wang *et al.*, 2012), e.g. in the components of the photomorphogenesis-repressing COP1-SPA complex (Hoecker, 2017), it might mediate the interaction with the MYB-bHLH complex. This hypothesis becomes interesting in regard to the alternative splicing that was documented for *EML1* in this study (**Figure 3**). While the coiled-coil domain is not directly affected by the sequence shortening of splice variant *EML1.1*, the function of this variant seems to be impaired, as it did not influence GSL accumulation in the OX experiment, in stark contrast to *EML1.3* (**Figure 12**). This suggests that the newly identified splice variants *EML1.4* and *EML1.5* are non-functional all the more. This novel "regulation of regulators" by alternative splicing highlights another level and the overall complexity of the regulatory machinery.

In this contest, it should be considered that EMLs are not the only factor interacting with the MYB-bHLH complex. While SDI1, BES1, JAZ inhibitors and other factors such as subgroup IIId bHLH

proteins (Goossens *et al.*, 2017) operate by interfering with the MYB-bHLH complex formation and binding, other proteins could rather cooperate with the transcription complex. This could not only be the case for MED25 and HAC1 with their roles in transcription and histone acetylation, but also for further TFs such as the AP2/ERF family, which was suggested for such a role by Li *et al.* (2014).

In conclusion, in this study I demonstrate that EML proteins likely function in the epigenetic regulation of JA-mediated GSL induction, while participating in the MYB-bHLH transcription complex. Therefore, I propose a model in which the MYB-bHLH complex is at the center of this regulation and EMLs directly interact with bHLHs, themselves mediating the function of further epigenetic components, possibly PRCs. Simultaneously, even more transcriptional and epigenetic factors such as MED25, HAC1 and hypothetically AP2/ERF take part in the complex. Inactivating signals such as low JA, high BR or low sulfur inhibit GSL biosynthesis via proteins that disturb MYB-bHLH interaction and DNA binding (**Figure 19**).

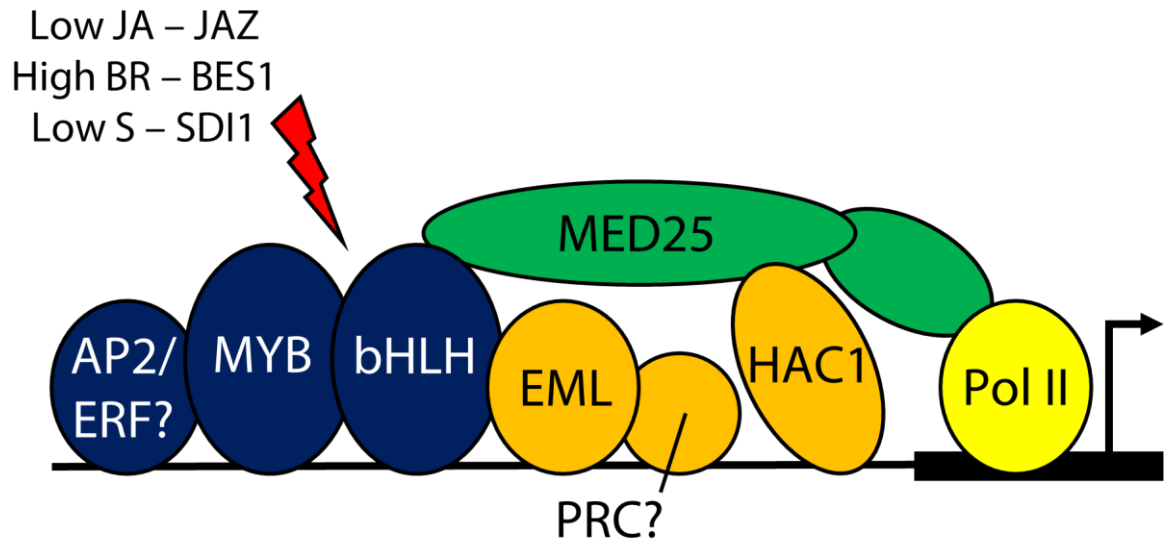


Figure 19: Model of the transcriptional complex in GSL biosynthesis.
Transcription of genes involved in GSL biosynthesis is specifically induced by TFs (blue): Subgroup 12 MYBs bind to subgroup IIIe bHLHs, with a hypothetical participation of AP2/ERF proteins. This core complex binds to the respective promoters and mediates the recruitment of additional components, such as epigenetic factors (orange) and Mediator proteins (green). EML histone readers interact with bHLHs and hypothetically mediate the recruitment of PRCs. The Mediator protein MED25 directly interacts with the histone writer HAC1, which activates the target promoter. It also recruits further Mediator components to initiate transcription by RNA polymerase II (yellow). The MYB-bHLH complex also integrates hormonal and other information via diverse factors that interfere with its formation and thus prevent its function, e.g. JAZ at low JA concentrations, BES1 at high BR concentrations and SDI1 under sulfur starvation. The thin black line indicates the promoter, while the thick black line indicates the transcribed sequence.

To further elucidate EML function, future research should include co-immunoprecipitation assays of MYB, bHLH and EML proteins as well as further OX assays, especially in the Arabidopsis *bhlh04/05/06* cell culture where the interfering feedback effects of the native bHLHs should be limited. Furthermore, studies on the role of EMLs in processes beyond pure GSL metabolism could shed light on the coordinated regulation of defense and development in Brassicaceae – for example in regard to ageing, the broader defense priming response or other JA-mediated processes including anthocyanin synthesis and root growth (Huang *et al.*, 2017; Chen and Wang, 2019).

4.2 Exploring the regulatory basis of GSL variation

Several studies have been conducted that address the contribution of biosynthetic genes to interspecific or inter-accession diversity in GSL profiles. These effects include secondary modifications of aliphatic GSLs by *AOP2* and *AOP3* (Kliebenstein *et al.*, 2001; Liu *et al.*, 2014), of indolic GSLs by *CYP81F4* (Czerniawski *et al.*, 2021) as well as chain elongation by *MAM* genes (Liu *et al.*, 2014; Czerniawski *et al.*, 2021). At the same time, regulatory genes have been suggested to play a significant role in creating this diversity (Windsor *et al.*, 2005; Kerwin *et al.*, 2015) – however, no studies have directly addressed this topic so far. Thus, my aim in this study was to explore to which extent *MYB* and *bHLH* genes could contribute to interspecies variation in GSL accumulation. Phylogenetic comparison and overexpression experiments suggest that they may be limited to a relatively unspecific influence – which is likely the case in *C. rubella*, where alternative splicing of *bHLH* transcripts and low expression of *MYBs* seem to play a part in an extremely low production of indolic GSLs.

4.2.1 The contribution of regulatory genes to interspecific GSL diversity

A phylogenetic analysis of subgroup 12 *MYB* and subgroup IIIe *bHLH* sequences from Arabidopsis, *A. alpina*, *C. hirsuta* and *C. rubella* reveals a strong conservation of *bHLHs*. While the same is generally true for *MYBs*, a somewhat higher degree of variation is visible in this group (**Figure 14**). First of all, *MYB76* is not present widely among the Brassicaceae, but only in plants of lineage I with a secondary loss in *Capsella* species. Taking this into account, all ortholog genes cluster together, with the distance between clusters reflecting the evolutionary history of the subgroup 12 *MYB* family: *MYB34* derives from a pre-At-α duplication, while the paralog pairs *MYB28/MYB29* and *MYB51/MYB122* were created during the At-α WGD (Hofberger *et al.*, 2013; Barco and Clay, 2019) and *MYB76* only arose recently from *MYB29* during the split of lineage I from the other Brassicaceae plants.

Nonetheless, *MYB29*, *MYB34* and *MYB122* do not cluster in complete accordance with the established Brassicaceae phylogeny. These differences could be explained by a single assumption, namely an increased divergence of *MYB* genes in *C. rubella*. This prompted us to focus in particular on the interspecific conservation of MYB function and the role of regulatory factors in *C. rubella* in the following experiments.

To complement the *in silico* sequence similarity with *in vitro* experimental data about functional conservation, various *MYBs* from all examined species were overexpressed in Arabidopsis Ler-0 cell culture (**Figure 15**). For indolic *MYBs*, no differential effect was observed, as all of them only induce a single GSL species, the main indolic GSL I3M. This accords with previous studies which reported the regulation of the synthesis of modified indolic GSLs to be independent from MYB activity (Wiesner *et al.*, 2013; Frerigmann and Gigolashvili, 2014; Zang *et al.*, 2015). A quantitative difference in I3M induction between the different indolic *MYBs* could in theory exist, but this question is beyond the scope of the experiment and would need to be addressed in the future.

Among the aliphatic *MYBs*, the individual orthologs yield qualitatively different GSL profiles. MYB28 seems to have a broader function than the others, inducing more long-chain Met-GSLs – in accordance with previous reports (Sønderby *et al.*, 2007; Beekwilder *et al.*, 2008; Sønderby *et al.*, 2010a; Li *et al.*, 2013) – and also a significant amount of the indolic GSL I3M. This effect is conserved between the two examined MYB28 orthologs from Arabidopsis and *A. alpina*, i.e. from the two most distantly related species. It could stem either from an induction of indolic GSL synthesis enzymes or from a certain affinity of aliphatic enzymes for indolic substrates, a question I addressed in more detail in *C. rubella*, as discussed in section 4.2.3. In comparison, OXs of *MYB29* and *MYB76* led to a relatively lower accumulation of long-chain Met-GSLs and no I3M accumulation at all. The functional difference between these two TFs seems to be notably smaller, as the evolutionary split between these two is much younger than between them and *MYB28*.

It is thought that the At-α WGD event enabled the structural diversification of GSLs, which in turn led to the diversification and radiation of the Brassicaceae family and a subsequent coradiation of the butterfly family Pieridae (Schranz *et al.*, 2011; Edger *et al.*, 2015; Edger *et al.*, 2018). Despite this evolutionary central role of GSL diversity and the emergence of most subgroup 12 *MYBs* as part of At-α, it seems that the individual MYB factors are highly conserved across the Brassicaceae. This is true both sequence-wise, with at least the four key TFs *MYB28*, *MYB29*, *MYB34* and *MYB51* present in all Brassicaceae species studied so far, and functionality-wise, with the respective orthologs generating similar GSL profiles.

This means that variation in subgroup 12 *MYBs* may only influence GSL diversity as a function of their expression strength, in two different ways: Either by promoting or repressing the entire aliphatic or indolic branches of GSL synthesis – or by conveying rather subtle differences in I3M or long-chain GSL induction. This corresponds well to the study by Windsor *et al.* (2005), who proposed a role of regulatory factors in broad changes to GSL metabolism, such as the absence of Met-derived GSLs from most tissues (except seeds) in some Brassicaceae species. The same is true for Kerwin *et al.* (2015), who demonstrated the evolutionary significance of *MYBs* by a comparison between Col-0 wildtype and *myb28* and *myb29* mutant plants. In contrast, studies in *B. rapa* reported that divergence between additional *MYB* copies of the polyploid genome might result in more significant functional divergences (Seo *et al.*, 2016; Seo *et al.*, 2017), although further research is needed here.

To further explore this topic, native and induced expression profiles of subgroup 12 *MYB* genes could be established in different Brassicaceae species, then correlated with GSL profiles. This approach would also be feasible for different accessions of one species or different tissues of one accession, where the *MYBs* have less potential for divergence. Alternatively, OX assays such as the ones performed in this study could be complemented with quantitative measurements such as qPCR to obtain a more precise measure of how strongly particular MYBs induce particular GSLs.

In summary, a slight degree of functional variation exists between some of the individual subgroup 12 MYB TFs. Overall however, these MYBs seem highly conserved between Brassicaceae species and thus may only contribute to GSL profile diversity in a rather broad fashion, i.e. by mediating very specific differences or by affecting entire branches of aliphatic or indolic GSL synthesis. It seems probable that the impact of bHLHs on GSL diversity is not too dissimilar from the impact of MYBs. While both can be assumed to primarily exert broad influences, the role of bHLHs is presumably even broader because they do not mediate specificity in regard to particular GSL synthesis genes. Additionally, subgroup IIIe *bHLHs* seem to be conserved even more strongly that subgroup 12 *MYBs* in the phylogenetic comparison (**Figure 14B**). This may easily be explained by their central importance for numerous regulatory processes besides GSL synthesis, which all depend on their correct functions (Kazan and Manners, 2013; Pireyre and Burow, 2015; Goossens *et al.*, 2017).

However, it is conceivable that different bHLH TFs could possess different affinities towards different MYBs, depending on the makeup of their JID. It is also possible that bHLHs can be modified post-transcriptionally or -translationally under specific circumstances, a topic that I partially explored in *C. rubella*, as discussed in section 4.2.3.

4.2.2 Indolic GSL accumulation is drastically reduced in *C. rubella*

The large range of variation in GSL profiles between different Brassicaceae species is well documented. To put the findings from Arabidopsis into context, I also aimed to study the contribution of regulatory genes to GSL diversity in other species, namely *A. alpina*, *C. rubella* and *C. hirsuta*. The first goal was to establish cell cultures of these plants. In parallel, an LC-MS analysis was conducted that generated detailed GSL profiles for root and shoot tissues.

As a basis for further experiments with the selected Brassicaceae species, root tissue-derived cell cultures of *A. alpina* and *C. rubella* were established. *C. rubella*, the closest relative to Arabidopsis, was successfully cultivated under the conditions established for Arabidopsis and subsequently used for overexpression assays in this study. *A. alpina*, a perennial plant, exhibited slower growth as well as significant oxidative browning. These problems required modifications to the protocol, but ultimately led to the successful establishment of callus and suspension cultures which can be used for future studies. Additionally, the oxidative browning, which is caused by the overproduction and subsequent oxidation of phenolic compounds as a result of the stress response, could be addressed by the use of antioxidants, adsorbants or the targeted inhibitor aminoindane-2-phosphonic acid (Jones and Saxena, 2013). A *C. hirsuta* cell culture could not be established as lower auxin concentrations led to dedifferentiation of the callus culture, while higher auxin concentrations led to an arrest of growth. In the future, it could be attempted if callus growth can be restored by an additional increase in cytokinin concentrations.

The LC-MS analysis underlined both interspecific and inter-tissue diversity in GSL profiles and provided novel data both in regard to the examined tissues and the analytic sensitivity (**Table 22**). As far as they can be compared, the results obtained in this study are in excellent agreement with previously published

data. Seeds of *A. alpina* and *C. hirsuta* have been analyzed twice in the past (Daxenbichler *et al.*, 1991; Bennett *et al.*, 2004). While all GSL species described there were also detected for root or shoot tissue in this study, a higher sensitivity enabled the detection of numerous additional GSL structures. Bednarek *et al.* (2011) quantified indolic GSLs (I3M, 1MO-I3M and 4MO-I3M) in leaves of numerous Brassicaceae species, including the three analyzed in this study. The trends described for these compounds were confirmed in this study, with all three GSL structures present in *C. hirsuta*, 1MO-I3M absent in *A. alpina* and overall low indolic GSL content in *C. rubella*. Czerniawski *et al.* (2021) revisited *C. rubella* with a detailed analysis of root and leaf tissue, the latter in various developmental stages. These results notably agree with this study insofar as in both, small amounts of I3M, 4OH-I3M and 4MO-I3M were detected. Additionally, all reported Met-GSL structures were confirmed in this study, with additional findings of alkyl and benzenic GSLs.

The total numbers of identified GSL structures in root and shoot tissues were 27 resp. 17 for *C. rubella*, 28 resp. 26 for *C. hirsuta* and in both cases 32 for *A. alpina*. In comparison with previous studies, the data agree with observed tendencies that root tissue produces a higher diversity of GSLs, but disagree with tendencies of a higher total concentration (Brown *et al.*, 2003; van Dam *et al.*, 2009; Zang *et al.*, 2015), which emphasizes that such trends cannot be regarded as universal across the Brassicaceae.

The majority of identified GSL structures in this study are well-described, including the responsible enzymes. Some however are completely unknown and seem to represent isomers to established GSLs. These structures could either represent novel GSLs or examples for the proposed iso-GSLs, molecules where the defining GSL backbone itself (**Figure 1**) exhibits one or more isomerisms compared to actual GSLs (Blažević *et al.*, 2020). These compounds would have to be products of so far undescribed enzymatic reactions or regulatory mechanisms. Identifying such compounds would however require tandem mass spectrometry (MS/MS) and possibly nuclear magnetic resonance (NMR) analysis.

The wealth of data generated by the LC-MS analysis provides possibilities for many further avenues of research, as also for the documented GSL structures, many pieces of the puzzle are still missing: In some cases, such as for the modified indolic GSLs, the responsible enzymes are known (Pfalz *et al.*, 2009; Pfalz *et al.*, 2011; Pfalz *et al.*, 2016), but their regulation remains unclear. In other cases, such as for benzenic and methylsulfonylalkyl GSLs, both synthesis and regulation are unknown to date. As all of these compounds are differentially accumulated across the examined species in this study, genomic and transcriptomic comparisons could provide information about candidates for the missing components. If more closely related species were to be used, this could also be approached by crossing and subsequent quantitative trait locus (QTL) mapping.

The most striking part of all analyzed GSL profiles is the almost complete lack of indolic GSLs in *C. rubella*. It represents a divergence whose broad nature corresponds well to what was postulated to potentially be mediated by variation in *MYB* and *bHLH* genes in the previous section. Several deregulations affecting both biosynthetic and regulatory components of the indolic GSL pathway have

already been identified in *C. rubella* and other members of Camelineae clade II (**Figure 13**). They include the loss of several genes, namely *MYB76*, the atypical myrosinase *PEN2* as well as the indolic secondary modification enzymes *CYP81F2*, *CYP81F4* and multiple copies of *IGMT* genes. In particular, the loss of *CYP81F4* has been suggested to cause the absence of 1MO-I3M (Bednarek *et al.*, 2011; Czerniawski *et al.*, 2021). Despite this, *C. rubella* fundamentally retains the biosynthetic capability to produce the indolic GSLs I3M, 4OH-I3M and 4MO-I3M, albeit their concentration is drastically reduced.

