FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA
Chemisch-Geowissenschaftliche
Fakultät

MASTERARBEIT

Production of new-to-nature benzoxazinoids in engineered *Nicotiana benthamiana*

Susan Schlüter
geboren am 17. November 1998
in Göttingen

Erstgutachter/in:	Prof. Dr. Sarah E. O‘Connor
Zweitgutachter/in:	Prof. Dr. Julie A. Z. Zedler

Jena, 15. Juli 2024

Acknowledgements

I would like to thank Prof. Dr. Sarah O'Connor and Dr. Tobias Köllner for assigning the topic of this thesis and enabling my work at the Department of Natural Product Biosynthesis at the Max Planck Institute for Chemical Ecology. Further, I would like to thank Prof. Dr. Julie Zedler for taking the position as second examiner of my thesis. I extend my special thanks to my supervisor Matilde Florean, who was a great help during the practical work in the laboratory as well as in the process of writing this thesis and was always open for my questions and problems. For the extensive support with the analytics, I would like to thank Dr. Maritta Kunert, Sarah Heinicke and Dr. Yoko Nakamura. Lastly, I am grateful to the whole SOC group for a pleasant, cooperative and welcoming atmosphere.

Abstract

Throughout human history, natural products have been used for crop protection or as medicine and up to today, they play an important role as lead structures in modern drug discovery and development of agrochemicals. Halogenation is a frequently found structural modification that can alter the biological activity of a compound, improve its stability, bioavailability, and selectivity. Benzoxazinoids are a group of indole-derived plant defense compounds with antifeedant, antimicrobial, and allelopathic activities. Compared to plant-derived benzoxazinoids, synthetic halogenated analogs showed improved antifungal and selective herbicidal activity. Chemical total synthesis and derivatization of natural products can be challenging due to poor regioselectivity, complex structural features, or instability of the compound. Production of structural analogs through biosynthetic approaches can help tackle these problems. In a precursor-directed biosynthesis using a heterologous host, analogs of an early pathway intermediate that are commercially available can be incorporated by the enzymes of the pathway. The biosynthesis of benzoxazinoids in plants evolved independently several times and the biosynthetic enzymes involved have been discovered in for several of these plant species. The variety of different enzymes catalyzing the same reaction increases chances for good incorporation of substrate analogs.

Starting from halogenated indole, the substrate specificity of the pathway enzymes was characterized individually and the precursor-directed biosynthesis of multiple chlorinated and fluorinated benzoxazinoids in the plant host *N. benthamiana* was successfully achieved. Furthermore, the combined transient expression of the pathway enzymes together with bacterial halogenases allowed for the complete de novo biosynthesis of halogenated benzoxazinoids in planta. The co-expression of different tryptophan halogenases with a tryptophanase provided chlorinated indole that subsequently was converted into benzoxazinoids. Directing the subcellular localization of the tryptophan halogenases to the chloroplast and thus adapting it to the plant host, led to a significant increase in yields of halogenated benzoxazinoids.

Zusammenfassung

Seit Jahrtausenden werden Naturstoffprodukte für den Pflanzenschutz in der Landwirtschaft oder als Medizin verwendet. Bis heute spielen sie als Leitstrukturen eine wichtige Rolle in der Entwicklung von modernen Arzneimitteln oder Agrochemikalien. Die Halogenierung ist dabei eine häufig eingesetzte strukturelle Modifikation, die die biologische Aktivität einer Verbindung verändern kann, um ihre Stabilität, Bioverfügbarkeit und Selektivität zu verbessern. Benzoxazinoide sind eine Gruppe pflanzlicher, von Indol abgeleiteter Abwehrstoffe, die zum Schutz vor Herbivoren und Mikroben produziert werden und zudem allelopathische Aktivität besitzen. Im Vergleich zu natürlich vorkommenden Benzoxazinoiden zeigten synthetische, halogenierte Analoga eine verbesserte antifungale Aktivität und eine selektive Herbizidwirkung. Die chemische Totalsynthese und Derivatisierung von Naturstoffen ist jedoch aufgrund mangelnder Regioselektivität, komplexer struktureller Merkmale oder Instabilität der Verbindung oftmals eine Herausforderung. Die Herstellung von Strukturanaloga durch biosynthetische Ansätze kann helfen, diese Probleme zu lösen. Bei der vorläufergesteuerten Biosynthese unter Verwendung eines heterologen Expressionssystems können Analoga von kommerziell erhältlichen Zwischenprodukten des Stoffwechselwegs von den Enzymen des Stoffwechselwegs eingebaut werden. Die Biosynthese von Benzoxazinoiden hat sich in den Poaceae und verschiedenen Arten der Eudikotyledonen unabhängig voneinander entwickelt und die beteiligten Enzyme sind für mehrere Pflanzenarten bekannt. Die Vielfalt der verschiedenen Enzyme, die dieselbe Reaktion katalysieren, erhöht die Chancen für einen guten Einbau von Substratanaloga.

Ausgehend von halogeniertem Indol wurde die Substratspezifität der einzelnen Enzyme des Stoffwechselwegs charakterisiert und die vorläufergesteuerte Biosynthese mehrerer chlorierter und fluorierter Benzoxazinoide in der Host-Pflanze *N. benthamiana* erfolgreich erreicht. Darüber hinaus konnte durch die kombinierte transiente Expression der Enzyme des Stoffwechselwegs zusammen mit bakteriellen Halogenasen die vollständige *de novo* Biosynthese halogenierter Benzoxazinoide *in planta* erreicht werden. Die Koexpression verschiedener Tryptophan-Halogenasen mit einer Tryptophanase erzeugte chloriertes Indol, das anschließend in Benzoxazinoide umgewandelt wurde. Die Verlagerung der subzellulären Lokalisierung der Tryptophan-Halogenasen in die Chloroplasten und die damit verbundene Anpassung an das pflanzliche Expressionssystem führte zu einer signifikanten Steigerung der Ausbeute an halogenierten Benzoxazinoiden.

Table of Contents

List of figures

List of tables

Introduction

Natural products as lead structures for the development of biologically active compounds

Plants as sessile organisms produce a plethora of functionally diverse metabolites to cope with a multitude of abiotic and biotic stresses [1,2]. These metabolites are usually referred to as natural products or specialized metabolites and play a role in defense [3], signaling [4], and competition for nutrients [5]. In human history, natural products have been used for crop protection or as medicine for millennia and up to today, they play an important role in modern drug discovery and development of agrochemicals [6,7]. In the field of crop protection, 18% of agrochemicals used are natural products and a further 38% of compounds are based on natural products that have served as lead structures for further development [7]. Natural products are especially of interest for the development of antimicrobial and anticancer agents [8–10]. In fact, by 2007, 70% of antimicrobial drugs and more than 60% of anticancer drugs that entered clinical trials within the past 30 years were based on natural products [11].

Using natural products as lead structures for drug or agrochemical development has several advantages. Since the biological activity of a natural product is the product of evolutionary selection processes, the probability to find a structure displaying activity and that can be used as a scaffold for further modification is increased [1]. The compound of interest can be isolated from the producing organisms or, if the biosynthetic pathway is already elucidated, the compound can also be produced in a heterologous host, allowing for higher yields and better stereospecificity. This is of relevance as natural products can have complex chemical structures including chiral centers, which can be challenging to generate in total chemical synthesis [9,12].

In some cases, natural products derived from isolation or total synthesis show sufficient activity and stability for application such as the anticancer polyketide salinosporamide A produced by the marine actinomycete bacterium *Salinispora tropica* and the antibiotic glycopeptide vancomycin isolated from the gram-positive bacterium *Amycolatopsis orientalis,* both of which are examples of naturally occurring chlorinated products [13–15]. However, structural modifications are often needed to improve clinically important properties like bioavailability, activity, photostability, application dose and rate, or to minimize unwanted side effects. Structural modification of natural products can be achieved through semisynthesis, which refers to the process of chemically modifying a compound isolated from natural sources [16,17]. However, this can be difficult due to insufficient stability of natural products under harsh reaction conditions or due to poor regio and stereospecificity of these reactions. Modification

of lead structures can also be obtained through precursor-directed biosynthesis (PDB). In this approach, analogues of the early pathway intermediates, are provided to the producing organism to accomplish the remaining, more complex, transformations using the biosynthetic machinery of the natural product pathway. This can be achieved by feeding chemically synthesized analogues of pathway intermediates to the native organism producing the compounds of interest [18,19], to a heterologous host where the pathway of interest has been reconstituted [20], or by genetically engineering the native producer to provide unnatural substrate [21]. Lastly, it is possible to transform the native or heterologous host harboring the pathway of interest with a gene that allows the installation of a chemical handle, which allows the subsequent chemical selective functionalization of the compound of interest [22].

Benzoxazinoids

Benzoxazinoids (BXDs) are a structurally diverse group of specialized plant metabolites found primarily in the grasses (Poaceae), but also in some species of the eudicotyledons [23]. BXDs possess antifeedant, antimicrobial, and insecticidal activity and are abundant in important crops like maize, wheat, and rye [24,25].

Biosynthesis of BXDs in *Zea mays*

The biosynthesis of BXDs was first elucidated in maize (*Zea mays*) (**Figure 1B**) [26]. The first committed step of the biosynthetic pathway consists of the cleavage of indole-3-glycerol phosphate (I3GP) to indole catalyzed by the indole-3-glycerol phosphate lyase (IGL) BX1 that is localized in the plastids [27]. The following oxidation steps are catalyzed by four cytochrome P450 monooxygenases (CYP) localized in the endoplasmic reticulum (ER) [28,29]. BX2 catalyzes the oxidation of indole to indolin-2-one (I2O), which is subsequently oxidized by BX3 to form 3-hydroxy-indolin-2-one (3HI2O). 3HI2O is oxidized to 2-hydroxy-3,4-dihydro-1,4-benzoxazin-3-one (HBOA) by BX4, which catalyzes an oxidative ring expansion of the heterocycle, forming a second 6-membered ring. Subsequent *N*-hydroxylation through BX5 yields 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA). Since DIBOA is both unstable and phytotoxic, it is glucosylated into DIBOA-Glc by the UDP-glucosyltransferases (UGT) BX8 or BX9 in the cytosol, which allows storage in the vacuole. DIBOA-Glc can be subject to further modifications [30]. The 2-oxoglutarate-dependent dioxygenase (2-ODD) BX6 can hydroxylate the aromatic ring at C7, forming 2-(2,4,7-trihydroxy-1,4-benzoxazin-3-one)-β-D-glucopyranose (TRIBOA-Glc). TRIBOA-Glc does not accumulate in plant tissue but is converted to 2-(2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one)-β-D-glucopyranose (DIMBOA-Glc) through the *O*-methyltransferase (OMT) BX7 [31]. Biotic stress through pathogen infestation and insect feeding has been shown to induce further modifications of DIMBOA-Glc in maize [32]. The 2-ODD BX13 was shown to catalyze the oxidation of DIMBOA-Glc to 2-(2,4,7-trihydroxy-8-

methoxy-1,4-benzoxazin-3-one)-β-D-glucopyranose (TRIMBOA-Glc). Notably, the oxidation involves hydroxylation and an ortho-rearrangement of the already existing methoxy group from C7 to C8 [33]. This reaction is followed by another methylation catalyzed by BX7, forming 2-(2,4-dihydroxy-7,8-dimethoxy-1,4-benzoxazin-3-one)-β-D-glucopyranose (DIM_2BOA-Glc). The *O*-methyltransferases BX10, BX11, BX12, and BX14 can methylate the *N*-hydroxyl group of DIMBOA-Glc to produce 2-(2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one)-β-D-glucopyranose (HDMBOA-Glc), while BX14 is the only enzyme able to methylate the *N*-hydroxyl group of DIM_2BOA-Glc, producing 2-(2-hydroxy-4,7,8-trimethoxy-1,4-benzoxazin-3-one)-β-D-glucopyranose (HDM_2BOA-Glc) [33].