4.2.3 Alternative splicing of *bHLH* transcripts and low *MYB* expression contribute to low indolic GSL levels in *C. rubella*

The unique GSL profile of *C. rubella*, together with the successful establishment of a cell culture and the potential divergences of *MYB* genes apparent in the phylogenetic data, led me to focus on the role of *MYBs* and *bHLHs* in the regulation of GSL biosynthesis in this species. In fact, this study identified a new regulatory mechanism in *C. rubella* which so far has not been described in other Brassicaceae species: the alternative splicing of subgroup IIIe *bHLHs* (**Table 23**). All three *bHLHs* considered here seem to be affected in similar ways, i.e. by truncating large parts of the transcript, often including the JID, TAD or even bHLH domains. The resulting transcripts are rendered non-functional, as demonstrated by OX assays with GSL quantification (**Figure 17**) and GUS staining (**Figure 16**) in Arabidopsis *bhlh04/05/06* cell culture.

The true relevance of alternative *bHLH* splicing as a regulatory mechanism needs to be assessed in the future by a quantitative analysis such as qPCR or RNA-Seq (which holds the power to detect further splice variants) *in vivo,* preferably under various stressed and unstressed conditions, in different tissues and developmental stages. Further information about the prevalence and potential role of non-functional splice variants would be interesting not least because such a regulation of bHLH functions would affect a plethora of JA-mediated mechanisms besides GSL synthesis. To date, there is no indication that JA signaling is generally disturbed in *C. rubella*. However, in this study, JA treatment did not induce GSL production in *C. rubella* cells. Thus, further research is needed in this direction.

As mentioned above, the phylogenetic sequence analysis provided evidence for a possible beginning divergence of *MYB* genes, especially regarding *CrMYB34* (**Figure 14A**). At the same time, expression of all indolic *CrMYBs* is drastically reduced in comparison to Arabidopsis, with *CrMYB34* transcripts practically completely absent (P. Bednarek, unpublished data). Consequentially, only *CrMYB51* could be cloned successfully, with its OX in Arabidopsis Ler-0 cell culture not revealing any obvious functional differences to its Brassicaceae orthologs (**Figure 15**). On the one hand, it would be highly interesting to test the functionality of the more divergent *CrMYB34*, but on the other hand, this functionality may not play any significant role *in vivo* if the gene is never expressed. In fact, the relative conservation of *CrMYB51* may result from its functional importance remaining relatively high compared to the other indolic *CrMYBs*. In any case, their low expression rates may play a decisive role in reducing indolic GSL synthesis in *C. rubella*.

Taken together, both the decreased expression of indolic *MYB* genes and the alternative splicing of *bHLH* transcripts could potentially diminish the expression of indolic GSL biosynthetic genes and thus negatively influence the production of these metabolites in *C. rubella*. To test this hypothesis, *MYBs* from all analyzed species and *bHLHs* from Arabidopsis were overexpressed in *C. rubella* cell culture with subsequent GSL quantification and GUS staining. Interestingly, ortholog *MYBs* elicited similar GSL profiles, and these profiles were in turn highly similar to the ones observed in Arabidopsis Ler-0 cell culture.

Generally, indolic MYBs were able to induce a minor accumulation of I3M at concentrations of 10–100 nmol/g DW, while AtMYB28 and AaMYB28 were able to induce a significantly higher amount of 1–2 µmol/g DW (**Figure 18**). At the same time, GUS staining indicated only a weak induction of *AtCYP79B3* expression under *MYB28* OX, while OX of indolic *MYBs* led to a mild, but notable staining (**Supplementary Figure 8**). This would tendentially suggest that MYB28 does not induce indolic GSL biosynthesis enzymes, but rather that the aliphatic enzymes have a residual affinity for indolic GSL substrates. However, the high similarity in binding site specificity exhibited between members of one MYB subgroup (Romero *et al.*, 1998; Jiang and Rao, 2020), as well as empiric data from Arabidopsis that potentially indicates a minor affinity of MYB28 for indolic GSL synthesis genes (Gigolashvili *et al.*, 2007b; Hirai *et al.*, 2007) point towards the second possibility, with the respective contributions of both mechanisms to I3M induction by MYB28 remaining to be elucidated.

Whereas orthologs of the aliphatic regulator MYB28 proved capable to induce indolic GSLs in this experiment, orthologs of the other aliphatic regulators MYB29 and MYB76 did not show any influence on indolic GSL levels, identical to what was also observed in Arabidopsis cell culture. This difference could stem from a different affinity towards either indolic GSL synthesis genes or particular aliphatic GSL synthesis enzymes that accept indolic substrates. The former hypothesis is supported by slightly lower *AtCYP79B3* induction displayed in the qualitative GUS assay under *MYB29* and *MYB76* OX than under *MYB28* OX (**Supplementary Figure 8**).

Most importantly, this experiment shows that low indolic *MYB* expression is at least partially the cause of low indolic GSL production in *C. rubella* and that *MYB* OX can induce the intact core pathway of indolic GSL synthesis to a certain extent. Additionally, *bHLH* OX seems to have a compounding effect on I3M concentrations (**Figure 18A**), but it is difficult to pinpoint due to the metabolic feedback mechanism which presumably is also active in *C. rubella*. The connection between this effect and the alternative splicing of *CrbHLH* transcripts needs to be studied in the future.

Notably, no combination of *MYB* and *bHLH* OX tested in this study could increase I3M production to concentrations observed in other species. A possible explanation for this is that other further genes could be involved in this phenomenon, considering that *C. rubella* shows multiple signs of a beginning deregulation of indolic GSL synthesis. Such candidates include *PEN2*, which is essential for the biological activation of 4MO-I3M (Bednarek *et al.*, 2009), but absent in the Camelineae clade II species (Bednarek

et al., 2011; Barco and Clay, 2019; Czerniawski *et al.*, 2021). The lack of *PEN2* hydrolysis products might disequilibrate the metabolic feedback mechanisms of GSL synthesis in *C. rubella* and prevent further indolic GSL production. However, *C. rubella* retains defense functions that are mediated via the PEN2-dependent hydrolysis of 4MO-I3M (Clay *et al.*, 2009; Bednarek *et al.*, 2011), which suggests that PEN2 functions could be taken over by the paralog *PEN2L* (Barco and Clay, 2019).

Hypothetically, transcriptional or functional divergences in enzymes of the indolic GSL synthesis pathway, e.g. reduced expression or substrate affinity of *CYP83B1*, could also contribute to low indolic GSL levels. Another possibility could be the reduced availability of tryptophan as amino acid precursor. This is unlikely, however, because such a change in primary metabolism would not only profoundly affect growth and development in general, but also auxin and camalexin metabolism in particular, both of which are apparently not deregulated in *C. rubella* (Bednarek et al., 2011). Eventually, future research should aim to determine the relative roles of reduced *MYB* expression, alterative *bHLH* splicing, absence of *PEN2* and potential changes to the biosynthetic machinery in causing the remarkable indolic GSL profile of *C. rubella*.

In conclusion, in light of the multiple deregulations both on the biosynthetic and regulatory level, it seems likely that the indolic branch of the GSL pathway is in the process of being abolished in *C. rubella* and its relatives of Camelineae clade II (Czerniawski *et al.*, 2021). This of course raises the question of which ecological context would render the loss of indolic GSLs into an evolutionary fitness benefit. Have these molecules become useless against the main pathogens and herbivores of these species? It could be speculated that the plants are especially targeted by specialist herbivores, who are much less strongly affected by GSLs than generalist herbivores (Mitreiter and Gigolashvili, 2020). These insects have various ways of circumventing the GSL-myrosinase system (Wittstock *et al.*, 2016a): For example, Pieridae butterflies produce endogenous nitrile specifier proteins to detoxify GSLs (Wittstock *et al.*, 2004; Wheat *et al.*, 2007). In fact, GSLs even serve as oviposition cues for specialist herbivores (Marazzi and Städler, 2004; Sun *et al.*, 2009; Schweizer *et al.*, 2013) and thus could lead to drawbacks in fitness.

Future research needs to determine if in Camelineae clade II plants, the functions of indolic GSLs are taken over by other compounds. This could mean either molecules that are related in synthesis and regulation, such as camalexin (Glawischnig *et al.*, 2004; Frerigmann *et al.*, 2015; Frerigmann *et al.*, 2016), 4-hydroxyindole-3-carbonylnitrile (Rajniak *et al.*, 2015; Barco and Clay, 2020) and indole-3-carboxylic acids (Böttcher *et al.*, 2009; Böttcher *et al.*, 2014; Frerigmann *et al.*, 2016) or completely novel molecules. In any case, such an evolutionary development could provide a fascinating insight into the ongoing biochemical arms race between plants and insects that has been the driving factor in developing GSL diversity in the first place (Schranz *et al.*, 2011; Edger *et al.*, 2015; Edger *et al.*, 2018).

4.3 Conclusion

In this study, I expanded the knowledge about the transcriptional regulation of GSL biosynthesis in two areas: First, I demonstrated that the EML protein family likely participates in the epigenetic control of

GSL biosynthesis, in particular in its JA-mediated induction. In their function as histone readers, EMLs recognize activating histone modifications and likely interact with further epigenetic components to contribute to an epigenetic shift in the activation of GSL defenses. This function is mediated by the direct interaction between EMLs and subgroup IIIe bHLHs, i.e. the participation of EMLs in the core MYB-bHLH regulatory complex of GSL biosynthesis. These findings shed light on the epigenetic regulation of GSLs in particular and secondary metabolism in general.

Second, the results presented in this thesis suggest that the role of *MYBs* and *bHLHs* in interspecific GSL diversity may be mostly limited to broad influences on the entire aliphatic and indolic GSL pathways. This effect is likely exemplified in *C. rubella*, where the alternative splicing of *bHLH* transcripts into non-functional variants and the reduced expression of indolic *MYBs* may play crucial roles in diminishing the accumulation of indolic GSLs. These findings enable a better understanding of how the biosynthetic and regulatory levels work together to generate diverse GSL profiles.

Ultimately, both topics can be interpreted as two sides of the same coin, as both contribute to the optimal coordination of defense with growth and development. The relationship between these two aspects of plant life is much more than a simple trade-off between resources – in fact, the metabolic requirements are not automatically identical for both. Instead, defense and development should be understood as interacting in a highly complex cross-talk, in which reactions to environmental stress are facilitated by adjustments in the developmental program (Kliebenstein, 2016; Burow and Halkier, 2017).

This coordination needs to integrate information about the plant's previous experiences as well as countless internal and external (i.e. abiotic and biotic) factors and consequently operates on many different regulatory levels. These levels do not only encompass transcription factors and hormones, but also a variety of epigenetic, posttranslational and posttranscriptional mechanisms (with the latter including alternative splicing). On an interspecific or inter-accession level, we can additionally study how these mechanisms interact with variation in the target genes themselves.

For all of this, the regulation of GSL biosynthesis can serve as a prominent example. Both the epigenetic regulation by EML proteins in response to stress and the alterative splicing of *EML* and *bHLH* transcripts documented in this study represent mechanisms whose importance, in particular in the context of ecology and evolution, is just beginning to be understood (Syed *et al.*, 2012; Richards *et al.*, 2017; Laloum *et al.*, 2018; Alonso *et al.*, 2019). Taken together, an extremely complex system of multi-layered regulation provides plants with a high degree of flexibility that enables them to continually adapt and maximize their fitness in response to an ever-changing environment – from an individual to a species-wide scale. This thesis opens up new perspectives for the study of this regulation, helping to fully understand all facets of these mechanisms in the future.

5 References

Aarabi F, Kusajima M, Tohge T, Konishi T, Gigolashvili T, Takamune M, Sasazaki Y, Watanabe M, Nakashita H, Fernie AR, Saito K, Takahashi H, Hubberten H-M, Hoefgen R, Maruyama-Nakashita A. 2016. Sulfur deficiency–induced repressor proteins optimize glucosinolate biosynthesis in plants. Science Advances **2**, e1601087.

Agerbirk N, Olsen CE. 2012. Glucosinolate structures in evolution. Phytochemistry **77**, 16-45.

Aller EST, Jagd LM, Kliebenstein DJ, Burow M. 2018. Comparison of the Relative Potential for Epigenetic and Genetic Variation To Contribute to Trait Stability. G3: Genes|Genomes|Genetics **8**, 1733-1746.

Alonso C, Ramos-Cruz D, Becker C. 2019. The role of plant epigenetics in biotic interactions. New Phytologist **221**, 731-737.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol **215**, 403-410.

Alvarez-Venegas R. 2010. Regulation by polycomb and trithorax group proteins in Arabidopsis. Arabidopsis Book **8**, e0128.

An C, Li L, Zhai Q, You Y, Deng L, Wu F, Chen R, Jiang H, Wang H, Chen Q, Li C. 2017. Mediator subunit MED25 links the jasmonate receptor to transcriptionally active chromatin. Proceedings of the National Academy of Sciences **114**, E8930-E8939.

Augustine R, Bisht N. 2016. Regulation of Glucosinolate Metabolism: From Model Plant Arabidopsis thaliana to Brassica Crops. 1-37.

Augustine R, Bisht NC. 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 NC. 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**, 4907-4921.

Bahlmann A-K. 2020. Understanding the molecular mechanisms of transcriptional regulation and metabolite sensing by the MYB–bHLH complex, University of Cologne, Cologne.

Bakhtiari M, Glauser G, Rasmann S. 2018. Root JA Induction Modifies Glucosinolate Profiles and Increases Subsequent Aboveground Resistance to Herbivore Attack in Cardamine hirsuta. Frontiers in Plant Science **9**, 1230.

Barco B, Clay NK. 2019. Evolution of Glucosinolate Diversity via Whole-Genome Duplications, Gene Rearrangements, and Substrate Promiscuity. Annual Review of Plant Biology **70**, 585-604.

Barco B, Clay NK. 2020. Hierarchical and Dynamic Regulation of Defense-Responsive Specialized Metabolism by WRKY and MYB Transcription Factors. Frontiers in Plant Science **10**, 1775.

Bartlet E, Parsons D, Williams IH, Clark SJ. 1994. The influence of glucosinolates and sugars on feeding by the cabbage stem flea beetle, Psylliodes chrysocephala. Entomologia Experimentalis et Applicata **73**, 77-83.

Bednarek P, Piślewska-Bednarek M, Svatoš A, Schneider B, Doubský J, Mansurova M, Humphry M, Consonni C, Panstruga R, Sanchez-Vallet A, Molina A, Schulze-Lefert P. 2009. A Glucosinolate Metabolism Pathway in Living Plant Cells Mediates Broad-Spectrum Antifungal Defense. Science **323**, 101-106.

Bednarek P, Piślewska-Bednarek M, Ver Loren van Themaat E, Maddula RK, Svatoš A, Schulze-Lefert P. 2011. Conservation and clade-specific diversification of pathogen-inducible tryptophan and indole glucosinolate metabolism in Arabidopsis thaliana relatives. New Phytologist **192**, 713-726.

Beekwilder J, van Leeuwen W, van Dam NM, Bertossi M, Grandi V, Mizzi L, Soloviev M, Szabados L, Molthoff JW, Schipper B, Verbocht H, de Vos RCH, Morandini P, Aarts MGM, Bovy A. 2008. The Impact of the Absence of Aliphatic Glucosinolates on Insect Herbivory in Arabidopsis. PLOS ONE **3**, e2068.

Beilstein MA, Al-Shehbaz IA, Kellogg EA. 2006. Brassicaceae phylogeny and trichome evolution. American Journal of Botany **93**, 607-619.

Bekaert M, Edger PP, Hudson CM, Pires JC, Conant GC. 2012. Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis. The New Phytologist **196**, 596-605.
Bell L. 2019. The Biosynthesis of Glucosinolates: Insights, Inconsistencies, and Unknowns. *Annual Plant Reviews online*: American Cancer Society, 969-1000.
Bell L, Oloyede OO, Lignou S, Wagstaff C, Methven L. 2018. Taste and Flavor Perceptions of Glucosinolates, Isothiocyanates, and Related Compounds. Molecular Nutrition & Food Research **62**, 1700990.
Bennett RN, Mellon FA, Kroon PA. 2004. Screening crucifer seeds as sources of specific intact glucosinolates using ion-pair high-performance liquid chromatography negative ion electrospray mass spectrometry. Journal of Agricultural and Food Chemistry **52**, 428-438.
Bennett RN, Wenke T, Freudenberg B, Mellon FA, Ludwig-Müller J. 2005. The tu8 mutation of Arabidopsis thaliana encoding a heterochromatin protein 1 homolog causes defects in the induction of secondary metabolite biosynthesis. Plant Biology (Stuttgart, Germany) **7**, 348-357.
Berr A, Ménard R, Heitz T, Shen W-H. 2012. Chromatin modification and remodelling: a regulatory landscape for the control of Arabidopsis defence responses upon pathogen attack. Cellular Microbiology **14**, 829-839.
Blažević I, Montaut S, Burčul F, Olsen CE, Burow M, Rollin P, Agerbirk N. 2020. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. Phytochemistry **169**, 112100.
Böttcher C, Chapman A, Fellermeier F, Choudhary M, Scheel D, Glawischnig E. 2014. The Biosynthetic Pathway of Indole-3-Carbaldehyde and Indole-3-Carboxylic Acid Derivatives in Arabidopsis. Plant Physiology **165**, 841-853.
Böttcher C, Westphal L, Schmotz C, Prade E, Scheel D, Glawischnig E. 2009. The Multifunctional Enzyme CYP71B15 (PHYTOALEXIN DEFICIENT3) Converts Cysteine-Indole-3-Acetonitrile to Camalexin in the Indole-3-Acetonitrile Metabolic Network of Arabidopsis thaliana. The Plant Cell **21**, 1830-1845.
Bourbousse C, Ahmed I, Roudier F, Zabulon G, Blondet E, Balzergue S, Colot V, Bowler C, Barneche F. 2012. Histone H2B monoubiquitination facilitates the rapid modulation of gene expression during Arabidopsis photomorphogenesis. PLOS Genetics **8**, e1002825-e1002825.
Boyko A, Kovalchuk I. 2008. Epigenetic control of plant stress response. Environmental and Molecular Mutagenesis **49**, 61-72.
Brasil JN, Cabral LM, Eloy NB, Primo LMF, Barroso-Neto IL, Grangeiro LPP, Gonzalez N, Inzé D, Ferreira PCG, Hemerly AS. 2015. AIP1 is a novel Agenet/Tudor domain protein from Arabidopsis that interacts with regulators of DNA replication, transcription and chromatin remodeling. BMC Plant Biology **15**, 270.
Brkljacic J, Grotewold E. 2017. Combinatorial control of plant gene expression. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1860**, 31-40.
Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J. 2003. Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. Phytochemistry **62**, 471-481.
Burow M, Halkier BA. 2017. How does a plant orchestrate defense in time and space? Using glucosinolates in Arabidopsis as case study. Current Opinion in Plant Biology **38**, 142-147.
Burow M, Halkier BA, Kliebenstein DJ. 2010. Regulatory networks of glucosinolates shape Arabidopsis thaliana fitness. Current Opinion in Plant Biology **13**, 348-353.
Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL. 2010. Genome-Wide Classification and Evolutionary Analysis of the bHLH Family of Transcription Factors in Arabidopsis, Poplar, Rice, Moss, and Algae. Plant Physiology **153**, 1398-1412.
Cavalli G, Heard E. 2019. Advances in epigenetics link genetics to the environment and disease. Nature **571**, 489-499.
Celenza JL, Quiel JA, Smolen GA, Merrikh H, Silvestro AR, Normanly J, Bender J. 2005. The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. Plant Physiology **137**, 253-262.