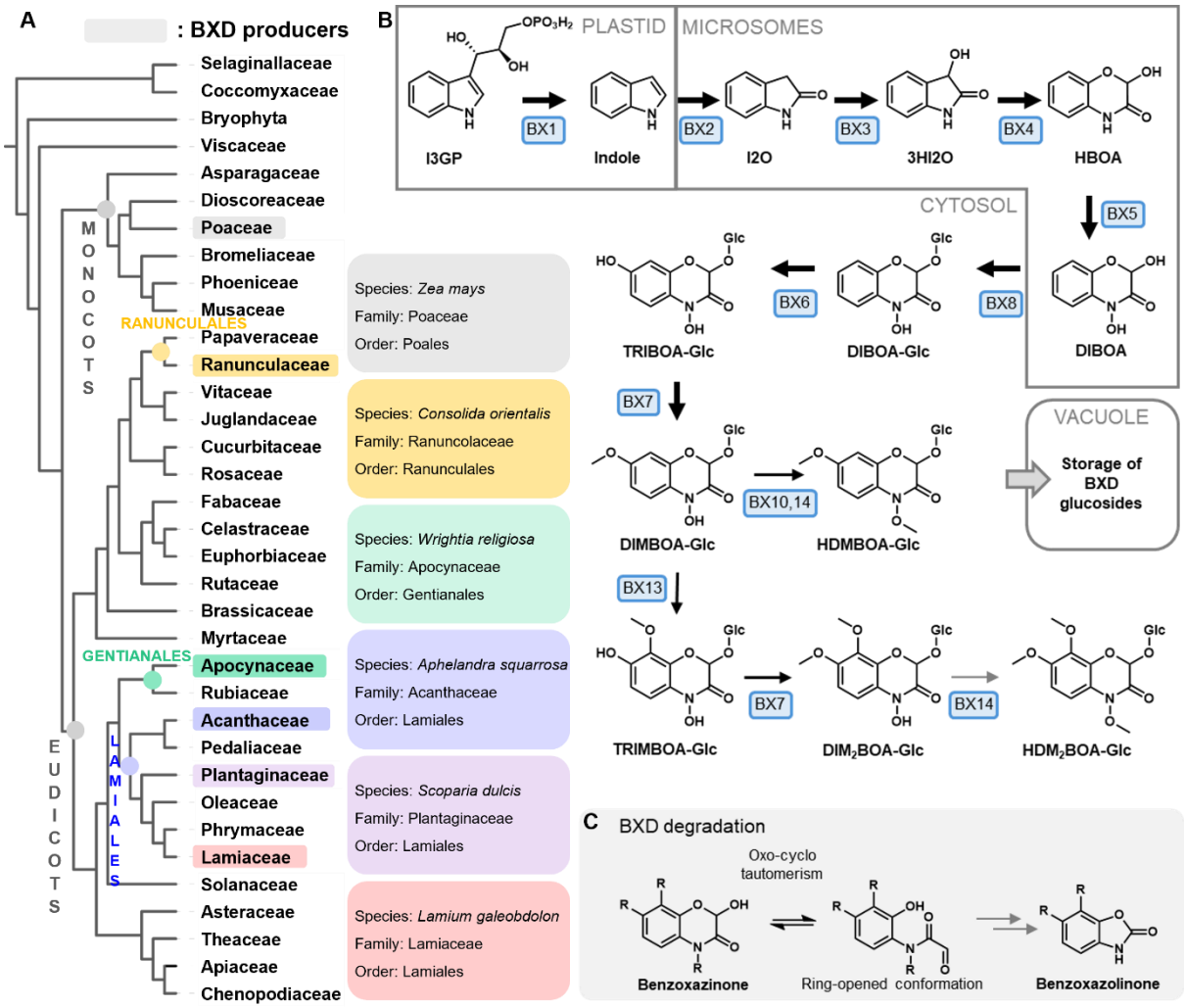

Figure 1: **Phylogenetic tree with highlighted BXD-producing species of which enzymes were used in this study (A). BXD biosynthesis and BX enzymes catalyzing the respective reactions. Subcellular localization of steps occurring outside of the cytosol are marked in boxes (B).** Thick black arrows: Main BXD biosynthesis present in all Poaceae. Thin black arrows: BXD biosynthesis in wheat (Triticum aestivum). Gray arrow: BXD biosynthesis only present in maize (Z. mays). **Benzoxazinone degradation to benzoxazolinones via the ring-opened conformation through oxo-cyclo tautomerism (C).** [23,24,33–35]

Independent evolution of BXD biosynthesis

BXDs are produced not only by the monocot family Poaceae, but also by several eudicot species belonging to the Acanthaceae, Apocynaceae, Lamiaceae, Ranunculaceae, and Scropulariaceae (**Figure 1A**) [34]. Phylogenetic analysis has shown that the biosynthesis of BXDs, at least up to DIMBOA-Glc, is of monophyletic origin in the Poaceae [35]. Prior to radiation of this family, BXD biosynthesis evolved through a duplication of the *tryptophan synthase α-subunit* (*TSA*) gene resulting in the evolution of *Bx1,* and a tandem gene duplication of a *Bx2-like CYP71,* which originated *Bx2*, *Bx3*, *Bx4,* and *Bx5* [35]. Independent evolution of the IGL activity in *Consolida orientalis* (Ranunculaceae – eudicots) led to first assumptions of independent evolution of BXDs in monocot and eudicot species [28]. Indeed, recent elucidation of the BXD pathway in different eudicot species has shown that although the biosynthesis occurs via the same intermediates, the pathway has evolved at least three times independently in eudicots [34].

Table 1: BX enzymes identified in maize and different eudicot species for each step of the biosynthesis up to DIMBOA-Glc. Indole-3-glycerolphosphate lyase – IGL; cytochrome P450 monooxygenase – CYP; flavin-containing monooxygenase – FMO; 2-oxoglutarate-dependent dioxygenase – 2-ODD; *O*-methyltransferase – OMT; UDP-glucosyltransferases – UGT (from [26,31,33,34], * Florean unpublished).

	Enzyme family for the respective step							
Species	**BX1**	**BX2**	**BX3**	**BX4**	**BX5**	**BX6**	**BX7**	**BX8**
Z. mays	IGL	CYP71	CYP71	CYP71	CYP71	2-ODD	OMT	UGT710
A. squarrosa	IGL		FMO	CYP92	CYP81	2-ODD	OMT	
L. galeobdolon	IGL		FMO	CYP92	CYP71			UGT85
C. orientalis	IGL		FMO*	CYP82*	CYP736*			UGT85*
W. religiosa				CYP92				
S. dulcis				CYP92				

The first two oxidation steps catalyzed by BX2 and BX3 in *Z. mays* are catalyzed by a bifunctional flavin-dependent monooxygenase (FMO) in the eudicot species *Aphelandra squarrosa*, *Lamium galeobdolon,* and *C. orientalis*, providing an example of convergent evolution. The subsequent oxidation steps of the BXD pathway also evolved independently in monocots and eudicots, as well as among the different families of eudicots [34]. While BX4 from *Z. mays* belongs to the CYP71 family, BX4s from *A. squarrosa*, *L. galeobdolon*, *Wrightia religiosa*, and *Scoparia dulcis* belong to the CYP92 family, and BX4 from *C. orientalis* belongs to the CYP82 family. Phylogenetic analysis revealed that although they belong to the CYP92 family, the BX4s from *A. squarrosa* and *L. galeobdolon* belong to family-specific clades and are more closely related to other CYP92 enzymes of their respective families than to each other [34]. In *Z. mays* and *L. galeobdolon,* the BX5 enzymes belong to the CYP71 family, while

A. squarrosa BX5 belongs to the CYP81 family and *C. orientalis* BX5 belongs to the CYP736 family. For the glucosylation catalyzed by BX8, UGTs were identified in *Z. mays* (UGT710), *L. galeobdolon* (UGT-85), and *C. orientalis* (UGT-85) [34,36]. As in maize, BX6 and BX7 from *A. squarrosa,* the only eudicot sp. where BXD biosynthesis proceeds until DIMBOA-Glc, are a 2-ODD and an OMT, respectively. However, low amino acid sequence identity of 29-33% and phylogenetic analysis showed that the UGTs, 2-ODDs, and the OMTs cluster in well-separated clades between monocots and eudicots [34].

Biological activity of BXDs

BXDs are stored as stable glucosides in the vacuole of plant cells. Upon cell damage and disruption of the cell compartmentalization, BXD glucosides can be cleaved into the respective aglucons by glucosidases present in plastids, cell walls, and the cytoplasm [37]. BXD aglucones released from roots can act as allelopathic compounds, mediating plant-plant interactions in the rhizosphere. Notably, BXD levels are especially high in the roots of young maize, wheat, and rye when plants are competing for light and nutrients in early stages of growth [29,38]. BXD aglucones released into the soil are also involved in iron uptake by forming complexes with Fe^{III} ions, which can be transported into the roots. Moreover, they contribute to metal detoxification by chelating e.g. toxic Al^{III} ions (**Figure 2A**) [24,39].

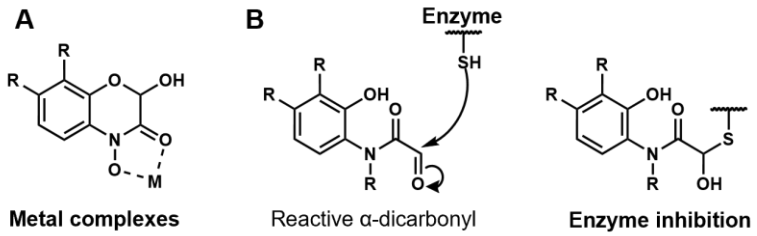

Figure 2: Biological activity of BXDs including chelating of metals like Fe^{III} or Al^{III} (A) and enzyme inhibition through the electrophilic α-dicarbonyl after ring opening of the heterocycle (B). [24,42]

Furthermore, BXD aglucones have antifeedant and antimicrobial activity. DIMBOA shows the highest toxicity against chewing herbivores, followed by DIBOA and DIM_2BOA. HDMBOA and HBOA had no effect [23]. BXD activity against herbivores and pathogens is linked to the fact that aglucones can undergo oxo-cyclo tautomerism due to their hemiacetalic nature, which favors the ring opening of the heterocycle. Upon ring opening, the resulting electrophile α-dicarbonyl moiety can react with amino acid residues, inhibiting enzymatic activity (**Figure 2B**) [23,24]. In the alkaline gut environment of insects, the spontaneous ring opening favors a ring contraction reaction that leads to the formation of the degradation products benzoxazolinones (1,3-benzoxazol-2-one) (**Figure 1C**) [23], which are believed to contribute to the biological activity observed [37]. Furthermore, BXDs exhibit antimicrobial activity against *Staphylococcus aureus, Escherichia coli,* and *Candica albicans* [40,41] as well as antifungal activity against *Aspergillus flavus* [24].

Halogenated Benzoxazinoids

Attempts to alter the biological activity of BXDs through structural modification have been made for decades. Bravo and Lazo (1993) compared the antimicrobial activity of DIMBOA, DIBOA, and DIBOA analogs with different C7 substituents against *S. aureus*, *E. coli,* and *C. albicans*. With increasing electron-donating capacity of the aromatic substituent, an increase in antimicrobial activity was reported [40]. Many broad-spectrum fungicides widely used today share the structure of the NHCO group with the open ring conformation of BXD aglucons, making BXDs a promising scaffold for further development of biologically active compounds [42,43].

The potential herbicidal activity of different chlorinated and fluorinated DIBOA-Glc derivatives has been investigated in the model system *Oryza sativa* (rice as crop) – *Echinochloa crus-galli* (barnyard grass as weed). Compared to the commercial herbicide propanil, 6-Cl-DIBOA and 6-F-DIBOA caused the same level of phototoxicity at 15 to 30 times lower concentrations. The difluorinated 6,7-diF-DIBOA showed good phytotoxicity as well, while 8-Cl-DIBOA additionally resulted in a good selectivity for the system tested [42].

Naturally occurring halogenated BXD were sporadically reported to be isolated from *Z. mays* as 5-Cl-DIMBOA-Glc and 4-Cl-6,7-dimethoxy-2-benzoxazolinone [44,45], and *Achantus ilicifolius* (7-Cl-DIBOA-Glc) [46], although no halogenase for BXDs is known. However, an extensive and careful screening of a large number of maize varieties in our laboratory did not reveal any halogenated BXDs, indicating that the halogenated BXDs reported in the literature are most likely artifacts of the isolation procedures (T. Köllner, personal communication).

Enzymatic halogenation reactions

Halogenation is a common structural modification of natural products, especially in many modern agrochemicals [47]. In general, the introduction of one or multiple halides can affect the chemical and biological properties of a molecule by changing its lipophilicity, size, dipole moment, and polarity. Furthermore, halides are stable leaving groups that can increase the reactivity of the molecule [42]. Unfortunately, synthetic halogenation of natural products often lacks regioselectivity and requires harsh and environmentally toxic reaction conditions, problems that do not occur when using enzymes [48]. In nature, the majority of organohalogen compounds are chlorinated and brominated, with bromination mainly observed for marine natural products [49]. Fluorinated natural products are rare, because of the high electronegativity and high enthalpic costs for the desolvation required for the addition of the halogen to occur [50]. Due to the low abundance of iodide, little organoiodides are observed [51].

For enzymatic halogenation, multiple strategies exist depending on the halide oxidation state. Besides hypohalites used by Heme-Fe, vanadium-dependent, and flavin-dependent halogenases, some halogenases catalyze reactions via a radical mechanism or in rare cases, transfer the halide without changing the oxidation state[14,51].

Heme-Fe and vanadium-dependent haloperoxidases

The first halogenating enzyme discovered belonged to the group of heme-Fe haloperoxidases. Heme-Fe and vanadium-dependent haloperoxidases both use hydrogen peroxide to oxidize chloride or bromide to form an electrophilic hypohalite species (**Figure 3A+B**). Hypohalous acid can then halogenate substrates bound in close proximity to the Fe or vanadium center or can be released to halogenate more distantly bound substrates [14]. Vanadium-dependent haloperoxidases were mainly found to catalyze bromination reactions in marine seaweeds, while chlorination reactions were reported for terrestrial fungi and bacteria [52,53]. These halogenases show a wide range of catalyzed reactions and high stability, however, they suffer from low regioselectivity due to the release of free hypohalous acid [14].