Çevik V, Kidd BN, Zhang P, Hill C, Kiddle S, Denby KJ, Holub EB, Cahill DM, Manners JM, Schenk PM, Beynon J, Kazan K. 2012. MEDIATOR25 Acts as an Integrative Hub for the Regulation of Jasmonate-Responsive Gene Expression in Arabidopsis. Plant Physiology **160**, 541-555.
Chen R, Jiang H, Li L, Zhai Q, Qi L, Zhou W, Liu X, Li H, Zheng W, Sun J, Li C. 2012. The Arabidopsis Mediator Subunit MED25 Differentially Regulates Jasmonate and Abscisic Acid Signaling through Interacting with the MYC2 and ABI5 Transcription Factors. The Plant Cell **24**, 2898-2916.
Chen S, Wang S. 2019. GLABRA2, a Common Regulator for Epidermal Cell Fate Determination and Anthocyanin Biosynthesis in Arabidopsis. International Journal of Molecular Sciences **20**.
Cheng K, Xu Y, Yang C, Ouellette L, Niu L, Zhou X, Chu L, Zhuang F, Liu J, Wu H, Charron J-B, Luo M. 2019. Histone tales: lysine methylation, a protagonist in Arabidopsis development. Journal of Experimental Botany **71**, 793-807.
Chezem WR, Clay NK. 2016. Regulation of plant secondary metabolism and associated specialized cell development by MYBs and bHLHs. Phytochemistry **131**, 26-43.
Chini A, Fonseca S, Fernández G, Adie B, Chico JM, Lorenzo O, García-Casado G, López-Vidriero I, Lozano FM, Ponce MR, Micol JL, Solano R. 2007. The JAZ family of repressors is the missing link in jasmonate signalling. Nature **448**, 666-671.
Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Krämer U, Merchant SS, Zhang X, Jacobsen SE, Pellegrini M. 2010. Relationship between nucleosome positioning and DNA methylation. Nature **466**, 388-392.
Clarke DB. 2010. Glucosinolates, structures and analysis in food. Analytical Methods **2**, 310-325.
Clay NK, Adio AM, Denoux C, Jander G, Ausubel FM. 2009. Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response. Science (New York, N.Y.) **323**, 95-101.
Conrath U, Beckers GJM, Langenbach CJG, Jaskiewicz MR. 2015. Priming for Enhanced Defense. Annual Review of Phytopathology **53**, 97-119.
Coursey T, Milutinovic M, Regedanz E, Brkljacic J, Bisaro DM. 2018. Arabidopsis Histone Reader EMSY-LIKE 1 Binds H3K36 and Suppresses Geminivirus Infection. Journal of Virology **92**.
Czerniawski P, Piasecka A, Bednarek P. 2021. Evolutionary changes in the glucosinolate biosynthetic capacity in species representing Capsella, Camelina and Neslia genera. Phytochemistry **181**, 112571.
Davis MW. 2020. ApE - A plasmid Editor. Salt Lake City, UT, USA: University of Utah.
Daxenbichler ME, Spencer GF, Carlson DG, Rose GB, Brinker AM, Powell RG. 1991. Glucosinolate composition of seeds from 297 species of wild plants. Phytochemistry **30**, 2623-2638.
Del Carmen Martínez-Ballesta M, Moreno DA, Carvajal M. 2013. The Physiological Importance of Glucosinolates on Plant Response to Abiotic Stress in Brassica. International Journal of Molecular Sciences **14**, 11607-11625.
Ding B, Wang G-L. 2015. Chromatin versus pathogens: the function of epigenetics in plant immunity. Frontiers in Plant Science **6**, 675-675.
Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010. MYB transcription factors in Arabidopsis. Trends in Plant Science **15**, 573-581.
Eberharter A, Becker PB. 2002. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. EMBO Reports **3**, 224-229.
Edger PP, Hall JC, Harkess A, Tang M, Coombs J, Mohammadin S, Schranz ME, Xiong Z, Leebens-Mack J, Meyers BC, Sytsma KJ, Koch MA, Al-Shehbaz IA, Pires JC. 2018. Brassicales phylogeny inferred from 72 plastid genes: A reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. Am J Bot **105**, 463-469.
Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, Hofberger JA, Smithson A, Hall JC, Blanchette M, Bureau TE, Wright SI, dePamphilis CW, Eric Schranz M, Barker MS, Conant GC, Wahlberg N, Vogel H, Pires JC, Wheat CW. 2015. The butterfly plant arms-race escalated by gene and genome duplications. Proc Natl Acad Sci U S A **112**, 8362-8366.
Feller A, Hernandez JM, Grotewold E. 2006. An ACT-like Domain Participates in the Dimerization of Several Plant Basic-helix-loop-helix Transcription Factors. Journal of Biological Chemistry **281**, 28964-28974.

Feller A, Machemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. The Plant Journal **66**, 94-116.
Fernández-Calvo P, Chini A, Fernández-Barbero G, Chico J-M, Gimenez-Ibanez S, Geerinck J, Eeckhout D, Schweizer F, Godoy M, Franco-Zorrilla JM, Pauwels L, Witters E, Puga MI, Paz-Ares J, Goossens A, Reymond P, Jaeger GD, Solano R. 2011. The Arabidopsis bHLH Transcription Factors MYC3 and MYC4 Are Targets of JAZ Repressors and Act Additively with MYC2 in the Activation of Jasmonate Responses. The Plant Cell **23**, 701-715.
Figueroa P, Browse J. 2015. Male sterility in Arabidopsis induced by overexpression of a MYC5-SRDX chimeric repressor. The Plant Journal **81**, 849-860.
Fonseca S, Chini A, Hamberg M, Adie B, Porzel A, Kramell R, Miersch O, Wasternack C, Solano R. 2009. (+)-7-iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. Nature Chemical Biology **5**, 344-350.
Frerigmann H, Berger B, Gigolashvili T. 2014. bHLH05 is an interaction partner of MYB51 and a novel regulator of glucosinolate biosynthesis in Arabidopsis. Plant Physiol **166**, 349-369.
Frerigmann H, Gigolashvili T. 2014. MYB34, MYB51, and MYB122 Distinctly Regulate Indolic Glucosinolate Biosynthesis in Arabidopsis thaliana. Molecular Plant **7**, 814-828.
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.
Frerigmann H, Piślewska-Bednarek M, Sánchez-Vallet A, Molina A, Glawischnig E, Gigolashvili T, Bednarek P. 2016. Regulation of Pathogen-Triggered Tryptophan Metabolism in Arabidopsis thaliana by MYB Transcription Factors and Indole Glucosinolate Conversion Products. Molecular Plant **9**, 682-695.
Gan X, Hay A, Kwantes M, Haberer G, Hallab A, Ioio RD, Hofhuis H, Pieper B, Cartolano M, Neumann U, Nikolov LA, Song B, Hajheidari M, Briskine R, Kougioumoutzi E, Vlad D, Broholm S, Hein J, Meksem K, Lightfoot D, Shimizu KK, Shimizu-Inatsugi R, Imprialou M, Kudrna D, Wing R, Sato S, Huijser P, Filatov D, Mayer KFX, Mott R, Tsiantis M. 2016. The Cardamine hirsuta genome offers insight into the evolution of morphological diversity. Nature Plants **2**, 1-7.
Gasperini D, Chételat A, Acosta IF, Goossens J, Pauwels L, Goossens A, Dreos R, Alfonso E, Farmer EE. 2015. Multilayered Organization of Jasmonate Signalling in the Regulation of Root Growth. PLOS Genetics **11**, e1005300.
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**, 3-13.
Gigolashvili T, Berger B, Mock H-P, Müller C, Weisshaar B, Flügge U-I. 2007a. The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana. The Plant Journal: For Cell and Molecular Biology **50**, 886-901.
Gigolashvili T, Engqvist M, Yatusevich 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**, 627-642.
Gigolashvili T, Yatusevich R, Berger B, Müller C, Flügge U-I. 2007b. The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in Arabidopsis thaliana. The Plant Journal: For Cell and Molecular Biology **51**, 247-261.
Glawischnig E, Hansen BG, Olsen CE, Halkier BA. 2004. Camalexin is synthesized from indole-3-acetaldoxime, a key branching point between primary and secondary metabolism in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **101**, 8245-8250.
Gomez-Cano F, Carey L, Lucas K, García Navarrete T, Mukundi E, Lundback S, Schnell D, Grotewold E. 2020. CamRegBase: a gene regulation database for the biofuel crop, Camelina sativa. Database **2020**.
Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2011. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40**, D1178-D1186.
Goossens J, Mertens J, Goossens A, Napier R. 2017. Role and functioning of bHLH transcription factors in jasmonate signalling. Journal of Experimental Botany **68**, 1333-1347.

Guo R, Qian H, Shen W, Liu L, Zhang M, Cai C, Zhao Y, Qiao J, Wang Q. 2013a. BZR1 and BES1 participate in regulation of glucosinolate biosynthesis by brassinosteroids in Arabidopsis. Journal of Experimental Botany **64**, 2401-2412.
Guo R, Shen W, Qian H, Zhang M, Liu L, Wang Q. 2013b. Jasmonic acid and glucose synergistically modulate the accumulation of glucosinolates in Arabidopsis thaliana. Journal of Experimental Botany **64**, 5707-5719.
Halkier BA, Gershenzon J. 2006. Biology and Biochemistry of Glucosinolates. Annual Review of Plant Biology **57**, 303-333.
He G, Elling AA, Deng XW. 2011. The Epigenome and Plant Development. Annual Review of Plant Biology **62**, 411-435.
Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. 2003. The Basic Helix–Loop–Helix Transcription Factor Family in Plants: A Genome-Wide Study of Protein Structure and Functional Diversity. Molecular Biology and Evolution **20**, 735-747.
Hernandez JM, Feller A, Morohashi K, Frame K, Grotewold E. 2007. The basic helix–loop–helix domain of maize R links transcriptional regulation and histone modifications by recruitment of an EMSY-related factor. Proceedings of the National Academy of Sciences of the United States of America **104**, 17222-17227.
Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K. 2007. Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proceedings of the National Academy of Sciences **104**, 6478-6483.
Hoecker U. 2017. The activities of the E3 ubiquitin ligase COP1/SPA, a key repressor in light signaling. Current Opinion in Plant Biology **37**, 63-69.
Hofberger JA, Lyons E, Edger PP, Chris Pires J, Eric Schranz M. 2013. Whole Genome and Tandem Duplicate Retention Facilitated Glucosinolate Pathway Diversification in the Mustard Family. Genome Biology and Evolution **5**, 2155-2173.
Howe GA, Major IT, Koo AJ. 2018. Modularity in Jasmonate Signaling for Multistress Resilience. Annual Review of Plant Biology **69**, 387-415.
Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, Mueller LA, Bhattacharyya D, Bhaya D, Sobral BW, Beavis W, Meinke DW, Town CD, Somerville C, Rhee SY. 2001. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Research **29**, 102-105.
Huang H, Liu B, Liu L, Song S. 2017. Jasmonate action in plant growth and development. Journal of Experimental Botany **68**, 1349-1359.
Huseby S, Koprivova A, Lee B-R, Saha S, Mithen R, Wold A-B, Bengtsson GB, Kopriva S. 2013. Diurnal and light regulation of sulphur assimilation and glucosinolate biosynthesis in Arabidopsis. Journal of Experimental Botany **64**, 1039-1048.
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**, 48-59.
Jensen L, Jepsen H, Halkier B, Kliebenstein D, Burow M. 2015a. Natural variation in cross-talk between glucosinolates and onset of flowering in Arabidopsis. Frontiers in Plant Science **6**.
Jensen LM, Kliebenstein DJ, Burow M. 2015b. Investigation of the multifunctional gene AOP3 expands the regulatory network fine-tuning glucosinolate production in Arabidopsis. Frontiers in Plant Science **6**.
Jiang C-K, Rao G-Y. 2020. Insights into the Diversification and Evolution of R2R3-MYB Transcription Factors in Plants. Plant Physiology **183**, 637-655.
Jones AMP, Saxena PK. 2013. Inhibition of phenylpropanoid biosynthesis in Artemisia annua L.: a novel approach to reduce oxidative browning in plant tissue culture. PLOS ONE **8**, e76802-e76802.
Justen VL, Fritz VA. 2013. Temperature-induced Glucosinolate Accumulation Is Associated with Expression of BrMYB Transcription Factors. HortScience **48**, 47-52.
Katz E, Nisani S, Chamovitz DA. 2018. Indole-3-carbinol: a plant hormone combatting cancer. F1000Research **7**, 689.

Katzen F. 2007. Gateway® recombinational cloning: a biological operating system. Expert Opinion on Drug Discovery **2**, 571-589.
Kazan K, Manners JM. 2013. MYC2: The Master in Action. Molecular Plant **6**, 686-703.
Kelemen Z, Sebastian A, Xu W, Grain D, Salsac F, Avon A, Berger N, Tran J, Dubreucq B, Lurin C, Lepiniec L, Contreras-Moreira B, Dubos C. 2015. Analysis of the DNA-Binding Activities of the Arabidopsis R2R3-MYB Transcription Factor Family by One-Hybrid Experiments in Yeast. PLOS ONE **10**, e0141044.
Kerwin R, Feusier J, Corwin J, Rubin M, Lin C, Muok A, Larson B, Li B, Joseph B, Francisco M, Copeland D, Weinig C, Kliebenstein DJ. 2015. Natural genetic variation in Arabidopsis thaliana defense metabolism genes modulates field fitness. eLife **4**, e05604.
Kim D-H, Sung S. 2014. Polycomb-mediated gene silencing in Arabidopsis thaliana. Molecules and cells **37**, 841-850.
Kim J-M, Sasaki T, Ueda M, Sako K, Seki M. 2015. Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. Frontiers in Plant Science **6**.
Kim J-M, To TK, Matsui A, Tanoi K, Kobayashi NI, Matsuda F, Habu Y, Ogawa D, Sakamoto T, Matsunaga S, Bashir K, Rasheed S, Ando M, Takeda H, Kawaura K, Kusano M, Fukushima A, Endo TA, Kuromori T, Ishida J, Morosawa T, Tanaka M, Torii C, Takebayashi Y, Sakakibara H, Ogihara Y, Saito K, Shinozaki K, Devoto A, Seki M. 2017. Acetate-mediated novel survival strategy against drought in plants. Nature Plants **3**, 17097.
Kim J, Felton GW. 2013. Priming of antiherbivore defensive responses in plants. Insect Science **20**, 273-285.
Kim JH, Durrett TP, Last RL, Jander G. 2004. Characterization of the Arabidopsis TU8 Glucosinolate Mutation,an Allele of TERMINAL FLOWER2. Plant Molecular Biology **54**, 671-682.
Kissen R, Eberl F, Winge P, Uleberg E, Martinussen I, Bones AM. 2016. Effect of growth temperature on glucosinolate profiles in Arabidopsis thaliana accessions. Phytochemistry **130**, 106-118.
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.
Kliebenstein DJ. 2016. False idolatry of the mythical growth versus immunity tradeoff in molecular systems plant pathology. Physiological and Molecular Plant Pathology **95**, 55-59.
Kliebenstein DJ, Kroymann J, Mitchell-Olds T. 2005. The glucosinolate–myrosinase system in an ecological and evolutionary context. Current Opinion in Plant Biology **8**, 264-271.
Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T. 2001. Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. Plant Cell **13**, 681-693.
Koç A, Markovic D, Ninkovic V, Martinez G. 2020. Chapter 16 - Molecular mechanisms regulating priming and stress memory. In: Hossain MA, Liu F, Burritt DJ, Fujita M, Huang B, eds. *Priming-Mediated Stress and Cross-Stress Tolerance in Crop Plants*: Academic Press, 247-265.
Koncz C, Schell J. 1986. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. Molecular and General Genetics MGG **204**, 383-396.
Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics **23**, 1289-1291.
Kouzarides T. 2007. Chromatin Modifications and Their Function. Cell **128**, 693-705.
Kranz HD, Denekamp M, Greco R, Jin H, Leyva A, Meissner RC, Petroni K, Urzainqui A, Bevan M, Martin C, Smeekens S, Tonelli C, Paz-Ares J, Weisshaar B. 1998. Towards functional characterisation of the members of the R2R3-MYB gene family from Arabidopsis thaliana. The Plant Journal: For Cell and Molecular Biology **16**, 263-276.
Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol **35**, 1547-1549.
Laloum T, Martín G, Duque P. 2018. Alternative Splicing Control of Abiotic Stress Responses. Trends in Plant Science **23**, 140-150.
Lämke J, Bäurle I. 2017. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biology **18**, 124.