Flavin-dependent tryptophan halogenases

A well investigated group of halogenases are bacterial flavin-dependent tryptophan halogenases. The first halogenase characterized from this group was PrnA from *Pseudomonas fluorescens*, chlorinating the C7 position of L-tryptophan [54]. Further regioselective tryptophan halogenases were discovered, introducing one halide at position C5-7 of tryptophan [55,56]. For regiospecific halogenation of tryptophan, the flavin-dependent halogenases have two different modules linked by a tunnel. At the flavin-binding module, molecular oxygen forms a flavin hydroperoxide with the cofactor, which is subsequently attacked by chloride that is released as hypochlorous acid (HOCl). This electrophilic chlorinating compound is guided to the substrate-binding module, where a chloramine is formed with a lysin residue conserved in all flavin-dependent halogenases. Formation of the chloramine enables the subsequent regioselective halogenation of tryptophan (**Figure 3C**) [57]. Notably, this enzyme can also catalyze bromination at high bromide concentrations. Two halogenases from *Kutzneria* sp. 744 were reported to produce dichlorinated tryptophan through sequential chlorination [58]. For co-factor regeneration, partner reductases were identified in the respective biosynthetic gene clusters and co-expression with flavin-reductases was reported to enhance yields of chlorinated tryptophan for some flavin-dependent halogenases when expressed in a heterologous host [14,59]. Due to low stability *in vitro*, these halogenases produce low yields of halogenated product [1]. However, flavin-dependent tryptophan halogenases have successfully been expressed in plant heterologous hosts and have even been combined with subsequent biosynthetic pathways [21,59]. Furthermore,

evaluation of the substrate scope is subject to many studies and showed halogenation of tryptophan derivatives [60–62]. Rational design has been used to further expand the substrate range from tryptophan derivatives to tryptamine and industrially important indole derivatives or to generate mutants with an increased thermal stability [16,61,63–67].

Halide methyltransferases

Many fluorinated compounds are used as pharmaceuticals or agrochemicals [47,68]. However, due to the high electronegativity and consequently large hydrate envelope of fluorine, enzymatic fluorination reactions are rare [69]. Halide methyltransferases (HMT) are found in plants, algae, bacteria, and fungi [70]. Peng and colleagues (2021) used a halide methyltransferase from *Burkholderia xenovorans* to produce fluorinated *S*-adenosylmethionine (F-SAM) from S-adenosyl-homocysteine (SAH) and fluoromethyliodide (F-MeI) (**Figure 3D**). Subsequently, F-SAM functioned as donor of a fluorinated methyl group for SAM-dependent methyltransferases. *In vitro* experiments showed that a variety of methyltransferases accept the modified F-SAM and were able to transfer it to substrates with C-, N-, and O-nucleophiles [71].

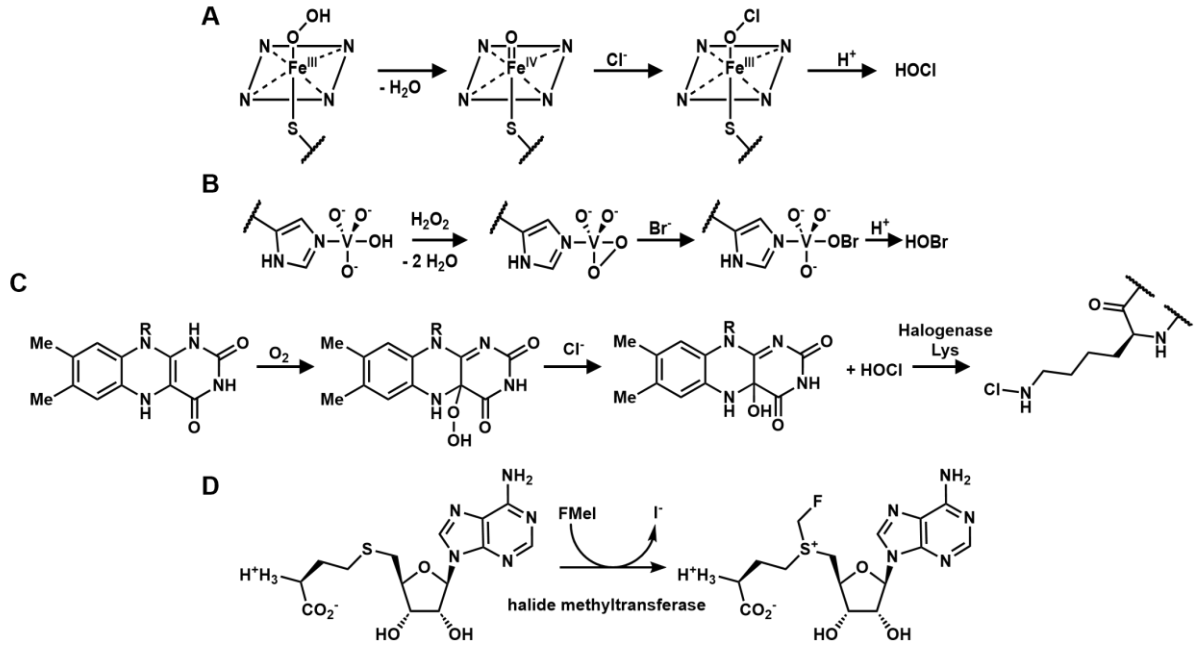

Figure 3: Mechanisms for the generation of an electrophilic hypohalite by Heme-dependent haloperoxidases (A), vanadium-dependent haloperoxidases (B), and flavin-dependent halogenases (C). Generation of F-SAM as fluoro-methyl donor through the halide methyltransferase from *B. xenovorans* (D). [14,72]

Nicotiana benthamiana as production platform for modified natural products

Due to the challenges associated with total chemical synthesis of modified natural products and the environmental impacts of those approaches, biosynthesis of natural products in heterologous hosts such as bakers yeast (*Saccharomices cerevisiae*), *Aspergillus* species, and the plant *Nicotiana benthamiana* are becoming more valuable strategies [72,73].

For the production of plant-derived natural products, a well-established platform is the *Solanaceae* species *N. benthamiana* [74]. The amenability of this species to genetic manipulation through *Agrobacterium tumefaciens*-mediated stable and transient transformation makes *N. benthamiana* a valuable chassis for the discovery of biosynthetic pathways and a biotechnological platform for high-level protein expression and product formation [75], providing a cost-efficient and scalable manufacturing system [76].

Reconstitution of pathways of interest through the expression of pathway genes in *N. benthamiana* can be combined with natural product modification *in planta*. Following a precursor-directed biosynthesis approach, substrate analogs can be infiltrated into transformed leaves [20]. Alternatively, enzymes modifying the starting substrate or pathway intermediates can be expressed together with pathway enzymes to produce *de novo* modified natural products [59].

Aim of the project

BXD derivatives have been shown to be promising candidates for the application as antimicrobial or herbicidal agents. The aim of this work was to produce halogenated BXDs in *N. benthamiana,* taking advantage of the numerous convergently evolved BX enzymes and independently evolved biosynthetic strategies utilized in nature. The first step was to investigate whether the availability of several evolutionarily unrelated enzymes, each with a slightly different catalytic site, could overcome bottlenecks in the acceptance of halogenated BXD analogs as substrates. Subsequently, two different biosynthetic strategies should be compared: precursor-directed biosynthesis (PDB) and enzymatic *de novo* biosynthesis of halogenated BXDs by expression of different types of bacterial halogenases. Finally, the yield of halogenated BXDs produced in *N. benthamiana* should be increased by optimizing the compartmentalization of bacterial halogenases.

Material & Methods

Chemicals

All chemicals used in this study were purchased molecular biology grade or higher from Sigma Aldrich, ThermoFisher Scientific or Tokyo Chemical Industry (TCI) unless otherwise stated.

Plant material

N. benthamiana plants were grown in a greenhouse on a 16-h-light/8-h-dark photoperiod at 20-24°C and 60% relative humidity.

Cloning

Table S 1 shows genes that were already provided in vectors [33,34,59]. Sequences of genes of interest that were not available already cloned in vectors were retrieved from NCBI and amplified by polymerase chain reaction (PCR) from complementary deoxyribonucleic acid (cDNA) of *Arabidopsis thaliana* or from synthetic genes ordered from Twist Bioscience, using the corresponding primers shown in Table S 3 (Table S 2: list of genes). Primers were designed in Genious Prime. If chloroplastic localization peptides (CLP) had to be removed primers were designed after the CLP predicted using DeepLoc - 2.0 (DTU Health Technology). PCR reactions of 25 µL were assembled, using Platinum SuperFi II 2x PCR Master Mix (ThermoFisher Scientific). Reactions were incubated in a thermocycler (Eppendorf) for 30 s at 98°C for denaturation, followed by 35 cycles of 10 s at 98°C, 30 s annealing at 55°C, and 1 min elongation at 72°C, and a final elongation step at 72°C for 5 min. To confirm the correct amplicon size, PCR products were run on a 1% agarose gel (in 0.5x TAE buffer (tris(hydroxymethyl)aminomethane (Tris) base, acetic acid, ethylenediaminetetraacetic acid (EDTA)) from TAE Buffer, 50x, VWR Life Science) with 1x Purple loading dye (New England Biolabs) alongside 5 µL of GeneRuler 1kb Plus (ThermoFisher Scientific) as DNA ladder. correct PCR products were purified using DNA Clean and Concentrator-5 (Zymo). DNA concentrations were measured spectrophotometrically on a NanoPhotometer® N60/N50 (Implen). 8 to 10 µg of p3Ω1 plasmid were digested in a total volume of 50 µL with 1x CutSmart Buffer with 1 µL BsaI-HF restriction enzyme (New England Biolabs). InFusion HD Cloning (Takara Bio) reactions of 10 µL total volume were prepared using 40-60 ng of digested p3Ω1 vector, and 40-60 ng of purified PCR product and were incubated at 50°C for 15 min. If more than one PCR fragment had to be assembled, as per genes fused with a CLP sequence, primers for each amplicon were designed with matching overhangs and the same protocol for InFusion cloning was also used.

Golden braid assembly

Golden braid assembly [77] was used for the assembly of the CLP sequence and the tryptophanase *TnaA*. T4 ligase (New England Biolab) and BsaI HF (New England Biolab) were added to 50 ng of empty α-vector backbone, 20 ng of plasmids carrying the p and terminator, respectively, and 20 ng of purified PCR products. Samples were incubated at 50 cycles of 37°C for 5 min and 16°C for 5 min followed by 10 min at 37°C, 10 min at 65°C, and 80°C for 10 min. *E. coli* TOP10 were transformed by heat shock, plated on LB-agar plates with kanamycin and incubated at 37°C for 24 h.

E. coli transformation

E. coli TOP10 was transformed using the heat shock method. Chemically competent cells were thawed on ice and incubated with 8 µL of InFusion or golden braid reaction for 15 min on ice. The cells were incubated at 42°C for 50 s, followed by 2 min on ice. For cell recovery, 500 µL of lysogeny broth (LB)-medium without antibiotics were added and cells were incubated at 37°C and 200 rpm for 1 h. Cells were plated on LB-agar plates with appropriate selection and incubated at 37°C overnight. For golden braid assembly, colonies were screened using white/blue screening: each plate was supplemented with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM and X-Gal (ThermoFisher Scientific) to the final concentration of 200 µg/mL.

Plasmid isolation

The presence of transformed colonies was checked by colony PCR (98°C, 5 min; 30x (98°C, 30 s; 55°C, 30 s; 72°C, 90 s); 72°C, 5 min) using single colonies as template for 15 µL reactions with Phire Hot Start II DNA polymerase (ThermoFischer Scientific) and sequencing primers for the respective vector. Two to five colonies were inoculated in 3-4 mL LB media with respective antibiotics and grown at 37°C and 200 rpm for 4-5 h. Plasmids were isolated from colonies showing inserts of the correct size in colony PCR using Wizard Plus SV Minipreps DNA Purification System kit (Promega).

Sanger sequencing

The correctness of cloned sequences was confirmed by Sanger sequencing at Genewiz, (Azenta; Prague, CZE) with p3Ω1 sequencing primers (Table S 3).

Agrobacterium tumefaciens-mediated transient transformation of *N. benthamiana*

Agrobacterium tumefaciens transformation

Agrobacterium tumefaciens GV3101 (Goldbio) cells were transformed by electroporation. Competent cells were thawed on ice and incubated with 20-700 ng of sequence-confirmed plasmid carrying the gene of interest for 20-30 min on ice. The mixture was transferred into an electroporation cuvette (path, 1 mm; BioRad) and, after electroporation (1.8 Volt (V); capacity = 250 micro-Farad; electrical resistance R = 200 Ohm), 1 mL of LB media without antibiotics was added and cells were transferred into an Eppendorf reaction tube. For cell recovery, transformed cells were incubated at 28°C, 200 rpm for 3 h followed by plating on LB-agar plates with appropriate selection at 28°C for 48 h. Rifampicin (10 µg/mL) and gentamicin (25 µg/mL) were always added when growing *A. tumefaciens* as they constitute constitutive resistances. Depending on the plasmid backbone, spectinomycin (200 µg/mL; p3Ω1) or kanamycin (50 µg/mL; α, pCambia, pTurbo) were used. Colonies were screened for successful transformation through colony PCR (98°C, 5 min; 30x (98°C, 30 s; 55°C, 30 s; 72°C, 90 s); 72°C, 5 min) using Phire Hot Start II DNA polymerase (ThermoFisher Scientific) in 15 µL reaction and the sequencing primers corresponding to the plasmid backbone (Table S *3*). The size of the products was controlled by gel electrophoresis (1% agarose gels) using GeneReuler 1kb Plus (ThermoFisher Scientific). Single colonies were grown overnight in LB media with rifampicin (10 µg/mL), gentamicin (25 µg/mL) and appropriate vector selection at 28°C, 200 rpm. Glycerol stocks were prepared by adding one volume of culture with one volume of 50% glycerol solution and stored at -70°C for later use. When needed, glycerol stocks were plated on LB-agar plates with antibiotics and grown overnight at 28°C.