Levy M, Wang Q, Kaspi R, Parrella MP, Abel S. 2005. Arabidopsis IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense. The Plant Journal: For Cell and Molecular Biology **43**, 79-96.
Li B, Gaudinier A, Tang M, Taylor-Teeples M, Nham NT, Ghaffari C, Benson DS, Steinmann M, Gray JA, Brady SM, Kliebenstein DJ. 2014. Promoter-Based Integration in Plant Defense Regulation. Plant Physiology **166**, 1803-1820.
Li F, Cui X, Feng Z, Du X, Zhu J. 2012. The effect of 2,4-D and kinetin on dedifferentiation of petiole cells in Arabidopsis thaliana. Biologia Plantarum **56**, 121-125.
Li N, Han X, Feng D, Yuan D, Huang L-J. 2019. Signaling Crosstalk between Salicylic Acid and Ethylene/Jasmonate in Plant Defense: Do We Understand What They Are Whispering? International Journal of Molecular Sciences **20**, 671.
Li Y, Sawada Y, Hirai A, Sato M, Kuwahara A, Yan X, Hirai MY. 2013. Novel insights into the function of Arabidopsis R2R3-MYB transcription factors regulating aliphatic glucosinolate biosynthesis. Plant & Cell Physiology **54**, 1335-1344.
Lian T-f, Xu Y-p, Li L-f, Su X-D. 2017. Crystal Structure of Tetrameric Arabidopsis MYC2 Reveals the Mechanism of Enhanced Interaction with DNA. Cell Reports **19**, 1334-1342.
Liao K, Peng Y-J, Yuan L-B, Dai Y-S, Chen Q-F, Yu L-J, Bai M-Y, Zhang W-Q, Xie L-J, Xiao S. 2020. Brassinosteroids Antagonize Jasmonate-Activated Plant Defense Responses through BRI1-EMS-SUPPRESSOR1 (BES1). Plant Physiology **182**, 1066-1082.
Liu C, Lu F, Cui X, Cao X. 2010. Histone Methylation in Higher Plants. Annual Review of Plant Biology **61**, 395-420.
Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang T-J, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park B-S, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee T-H, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH. 2014. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nature Communications **5**, 1-11.
Lohar DP, Schuller K, Buzas DM, Gresshoff PM, Stiller J. 2001. Transformation of Lotus japonicus using the herbicide resistance bar gene as a selectable marker. Journal of Experimental Botany **52**, 1697-1702.
Ludwig-Müller J, Pieper K, Ruppel M, Cohen JD, Epstein E, Kiddle G, Bennett R. 1999. Indole glucosinolate and auxin biosynthesis in Arabidopsis thaliana (L.) Heynh. glucosinolate mutants and the development of clubroot disease. Planta **208**, 409-419.
Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature **389**, 251-260.
Maeda T, Kami D, Kido S, Nakamura I, Otokita K, Sato T, Suzuki T, Oosawa K, Suzuki M. 2008. Development of Asexual Propagation System via *in Vitro* Culture in *Cardamine yezoensis Maxim.* and its Application to Hydroponic Cultivation. Journal of the Japanese Society for Horticultural Science **77**, 270-276.
Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A. 2008. The Transcript and Metabolite Networks Affected by the Two Clades of Arabidopsis Glucosinolate Biosynthesis Regulators. Plant Physiology **148**, 2021-2049.
Marazzi C, Städler E. 2004. Arabidopsis thaliana leaf-surface extracts are detected by the cabbage root fly (Delia radicum) and stimulate oviposition. Physiological Entomology **29**, 192-198.
Martinez-Medina A, Flors V, Heil M, Mauch-Mani B, Pieterse CMJ, Pozo MJ, Ton J, van Dam NM, Conrath U. 2016. Recognizing Plant Defense Priming. Trends in Plant Science **21**, 818-822.
Maruyama-Nakashita A. 2017. Metabolic changes sustain the plant life in low-sulfur environments. Current Opinion in Plant Biology **39**, 144-151.
Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nature Reviews Genetics **15**, 394-408.

Mauch-Mani B, Baccelli I, Luna E, Flors V. 2017. Defense Priming: An Adaptive Part of Induced Resistance. Annual Review of Plant Biology **68**, 485-512.
Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. 2003. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci **28**, 69-74.
Miao H-y, Wang M-y, Chang J-q, Tao H, Sun B, Wang Q-m. 2017. Effects of glucose and gibberellic acid on glucosinolate content and antioxidant properties of Chinese kale sprouts. Journal of Zhejiang University. Science. B **18**, 1093-1100.
Mikkelsen MD, Petersen BL, Glawischnig E, Jensen AB, Andreasson E, Halkier BA. 2003. Modulation of CYP79 Genes and Glucosinolate Profiles in Arabidopsis by Defense Signaling Pathways. Plant Physiology **131**, 298-308.
Millard PS, Weber K, Kragelund BB, Burow M. 2019. Specificity of MYB interactions relies on motifs in ordered and disordered contexts. Nucleic Acids Research **47**, 9592-9608.
Milutinovic M, Lindsey III BE, Wijeratne A, Hernandez JM, Grotewold N, Fernández V, Grotewold E, Brkljacic J. 2019. Arabidopsis EMSY-like (EML) histone readers are necessary for post-fertilization seed development, but prevent fertilization-independent seed formation. Plant Science **285**, 99-109.
Mitreiter S, Gigolashvili T. 2020. Regulation of glucosinolate biosynthesis. Journal of Experimental Botany **72**, 70-91.
Mitsui Y, Shimomura M, Komatsu K, Namiki N, Shibata-Hatta M, Imai M, Katayose Y, Mukai Y, Kanamori H, Kurita K, Kagami T, Wakatsuki A, Ohyanagi H, Ikawa H, Minaka N, Nakagawa K, Shiwa Y, Sasaki T. 2015. The radish genome and comprehensive gene expression profile of tuberous root formation and development. Scientific Reports **5**, 10835.
Morris EK, Fletcher R, Veresoglou SD. 2020. Effective methods of biofumigation: a meta-analysis. Plant and Soil **446**, 379-392.
Müller R, de Vos M, Sun JY, Sønderby IE, Halkier BA, Wittstock U, Jander G. 2010. Differential Effects of Indole and Aliphatic Glucosinolates on Lepidopteran Herbivores. Journal of Chemical Ecology **36**, 905-913.
Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T. 2007. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. Journal of bioscience and bioengineering **104**, 34-41.
Neal CS, Fredericks DP, Griffiths CA, Neale AD. 2010. The characterisation of AOP2: a gene associated with the biosynthesis of aliphatic alkenyl glucosinolates in Arabidopsis thaliana. BMC Plant Biology **10**, 170.
Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, Bailey CD, Tsiantis M. 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. New Phytologist **222**, 1638-1651.
Niu Y, Figueroa P, Browse J. 2011. Characterization of JAZ-interacting bHLH transcription factors that regulate jasmonate responses in Arabidopsis. Journal of Experimental Botany **62**, 2143-2154.
Pandey G, Sharma N, Sahu PP, Prasad M. 2016. Chromatin-Based Epigenetic Regulation of Plant Abiotic Stress Response. Current genomics **17**, 490-498.
Pattanaik S, Xie CH, Yuan L. 2008. The interaction domains of the plant Myc-like bHLH transcription factors can regulate the transactivation strength. Planta **227**, 707-715.
Pfalz M, Mikkelsen MD, Bednarek P, Olsen CE, Halkier BA, Kroymann J. 2011. Metabolic Engineering in *Nicotiana benthamiana* Reveals Key Enzyme Functions in *Arabidopsis* Indole Glucosinolate Modification. The Plant Cell **23**, 716-729.
Pfalz M, Mukhaimar M, Perreau F, Kirk J, Hansen CIC, Olsen CE, Agerbirk N, Kroymann J. 2016. Methyl Transfer in Glucosinolate Biosynthesis Mediated by Indole Glucosinolate *O*-Methyltransferase 5. Plant Physiology **172**, 2190-2203.
Pfalz M, Vogel H, Kroymann J. 2009. The Gene Controlling the *Indole Glucosinolate Modifier1* Quantitative Trait Locus Alters Indole Glucosinolate Structures and Aphid Resistance in *Arabidopsis*. The Plant Cell **21**, 985-999.

Pieterse CMJ, Van der Does D, Zamioudis C, Leon-Reyes A, Van Wees SCM. 2012. Hormonal Modulation of Plant Immunity. Annual Review of Cell and Developmental Biology **28**, 489-521.
Pires N, Dolan L. 2010. Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants. Molecular Biology and Evolution **27**, 862-874.
Pireyre M, Burow M. 2015. Regulation of MYB and bHLH Transcription Factors: A Glance at the Protein Level. Molecular Plant **8**, 378-388.
Qi T, Huang H, Song S, Xie D. 2015. Regulation of Jasmonate-Mediated Stamen Development and Seed Production by a bHLH-MYB Complex in Arabidopsis. The Plant Cell **27**, 1620-1633.
Qi T, Song S, Ren Q, Wu D, Huang H, Chen Y, Fan M, Peng W, Ren C, Xie D. 2011. The Jasmonate-ZIM-Domain Proteins Interact with the WD-Repeat/bHLH/MYB Complexes to Regulate Jasmonate-Mediated Anthocyanin Accumulation and Trichome Initiation in Arabidopsis thaliana. The Plant Cell **23**, 1795-1814.
R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
Rajniak J, Barco B, Clay NK, Sattely ES. 2015. A new cyanogenic metabolite in Arabidopsis required for inducible pathogen defence. Nature **525**, 376-379.
Ramirez-Prado JS, Abulfaraj AA, Rayapuram N, Benhamed M, Hirt H. 2018. Plant Immunity: From Signaling to Epigenetic Control of Defense. Trends in Plant Science **23**, 833-844.
Ran X, Zhao F, Wang Y, Liu J, Zhuang Y, Ye L, Qi M, Cheng J, Zhang Y. 2020. Plant Regulomics: a data-driven interface for retrieving upstream regulators from plant multi-omics data. The Plant Journal **101**, 237-248.
Rasmann S, Vos MD, Casteel CL, Tian D, Halitschke R, Sun JY, Agrawal AA, Felton GW, Jander G. 2012. Herbivory in the Previous Generation Primes Plants for Enhanced Insect Resistance. Plant Physiology **158**, 854-863.
Reichelt M, Brown PD, Schneider B, Oldham NJ, Stauber E, Tokuhisa J, Kliebenstein DJ, Mitchell-Olds T, Gershenzon J. 2002. Benzoic acid glucosinolate esters and other glucosinolates from Arabidopsis thaliana. Phytochemistry **59**, 663-671.
Richards CL, Alonso C, Becker C, Bossdorf O, Bucher E, Colomé-Tatché M, Durka W, Engelhardt J, Gaspar B, Gogol-Döring A, Grosse I, van Gurp TP, Heer K, Kronholm I, Lampei C, Latzel V, Mirouze M, Opgenoorth L, Paun O, Prohaska SJ, Rensing SA, Stadler PF, Trucchi E, Ullrich K, Verhoeven KJF. 2017. Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. Ecology Letters **20**, 1576-1590.
Roberts MR, López Sánchez A. 2019. Plant Epigenetic Mechanisms in Response to Biotic Stress. In: Alvarez-Venegas R, De-la-Peña C, Casas-Mollano JA, eds. *Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications: Transcriptional Regulation and Chromatin Remodelling in Plants*. Cham: Springer International Publishing, 65-113.
Romero I, Fuertes A, Benito MJ, Malpica JM, Leyva A, Paz-Ares J. 1998. More than 80R2R3-MYB regulatory genes in the genome of Arabidopsis thaliana. Plant J **14**, 273-284.
Rosa EAS, Heaney RK, Rego FC, Fenwick GR. 1994. The variation of glucosinolate concentration during a single day in young plants of Brassica oleracea var Acephala and Capitata. Journal of the Science of Food and Agriculture **66**, 457-463.
Roux CL, Prete SD, Boutet-Mercey S, Perreau F, Balagué C, Roby D, Fagard M, Gaudin V. 2014. The hnRNP-Q Protein LIF2 Participates in the Plant Immune Response. PLOS ONE **9**, e99343.
RStudio Team. 2020. RStudio: Integrated Development Environment for R. Boston, MA, USA: RStudio, PBC.
Salehin M, Li B, Tang M, Katz E, Song L, Ecker JR, Kliebenstein DJ, Estelle M. 2019. Auxin-sensitive Aux/IAA proteins mediate drought tolerance in Arabidopsis by regulating glucosinolate levels. Nature Communications **10**, 1-9.
Schillheim B, Jansen I, Baum S, Beesley A, Bolm C, Conrath U. 2018. Sulforaphane Modifies Histone H3, Unpacks Chromatin, and Primes Defense. Plant Physiology **176**, 2395-2405.
Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative CT method. Nature Protocols **3**, 1101-1108.

Schranz E, Edger P, Pires J, Dam N, Wheat C. 2011. Comparative Genomics in the Brassicales: Ancient Genome Duplications, Glucosinolate Diversification and Pierinae Herbivore Radiation. *Genetics, Genomics and Breeding of Oilseed Brassicas*, 206-218.
Schweizer F, Fernández-Calvo P, Zander M, Diez-Diaz M, Fonseca S, Glauser G, Lewsey MG, Ecker JR, 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**, 3117-3132.
Seo M-S, Jin M, Chun J-H, Kim S-J, Park B-S, Shon S-H, Kim JS. 2016. Functional analysis of three BrMYB28 transcription factors controlling the biosynthesis of glucosinolates in Brassica rapa. Plant Molecular Biology **90**, 503-516.
Seo M-S, Jin M, Sohn S-H, Kim JS. 2017. Expression profiles of BrMYB transcription factors related to glucosinolate biosynthesis and stress response in eight subspecies of Brassica rapa. FEBS Open Bio **7**, 1646-1659.
Seo M-S, Kim JS. 2017. Understanding of MYB Transcription Factors Involved in Glucosinolate Biosynthesis in Brassicaceae. Molecules **22**, 1549.
Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C. 2014. The Functional Topography of the *Arabidopsis* Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. The Plant Cell **26**, 2351-2366.
Singh P, Yekondi S, Chen PW, Tsai CH, Yu CW, Wu K, Zimmerli L. 2014. Environmental History Modulates Arabidopsis Pattern-Triggered Immunity in a HISTONE ACETYLTRANSFERASE1-Dependent Manner. Plant Cell **26**, 2676-2688.
Skirycz A, Reichelt M, Burow M, Birkemeyer C, Rolcik J, Kopka J, Zanor MI, Gershenzon J, Strnad M, Szopa J, Mueller-Roeber B, Witt I. 2006. DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in Arabidopsis. The Plant Journal: For Cell and Molecular Biology **47**, 10-24.
Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Newman LK, Wang W, Mandáková T, Vello E, Smith LM, Henz SR, Steffen J, Takuno S, Brandvain Y, Coop G, Andolfatto P, Hu TT, Blanchette M, Clark RM, Quesneville H, Nordborg M, Gaut BS, Lysak MA, Jenkins J, Grimwood J, Chapman J, Prochnik S, Shu S, Rokhsar D, Schmutz J, Weigel D, Wright SI. 2013. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nature Genetics **45**, 831-835.
Sønderby IE, Burow M, Rowe HC, Kliebenstein DJ, Halkier BA. 2010a. A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis. Plant Physiology **153**, 348-363.
Sønderby IE, Geu-Flores F, Halkier BA. 2010b. Biosynthesis of glucosinolates--gene discovery and beyond. Trends in Plant Science **15**, 283-290.
Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ. 2007. A Systems Biology Approach Identifies a R2R3 MYB Gene Subfamily with Distinct and Overlapping Functions in Regulation of Aliphatic Glucosinolates. PLOS ONE **2**, e1322.
Song S, Huang H, Wang J, Liu B, Qi T, Xie D. 2017. MYC5 is Involved in Jasmonate-Regulated Plant Growth, Leaf Senescence and Defense Responses. Plant & Cell Physiology **58**, 1752-1763.
Stracke R, Werber M, Weisshaar B. 2001. The R2R3-MYB gene family in Arabidopsis thaliana. Current Opinion in Plant Biology **4**, 447-456.
Sugiyama R, Hirai MY. 2019. Atypical Myrosinase as a Mediator of Glucosinolate Functions in Plants. Frontiers in Plant Science **10**, 1008.
Sun J, Zhang M, Chen P. 2016. GLS-Finder: A Platform for Fast Profiling of Glucosinolates in Brassica Vegetables. J Agric Food Chem **64**, 4407-4415.
Sun JY, Sønderby IE, Halkier BA, Jander G, Vos Md. 2009. Non-Volatile Intact Indole Glucosinolates are Host Recognition Cues for Ovipositing Plutella xylostella. Journal of Chemical Ecology **35**, 1427-1436.
Syed NH, Kalyna M, Marquez Y, Barta A, Brown JWS. 2012. Alternative splicing in plants – coming of age. Trends in Plant Science **17**, 616-623.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol **10**, 512-526.
Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA, Browse J. 2007. JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. Nature **448**, 661-665.
Thiruvengadam M, Baskar V, Kim S-H, Chung I-M. 2016. Effects of abscisic acid, jasmonic acid and salicylic acid on the content of phytochemicals and their gene expression profiles and biological activity in turnip (Brassica rapa ssp. rapa). Plant Growth Regulation **80**, 377-390.
Tsuchiya T, Eulgem T. 2011. EMSY-Like Genes Are Required for Full RPP7-Mediated Race-Specific Immunity and Basal Defense in Arabidopsis. Molecular Plant-Microbe Interactions **24**, 1573-1581.
Turck F, Roudier F, Farrona S, Martin-Magniette ML, Guillaume E, Buisine N, Gagnot S, Martienssen RA, Coupland G, Colot V. 2007. Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genet **3**, e86.
van Dam NM, Tytgat TOG, Kirkegaard JA. 2009. Root and shoot glucosinolates: a comparison of their diversity, function and interactions in natural and managed ecosystems. Phytochemistry Reviews **8**, 171-186.
van den Bergh E, Hofberger JA, Schranz ME. 2016. Flower power and the mustard bomb: Comparative analysis of gene and genome duplications in glucosinolate biosynthetic pathway evolution in Cleomaceae and Brassicaceae. American Journal of Botany **103**, 1212-1222.
van der Fits L, Deakin EA, Hoge JH, Memelink J. 2000. The ternary transformation system: constitutive virG on a compatible plasmid dramatically increases Agrobacterium-mediated plant transformation. Plant Mol Biol **43**, 495-502.
Variyar PS, Banerjee A, Akkarakaran JJ, Suprasanna P. 2014. Chapter 12 - Role of Glucosinolates in Plant Stress Tolerance. In: Ahmad P, Rasool S, eds. *Emerging Technologies and Management of Crop Stress Tolerance*. San Diego: Academic Press, 271-291.
Verma V, Ravindran P, Kumar PP. 2016. Plant hormone-mediated regulation of stress responses. BMC Plant Biology **16**, 86.
Voinnet O, Pinto YM, Baulcombe DC. 1999. Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci U S A **96**, 14147-14152.
Voinnet O, Rivas S, Mestre P, Baulcombe D. 2003. Retracted: An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. The Plant Journal **33**, 949-956.
Wang Y, Zhang X, Zhang H, Lu Y, Huang H, Dong X, Chen J, Dong J, Yang X, Hang H, Jiang T. 2012. Coiled-coil networking shapes cell molecular machinery. Molecular Biology of the Cell **23**, 3911-3922.
Wentzell AM, Kliebenstein DJ. 2008. Genotype, Age, Tissue, and Environment Regulate the Structural Outcome of Glucosinolate Activation. Plant Physiology **147**, 415-428.
Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007. The genetic basis of a plant–insect coevolutionary key innovation. Proceedings of the National Academy of Sciences **104**, 20427-20431.
Wieczorek MN, Walczak M, Skrzypczak-Zielińska M, Jeleń HH. 2018. Bitter taste of Brassica vegetables: The role of genetic factors, receptors, isothiocyanates, glucosinolates, and flavor context. Critical Reviews in Food Science and Nutrition **58**, 3130-3140.
Wiesner M, Hanschen FS, Schreiner M, Glatt H, Zrenner R. 2013. Induced production of 1-methoxy-indol-3-ylmethyl glucosinolate by jasmonic acid and methyl jasmonate in sprouts and leaves of pak choi (Brassica rapa ssp. chinensis). International Journal of Molecular Sciences **14**, 14996-15016.
Willing E-M, Rawat V, Mandáková T, Maumus F, James GV, Nordström KJV, Becker C, Warthmann N, Chica C, Szarzynska B, Zytnicki M, Albani MC, Kiefer C, Bergonzi S, Castaings L, Mateos JL, Berns MC, Bujdoso N, Piofczyk T, de Lorenzo L, Barrero-Sicilia C, Mateos I, Piednoël M, Hagmann J, Chen-Min-Tao R, Iglesias-Fernández R, Schuster SC, Alonso-Blanco C, Roudier F, Carbonero P, Paz-Ares J, Davis SJ, Pecinka A, Quesneville H, Colot V, Lysak MA, Weigel D, Coupland G, Schneeberger K. 2015. Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. Nature Plants **1**, 1-7.