N. benthamiana transient transformation

For transient transformation of *N. benthamiana*, liquid LB media with rifampicin (10 µg/mL), gentamicin (25 µg/mL) and appropriate vector selection was inoculated from *A. tumefaciens* plates and grown at 28°C, 200 rpm shaking overnight. Cells were pelleted by centrifugation at 4000 rpm and 14°C for 10 min, resuspended in infiltration media (10 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 10 millimolar (mM) magnesium chloride ($MgCl_2$), pH 5.7; 100 µM acetosyringone freshly added) to an optical density measured at 600 nm (OD_{600}) of 1.0, and incubated at 28°C and 200 rpm for 2-2.5 h. Isovolumes of the solutions were mixed to obtain the desired co-infiltration mixtures and 1-mL needless syringes were used to infiltrate 3 to 4 week-old *N. benthamiana* leaves from the abaxial side. A construct encoding the silencing repressor P19 was added to each co-infiltration mixture to increase expression. When testing exogenous substrates *in planta*, 2 to 3 days after transformation,

substrates (indole derivatives 30 mM in methanol (MeOH)) were added to infiltration media without acetosyringone at a final concentration of 500 µM and infiltrated using a needleless 1-mL syringe. Salt solutions (potassium chloride (KCl), potassium bromide (KBr)) were infiltrated at a final concentration of 500 µM or 1 mM. After transformation, plants were maintained in the lab on a 16-h-light/8-h-dark photoperiod at 20-24°C.

Metabolic extraction

Two to three days after substrate infiltration and five days after transient transformation, 100 mg fresh weight (FW) of transformed *N. benthamiana* leaves were harvested using a leaf borer. The plant material was collected in 2 mL Eppendorf tubes and snap-frozen in liquid nitrogen. The tissue was ground to a fine powder by adding two 3 mm Tungsten Carbide Beads using a TissueLyser II (Quiagen) at 25 s^{-1} for 1 min. For liquid chromatography-quadrupole time-of-flight mass spectrometry (LC-qTOF) analysis, metabolites were extracted by adding 500 µL MeOH (LC-MS grade), followed by shaking at room temperature for 15 min. The samples were centrifuged at 14000 rpm for 10-15 min and the supernatant was filtered with a polytetrafluorethylene filter (PTFE; 0.2 µm; ThermoFisher Scientific). Samples were stored at -20°C until analysis. For analysis using gas chromatography-mass spectrometry (GC-MS), samples were collected identically, but extracted with 200 µL hexane followed by sonication for 10 min. Samples were centrifuged at 14000 rpm for 10 min and 100 µL of supernatant were stored at -20°C in glass vials until analysis.

Leaf disk assay

A similar workflow as described above was performed for the preparation of leaf disk assays according to Hong et al.[78] The transient transformation of single *N. benthamiana* leaves using *A. tumefaciens* co-infiltration mixtures was performed identically. Three to four days after, leaf disks were collected using a leaf borer (9 mm diameter) and placed in a 48-well plate. Unlike the substrate infiltration *in planta*, the substrates were added to 50 mM HEPES buffer at pH 7.5 to the final concentrations described above. A volume of 250 µL of buffer containing the substrate was added in each well. Leaf disks were stabbed twice using a pipet tip to prevent floating on the liquid surface. The substrate was infiltrated by placing the open plate in a desiccator connected to a vacuum line and by applying three to five cycles of 30 s vacuum and 10 s release. The 48-well plates were sealed with Parafilm and placed in a growth chamber at 28°C for 18 h. Three leaf disks were combined, snap frozen and pulverized. For LC-qTOF analysis, the ground tissue and 100 µL of the HEPES buffer in which the leaf disks were incubated, were extracted with 300 µL MeOH. For GC-MS analysis, 200 µL hexane were used and samples prepared as described above.

Metabolic analysis using liquid chromatography-quadrupole time-of-flight mass spectrometry (LC-qTOF)

Liquid chromatography-quadrupole time-of-flight mass spectrometry (LC-qTOF) analysis was performed using a Thermo Scientific UltiMate 3000 ultra-high performance liquid chromatography (UHPLC) system coupled to an Impact II UHR-Q-ToF (Ultra-High Resolution Quadrupole Time-of-Flight) mass spectrometer (Bruker Daltonics). The samples were separated by reverse-phase liquid chromatography using a Phenomenex Kinetex XB-C18 column (100 x 2.1 mm, 2.6 µm; 100 Å) at 35°C. Water with 0.1% formic acid (A) and acetonitrile (B) were used as mobile phases with a flow rate of 0.3 mL/min. The injection volume was set to 2 µL. The chromatographic separation was performed starting at 5% B for 1 min and increasing from 5% to 50% of solvent B in 7 min, 100% B for 2.5 min, and 5% B for 2.5 min. Mass spectrometry acquisition was performed in negative electron spray ionization (ESI) mode, except for samples containing HDMBOA-Glc, HDM_2BOA-Glc, and DIM_2BOA-Glc for which positive ESI mode was performed additionally. Non-halogenated authentic BXD standards were prepared as 20 µM solutions. For the negative ionization mode, a capillary voltage of 3500 V and an end plate offset of 500 V were used. A nebulizer pressure of 2.0 bar was used, with nitrogen at 250°C and a flow of 10 L/min as the drying gas. For positive ionization mode, a capillary voltage of 3500 V and an end plate offset of 500 V were used. A nebulizer pressure of 2.8 bar was used, with nitrogen at 280°C and a flow of 8 L/min as the drying gas. In both modes, acquisition was performed at 12 hertz (Hz) in the mass to charge ratio (*m/z*) range from *m/z* 100 to 1000, with data dependent MS2 and an active exclusion window of 1 min. For collision energy, the stepping option model (from 20 to 50 electronvolt (eV)) was used. For calibration of the MS spectra, each sample run started with the injection of a sodium formate-isopropanol calibration solution in the source at 0.18 mL/h.

Metabolic analysis using gas chromatography-MS/MS (GC-MS)

For the detection of indole and its derivatives, gas chromatography-mass spectrometry analysis (GC-MS) was conducted using an Agilent 8890 Series gas chromatograph coupled to an Agilent 5977B quadrupole mass selective detector (Agilent Technologies; injection temperature, 230°C; interface temperature, 280°C; quadrupole temperature, 150°C; source temperature, 230°C; electron energy, 70eV). Samples were separated using a ZB5 column (Phenomenex; 30 m x 0.25 mm x 0.25 µm) and He as a carrier gas. The sample was injected at an initial oven temperature of 45°C and held at this temperature for 2 min. The temperature was increased to 240°C with a gradient of 15°C/min. Then increased to 280°C with a gradient of 30°C/min and held at this temperature for 3 min. Plant samples extracted in hexane were analyzed with a gas flow of 1 mL/min and splitless injection of 1 µL. Compounds were

identified by comparison of retention times and mass spectra to those measured of authentic standards.

Halogenated BXDs biosynthesis scale-up

Vacuum infiltration of *N. benthamiana*

For vacuum infiltration of whole *N. benthamiana* plants, two 5 mL precultures of *A. tumefaciens* were inoculated from glycerol stocks and grown at 28°C and 200 rpm overnight. 100 mL LB media were inoculated and grown overnight in the same conditions. Cells were pelleted by centrifugation at 4000 rpm and 14°C for 20 min. Each strain was resuspended in infiltration media to OD_{600} = 1.0 and incubated at 28°C and 200 rpm for 1.5 h. Isovolumes of the solutions were mixed to obtain a total of 1 L co-infiltration mixtures. A beaker was placed inside the vacuum desiccator and filled with the co-infiltration mixture. Plants, equipped with a plastic disk that prevented soil spillage, were submerged inverted into the co-infiltration mixtures. Two cycles of vacuum for 30 s and 10 s of release were applied. Leaves that did not get fully infiltrated were further infiltrated using a needleless 1 mL syringe. Two to three days after, the substrate was infiltrated using the same method.

Preparative scale metabolic extraction

For the BXD extraction, whole leaves were collected, wrapped in aluminum foil and snap frozen in liquid nitrogen. An electric hammermill was used to pulverize the collected tissue. The plant tissue was extracted in 100% MeOH (3-fold excess *v/w*) and incubated at room temperature stirring for 30 min. The extract was filtered twice through two layers of Miracloth (Merck-Millipore). The extract was reduced using a rotary evaporator and resuspended in 5 mL 60% MeOH. Extracts were prepared by solid phase extraction (SPE), using an Oasis PRiME HLB 6 cc Vac Cartridge (Wicom) which was equilibrated first with 100% MeOH and then 60% MeOH. After the sample was passed on the resin, the cartridge was washed using 2 mL 60% MeOH and the flowthrough stored at 4°C for further use.

Preparative high performance liquid chromatography (HPLC)

Preparative high performance liquid chromatography (HPLC) was performed using Agilent 1290 Infinity II HPLC system coupled with a DAAD UV-detector. The samples were separated by reverse-phase liquid chromatography using a Phenomenex Kinetex XB-C18 column (250 x 10 mm, 5 µm; 100 Å) at 35°C. Water with 0.1% formic acid (A) and acetonitrile (B) were used as mobile phases with a flow rate of 6 mL/min. The injection volume was 600 µL. For purification of 8-Cl-HBOA-Glc, the chromatographic separation was performed starting at 5% B and increasing from 5% to 35% of solvent B in 24 min, followed by 100% B for 6 min, and 5% B for 5 min. For purification of 8-F-HBOA-Glc, the chromatographic separation was

performed starting at 10% B and increasing from 10% to 25% of solvent B in 13 min, followed by 100% B for 5 min, and 5% B for 5 min. Collected fractions were chosen based on photometric analysis at 280 nm and 254 nm and subsequent analysis by LC-qTOF. Pure fractions containing the wanted halogenated HBOA-Glc analogues were combined, and solvents were evaporated using a rotatory-evaporator.

NMR analysis

NMR measurements were carried out on a 700 MHz Bruker Avance III HD spectrometer and a 500 MHz Bruker Advance III HD spectrometer (Bruker Biospin GmbH, Rheinstetten, Germany), equipped with a TCI cryoprobe using standard pulse sequences as implemented in Bruker Topspin ver. 3.6.1. (Bruker Biospin GmbH, Rheinstetten, Germany). Chemical shifts were referenced to the residual solvent signals of MeOH-d_3 (δ_H 3.31/δ_C 49.0) respectively. NMR analysis was kindly carried out by Yoko Nakamura (NMR service group, MPI-CE).

Purification of His-tagged proteins expressed in *N. benthamiana*

A C-terminal 6xHis-tag was cloned in frame to *RebH-3LRS*, *KtzR*, *KtzQ*, *RebF*, and *TnaA* gene sequences through PCR, using a normal forward primer and a reverse primer including the additional *6His* sequence (Table S *3*). InFusion reaction, transformation of *E. coli* and of *A. tumefaciens* with sequence confirmed p3Ω1 plasmids were performed as described above. 3-week-old *N. benthamiana* were transiently transformed with *P19* and one bacterial gene each. As negative control, *N. benthamiana* was transiently transformed with *A. tumefaciens* strains carrying *P19* and an empty vector (p3Ω1 EV). Four days after infiltration, 8 leaves were collected, the central vascular tissue of the leaves was removed, and the remaining tissue was snap-frozen in liquid nitrogen. The tissue was ground with a spoonful of polyvinylpolypyrrolidone (PVPP) using a pre-cooled mortar and pestle. The powdered tissue was extracted with 5 mL ice-cold extraction buffer (100 mM Tris-HCl pH 7.4, 10% glycerol and 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM β-mercaptoethanol, 2% *w/v* PVPP added directly before use) and incubated at 4°C for 30 min shaking. The extracted homogenate was centrifuged at 4000 rpm for 10 min at 4°C and the supernatant filtered through Miracloth (Merck-Millipore). The filtered extract was centrifuged at 35000 x g for 30 min at 4°C. The samples were concentrated using 10 kilo Dalton (kDa) or 30 kDa Amicon Ultra-15 centrifugal filter units (Sigma Aldrich) at 4000 rpm and 4°C for 10 min. Same volumes of concentrated protein and 2x Laemmli Sample Buffer (Sigma Aldrich) were combined and 10 µL of samples were used for SDS-PAGE. Proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with 6 µL of Color protein Standard Broad range protein ladder (New England Biolabs) on a 12% denaturing polyacrylamide gel (Novex™ 12%, Tris-Glycine Plus WedgeWell™ Gel, ThermoFisher Scientific) at 225 V for 35 min. For protein

staining using Quick Coomassie Stain (Calibre Scientific UK), the gel was rinsed with distilled water (dH_2O), stained for 20 min with Coomassie solution shaking at room temperature and destained in dH_2O shaking overnight. For Semi-dry transfer Western-blot, the gel was washed with dH_2O, equilibrated in 1x Power Blotter 1-step transfer buffer (ThermoFisher Scienti) for 2-3 min. A pre-assembled Power Blotter Select Transfer Stack with a polyvinylidene difluoride membrane (PDVF; ThermoFisher Scientific) and the Power Blotter System (ThermoFisher Scientific) were used for the protein transfer at 25 V, 1.3 ampere (A) for 7 min. After the transfer, the membrane was washed in TBST buffer (6.05 g/L Tris, 8.76 g/L sodium chloride (NaCl), pH7.5, 1 mL/L Tween), and blocked in TBST buffer + 5% *w/v* milk powder shaking at 4°C overnight. The blocking solution was removed and the membrane was incubated in TBST buffer + 3% *w/v* skimmed milk + Mouse Anti Penta Histidine Tag:Horseradish Peroxidase antibody (Biorad) shaking at room temperature for 2 h. The antibody solution was removed, the membrane rinsed with water and washed with TBST buffer 6 times for 2-3 min. Prior to imaging, the membrane was incubated in Clarity Western ECL Susbtrate (Biorad). The image of Coomassi-stained polyacrylamide gel and PDVF membrane were incubated using the iBright 1500 system (ThermoFisher Scientific).