Windsor AJ, Reichelt M, Figuth A, Svatoš A, Kroymann J, Kliebenstein DJ, Gershenzon J, Mitchell-Olds T. 2005. Geographic and evolutionary diversification of glucosinolates among near relatives of Arabidopsis thaliana (Brassicaceae). Phytochemistry **66**, 1321-1333.
Wittstock U, Agerbirk N, Stauber EJ, Olsen CE, Hippler M, Mitchell-Olds T, Gershenzon J, Vogel H. 2004. Successful herbivore attack due to metabolic diversion of a plant chemical defense. Proceedings of the National Academy of Sciences of the United States of America **101**, 4859-4864.
Wittstock U, Burow M. 2010. Glucosinolate Breakdown in Arabidopsis: Mechanism, Regulation and Biological Significance. The Arabidopsis Book / American Society of Plant Biologists **8**, e0134.
Wittstock U, Halkier BA. 2000. Cytochrome P450 CYP79A2 from Arabidopsis thaliana L. Catalyzes the Conversion of l-Phenylalanine to Phenylacetaldoxime in the Biosynthesis of Benzylglucosinolate. Journal of Biological Chemistry **275**, 14659-14666.
Wittstock U, Kurzbach E, Herfurth A-M, Stauber EJ. 2016a. Chapter Six - Glucosinolate Breakdown. In: Kopriva S, ed. *Advances in Botanical Research*, Vol. 80: Academic Press, 125-169.
Wittstock U, Meier K, Dörr F, Ravindran BM. 2016b. NSP-Dependent Simple Nitrile Formation Dominates upon Breakdown of Major Aliphatic Glucosinolates in Roots, Seeds, and Seedlings of Arabidopsis thaliana Columbia-0. Frontiers in Plant Science **7**, 1821.
Xiao J, Wagner D. 2015. Polycomb repression in the regulation of growth and development in Arabidopsis. Current Opinion in Plant Biology **23**, 15-24.
Xue M, Long J, Jiang Q, Wang M, Chen S, Pang Q, He Y. 2015. Distinct patterns of the histone marks associated with recruitment of the methionine chain-elongation pathway from leucine biosynthesis. Journal of Experimental Botany **66**, 805-812.
Zang Y-x, Ge J-l, Huang L-h, Gao F, Lv X-s, Zheng W-w, Hong S-b, Zhu Z-j. 2015. Leaf and root glucosinolate profiles of Chinese cabbage (Brassica rapa ssp. pekinensis) as a systemic response to methyl jasmonate and salicylic acid elicitation. Journal of Zhejiang University. Science. B **16**, 696-708.
Zhang F, Yao J, Ke J, Zhang L, Lam VQ, Xin X-F, Zhou XE, Chen J, Brunzelle J, Griffin PR, Zhou M, Xu HE, Melcher K, He SY. 2015. Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling. Nature **525**, 269-273.
Zhang L, Zhang F, Melotto M, Yao J, He SY. 2017. Jasmonate signaling and manipulation by pathogens and insects. Journal of Experimental Botany **68**, 1371-1385.
Zhao S, Zhang B, Yang M, Zhu J, Li H. 2018. Systematic Profiling of Histone Readers in Arabidopsis thaliana. Cell Reports **22**, 1090-1102.
Zhou Y, Romero-Campero FJ, Gómez-Zambrano Á, Turck F, Calonje M. 2017. H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. Genome Biology **18**, 69.
Zhu Y, Dong A, Shen W-H. 2012. Histone variants and chromatin assembly in plant abiotic stress responses. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1819**, 343-348.
Zhu Z, An F, Feng Y, Li P, Xue L, A M, Jiang Z, Kim J-M, To TK, Li W, Zhang X, Yu Q, Dong Z, Chen W-Q, Seki M, Zhou J-M, Guo H. 2011. Derepression of ethylene-stabilized transcription factors (EIN3/EIL1) mediates jasmonate and ethylene signaling synergy in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **108**, 12539-12544.
Zhurov V, Navarro M, Bruinsma KA, Arbona V, Santamaria ME, Cazaux M, Wybouw N, Osborne EJ, Ens C, Rioja C, Vermeirssen V, Rubio-Somoza I, Krishna P, Diaz I, Schmid M, Gómez-Cadenas A, Peer YVd, Grbić M, Clark RM, Leeuwen TV, Grbić V. 2014. Reciprocal Responses in the Interaction between Arabidopsis and the Cell-Content-Feeding Chelicerate Herbivore Spider Mite. Plant Physiology **164**, 384-399.
Züst T, Heichinger C, Grossniklaus U, Harrington R, Kliebenstein DJ, Turnbull LA. 2012. Natural enemies drive geographic variation in plant defenses. Science (New York, N.Y.) **338**, 116-119.

6 Supplementary Data

6.1 Supplementary Tables

Supplementary Table 1: Gateway entry clones used and/or generated in this study.

Plasmid	Source
pDONR201-AabHLH05	this study
pDONR201-AabHLH06	this study
pDONR201-ChbHLH05	this study
pDONR201-ChbHLH06	this study
pDONR207-CrbHLH04 (full length)	this study
pDONR207-CrbHLH04 (short A)	this study
pDONR207-CrbHLH04 (short B)	this study
pDONR207-CrbHLH05 (short A)	this study
pDONR207-CrbHLH05 (short B)	this study
pDONR207-CrbHLH05 (short C)	this study
pDONR201-CrbHLH05 (short D)	this study
pDONR201-CrbHLH05 (short E)	this study
pDONR201-CrbHLH06 (full length)	this study
pDONR201-CrbHLH06 (short A)	this study
pDONR201-CrbHLH06 (short B)	this study
pENTR-/D-TOPO-EML1.1	Brkljacic Lab (Ohio State University, USA)
pDONR207-EML1.3	this study
pDONR207-EML1.4	this study
pDONR207-EML1.5	this study
pDONR207-EML2	this study
pDONR207-AaMYB28	this study
pDONR207-AaMYB29	this study
pDONR201-AaMYB34	this study
pDONR201-AaMYB51	this study
pDONR207-ChMYB29	this study
pDONR201-ChMYB34	this study
pDONR201-ChMYB51	this study
pDONR201-CrMYB51	this study

Supplementary Table 2: Gateway expression clones used and/or generated in this study.

Plasmid	Source
pAUBERGINE-AtbHLH05	Gigolashvili Lab
pGWB2-AtbHLH05	Gigolashvili Lab
pGWB5-AtbHLH05	Gigolashvili Lab
pGWB6-AtbHLH05	this study
pGWB2-CrbHLH04 (full length)	this study
pGWB2-CrbHLH04 (short A)	this study
pGWB2-CrbHLH05 (short A)	this study
pGWB2-CrbHLH05 (short B)	this study
pGWB2-CrbHLH06 (full length)	this study
pGWB2-CrbHLH06 (short A)	this study
pGWB3-proCYP79B3	Gigolashvili Lab
pGWB2-EML1.1	this study
pGWB2-EML1.3	this study
pGWB5-EML1.3	this study
pGWB6-EML1.3	this study
pGWB6-EML1.5	this study
pGWB2-EML2	this study
pGWB5-EML2	this study
pGWB2-AtMYB28	Gigolashvili Lab
pGWB2-AtMYB29	Gigolashvili Lab
pGWB2-AtMYB76	Gigolashvili Lab
pGWB2-AtMYB34	Gigolashvili Lab
pGWB2-AtMYB51	Gigolashvili Lab
pGWB5-AtMYB51	Gigolashvili Lab
pGWB6-AtMYB51	this study
pGWB2-AaMYB28	this study
pGWB2-AaMYB29	this study
pGWB5-AaMYB34[1]	this study
pGWB5-AaMYB51[1]	this study
pGWB2-ChMYB29	this study
pGWB5-ChMYB34[1]	this study
pGWB5-ChMYB51[1]	this study
pGWB5-CrMYB51[1]	this study

[1]Contains a stop codon, thus functionally identical to a pGWB2 construct

Supplementary Table 3: Primers used for genotyping of plant lines.
All oligonucleotides were purchased from Sigma-Aldrich (Steinheim, Germany).

Name	ID	Target	Sequence	Reference
NGA139_fw	170	NGA139 (SSLP)	GGTTTCGTTTC ACTATCCAGG	TAIR
NGA139_rv	171	NGA139 (SSLP)	AGAGCTACCA GATCCGATGG	TAIR
CIW5_fw	172	CIW5 (SSLP)	GGTTAAAAATT AGGGTTACGA	TAIR
CIW5_rv	173	CIW5 (SSLP)	AGATTTACGT GGAAGCAAT	TAIR
SAIL LB2	-	pCSA110 (T-DNA)	GCTTCCTATTATATCTT CCCAAATTACCAATACA	SIGnAL, Salk Institute (San Diego, CA, USA)
SALK LBb1.3	117	pROK2 (T-DNA)	ATTTTGCCGA TTTCGGAAC	SIGnAL, Salk Institute (San Diego, CA, USA)
Ds3-1 RB	175	Ds (transp.)	ACCCGACCGG ATCGTATCGGT	SIGnAL, Salk Institute (San Diego, CA, USA)
EML1 CS69823_LP[1]	174	*AT3G12140*	AATCTGCGAAT TGAGCTTGAG	Milutinovic *et al.* (2019)
EML1 CS69823_RP[1]	176	*AT3G12140*	TTTGTCCACAC TTTTCTTCCG	Milutinovic *et al.* (2019)
EML2 CS69824_LP[2]	178	*AT5G06780*	GTTTCTCTCTC CTCCAACTTG	Milutinovic *et al.* (2019)
EML2 CS69824_RP[2]	179	*AT5G06780*	AAATTCATCAG CCTGTGCTTG	Milutinovic *et al.* (2019)
EML3 CS69825_LP[2]	181	*AT5G13020*	ATGGATTACC GACCTTCTGAT	Milutinovic *et al.* (2019)
EML3 CS69825_RP[2]	182	*AT5G13020*	TCATTTCATCA GCGTTAACCC	Milutinovic *et al.* (2019)
EML4 SAIL_202_D12_LP[3]	183	*AT2G44440*	ATTTTGGCTTG CTGTCATCAG	SIGnAL, Salk Institute (San Diego, CA, USA)
EML4 SAIL_202_D12_RP[3]	184	*AT2G44440*	GTGGATAGGA GAGGATCCTGG	SIGnAL, Salk Institute (San Diego, CA, USA)
EML1 SALK_114038_LP[2]	185	*AT3G12140*	TTTTAAGGGC TTTTAAAGCGC	Tsuchiya and Eulgem (2011)
EML1 SALK_114038_RP[2]	186	*AT3G12140*	ATCCCATCTTA TGTCTTCCGG	Tsuchiya and Eulgem (2011)
EML1 SALK_077088_LP[2]	187	*AT3G12140*	TAGTCCCACTT TCTCAGCTGC	Tsuchiya and Eulgem (2011)
EML1 SALK_077088_RP[2]	188	*AT3G12140*	ATGAGTGCCT GTTCATGTTCC	Tsuchiya and Eulgem (2011)
BAR fw	189	*bar* (BASTA resist.)	TGCACCATCG TCAACCACTA	Lohar *et al.* (2001)
BAR rv	190	*bar* (BASTA resist.)	ACAGCGACCA CGCTCTTGAA	Lohar *et al.* (2001)

[1]used with Ds3-1 RB
[2]used with SALK LBb1.3
[3]used with SAIL LB2

Supplementary Table 4: Primers used for sequencing and sequence verification of constructs.
All oligonucleotides were purchased from Sigma-Aldrich (Steinheim, Germany).

Name	ID	Sequence	Reference
Actin_2_long_fw	-	TAACTCTCCCGCTATGTATGTCGC	Gigolashvili Lab
Actin_2_long_rv	-	CCACTGAGCACAATGTTACCGTAC	Gigolashvili Lab
35S fw	-	GCAAGACCCTTCCTCTATATAAG	Gigolashvili Lab
GFP rv	-	GAACTCCAGCAGGACCATGTG	Gigolashvili Lab
Nos rv	-	CGGCAACAGGATTCAATCTTAAG	Gigolashvili Lab
pDONR fw	-	TCGCGTTAACGCTAGCATGGATCTC	Invitrogen
pDONR rv	-	GTAACATCAGAGATTTTGAGACAC	Invitrogen
Aa_bHLH05_middle_fw	95	ACTATGGTTTGTATCGCTACGGAG	this study
Aa_bHLH05_middle_rv	96	CATGATTCAACGGCTCTTCTCTTC	this study
Ch_bHLH05_middle_fw	97	GAGCTAAATTCGTTAATCTCCGGC	this study
Ch_bHLH05_middle_rv	98	CTCTTCTTATTACTCTCCTCCGCC	this study
Ch_bHLH06_middle_fw	99	GAAGGCTGGACTTACGCTATTTTC	this study
Ch_bHLH06_middle_rv	100	CGGATTTTGGGTCTGAGAATGAAC	this study
Cr_bHLH06_middle_fw	101	GGTTGATGAGGAGGTTACGGATAC	this study
Cr_bHLH06_middle_rv	102	TTCTCTACCGCTACTTCTTTCACG	this study
EML1_fw	147	GAGACACAAATTCATCAACTTGAGC	this study
EML1_rv	132	CCTCCCTGTCTCCAATCTCTTATC	this study
EML2_fw	150	ACGTTCCAATCCTATCCGTC	this study
EML2_rv	151	CATCTTCCTCTCCATCCCATC	this study

Supplementary Table 5: Primers used for cloning.
All oligonucleotides were purchased from Sigma-Aldrich (Steinheim, Germany).

Name	ID	Sequence	Reference
Aa_bHLH05_fw	63	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGAACGACTACTTCATCAA	this study
Aa_bHLH05_rv	64	GGGGACCACTTTGTACAAGAAAGCT GGGTCTTATCCGACTTTCGCCATCA	this study
Aa_bHLH06_fw	65	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGAATCTCTGGACCACCGA	this study
Aa_bHLH06_rv	66	GGGGACCACTTTGTACAAGAAAGCT GGGTCTTAACCGATCTTCGAAATCA	this study
Aa_MYB28_fw	191	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGTCAAGAAAGCCATGTTG	this study
Aa_MYB28_rv	192	GGGGACCACTTTGTACAAGAAAGCT GGGTCCATGGAATGCTTTTCGAGGG	this study
Aa_MYB29_fw	193	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGTCAAGAAAGCCATGTTG	this study
Aa_MYB29_rv	194	GGGGACCACTTTGTACAAGAAAGCT GGGTCTATGAAGAAGTCCTTGGCGT	this study
Aa_MYB34_fw	59	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGGTGAGAACACCATGTTG	this study
Aa_MYB34_rv	60	GGGGACCACTTTGTACAAGAAAGCT GGGTCTTAGAAAAAGACACCTACCT	this study
Aa_MYB51_fw	61	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGGTACGAACACCATGTTG	this study
Aa_MYB51_rv	62	GGGGACCACTTTGTACAAGAAAGCT GGGTCTCATCCAAAATAGTTATCAA	this study
Ch_bHLH05_fw	55	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGAACGGCACAACATCATC	this study
Ch_bHLH05_rv	56	GGGGACCACTTTGTACAAGAAAGCT GGGTCTCAATAGTTTTCTCCGACTT	this study
Ch_bHLH06_fw	57	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGAATCTCTGGACCACCGA	this study
Ch_bHLH06_rv	58	GGGGACCACTTTGTACAAGAAAGCT GGGTCTTAACCGATCCTTGAAATCA	this study
Ch_MYB28_fw	195	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGTCAAGAAAGCCATGTTG	this study
Ch_MYB28_rv	196	GGGGACCACTTTGTACAAGAAAGCT GGGTCTATAAAATGCTTTTCGAGGG	this study
Ch_MYB29_fw	197	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGTCAAGAAAACCATGTTG	this study
Ch_MYB29_rv	198	GGGGACCACTTTGTACAAGAAAGCT GGGTCACTAAAATCGGAATGGTCAA	this study
Ch_MYB34_fw	51	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGGTGAGGACACCATGTTG	this study
Ch_MYB34_rv	52	GGGGACCACTTTGTACAAGAAAGCT GGGTCTCAGACAAAGACTCCAACCA	this study
Ch_MYB51_fw	53	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGGTGCGGACACCGTGCTG	this study
Ch_MYB51_rv	54	GGGGACCACTTTGTACAAGAAAGCT GGGTCTCATCCAAAATAGTTATCAA	this study

Name	ID	Sequence	Reference
Cr_bHLH06_fw	49	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGACTGATTACCGGCTACA	this study
Cr_bHLH06_rv	50	GGGGACCACTTTGTACAAGAAAGCT GGGTCTTAACTGATTTTTGAAATCA	this study
Cr_MYB28_fw	199	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGTCAAGAAAGCCATGTTG	this study
Cr_MYB28_rv	200	GGGGACCACTTTGTACAAGAAAGCT GGGTCTATGAAATGCTTTTCGAGGG	this study
Cr_MYB29_fw	201	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGTCGAGAAAGCCATGTTG	this study
Cr_MYB29_rv	202	GGGGACCACTTTGTACAAGAAAGCT GGGTCTACGGAGTTCTTGTCGTCGC	this study
Cr_MYB34_fw	43	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGGTGAGGACACCATGTTG	this study
Cr_MYB34_rv	44	GGGGACCACTTTGTACAAGAAAGCT GGGTCTTATTCGTTCCAACAACCAA	this study
Cr_MYB51_fw	45	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGGTGCGAACGCCGTGTTG	this study
Cr_MYB51_rv	46	GGGGACCACTTTGTACAAGAAAGCT GGGTCTCAGGAAAAATATTTATCAA	this study
CrbHLH04_attB_fw	103	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGTCTCCGACGAGTGTTCA	this study
CrbHLH04_attB_rv	104	GGGGACCACTTTGTACAAGAAAGCT GGGTCTAGGCATTCTCCGACTTTCT	this study
CrbHLH05_attB_fw	115	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGAACGGCACAGCATCATC	this study
CrbHLH05_attB_rv	116	GGGGACCACTTTGTACAAGAAAGCT GGGTCATAGGTTTCTCCGACTTTCG	this study
EML1_attB_fw	139	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGGAGACACAAATTCATCA	this study
EML1_attB_rv	140	GGGGACCACTTTGTACAAGAAAGCT GGGTCAGAAGAAGACCTATATTTCA	this study
EML2_attB_fw	143	GGGGACAAGTTTGTACAAAAAAGCA GGCTTCATGGTAGGTCTACACATTAA	this study
EML2_attB_rv	144	GGGGACCACTTTGTACAAGAAAGCT GGGTCCCCCAGCAGCATTGGAAGCT	this study

Supplementary Table 6: Primers used for qPCR.
All oligonucleotides were purchased from Sigma-Aldrich (Steinheim, Germany).