In silico docking PyMOL

Three dimensional (3D) models of BX4s from *Z. mays*, *A. squarrosa*, *L. galeobdolon*, *C. orientalis, W. Religiosa*, and *S. dulcis* were generated using AlphaFold 2.0. The heme prosthetic group was docked into *Zm*BX4 through alignment with the 3D model of the experimentally crystallized allene oxidase synthase cytochrome P450 74A from *A. thaliana,* in PyMOL. Insertion into all other BX4s was obtained through alignment of the 3D structure with *Zm*BX4. Potential docking poses of the substrate 3HI2O into the active site of the BX4s were generated using Webina 1.0.5. Protein images were generated in Chimera X.

Results

Precursor-directed biosynthesis of halogenated benzoxazinoids in *N. benthamiana*

Characterization of the substrate specificities of BX enzymes

The bottleneck of producing halogenated natural products enzymatically is often that the enzyme catalyzing the reaction on the canonical substrate is does not turn over the halogenated form of the substrate effectively. Since BXD convergently evolved among different plant species, several BX enzymes are available that catalyze the same reaction but belong to different protein families or enzyme classes. Thus, each of the BX enzymes available was characterized for substrate specificity to identify the ones that best accept each halogenated BXD precursor. For this purpose, BX enzymes or combinations of BX enzymes were transiently expressed in *N. benthamiana* and the leaves were subsequently infiltrated with various halogenated indoles. BX enzymes were selected so that all indole derivatives were converted as efficiently as possible by the enzyme combination. In all experiments, the silencing suppressor protein P19 as well as the UDP-glucosyltransferases *ZmBx8* and *LgBx8* were co-expressed with the tested genes. P19 enhances transient expression in *N. benthamiana*, while *Zm*BX8 and *Lg*BX8 form glucosylated BXDs that can be detected with a higher sensitivity compared to the respective aglucons. Peak areas of halogenated BXD glucosides per 100 mg fresh weight (FW) were calculated and the results were internally normalized to the peak area obtained with the BX enzymes from *Zea mays* (reference BXD pathway) for each substrate. The identity of the newly biosynthesized halogenated compounds was confirmed by comparison of the MS/MS spectra to available standards of the non-halogenated species.

Halogenated 3HI2O-Glc – BX2 and BX3

The oxidation of indole to I2O and the subsequent oxidation of I2O to 3HI2O are respectively catalyzed by the two P450 monooxygenases BX2 and BX3 in maize and other grasses, while the eudicot species *A. squarrosa*, *L. galeobdolon*, and *C. orientalis* each possess a bifunctional FMO that catalyzes both oxidations. To determine the substrate specificity of the BX enzymes, maize *ZmBx2* was expressed individually and in combination with *ZmBx3*, while the FMOs (*AsFMO*, *LgFMO*, *CoFMO*) were expressed together, in batch. All enzymes and enzyme combinations were tested with different chlorinated and fluorinated indole derivatives. To correct the yield for the endogenous BX2/3 activity of *N. benthamiana*, the peak area obtained for halogenated 3HI2O-Glc was subtracted from the yield of halogenated product obtained when expressing *ZmBx8*, *LgBx8*, and *P19* only (negative control).

The resulting relative enzyme activities are shown in **Table 2**. The combination of *Zm*BX2 and *Zm*BX3 resulted in high yields of halogenated 3HI2O-Glc for all substrates. Expressing *ZmBx2* individually showed high yields for 4-Cl- and 7-Cl-indole, while the FMO batch gave high yields of halogenated 3HI2O-Glc only with 7-Cl-I2O as a substrate and medium yields with 5-F-, 6-F-, and 7-F-indole. Since the co-expression of *ZmBx2* and *ZmBx3* resulted in high yields for all substrates tested, they were used for further screening of downstream enzymes of the BXD pathway.

Table 2: Relative yields of halogenated 3HI2O-Glc when using 4-Cl-, 7-Cl-, 5-Cl-, 4-F-, 7-F-, 5-F-, and 6-F-indole, as well as 7-Cl-I2O as substrates for *Zm*BX2 individually, *Zm*BX2 and *Zm*BX3 in combination, or *As*FMO, *Lg*FMO, and *Co*FMO combined normalized to the yield of *Zm*BX2+3. The endogenous substrate conversion by *N. benthamiana* was subtracted from peak areas per 100 mg fresh weight, which were normalized to the yield obtained when expressing *ZmBx2* and *ZmBx3* for each substrate. Results are given in percent.

BX2+3 / FMO	Indole			I2O	Indole			
	4-Cl	5-Cl	7-Cl	7-Cl	4-F	5-F	6-F	7-F
***Zm*BX2**	112	30	81	34	28	29	28	35
***Zm*BX2+3**	100	100	100	100	100	100	100	100
As* + *Lg* + *Co	44	18	34	112	35	51	61	49

Halogenated HBOA-Glc – BX4

The conversion of 3HI2O to HBOA is catalyzed by BX4, which expands the heterocycle by introducing an oxygen atom. This leads to a change in nomenclature since, according to IUPAC rules, the newly introduced oxygen has a higher priority than the nitrogen atom of the indole ring. Therefore, the oxygen atom is the new starting point for numbering the position of potential substituents (**Figure 5**).

Table 3: Relative yields of halogenated HBOA-Glc when using 4-Cl-, 7-Cl-, 5-Cl-, 4-F-, 7-F-, 5-F-, and 6-F-indole, as well as 7-Cl-I2O as substrates for *Zm*BX2/3 and *Zm*BX4, *As*BX4, *Lg*BX4, *Co*BX4, *Wr*BX4, and *Sd*BX4 normalized to the yield of *Zm*BX4. The peak areas per 100 mg fresh weight were normalized to the yield obtained with *Zm*BX4 for each substrate. Results are given in percent. N.d.: not detected.

BX4	Indole			I2O	Indole			
	4-Cl	5-Cl	7-Cl	7-Cl	4-F	5-F	6-F	7-F
***Zm*BX4**	100	100	100	100	100	100	100	100
***As*BX4**	697	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
***Lg*BX4**	N.d.	N.d.	291	34	83	17	4	14
***Co*BX4**	4327	47	N.d.	N.d.	176	134	77	24
***Wr*BX4**	406	99	110	142	361	42	6	32
***Sd*BX4**	1768	15	20	N.d.	N.d.	15	11	N.d.

To test the ability of the different BX4s to accept halogenated 3HI2O as substrate for the production of chlorinated or fluorinated HBOA-Glc, *ZmBx2* and *ZmBx3* were co-expressed with each *Bx4* individually. The results are shown in **Table 3**. BX4 of *A. squarrosa* and *S. dulcis* led to little or no halogenated HBOA-Glc, except for *As*BX4 when infiltrating 4-Cl-indole as a substrate. Similarly, expression of *LgBx4* generally resulted in low yields, except for 7-Cl-indole, producing the highest amount of chlorinated HBOA-Glc for this substrate. Since expressing *ZmBx4*, *LgBx4*, and *CoBx4* resulted in high yields for multiple substrates and covered all chlorinated and fluorinated substrates tested, these three BX4s were chosen for further screenings.

Halogenated DIBOA-Glc – BX5

The next step in BXD biosynthesis is the *N*-hydroxylation of HBOA to DIBOA, which is catalyzed by BX5. *ZmBx2*, *ZmBx3*, *ZmBx4*, *LgBx4*, and *CoBx4* were co-expressed with each *Bx5* individually and the relative yields were calculated (**Table 4**). *Zm*BX5 and *Lg*BX5 were able to produce chlorinated DIBOA-Glc when 4-Cl-indole was used as substrate, while *Lg*BX5 showed activity also on 5-Cl-indole. All substrates substituted at position 7 (7-Cl-indole, 7-Cl-I2O, and 7-F-indole) did not yield halogenated DIBOA-Glc. When using 4-F- and 6-F-indole as substrates, all enzymes except for *As*BX5 gave relatively high yields. Only *Lg*BX5 led to the incorporation of 5-F-indole to form fluorinated DIBOA-Glc. *Zm*BX5 and *Lg*BX5 were therefore chosen for further screening.

Table 4: Relative yields of DIBOA-Glc when using 4-Cl-, 7-Cl-, 5-Cl-, 4-F-, 7-F-, 5-F-, and 6-F-indole, as well as 7-Cl-I2O as substrates for *Zm*BX2/3, *Zm*BX4, *Lg*BX4, *Co*BX4, and *Zm*BX5, *As*BX5, *Lg*BX5, and *Co*BX5 normalized to the yield of *Zm*BX5. The peak areas per 100 mg fresh weight were normalized to the yield obtained with ZmBX5 for each substrate. Results are given in percent. N.d.: not detected.

BX5	Indole			I2O	Indole			
	4-Cl	5-Cl	7-Cl	7-Cl	4-F	5-F	6-F	7-F
***Zm*BX5**	100	N.d.	N.d.	N.d.	100	100	100	N.d.
***As*BX5**	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.	N.d.
***Lg*BX5**	123	100	N.d.	N.d.	84	225	71	N.d.
***Co*BX5**	N.d.	N.d.	N.d.	N.d.	101	57	73	N.d.

DIMBOA-Glc – BX6 and BX7

DIBOA-Glc can be further hydroxylated on C7 by BX6 to form TRIBOA-Glc, followed by *O*-methylation, catalyzed by BX7 at the same position, to form DIMBOA-Glc. Since TRIMBOA-Glc is not stable, *Bx6* and *Bx7* were co-expressed to obtain the more stable DIMBOA-Glc. Potential combinations of *Zm*BX6, *Zm*BX7, *As*BX6, and *As*BX7 were tested for the

incorporation of halogenated indole derivatives. 4-Cl-indole was the only chlorinated substrate that led to the formation of chlorinated DIMBOA-Glc and was incorporated by all combinations except for *As*BX6 and *Zm*BX7. No formation of chlorinated DIMBOA-Glc was observed when starting from 7-Cl-indole and 7-Cl-I2O. The substrate 4-F-indole resulted in similar yields for all combinations tested. In addition, 6-F-indole was also incorporated when co-expressing *AsBx6* and *ZmBx7*. Therefore, *Zm*BX6, *As*BX6, and *Zm*BX7 were chosen as candidates for these steps of the biosynthesis of BXDs up to DIMBOA-Glc *in planta*.

Starting from indole halogenated at position 5, no production of DIMBOA-Glc was expected since the methoxylation catalyzed by BX6 and BX7 would normally occur at the same position. However, a small peak was observed, when co-expressing *AsBx6* and *AsBx7* and using 5-F-indole as substrate. The retention time and MS fragmentation were similar to other F-DIMBOA-Glc derivatives (Figure S 2). Despite the presence of the fluoride atom, *AsBx6* and *AsBx7* may catalyze the methoxylation at a nearby position.

Table 5: Relative yields of DIMBOA-Glc when using 4-Cl-, 7-Cl-, 5-Cl-, 4-F-, 7-F-, 5-F-, and 6-F-indole, as well as 7-Cl-I2O as substrates for *Zm*BX2/3, *Zm*BX4, *Lg*BX4, *Co*BX4, and *Zm*BX5, *Lg*BX5, *Zm*BX6 and *Zm*BX7 and *As*BX6 and *As*BX7 on halogenated indole derivatives normalized to the yield of *Zm*BX6+7. The peak areas per 100 mg fresh weight were normalized to the yield obtained with *Zm*BX6 and *Zm*BX7 for each substrate. Results are given in percent. N.d.: not detected. *: Peak observed, although halogen atom is present at the same position as the methoxylation should occur.

BX 6 + 7	Indole			I2O	Indole			
	4-Cl	5-Cl	7-Cl	7-Cl	4-F	5-F	6-F	7-F
***Zm*BX6-7**	100	N.d.	N.d.	N.d.	100	N.d.	N.d.	N.d.
***As*BX6-7**	74	N.d.	N.d.	N.d.	61	*	N.d.	N.d.
***Zm*BX6+*As*BX7**	77	N.d.	N.d.	N.d.	97	N.d.	N.d.	N.d.
***As*BX6+*Zm*BX7**	N.d.	N.d.	N.d.	N.d.	81	N.d.	100	N.d.

Relative quantification of halogenated BXDs produced by the optimal combination of BX enzymes

N. benthamiana was transiently transformed (three biological replicates) with the optimal BX enzyme combination for all tested substrates previously identified for each step. Halogenated indole derivatives were infiltrated as substrates and BXDs obtained when infiltrating the natural substrate indole served as positive control. Methanolic extracts of the samples were analyzed by LC-qTOF. The identity of the halogenated products was assessed by comparing retention time (RT) to non-halogenated references and by evaluating the mass shift on parent and fragment ions due to the addition of the halogen. Halogenated compounds were previously shown to have a later RT than the non-halogenated derivatives [20]. For chlorinated

compounds, an increase in mass to charge ratio (*m/z*) of 34 (or 34 Da) and for fluorinated compounds an increase of *m/z* 18 was expected. The peak areas of each halogenated BXD produced were compared to that of non-halogenated BXD obtained by infiltrating indole. The silencing suppressor protein P19 as well as the UDP-glucosyltransferases *ZmBx8* and *LgBx8* were co-expressed in every step of BXD biosynthesis.

Halogenated 3HI2O-Glc – BX2 and BX3

*Zm*BX2 and *Zm*BX3 were chosen as the best enzymes for the biosynthesis of halogenated 3HI2O-Glc. The negative control consisted of *N. benthamiana* transiently expressing *P19* together with *ZmBx8* and *LgBx8*, infiltrated with the halogenated indole derivative. The highest yield of chlorinated 3HI2O-Glc was obtained starting from 5-Cl-indole and the yield was comparable to that of 3HI2O-Glc obtained by infiltration with indole. The other chlorinated 3HI2O-Glc derivatives gave about one fifth of the yield observed for the positive control (**Figure 4A**). The yields of all fluorinated 3HI2O-Glc derivatives were in the range of the indole control (**Figure 4B**). The triplicates for 5-Cl-indole and 6-F-indole were performed with a different batch of *N. benthamiana* than the other substrates. Nevertheless, the yields of 3HI2O-Glc obtained when using indole as a positive control were comparable for both batches. Due to endogenous BX2/3 activity in *N. benthamiana*, low but still detectable levels of halogenated 3HI2O-Glc were detected for all substrates when only *P19*, *ZmBx8,* and *LgBx8* were expressed as negative control and infiltrated with the halogenated substrate. Starting from 7-Cl-I2O as substrate could have led to higher yields as the product of BX2 was already provided. The yields observed were slightly higher, although not significantly, than the co-expression of *ZmBx2* and *ZmBx3* supplemented with 7-Cl-indole. On the other hand, infiltration of 7-Cl-I2O also resulted in a higher endogenous *N. benthamiana* BX3 activity in the respective negative control (**Figure 4A**).

Since no reference standard was available for 3HI2O-Glc, the shifts in RT and *m/z* of the produced halogenated BXDs were compared with the RT and *m/z* of 3HI2O-Glc produced by *N. benthamiana* transiently transformed with *ZmBx2-3* and infiltrated with the natural substrate indole. As expected, when chlorinated indole derivatives were infiltrated, peaks for chlorinated 3HI2O-Glc were observed with *m/z* 344.05 (3HI2O-Glc *m/z* 310.09 + Cl^- *m/z* 34). Additionally, a less abundant *m/z* of 346 was observed, which corresponded to the naturally occurring isotopic pattern of chlorine. The typical fragment ions detected for 3HI2O-Glc (*m/z* 147.03; *m/z* 160.04) showed the same increase by *m/z* 34 for chlorinated 3HI2O-Glc (*m/z* 180.99; *m/z* 194.00) (**Figure 4C**). A change in retention time was observed. Non-halogenated 3HI2O-Glc eluted at 3.07 min, while the chlorinated derivatives eluted between 3.75-4.25 min, depending on the position of the chlorine. The extracted ion chromatograms (EIC) of 4-Cl-3HI2O-Glc showed two peaks with slightly different retention times, but identical

MS/MS fragmentation. The same was observed for the EIC of non-halogenated 3HI2O-Glc. For the other chlorinated compounds, peaks that could not be fully chromatographically resolved were observed (Figure S 2A).

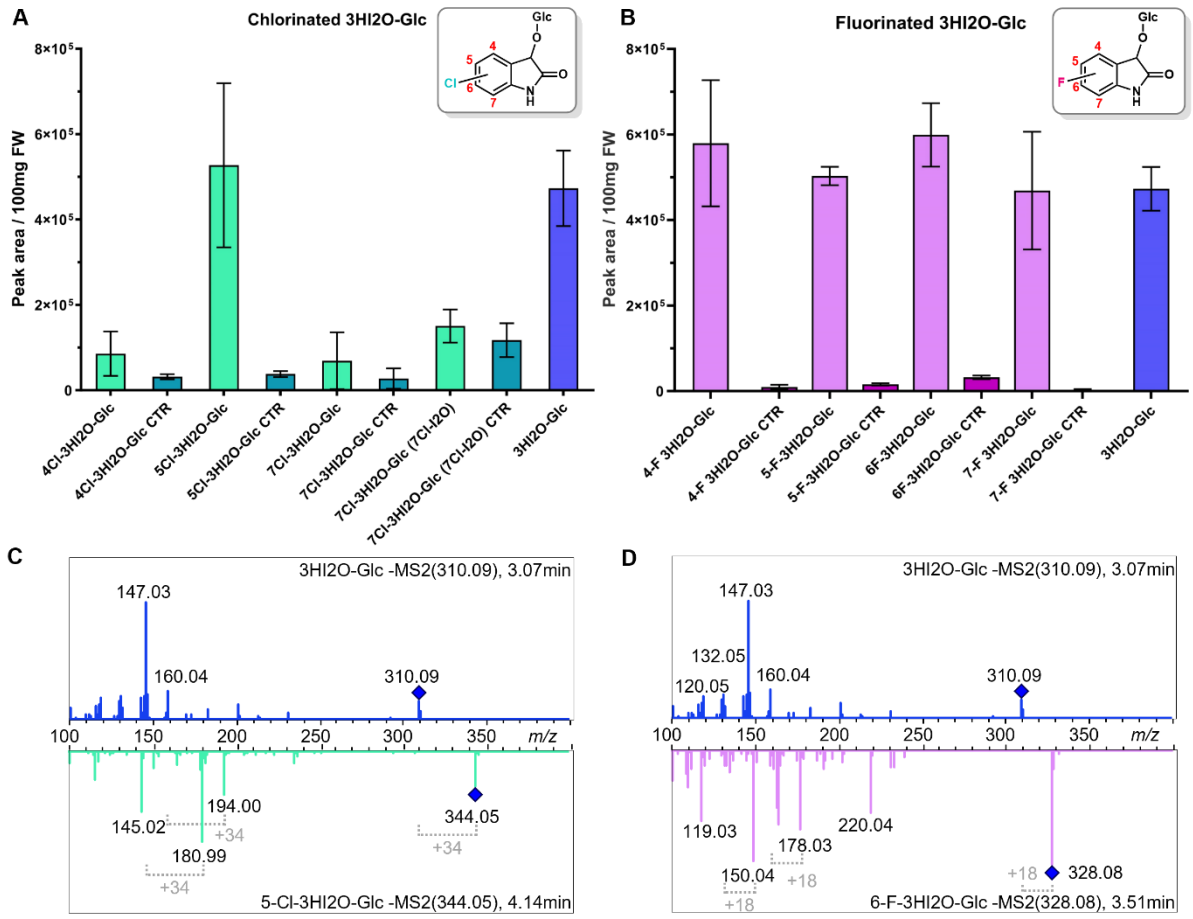

Figure 4: Yields of halogenated or non-halogenated 3HI2O-Glc produced in N. benthamiana starting from halogenated indole derivatives or indole (A+B) and MS/MS fragmentation of 5-Cl-3HI2O-Glc (C) and 6-F-3HI2O-Glc (D) compared to non-halogenated 3HI2O-Glc. Yields represent the peak area per 100 mg fresh weight of plants expressing *ZmBx2*, *ZmBx3*, *ZmBx8*, *LgBx8*, and *P19* using indole, 4-Cl-, 7-Cl-, 5-Cl-, 4-F-, 7-F-, 5-F-, and 6-F-indole, as well as 7-Cl-I2O as substrates. The MS/MS fragmentations shown are representative for chlorinated or fluorinated 3HI2O-Glc derivatives.

Peaks for fluorinated 3HI2O-Glc were detected at *m/z* of 328.08 (3HI2O-Glc *m/z* 310.09 + F^- *m/z* 18). The fragment ions observed for non-halogenated 3HI2O-Glc (*m/z* 147.03; *m/z* 160.04) showed the same increase by *m/z* 18 for the fluorinated derivatives (*m/z* 165.02; *m/z* 178.03) (**Figure 4D**). Fluorinated 3HI2O-Glc eluted between 3.0-3.75 min, after the non-halogenated and before the chlorinated compounds. Two distinct peaks were only seen for 4-F-3HI2O-Glc, while peaks that could not be fully chromatographically resolved were detected for the other derivatives (Figure S 2A).

Halogenated HBOA-Glc – BX4

For relative quantification of halogenated HBOA-Glc, the genes *ZmBx2-4*, *WrBx4*, and *CoBx4* were co-expressed in biological triplicates. The negative control consisted of *N. benthamiana* transiently expressing *ZmBx2*, *ZmBx3*, *ZmBx8*, *LgBx8*, and *P19*. The highest yield of chlorinated HBOA-Glc was obtained for 8-Cl-HBOA-Glc, starting from 4-Cl-indole. 4-Cl-indole was the only chlorinated substrate that yielded similar amounts to the positive control. Low yields were obtained when starting from 5-Cl-indole or 7-Cl-I2O and 7-Cl-indole, resulting in 7-Cl-HBOA-Glc or 5-Cl-HBOA-Glc, respectively. The observed yields of fluorinated HBOA-Glc were similar to the positive control, except for 5-F-HBOA-Glc starting from 7-F-indole. The negative control samples resulted in no production of (non-) halogenated HBOA-Glc (**Figure 5A**).

The results of the LC-qTOF analysis of the halogenated HBOA-Glc derivatives were compared with a reference standard of HBOA-Glc. Peaks for chlorinated HBOA-Glc were observed with

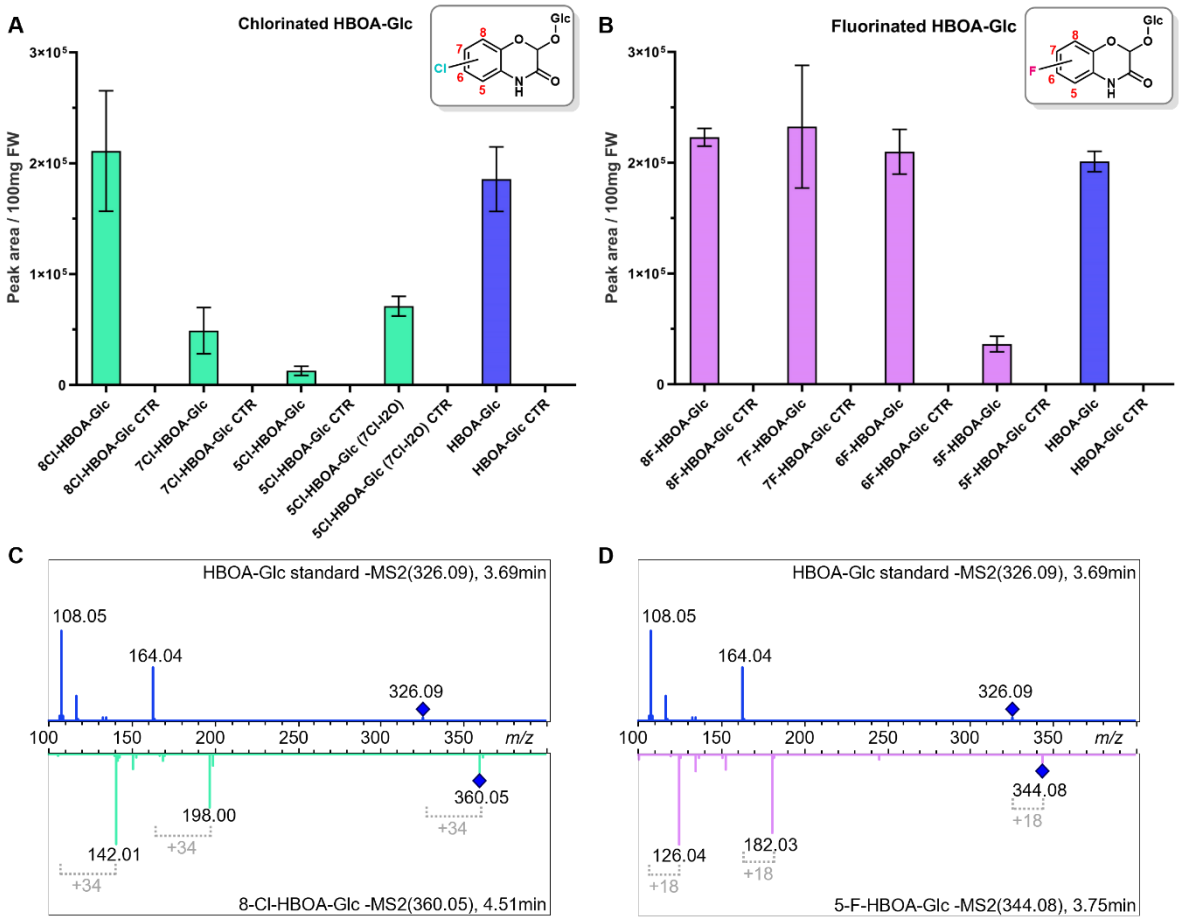

Figure 5: Yields of halogenated or non-halogenated HBOA-Glc produced in *N. benthamiana* starting from halogenated indole derivatives or indole (A+B) and MS/MS fragmentation of 8-Cl-HBOA-Glc (C) and 5-F-HBOA-Glc (D) compared to a HBOA-Glc reference. Yields represent the peak area per 100 mg fresh weight of plants expressing *ZmBx2-4*, *LgBx4*, *CoBx4*, *ZmBx8*, *LgBx8*, and *P19* using indole, 4-Cl-, 7-Cl-, 5-Cl-, 4-F-, 7-F-, 5-F-, and 6-F-indole, as well as 7-Cl-I2O as substrates. The MS/MS fragmentations shown are representative for chlorinated or fluorinated HBOA-Glc derivatives.

m/z 360.05 (HBOA-Glc *m/z* 326.09 + *m/z* 34). The fragment ions observed for HBOA-Glc (*m/z* 108.05; *m/z* 164.04) showed the same increase of *m/z* 34 for the chlorinated compounds (*m/z* 142.01; *m/z* 198.00). According to the isotopic pattern of chlorine, signals of lower intensity were observed, corresponding to the increased mass of the parental or fragment ions by *m/z* 2 (**Figure 5C**). While non-halogenated HBOA-Glc eluted at 3.7 min, chlorinated HBOA-Glc eluted between 4.3-4.7 min (Figure S 2B).