Name	ID	Target	Sequence	Reference
CYP79B3_RT_fw	124	*AT2G22330*	CTCCTTCTTCC TTGCAAATGGA	Frerigmann *et al.* (2014)
CYP79B3_RT_rv	125	*AT2G22330*	GAGAATCATCAA GAAGCAAAGGG	Frerigmann *et al.* (2014)
CYP79F2_RT_fw	126	*AT1G16400*	CATGCTTTCAAATCT TACTAGGATTTATCG	Frerigmann *et al.* (2014)
CYP79F2_RT_rv	127	*AT1G16400*	GTAGATTGCC GAGGATGGGC	Frerigmann *et al.* (2014)
CYP83B1_RT_fw	137	*AT4G31500*	GGCAACAAACCA TGTCGTATCAAG	Frerigmann *et al.* (2014)
CYP83B1_RT_rv	138	*AT4G31500*	CGTTGACACTCTT CTTCTCTAACCG	Frerigmann *et al.* (2014)
PP2A_RT_fw	128	*AT1G69960*	CAAGAGGTTCC ACACGAAGGA	Frerigmann *et al.* (2014)
PP2A_RT_rv	129	*AT1G69960*	TGTAACCAGC ACCACGAGGA	Frerigmann *et al.* (2014)

Supplementary Table 7: Lists of aliphatic and indolic GSL synthesis genes used for binding factor enrichment analysis.

Aliphatic GSL synthesis		**Indolic GSL synthesis**	
Gene Locus	**Name**	**Gene Locus**	**Name**
AT3G19710	*BCAT4*	*AT5G05730*	*ASA1*
AT5G23010	*MAM1*	*AT5G54810*	*TSB1*
AT5G23020	*MAM3*	*AT4G39950*	*CYP79B2*
AT1G16410	*CYP79F1*	*AT2G22330*	*CYP79B3*
AT1G16400	*CYP79F2*	*AT4G31500*	*CYP83B1*
AT4G13770	*CYP83A1*	*AT2G30860*	*GSTF9*
AT3G03190	*GSTF11*	*AT2G30870*	*GSTF10*
AT1G78370	*GSTU20*	*AT4G30530*	*GGP1*
AT4G30530	*GGP1*	*AT2G20610*	*SUR1*
AT2G20610	*SUR1*	*AT1G24100*	*UGT74B1*
AT1G24100	*UGT74B1*	*AT1G74100*	*SOT16*
AT1G18590	*SOT17*	*AT4G37430*	*CYP81F1*
AT1G74090	*SOT18*	*AT5G57220*	*CYP81F2*
AT1G65860	*FMO-GSOX1*	*AT4G37400*	*CYP81F3*
AT1G62540	*FMO-GSOX2*	*AT4G37410*	*CYP81F4*
AT1G62560	*FMO-GSOX3*	*AT1G21100*	*IGMT1*
AT1G62570	*FMO-GSOX4*	*AT1G21120*	*IGMT2*
AT1G12140	*FMO-GSOX5*		
AT1G12130	*FMO-GSOX6*		
AT1G12160	*FMO-GSOX7*		
AT4G03060	*AOP2*		
AT4G03050	*AOP3*		
AT2G25450	*GS-OH*		
AT1G65880	*BZO1*		

Supplementary Table 8: Epigenetic binding factor enrichment of aliphatic GSL biosynthesis genes.
Data from Plant Regulomics was filtered for enrichment in H2AK121, H3K27 and H3K4 modifications. All such entries that were generated for wildtype plants under regular or under stress conditions and overlapped with at least 50% of the genes of interest were considered.

Marker	GEO Dataset	Description	Overlap %	p-value (FDR)	Fold change
H2AK121ub	GSM2367150	WT	94	1.97e-4	2,50
H2AK121ub	GSM2367148	WT	94	2.22e-4	2,41
H2AK121ub	GSM2367146	WT	94	2.22e-4	2,35
H2AK121ub	GSM2367152	WT	94	2.43e-4	2,29
H2AK121ub	GSM2367140	WT	94	2.51e-4	2,25
H2AK121ub	GSM2367138	WT	88	7.48e-4	2,13
H3K27me3	GSM1530047	WT_Col/C24	59	3.84e-4	4,59
H3K27me3	GSM1625911	F1-Endosperm Columbia x Landsberg	53	7.58e-4	4,46
H3K27me3	GSM1625913	F1-Endosperm Landsberg x Columbia	53	1.41e-3	3,87
H3K27me3	GSM624616	Col whole seedling	71	3.84e-4	3,46
H3K27me3	GSM1625912	F1-Leaf Columbia x Columbia	71	3.84e-4	3,46
H3K27me3	GSM1225092	Col-0; young leaves	77	2.22e-4	3,43
H3K27me3	GSM1644813	Col; 2 week-old seedlings without roots	65	7.48e-4	3,32
H3K27me3	GSM1625914	F1-Endosperm Columbia x Landsberg	53	3.48e-3	3,29
H3K27me3	GSM1146786	WT; 14 day-old seedlings	71	4.37e-4	3,23
H3K27me3	GSM2027816	Col-0 leaves	65	9.02e-4	3,19
H3K27me3	GSM1146785	WT; 14 day-old seedlings	71	5.05e-4	3,16
H3K27me3	GSM1955349	2-week old wt seedlings without roots	65	9.60e-4	3,14
H3K27me3	GSM1225093	Col-0; young leaves	77	3.84e-4	3,09
H3K27me3	GSM1296933	14 day-old wt seedlings	65	1.04e-3	3,09
H3K27me3	GSM2367142	WT	71	6.22e-4	3,06
H3K27me3	GSM1625915	F1-Leaf Columbia x Columbia	77	3.84e-4	3,04
H3K27me3	GSM1296932	14 day-old wt seedlings	65	1.28e-3	2,98
H3K27me3	GSM1530043	WT_C24	77	4.10e-4	2,93
H3K27me3	GSM1048079	H3K27me3 of 14 day-old seedlings	77	4.32e-4	2,9
H3AK27me3	GSM2498439	WT	71	8.13e-4	2,88
H3K27me3	GSM1530044	WT_Col/C24	77	4.37e-4	2,87
H3K27me3	GSM2367144	WT	65	1.65e-3	2,87
H3K27me3	GSM2079966	Col-0 seeds	77	4.72e-4	2,84
H3K27me3	GSM2366608	WT; 14 day-old seedlings	77	4.73e-4	2,83
H3AK27me3	GSM2498437	WT	71	1.01e-3	2,77
H3K27me3	GSM1530042	WT_Col	77	5.87e-4	2,76
H3K27me3	GSM1592579	Col; seedlings at 12 DAG	71	1.01e-3	2,76
H3K27me3	GSM1625916	F1-Endosperm Landsberg x Columbia	65	2.57e-3	2,71
H3K27me3	GSM2028109	Wt seeds	77	9.27e-4	2,52

Marker	GEO Dataset	Description	Overlap %	p-value (FDR)	Fold change
H3K27me3	GSM2090072	C24; 10 day-old seedlings	77	9.60e-4	2,50
H3K27me3	GSM2028111	Wt seeds	77	1.26e-3	2,41
H3K27me3	GSM2101793	C24; 10 day-old seedlings	65	1.35e-2	2,14
H3K4me1	GSM1669404	Col-0; 21 day-old rosette leaves in drought stress state	65	7.99e-3	2,31
H3K4me1	GSM1669402	Col-0; 21 day-old rosette leaves in drought stress state	83	1.65e-3	2,11
H3K4me1	GSM1669418	Col-0; 21 day-old rosette leaves in drought stress state	83	2.07e-3	2,07
LHP1	GSM2028108	ProLHP1:LHP1:GFP	77	3.95e-4	2,95

Supplementary Table 9: Epigenetic binding factor enrichment of indolic GSL biosynthesis genes.
Data from Plant Regulomics was filtered for enrichment in H2AK121, H3K27 and H3K4 modifications. All such entries that were generated for wildtype plants under regular or under stress conditions and overlapped with at least 50% of the genes of interest were considered.

Marker	GEO Dataset	Description	Overlap %	p-value (FDR)	Fold change
H2AK121ub	GSM2367150	WT	100	1.96e-4	2,65
H2AK121ub	GSM2367148	WT	100	1.96e-4	2,56
H2AK121ub	GSM2367146	WT	100	1.96e-4	2,5
H2AK121ub	GSM2367152	WT	94	1.64e-3	2,27
H2AK121ub	GSM2367138	WT	94	1.64e-3	2,26
H2AK121ub	GSM2367140	WT	94	1.64e-3	2,23
H3K27me3	GSM2028113	WT seeds	87	6.73e-3	2,03
H3K4me1	GSM1669404	Col-0; 21 day-old rosette leaves in drought stress state	67	1.68e-2	2,38
H3K4me1	GSM1669390	Col-0; 21 day-old rosette leaves in normal watered state	67	2.32e-2	2,22
H3K4me1	GSM1669405	Col-0; 21 day-old rosette leaves in drought stress state	67	2.35e-2	2,21
H3K4me3	GSM1669407	Col-0; 21 day-old rosette leaves in drought stress state	94	1.64e-3	2,25
H3K4me3	GSM1669408	Col-0; 21 day-old rosette leaves in drought stress state	87	4.21e-3	2,22
H3K4me3	GSM1669392	Col-0; 21 day-old rosette leaves in normal watered state	94	2.67e-3	2,13
H3K4me3	GSM1669406	Col-0; 21 day-old rosette leaves in drought stress state	87	5.42e-3	2,1
H3K4me3	GSM1669391	Col-0; 21 day-old rosette leaves in normal watered state	87	7.48e-3	2

Supplementary Table 10: Statistical significances of prolonged stress experiments.
Statistical significances of the isolated factors genotype (wildtype or mutant) and jasmonate presence as well as the overall interaction between the two (GT-JA). **(A)** corresponds to the GSL quantification in set I *eml* mutants (**Figure 6**), **(B)** to the GSL quantification in set II *eml* mutants (**Figure 7**) and **(C)** to the qPCR of GSL biosynthesis genes in set II *eml* mutants (**Figure 8**). p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

A

	Genotype	**Jasmonate**	**GT-JA**
Aliphatic	*** (0.00059)	*** (< 2e-16)	*** (0.00020)
Indolic	* (0.011)	*** (< 2e-16)	- (0.19)

B

	Genotype	**Jasmonate**	**GT-JA**
Aliphatic	*** (< 2e-16)	*** (< 2e-16)	*** (4.3e-07)
Indolic	*** (2.6e-11)	*** (< 2e-16)	*** (4.19e-07)

C

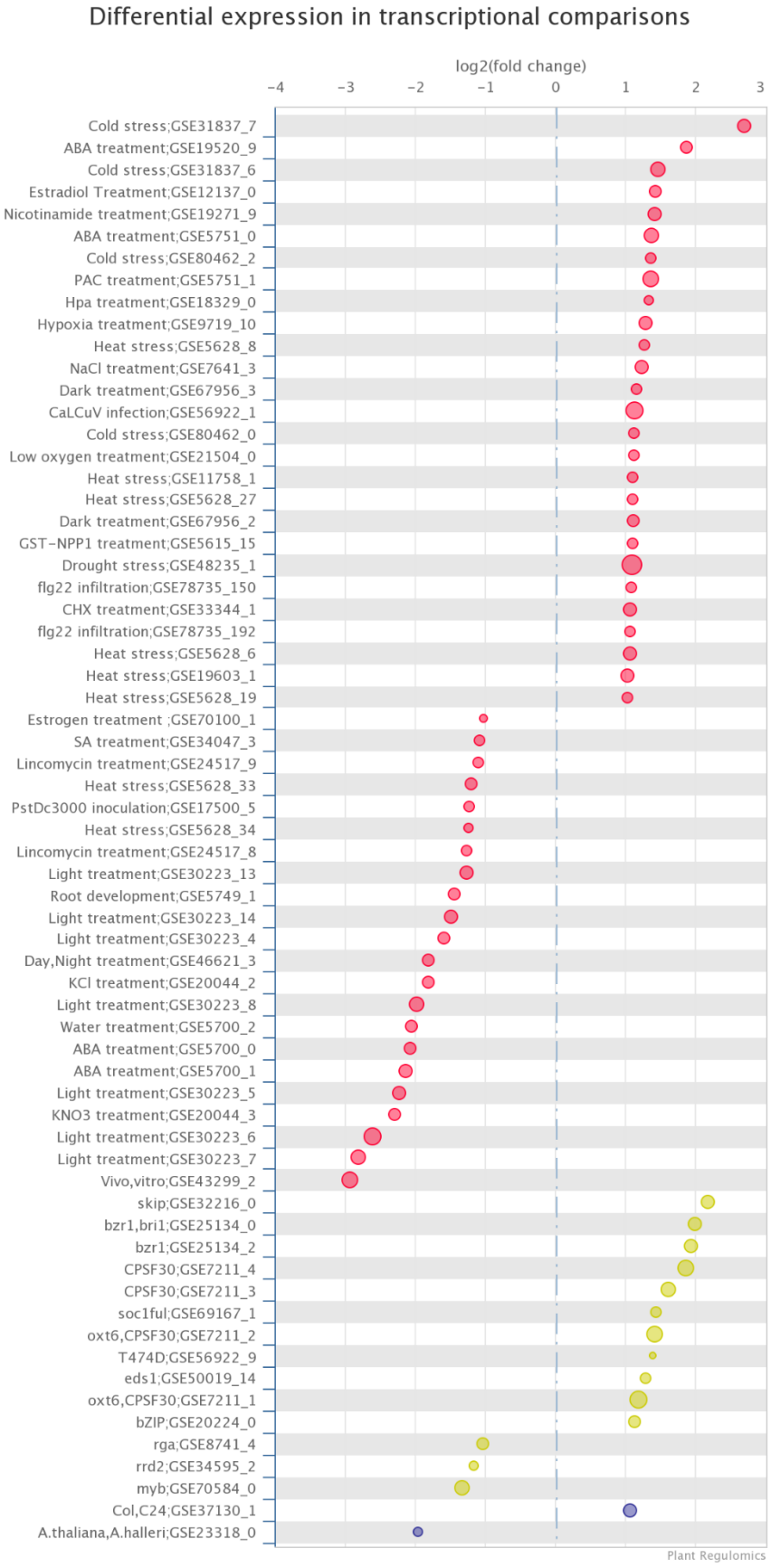
	Genotype	**Jasmonate**	**GT-JA**
CYP79F2	- (0.35)	*** (1.2e-05)	- (0.12)
CYP79B3	. (0.074)	*** (4.7e-14)	*** (8.1e-08)
CYP83B1	- (0.49)	** (0.0017)	* (0.028)

Supplementary Table 11: Standard deviations of GSL concentrations in Brassicaceae root and shoot tissue.