Fluorinated HBOA-Glc was observed with *m/z* 344.08 (HBOA-Glc *m/z* 326.09 + *m/z* 18). Fragment ions with *m/z* 126.03 and *m/z* 182.04 showed the same increase in mass (**Figure 5D**). Fluorinated HBOA-Glc eluted between 3.7-4.0 min, after the non-halogenated reference and before the chlorinated compound. Two peaks were observed in the EIC of 5-Cl-, 8-F-, and 5-F-HBOA-Glc, which could not be fully chromatographically resolved, but with identical MS/MS fragmentation (Figure S 2B).

Halogenated DIBOA-Glc – BX5

Biosynthesis of halogenated DIBOA-Glc was performed by co-expression of *ZmBx2-5*, *WrBx4*, *CoBx4*, and *LgBx5* in biological triplicates. The negative control consisted of *N. benthamiana* transiently expressing *ZmBx2-4*, *WrBx4*, *CoBx4*, *ZmBx8*, *LgBx8*, and *P19*. 8-Cl- and 8-F-DIBOA-Glc derived from indole halogenated at position 4 resulted in similar yields as the indole-derived positive control. No DIBOA-Glc chlorinated or fluorinated at position 5 was observed when starting from 7-Cl-indole and 7-Cl-I2O or 7-F-indole. 7-Cl-DIBOA-Glc yielded less product than the positive control and yields of 7-F- and 6-F-DIBOA-Glc were approximately double the control. No DIBOA-Glc derivatives were detected in the negative controls (**Figure 6A+B**).

The identity of the compounds was confirmed by comparing the LC-qTOF data of the halogenated compounds with a DIBOA-Glc reference standard. Chlorinated DIBOA-Glc was detected with *m/z* 376.05 (DIBOA-Glc *m/z* 342.09 + *m/z* 34). The fragment ions of DIBOA-Glc (*m/z* 134.03; *m/z* 162.03) were observed with an increase of *m/z* 34 for chlorinated DIBOA-Glc (*m/z* 167.99; *m/z* 196.00) (**Figure 6C**). The RT of the DIBOA-Glc reference at 3.7 min was shifted to 4.5-4.7 min for the chlorinated derivatives (Figure S 2C).

Fluorinated DIBOA-Glc was detected with *m/z* 360.07 and fragment ions with *m/z* 152.01 and *m/z* 180.01, which showed the expected shift in mass by *m/z* 18 (**Figure 6D**). The fluorinated derivatives eluted at 4.0 min, after the non-halogenated reference and before the chlorinated DIBOA-Glc derivatives (Figure S 2C).

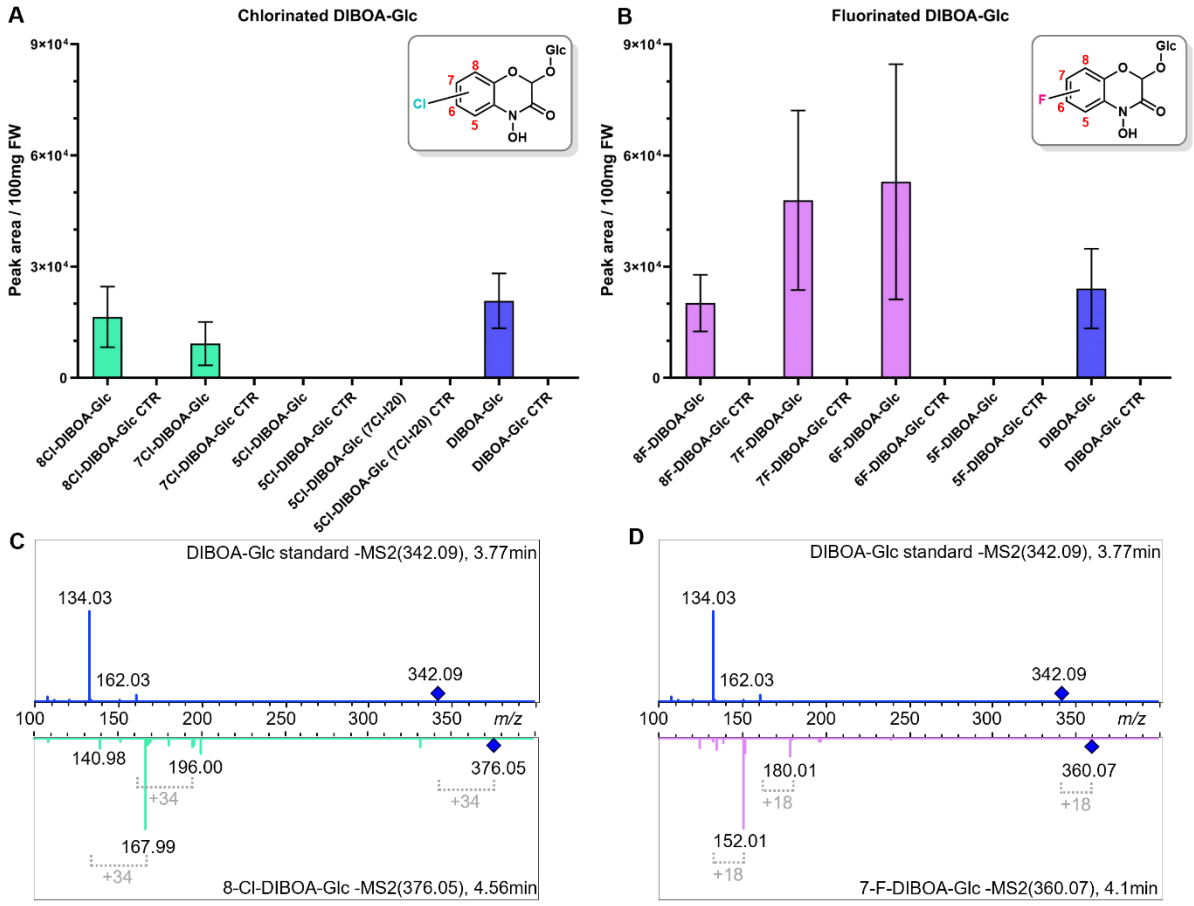

Figure 6: Yields of halogenated or non-halogenated DIBOA-Glc produced in *N. benthamiana* starting from halogenated indole derivatives or indole (A+B) and MS/MS fragmentation of 8-Cl-DIBOA-Glc (C) and 7-F-DIBOA-Glc (D) compared to a DIBOA-Glc reference. Yields represent the peak area per 100 mg fresh weight of plants expressing *ZmBx2-5*, *CoBx4*, *LgBx5*, *ZmBx8*, *LgBx8*, and *P19* using indole, 4-Cl-, 7-Cl-, 5-Cl-, 4-F-, 7-F-, 5-F-, and 6-F-indole, as well as 7-Cl-I2O as substrates. The MS/MS fragmentations shown are representative for chlorinated or fluorinated DIBOA-Glc derivatives.

Halogenated DIMBOA-Glc – BX6 and BX7

For the biosynthesis of DIMBOA-Glc, the genes for *ZmBx1-7*, *WrBx4*, *CoBx4*, *LgBx5*, and *AsBx6* were expressed in *N. benthamiana* in biological triplicates for each substrate. The negative control consisted of *N. benthamiana* transiently expressing *ZmBx2-5*, *WrBx4*, *CoBx4*, *LgBx5*, *ZmBx8*, *LgBx8*, and *P19*. 8-Cl-DIMBOA-Glc, starting from 4-Cl-indole, was the only chlorinated derivative produced and yielded approximately half of the positive control. With exception of 7-F-indole, all fluorinated substrates were incorporated and resulted in the formation of fluorinated DIMBOA-Glc. 8-F-DIMBOA-Glc gave the highest yield, similar to the control starting from indole. The yield of 6-F-DIMBOA-Glc was one third of the control (**Figure 7A+B**). As mentioned before, the conversion of DIBOA-Glc to DIMBOA-Glc is expected to be hindered when the starting substrate indole carries a halogen atom at position 7. Therefore, no production of DIMBOA-Glc halogenated at position 5 was expected. However, small

amounts of F-DIMBOA-Glc were detected when starting from 7-F-indole. This was also observed in the screening of *AsBx6* and *AsBx7* (**Table 5**). It is possible that despite the fluoride atom, BX6 and BX7 may be able to hydroxylate 7-F-DIBOA-Glc at a nearby position. The position of the methoxylation could only be elucidated by isolation of the compound and analysis by NMR (nuclear magnetic resonance). No formation of DIMBOA-Glc or the halogenated derivatives was observed in the negative controls.

Chlorinated DIMBOA-Glc was detected with *m/z* 406.05 (DIMBOA-Glc *m/z* 372.09 + *m/z* 34) and fragment ions with *m/z* 182.97 and *m/z* 198.00 including the expected shift in mass (**Figure 7C**). The reference of DIMBOA-Glc had an RT of 4.2 min, while 8-Cl-DIMBOA-Glc eluted at 4.6 min (Figure S 2D).

Fluorinated DIMBOA-Glc was detected with *m/z* 390.08, showing the expected increase by *m/z* 18. The typically observed fragment ions of DIMBOA-Glc (*m/z* 149.01; *m/z* 164.03)

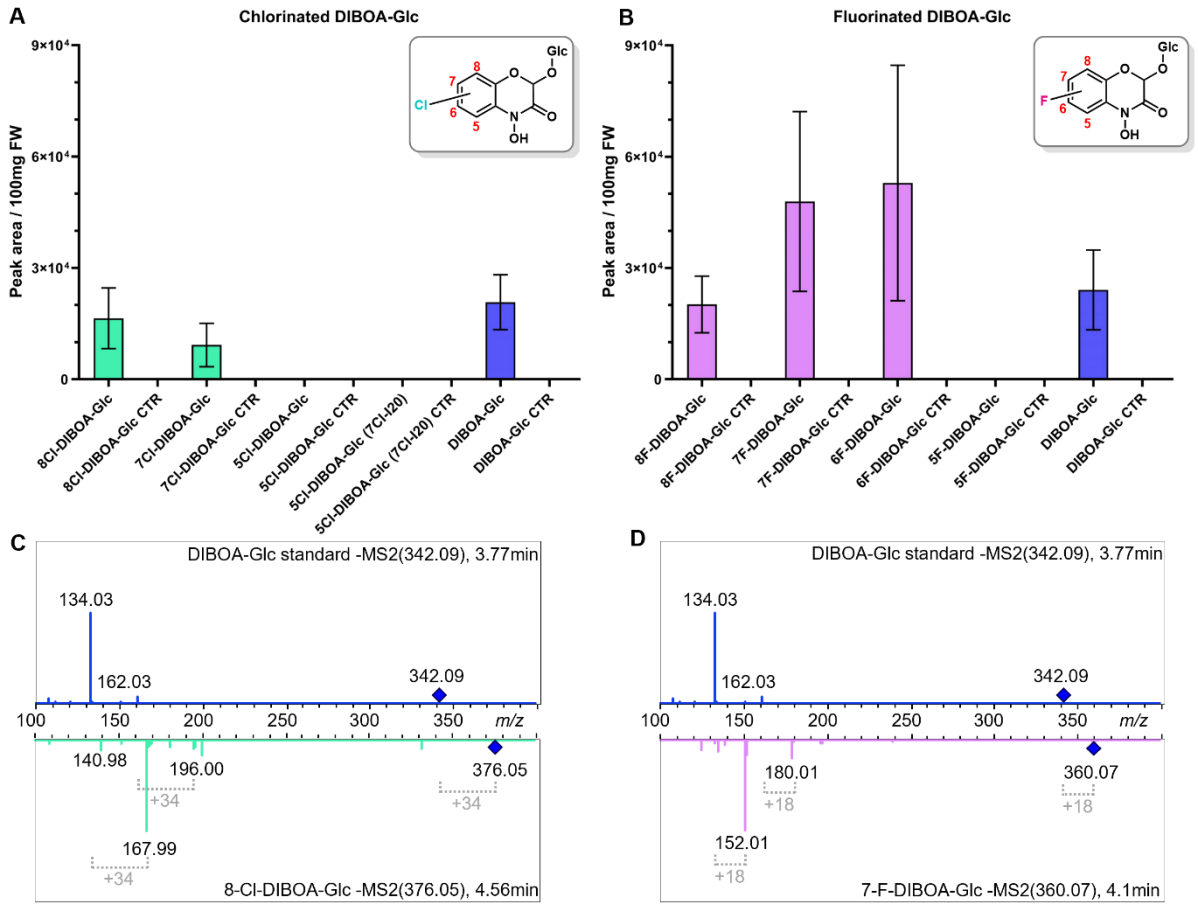

Figure 7: Yields of halogenated or non-halogenated DIMBOA-Glc produced in *N. benthamiana* starting from halogenated indole derivatives or indole (A+B) and MS/MS fragmentation of 8-Cl-DIMBOA-Glc (C) and 8-F-DIMBOA-Glc (D) compared to a DIMBOA-Glc reference. Yields represent the peak area per 100 mg fresh weight of plants expressing *ZmBx2-8*, *LgBx4*, *CoBx4*, *LgBx5*, *LgBx8*, and *P19* using indole, 4-Cl-, 7-Cl-, 5-Cl-, 4-F-, 7-F-, 5-F-, and 6-F-indole, as well as 7-Cl-I2O as substrates. The MS/MS fragmentations shown are representative for chlorinated or fluorinated DIBOA-Glc derivatives.

contained the same shift in mass by *m/z* 18 for the fluorinated derivatives (*m/z* 167.00; *m/z* 182.02) (**Figure 7D**). The RT of fluorinated DIMBOA-Glc was 4.3 min and slightly longer than the RT of the non-halogenated reference. A similar RT of 4.4 min and the same fragment ions were observed for the fluorinated DIMBOA-Glc starting from 5-F-indole that was not expected to be produced (Figure S 2E).