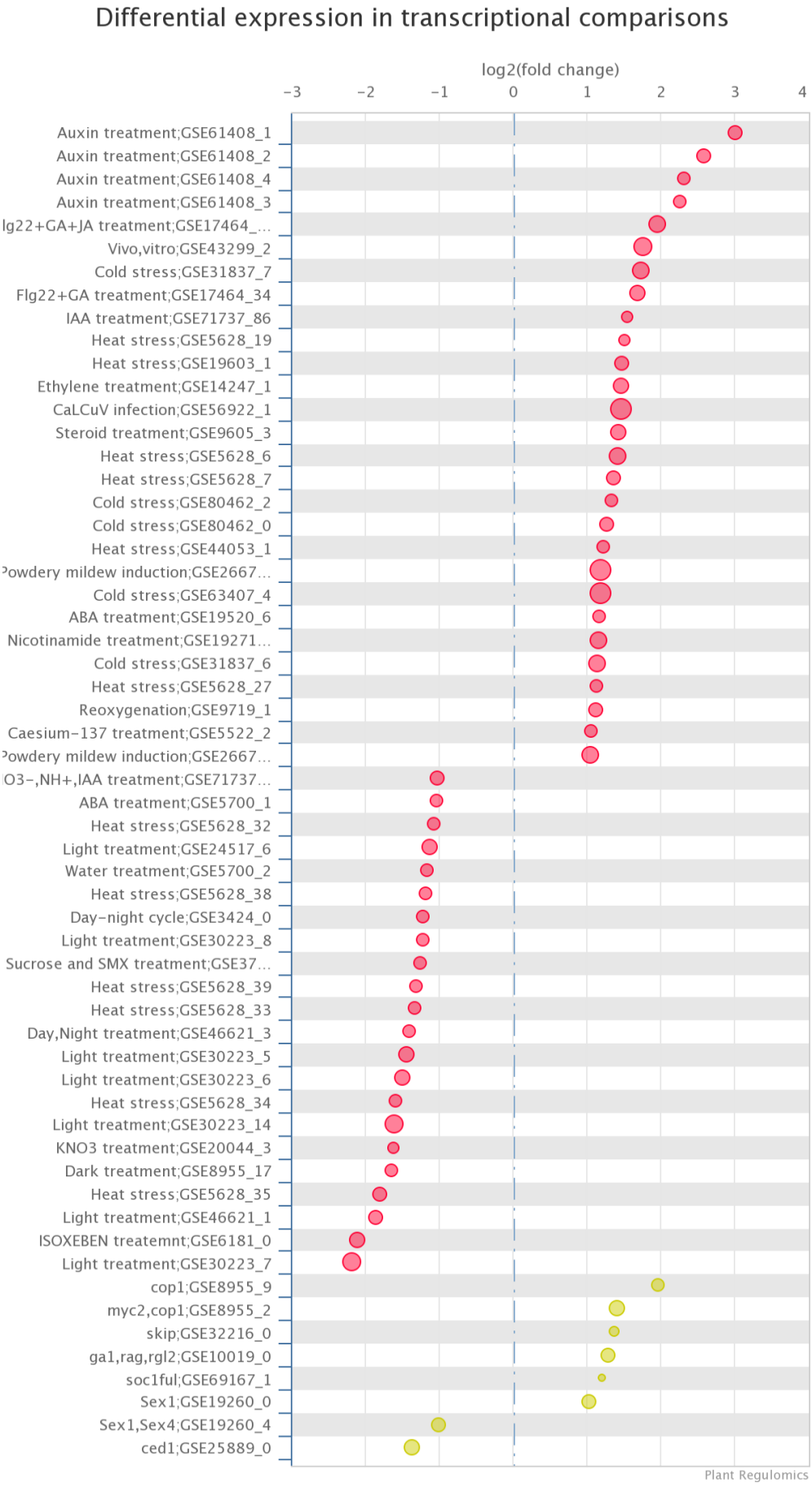
m/z ratio	RT (min)	Possible Identities	Number	Type	Derived from	Standard Deviations					
						Aa root	Aa shoot	Ch root	Ch shoot	Cr root	Cr shoot
360.0429	6	1-Methylethyl	56	A-Alkyl	Val	1,5E-02	2,4E-01	8,6E-05	1,4E-04	1,5E-04	2,0E-05
374.0585 a	6,8	1-Methylpropyl; 2-Methylpropyl; n-Butyl[1]	61, 62, 13[1]	A-Alkyl	Ile, Leu, ?	7,0E-02	2,4E-01	1,7E-04	2,6E-04	1,6E-04	5,2E-05
374.0585 b	7	1-Methylpropyl; 2-Methylpropyl; n-Butyl[1]	61, 62, 13[1]	A-Alkyl	Ile, Leu, ?	7,4E-03	2,7E-01	8,0E-05	1,3E-04	1,1E-04	4,2E-05
402.0898	9,05	n-Hexyl[1]; 3-Methypentyl[1]; 4-Methylpentyl[1]	20[1], 58[1], 59[1]	A-Alkyl	?, Ile, Leu	1,6E-03	1,1E-02	3,1E-03	8,4E-02	6,8E-05	0
416.1050	10	n-Heptyl	–[3]	A-Alkyl	?	1,9E-04	4,8E-03	2,1E-03	3,7E-02	5,2E-04	0
372.0429	6,4	But-3-enyl	12	A-Alkenyl	Met	2,0E-03	1,0E-01	5,5E-03	2,8E-01	1,6E-04	1,0E-04
386.0585	7,4	Pent-4-enyl; 3-Methylbut-3-enyl	101, 52[1]	A-Alkenyl	Met, Leu	5,5E-05	3,2E-04	8,8E-04	3,0E-02	0	0
494.0796	9,7	4-(Benzoyloxy)butyl; 1-(Benzoyloxymethyl)propyl[1]	5, 7[1]	A-BOA	Met, Ile	0	0	1,7E-04	8,8E-04	0	0
390.0534 a	5,19	1-(Hydroxymethyl)propyl; 2-Hydroxy-2-methylpropyl; 3-Hydroxybutyl[1]; 4-Hydroxybutyl[1]	30, 31, 25[1], 26[1]	A-HOA	Ile, Leu, Met, Met	0	0	0	4,3E-04	0	0
390.0534 b	5,4	1-(Hydroxymethyl)propyl; 2-Hydroxy-2-methylpropyl; 3-Hydroxybutyl[1]; 4-Hydroxybutyl[1]	30, 31, 25[1], 26[1]	A-HOA	Ile, Leu, Met, Met	1,1E-01	1,2E-02	0	0,	0,	0
406.0306	6,79	3-(Methylsulfanyl)propyl[2]	95	A-MSAA	Met	3,4E-05	4,8E-05	4,9E-05	5,1E-05	7,1E-05	7,2E-05
420.0457	7,45	**4-(Methylsulfanyl)butyl**	84	A-MSAA	Met	5,1E-04	9,4E-04	1,3E-04	1,4E-04	3,4E-05	2,7E-05
434.0619	8,33	**5-(Methylsulfanyl)pentyl**	94	A-MSAA	Met	1,5E-04	1,7E-04	9,0E-05	1,5E-04	0	0
448.0775	9,2	6-(Methylsulfanyl)hexyl	88	A-MSAA	Met	9,0E-05	9,0E-05	0	6,8E-05	0	0
462.0932	10,05	7-(Methylsulfanyl)heptyl	87	A-MSAA	Met	1,2E-04	1,2E-04	0	0	7,6E-05	0
476.1088	10,82	8-(Methylsulfanyl)octyl	92	A-MSAA	Met	5,3E-03	8,2E-04	0	0	2,6E-03	0
490.1245	11,6	9-(Methylsulfanyl)nonyl[1]	89[1]	A-MSAA	Met	3,9E-01	1,2E-01	1,2E-04	0	1,7E-01	7,3E-04
504.1401	12,5	10-(Methylsulfanyl)decyl[1]	85[1]	A-MSAA	Met	9,0E-02	9,7E-03	5,5E-06	0	1,4E-02	2,5E-04
422.0255	4,68	**3-(Methylsulfinyl)propyl**	73	A-MSIA	Met	4,4E-05	5,1E-05	3,2E-05	2,6E-05	5,4E-05	4,3E-05
436.0411	5,11	**4-(Methylsulfinyl)butyl**	64	A-MSIA	Met	9,2E-05	1,1E-03	1,6E-04	1,5E-03	3,3E-04	6,2E-05
450.0568	5,65	**5-(Methylsulfinyl)pentyl**	72	A-MSIA	Met	8,7E-05	6,1E-04	1,6E-04	2,1E-03	2,0E-05	0
464.0724	6,26	6-(Methylsulfinyl)hexyl[4]	67	A-MSIA	Met	4,2E-05	2,7E-04	5,6E-05	1,6E-04	0	0
478.0881	6,92	7-(Methylsulfinyl)heptyl	66	A-MSIA	Met	2,5E-05	2,3E-04	4,8E-06	0,0E+00	3,7E-05	0

492.1037	7,7	8-(Methylsulfinyl)octyl	69	A-MSIA	Met	1,3E-03	5,7E-03	2,6E-05	2,1E-05	2,3E-03	2,7E-04
506.1194 a	7,95	9-(Methylsulfinyl)nonyl[5]	68[5]	A-MSIA	Met	0	0	6,6E-04	1,0E-04	0	0
506.1194 b	8,45	9-(Methylsulfinyl)nonyl[5]	68[5]	A-MSIA	Met	9,6E-02	3,7E-01	0	0	2,0E-01	1,2E-02
520.1350	9,18	10-(Methylsulfinyl)decyl	65	A-MSIA	Met	1,2E-01	2,1E-01	6,8E-05	0	2,3E-01	1,4E-02
522.1143 a	8,22	9-(Methylsulfonyl)nonyl[5]	79[5]	A-MSOA	Met	6,1E-04	2,9E-04	0	0	4,0E-04	0
522.1143 b	8,55	9-(Methylsulfonyl)nonyl[5]	79[5]	A-MSOA	Met	4,2E-04	1,2E-03	1,6E-04	6,9E-05	5,9E-04	4,2E-05
522.1143 c	9	9-(Methylsulfonyl)nonyl[5]	79[5]	A-MSOA	Met	2,1E-04	1,2E-02	0	0	9,0E-05	1,2E-04
408.0429	7,4	**Benzyl**	11	B	Phe	2,8E-04	1,6E-04	1,6E-01	2,2E-01	5,6E-04	1,7E-04
422.0585	8,3	**2-Phenylethyl**	105	B	Phe	5,1E-03	3,0E-03	2,1E-02	2,4E-01	8,7E-05	1,8E-05
447.0538	7,87	**Indol-3-ylmethyl**	43	I	Trp	9,7E-04	8,6E-04	8,4E-02	6,0E-02	2,5E-04	0
463.0487	7	**4-Hydroxyindol-3-ylmethyl**	28	I	Trp	5,8E-03	4,1E-03	5,8E-04	5,8E-04	2,4E-04	0
477.0643 a	8,5	**4-Methoxyindol-3-ylmethyl**	48	I	Trp	5,3E-03	6,1E-02	3,8E-02	1,1E-01	7,5E-05	0
477.0643 b	9,05	**1-Methoxyindol-3-ylmethyl**	47	I	Trp	0	0	1,6E-02	1,7E-03	0	0

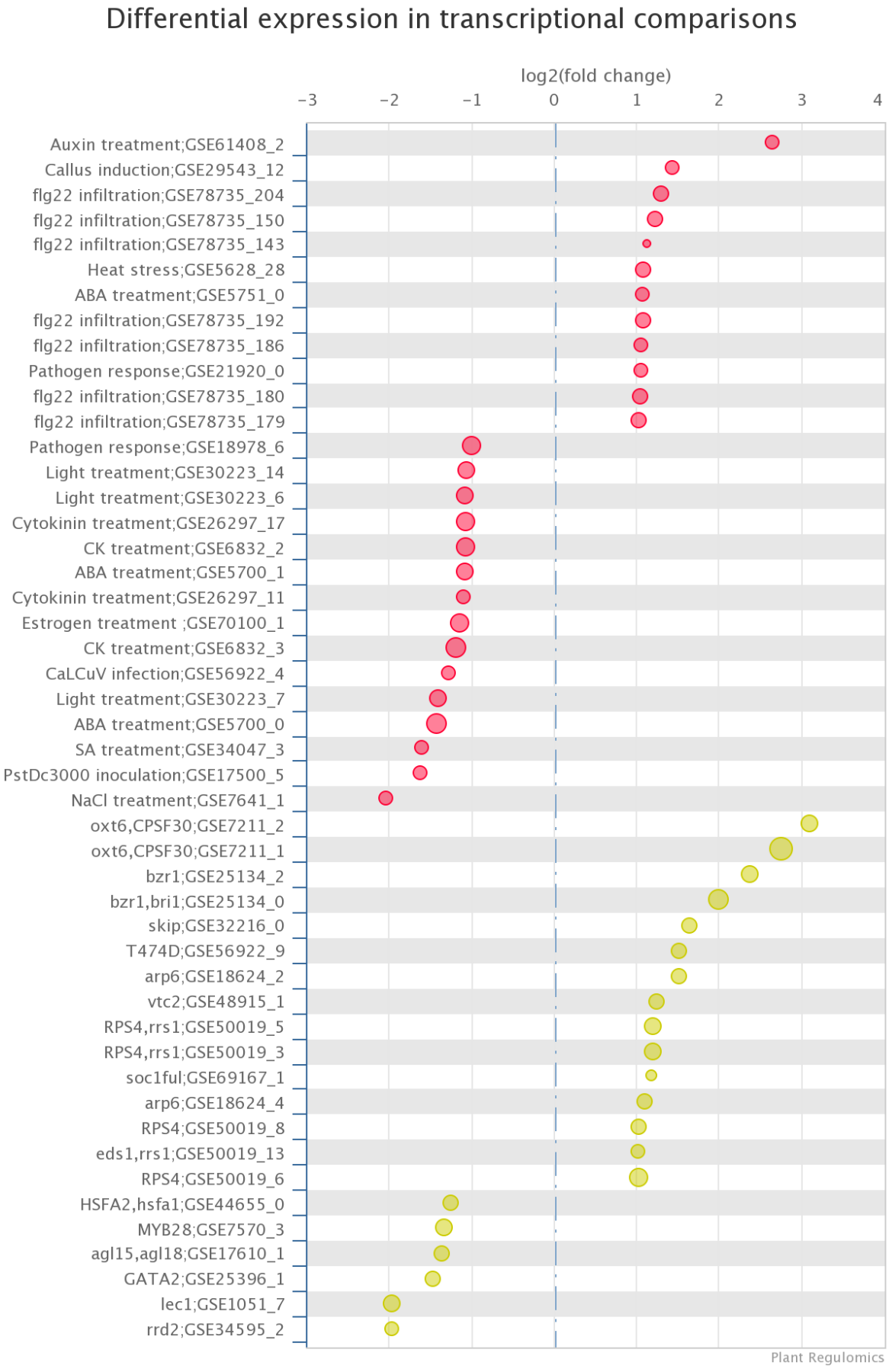
6.2 Supplementary Figures



Supplementary Figure 1: Transcriptomic enrichment analysis for *EML1*.



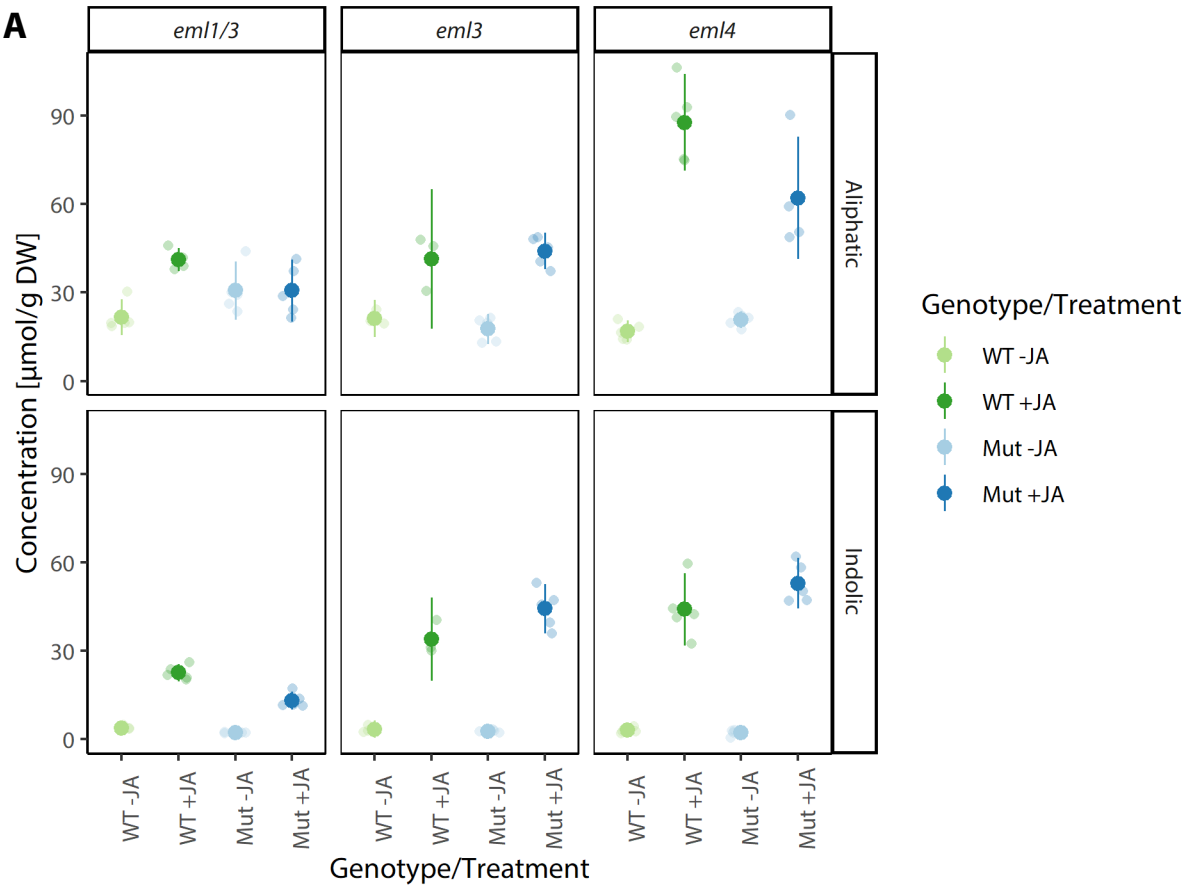
Supplementary Figure 2: Transcriptomic enrichment analysis for *EML2*.



Supplementary Figure 3: Transcriptomic enrichment analysis for *EML3*.

Supplementary Figure 4: Transcriptomic enrichment analysis for *EML4*.

A



B

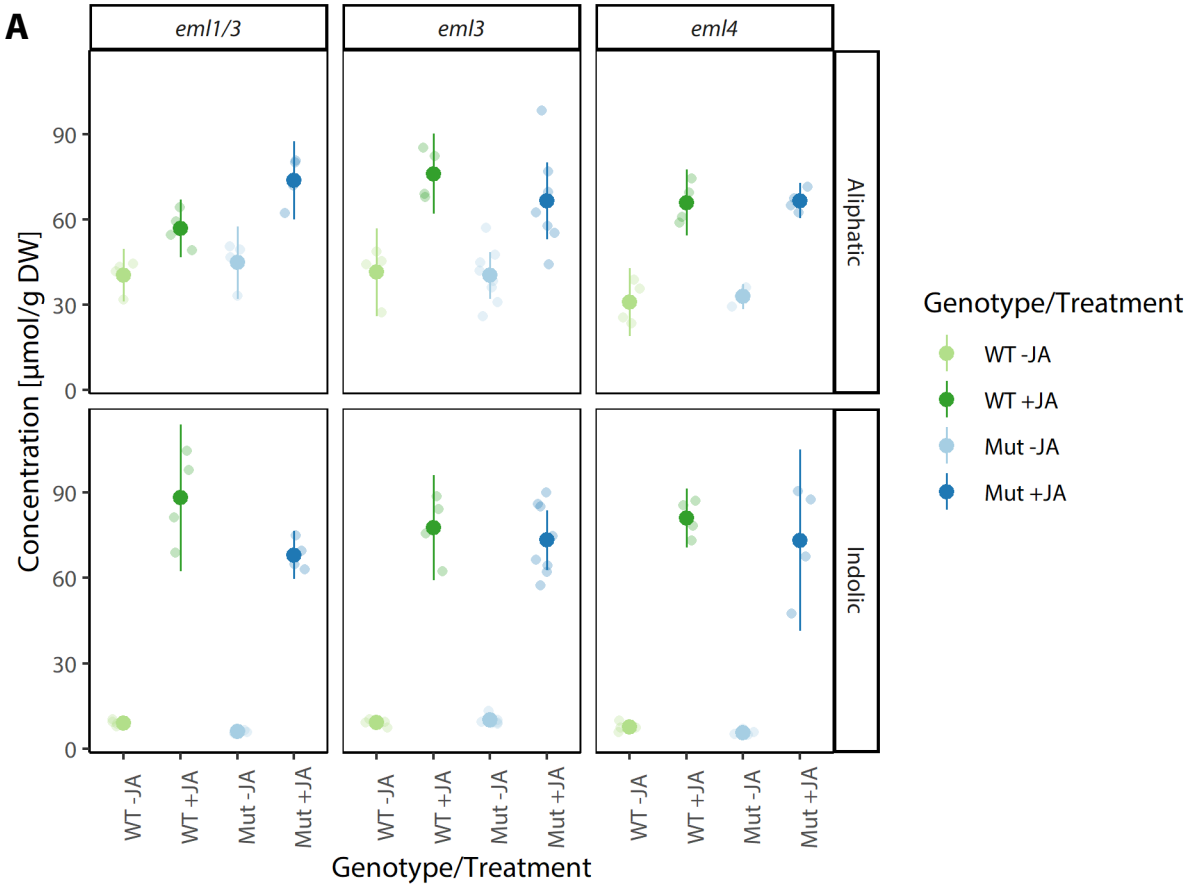
	Genotype	**Jasmonate**	**GT-JA**
Aliphatic	*** (3.7e-07)	*** (< 2e-16)	*** (6.0e-11)
Indolic	*** (1.5e-12)	*** (< 2e-16)	*** (8.7e-13)

C

	eml1/eml3	***eml3***	***eml4***
Aliphatic	** (0.0038)	- (0.29)	** (0.0073)
Indolic	*** (0.00014)	* (0.037)	. (0.087)

Supplementary Figure 5: The set II *eml1/eml3* double mutant is impaired in JA-mediated GSL induction (Repetition 2). **(A)** GSL quantification of set II *eml* mutants grown under JA presence. Error bars represent the 95% confidence interval. *eml1/eml3* is abbreviated to *eml1/3*. **(B)** Statistical significance of the isolated factors genotype (wildtype or mutant) and jasmonate presence as well as the overall interaction between the two (GT-JA). **(C)** Statistical significance of differences in GSL concentration between mutant plants and their respective wildtypes, regarding the interaction between and jasmonate presence. p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

A



B

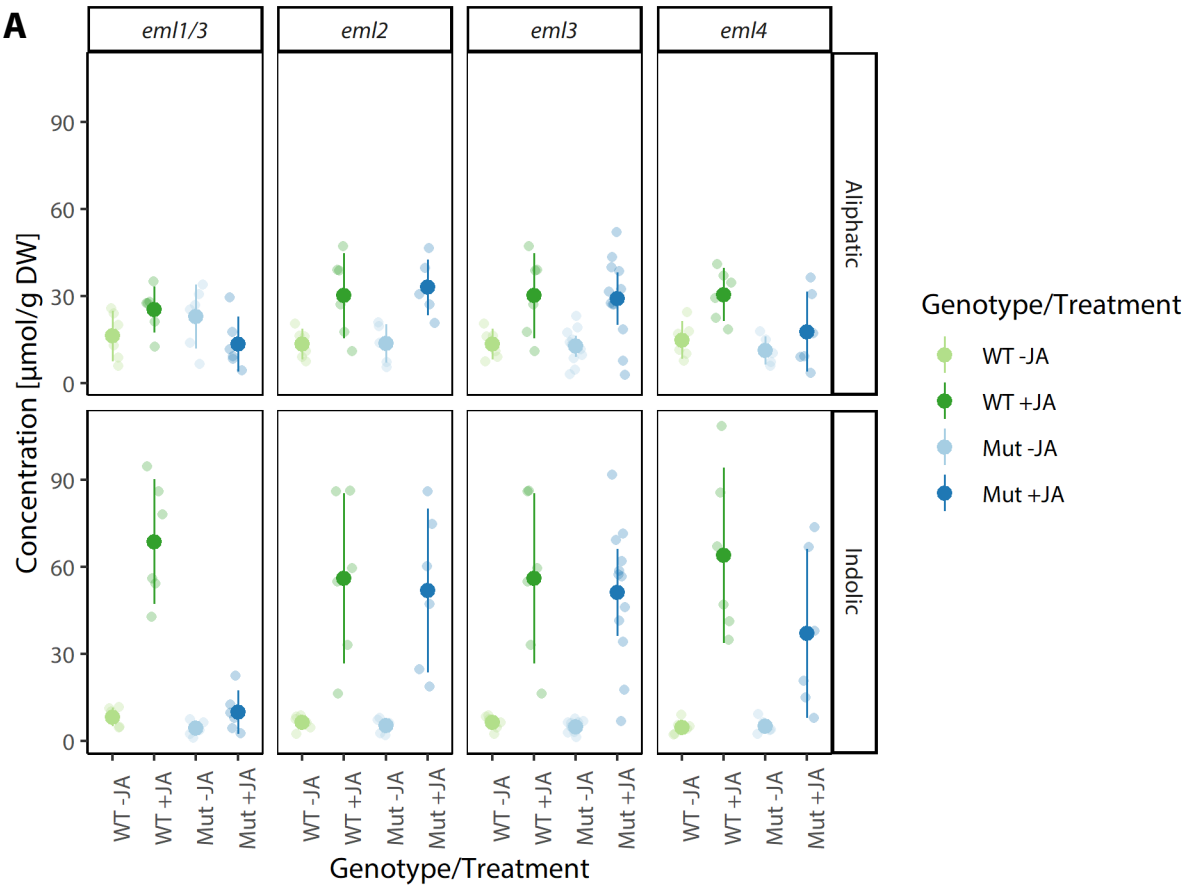
	Genotype	**Jasmonate**	**GT-JA**
Aliphatic	* (0.028)	*** (< 2e-16)	- (0.43)
Indolic	. (0.064)	*** (< 2e-16)	- (0.21)

C

	eml1/eml3	***eml3***	***eml4***
Aliphatic	- (0.12)	- (0.43)	- (0.81)
Indolic	. (0.071)	- (0.48)	- (0.59)

Supplementary Figure 6: The set II *eml1/eml3* double mutant is impaired in JA-mediated GSL induction (Repetition 3). **(A)** GSL quantification of set II *eml* mutants grown under JA presence. Error bars represent the 95% confidence interval. *eml1/eml3* is abbreviated to *eml1/3*. **(B)** Statistical significance of the isolated factors genotype (wildtype or mutant) and jasmonate presence as well as the overall interaction between the two (GT-JA). **(C)** Statistical significance of differences in GSL concentration between mutant plants and their respective wildtypes, regarding the interaction between and jasmonate presence. p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.

A



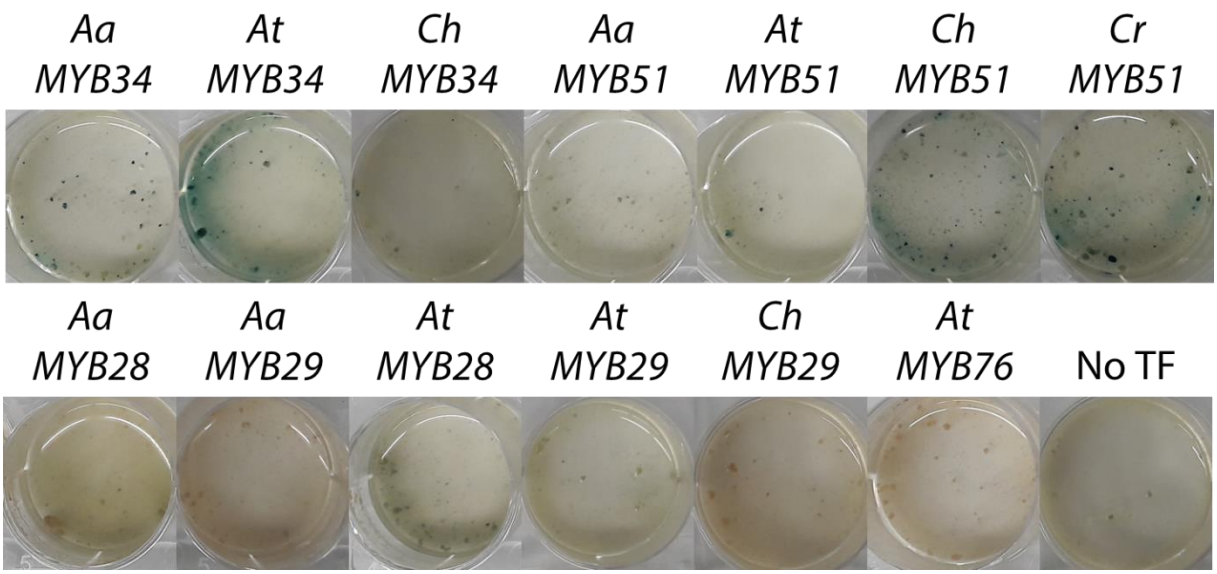
B

	Genotype	**Jasmonate**	**GT-JA**
Aliphatic	- (0.41)	*** (1.1e-07)	* (0.013)
Indolic	** (0.0015)	*** (< 2e-16)	** (0.0054)

C

	eml1/eml3	***eml2***	***eml3***	***eml4***
Aliphatic	* (0.019)	- (0.72)	- (0.97)	- (0.21)
Indolic	*** (5.8e-06)	- (0.85)	- (0.79)	- (0.11)

Supplementary Figure 7: The set II *eml1/eml3* double mutant is impaired in JA-mediated GSL induction (Repetition 4). **(A)** GSL quantification of set II *eml* mutants grown under JA presence. Error bars represent the 95% confidence interval. *eml1/eml3* is abbreviated to *eml1/3*. **(B)** Statistical significance of the isolated factors genotype (wildtype or mutant) and jasmonate presence as well as the overall interaction between the two (GT-JA). **(C)** Statistical significance of differences in GSL concentration between mutant plants and their respective wildtypes, regarding the interaction between and jasmonate presence. p-values were calculated using two-way ANOVA. Significance codes represent the p-value: "." < 0.1, "*" < 0.05, "**" < 0.01, "***" < 0.001.



Supplementary Figure 8: Indolic MYBs can induce transcription of *AtCYP79B3* in *C. rubella* cell culture.
Qualitative GUS staining of *C. rubella* cell culture overexpressing *MYB* genes from different Brassicaceae species and the *proAtCYP79B3:GUS* promoter-reporter construct, six days after transfection.

7 List of Abbreviations

1OH-I3M	1-Hydroxyindol-3-ylmethyl GSL
1MO-I3M	1-Methoxyindol-3-ylmethyl GSL
2,4-D	2,4-Dichlorophenoxyacetic acid
4OH-I3M	4-Hydroxyindol-3-ylmethyl GSL
4MO-I3M	4-Methoxyindol-3-ylmethyl GSL
Aa	*Arabis alpina*
ABA	Abscisic acid
ACT domain	Aspartate kinase, chorismate mutase and TyrA protein domain
AIP1	ABAP1-INTERACTING PROTEIN 1
amiRNA	Artificial microRNA
Amp	Ampicilin
ANOVA	Analysis of variance
AOP	ALKENYL/HYDROXYALKYL-PRODUCING
AP2/ERF	APETALA2/Ethylene-responsive factor
At	*Arabidopsis thaliana*
bar	Bialaphos resistance
BES1	BRI1-EMS-SUPPRESSOR 1
BGS	Benzyl GSL
bHLH	basic helix-loop-helix
bp	Base pairs
BR	Brassinosteroids
BRI1	BRASSINOSTEROID-INSENSITIVE 1
BZR1	BRASSINAZOLE-RESISTANT 1
Carb	Carbencilin
CDS	Coding DNA sequence
Ch	*Cardamine hirsuta*
Chl	Chloramphenicol
Col	Columbia
COP1-SPA	CONSTITUTIVE PHOTOMORPHOGENIC 1/SUPPRESSOR OF PHYA-105
Cr	*Capsella rubella*
CYP	CYTOCHROME P450
DEAE	Diethylaminoethyl
DAPI	4′,6-Diamidino-2-phenylindole
DMF	Dimethylformamide
DOF	DNA-binding-with-one-finger
DTT	Dithiothreitol
DW	Dry weight
EDTA	Ethylenediaminetetraacetic acid
EML	EMSY-LIKE
ENT domain	EMSY N-terminal domain
ET	Ethylene
FDR	False discovery rate
flg22	22-amino acid flagellin peptide
FW	Fresh weight
Gent	Gentamycin
GFP	Green fluorescent protein

GSL	Glucosinolate
GT	Genotype
GUS	β-glucuronidase
HAC1	HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY 1
HDA6	HISTONE DEACETYLASE 6
HPLC	High performance liquid chromatography
Hyg	Hygromycin
I3M	Indole-3-methyl GSL
IGMT	INDOLE GLUCOSINOLATE O-METHYLTRANSFERASE
IPMI	ISOPROPYLMALATE ISOMERASE
IPMDH	ISOPROPYLMALATE DEHYDROGENASE
IQD1	IQ-DOMAIN 1
JA	Jasmonic acid or jasmonate
JAR1	JASMONATE-RESISTANT 1
Jas domain	Jasmonate-associated domain
JAZ	JASMONATE ZIM-DOMAIN
JID	JAZ-interacting domain
Kan	Kanamycin
LB	Left border primer; Lysogeny broth
LC-MS	Liquid chromatography-mass spectrometry
Ler	Landsberg erecta
LIF2	LHP1-INTERACTING FACTOR 2
LHP1	LIKE HETEROCHROMATIN PROTEIN 1
LN_2	Liquid nitrogen
LP	Left primer
MAM	METHYLTHIOALKYLMALATE SYNTHASE
MED25	MEDIATOR 25
MeJA	Methyl jasmonate
MOPS	3-Morpholinopropanesulfonic acid
MPIPZ	Max-Planck-Institut für Pflanzenzüchtungsforschung
MS	Murashige and Skoog
MYB	MYELOBLASTOSIS
MYC	MYELOCYTOMATOSIS
NAA	1-Naphthaleneacetic acid
NINJA	NOVEL INTERACTOR OF JAZ
OBP2	OBF-BINDING PROTEIN
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
OX	Overexpression
PCR	Polymerase chain reaction
PEN2	PENETRATION 2
PEN2L	PEN2-LIKE
PRC	Polycomb repressive complex
qPCR	Quantitative real-time PCR
RB	Right border primer
Rif	Rifampicin
RIF1	R-INTERACTING FACTOR 1
RP	Right primer

rpm	Revolutions per minute
RT	Retention time; Room temperature
SA	Salicylic acid
SCF^{COI1}	Skp1-Cullin-F-box/CORONATINE INSENSITIVE1
SDI1	SULFUR DEFICIENCY-INDUCED 1
SDS	Sodium dodecyl sulfate
SSLP	Simple sequence length polymorphism
TAD	Transcriptional activation domain
TAE	Tris-acetate-EDTA
TAI	Time after induction
T-DNA	Transfer DNA
TF	Transcription factor
Tfb	Transformation buffer
TOF	Time of flight
TP	Time point
TPL	TOPLESS
TPR	TOPLESS-RELATED
UTR	Untranslated region
VM	Vermehrungserde
WGD	Whole-genome duplication
WT	Wildtype
X-Gluc	5-bromo-4-chloro-3-indolyl glucuronide
YEB	Yeast extract broth

8 List of Figures

9 List of Tables

10 Acknowledgments

First and foremost, I would like to thank PD Dr. Tamara Gigolashvili not only for providing me with the opportunity for this PhD project, but for supporting it with heart and soul every step of the way and making it a profoundly positive experience.

Moreover, I would like to thank Prof. Dr. Ute Höcker, Prof. Dr. Karin Schnetz and Dr. Jathish Ponnu for joining my thesis defense committee, as well as Prof. Dr. Juliette de Meaux and Dr. Franziska Turck for joining my thesis advisory committee and bringing their fresh ideas and new perspectives to the table.

A big thank you also goes to the members of AG Gigolashvili. I would like to thank Claudia Nothelle for her invaluable practical help and her relentless effort in keeping the lab running, as well as all the students who contributed to this project in one way or another: Philipp Katzy, Lara Krumbein, Henni Läßle, Jan Mandelkow, Daniela Porsch and Jill Voelkel. Special thanks go to Katharina Miethke and Jessica Wirtz who did their bachelor theses on this project.

I also would like to thank Dr. Sabine Metzger and Dr. Vera Wewer of the Biocenter MS platform for their analysis of the Brassicaceae plants, as well as our cooperation partners Prof. Dr. Paweł Bednarek, Dr. Jelena Brkljacic, Prof. Dr. Thomas Eulgem and Prof. Dr. Erich Grotewold for providing us with plasmids, seeds and information.

Many thanks go to AG Kopriva for supporting our little work group, especially to Prof. Dr. Stanislav Kopriva and Dr. Anna Koprivova for contributing with their expertise, as well as Irene Klinkhammer and Bastian Welter for their technical support. Moreover, I would like to warmly thank all former and current members of AG Kopriva for shared seminars and conferences, equipment and chemicals, beers and barbecues. Special thanks go to my friends and colleagues Dorian Baumann, Li Chen, Christof Dietzen, Süleyman Günal, Richard Jacoby, Tim Jobe, José López Ramos, Melina Schwier and Ivan Zenzen for the great times we spent together on this journey. I also would like to thank our dear neighbors from AGs Bucher, Höcker, Hülskamp and Zuccaro for providing us with feedback and material support.

Last but not least, I would like to thank my friends and family for their wonderful encouragement and support over the years. In particular, I would like to thank my parents who have always been there for me and have made all of this possible in the first place.

11 Data Availability Statement

The written records, datasets and biological materials generated and analyzed during this study are available from PD Dr. Tamara Gigolashvili upon request.

12 Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten – noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Köln, den 15.02.2021 Simon Mitreiter