Incorporation of halogenated substrates

To compare the incorporation of the different halogenated substrates in the biosynthesis of halogenated BXDs in *N. benthamiana*, the average peak areas of the halogenated products were divided by the yields of the respective non-halogenated BXD (**Table 6**).

Starting from 4-Cl-indole, high relative yields were obtained throughout the entire pathway with 47% of final product 8-Cl-DIMBOA-Glc compared to DIMBOA-Glc starting from indole. The bottleneck of 8-Cl-DIMBOA-Glc biosynthesis consisted of 4-Cl-2HI2O formation. The yield of 4-Cl-3HI2O-Glc was only 15%. 5-Cl-indole was incorporated up to 7-Cl-DIBOA-Glc with a relative yield of 38%. Both 7-Cl-indole derivatives gave low relative yields and were not incorporated further than 5-Cl-HBOA-Glc. However, infiltrating 7-Cl-I2O as a substrate led to a higher yield than when starting form 7-Cl-indole. For halogenated BXDs starting from fluorinated indole, an overall high relative incorporation was observed compared to the natural substrate indole. Similar to 4-Cl-indole, 4-F-indole was best incorporated by BX enzymes. A high relative yield of 93% was obtained for the final product 8-F-DIMBOA-Glc. For the substrates 5-F- and 6-F-indole, the relative yields up to F-DIBOA-Glc were similar to or even higher than DIBOA-Glc starting from indole. For the following conversion to F-DIMBOA-Glc, almost no yield (6%) was obtained for 7-F-DIMBOA-Glc, while 30% was obtained for 6-F-DIMBOA-Glc. Analogous to the 7-Cl-indole derivates, 7-F-indole only led to the biosynthesis up to 5-Cl-HBOA-Glc. The resulting relative yield of 18% was the lowest among fluorinated HBOA-Glc.

Table 6: Relative abundance of halogenated products obtained by infiltrating halogenated indole derivatives compared to the abundance of non-halogenated products when infiltrating indole. Values marked with an asterisk were obtained with another batch of *N. benthamiana* than experiments performed for the other substrates of this step. However, they were normalized to a control infiltrating indole performed with the same batch of plants.

	Indole			I2O	Indole			
	4-Cl	5-Cl	7-Cl	7-Cl	4-F	5-F	6-F	7-F
3HI2O-Glc	15	142*	12	26	101	88	161*	82
HBOA-Glc	114	27	7	38	111	116	104	18
DIBOA-Glc	68	38	-	-	84	199	220	-
DIMBOA-Glc	47	-	-	-	93	6	30	-

Identification of compounds through NMR

To confirm the biosynthesis of the expected halogenated BXD, 8-Cl- and 8-F-HBOA-Glc were structurally characterized through NMR. 8-Cl- and 8-F-HBOA-Glc were chosen as they accumulated in highest absolute amount, allowing isolation from plant material in quantities sufficient for NMR analysis. Thirty plants were transiently transformed with the optimal gene combination determined for BX1-4 and both, *ZmBx8* and *LgBx8*, for each compound, and 4-Cl- and 4-F-indole were used as substrates. The products were purified by preparative HPLC for subsequent NMR-analysis. Yields of purified 8-Cl-HBOA-Glc and F-HBOA-Glc were 5.25 mg and 0.59 mg, respectively. **Table 7** and **Table 8** show the chemical shift (δ), multiplicity and coupling constant (*J*) obtained by proton (^{1}H) and carbon-13 (^{13}C) NMR. The respective positions of the atoms referred to are marked in **Figure 8C**.

Further analysis was conducted to assign both compounds. Selective total correlation spectroscopy (SELTOCSY) was performed to selectively obtain signals of the glucoside coupling system. Distorsionless enhancement by polarization transfer including the detection of quaternary nuclei (DEPTQ) allowed for the identification of quaternary carbon atoms (position 3, 8, 9, and 10). Phase sensitive heteronuclear single quantum coherence (HSQC) was used to assign signals from ^{1}H NMR to the respective carbon atom the H is bound to and for the identification of the CH_2-group of the glucoside. Using heteronuclear multiple bond correlation (HMBC), it was possible to confirm the positions of quaternary carbons and the sugar, since coupling to the neighboring atom of the binding partner was observed. ^{1}H-^{1}H double-quantum filtered correlation spectroscopy (DQF COSY) was used to assign signals obtained from Hs near each other (Figure S 5, Figure S *8*).

Using the methods listed above, it was possible to assign the structure of 8-Cl-HBOA-Glc to the first compound isolated. The three aromatic protons were detected with a δ_H of 6.87-7.09, two peaks for H-5 and H-7 and three peaks for the central H-6 of that coupling system. No signal of a proton situated at carbon 8 was detected, as this is replaced by the Cl. The Hs bound to the sugar could be assigned by SELTOCSY, although signals were partly overlapped by impurities in the ^{1}H-NMR (**Table 7, Figure 8A , Figure S 5Figure S 6**).

The assignment of 8-F-HBOA-Glc was possible as well. Three aromatic protons were detected with a δ_H of 6.73-6.97. Because of the ^{19}F isotope, coupling of C and H to F was observed as well. Thus, an increased number of signals is seen for the three aromatic Hs and the quaternary Cs. The aromatic protons showed a decrease in J_{HF} with increasing distance to the fluoride at carbon 8 (**Figure 8B, Figure S 7, Figure S 8).** Furthermore, carbon 8 had by far the highest J_{CF}, confirming the C-F bound (**Table 8**).

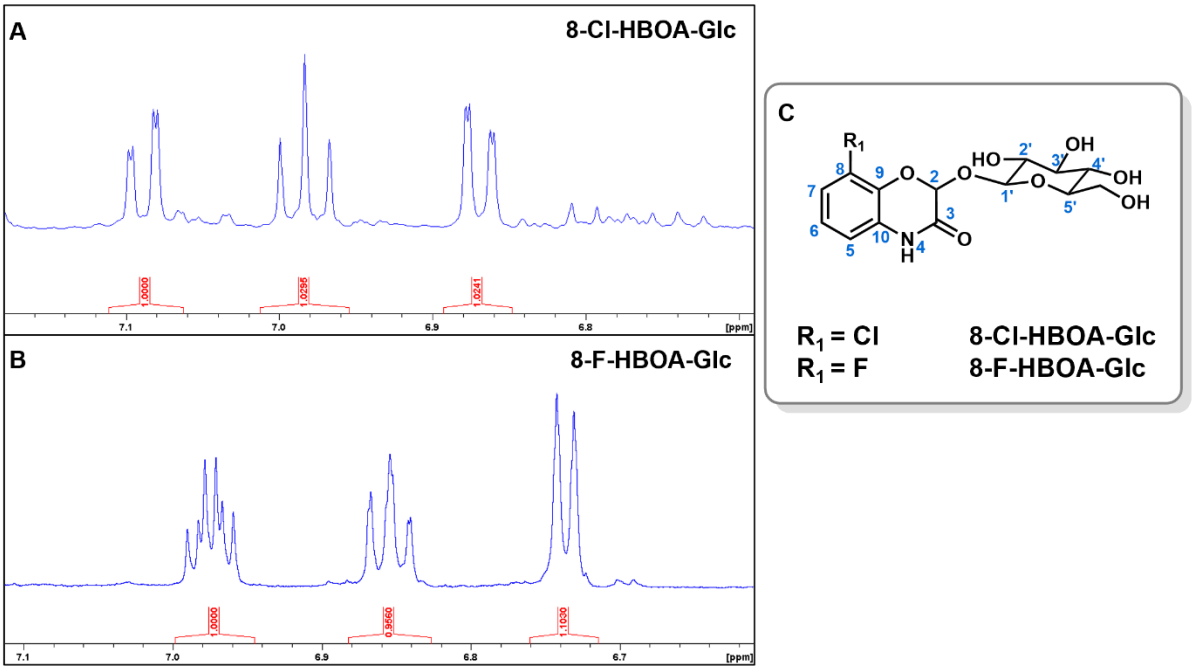

Figure 8: ^{1}H NMR of the aromatic range in MeOH-d_3 of 8-Cl-HBOA-Glc (A) and 8-F-HBOA-Glc (B). Chemical structure of halogenated HBOA-Glc with numbered positions for reference of the NMR results in **Table 7 and Table 8** (C).

Table 7: NMR analysis of 8-Cl-HBOA-Glc isolated from *N. benthamiana* when co-expressing *ZmBx2-4*, *LgBx4*, *CoBx4*, *ZmBx8*, *LgBx8* and *P19* and infiltrating 4-Cl-indole. Position, chemical shift (δ), multiplicity and coupling constant (*J*) are shown for ^{1}H- and ^{13}C-NMR.

Position	δ_H	Multiplicity	J_{HH}	δ_C
2	5.83	s	-	96.5
3	-	-	-	162.5
5	6.87	dd	7.9/1.2	115.4
6	6.98	dd	7.9/7.9	124.5
7	7.09	dd	7.9/1.2	125.5
8	-	-	-	124.0
9	-	-	-	138.3
10	-	-	-	128.8
1'	4.72	d	7.9	103.6
2'	3.20	dd	8.3/7.8	74.9
3'	3.37	m*	-	77.9
4'	3.38	m*	-	70.8
5'	3.32	m	-	78.3
6'a	3.82	dd	11.8/2.4	62.2
6'b	3.73	dd	11.8/4.1	62.2

500 MHz in MeOH-d_3 * overlapped signals *J* unresolved

Table 8: NMR analysis of 8-F-HBOA-Glc isolated from *N. benthamiana* when co-expressing *ZmBx2-4*, *LgBx4*, *CoBx4*, *ZmBx8*, *LgBx8* and *P19* and infiltrating 4-F-indole. Position, chemical shift (δ), multiplicity and coupling constant (*J*) are shown for ^{1}H- and ^{13}C-NMR.

Position	δ_H	Multiplicity	J_{HH}, J_{HF}	δ_C	J_{CF}
2	5.80	s	-	96.4	-
3	-	-	-	162.7	-
5	6.73	ddd	8.1/1.1, 1.0	112.2	3.2
6	6.97	ddd	8.4/8.1, 5.2	124.1	8.7
7	6.85	ddd	8.4/1.1, 8.9	112.0	18.7
8	-	-	-	153.6	245.9
9	-	-	-	130.3	14.9
10	-	-	-	129.5	3.7
1'	4.70	d	8.0	104.0	-
2'	3.19	dd	8.2/8.0	74.8	-
3'	3.36	m*	-	77.9	-
4'	3.35	dd	8.2/8.2	70.9	-
5'	3.33	m*	-	78.2	-
6'a	3.82	dd	11.7/2.0	62.3	-
6'b	3.72	dd	11.7/4.3	62.3	-

700 MHz in MeOH-d_3 * overlapped signals *J* unresolved

Expression of an extended BXD pathway in *N. benthamiana*

To further expand the BXD spectrum produced in *N. benthamiana*, plants were transiently transformed with *ZmBx1-8*, as well as either *Bx10* and *Bx14* or *Bx13* and *Bx14* from *Z. mays*. BX10 and BX14 catalyze the methylation of the *N*-hydroxylation of DIMBOA-Glc, resulting in HDMBOA-Glc. BX13 catalyzes a further hydroxylation of DIMBOA-Glc at the aromatic ring, leading to TRIMBOA-Glc, which is further methylated by BX7 forming DIM_2BOA-Glc. Subsequent methylation of the *N*-hydroxyl group by BX14 leads to HDM_2BOA-Glc (**Figure 1B, Figure 9C**).

HDMBOA-Glc (red) and DIM_2BOA-Glc (yellow) were successfully produced in *N. benthamiana* and confirmed by comparison of retention time and MS/MS fragmentation with available reference standards (blue). No standard was available for HDM_2BOA-Glc. However, the presence of the precursor DIM_2BOA-Glc was confirmed, and a peak of the expected *m/z* 418.13 was detected when co-expressing *ZmBx13* and *ZmBx14*. Furthermore, no peak of *m/z* 418.1 was detected in the control missing *ZmBx13* and *ZmBx14*. All samples were analyzed by LC-qTOF in negative and positive ionization mode. The results shown for HDMBOA-Glc and HDM_2BOA-Glc were obtained by analysis in positive mode. (**Figure 9A+B**).

When measured in negative ionization mode, an adduct formed with formic acid, present in the mobile phase, was observed which increased the mass by *m/z* 45.

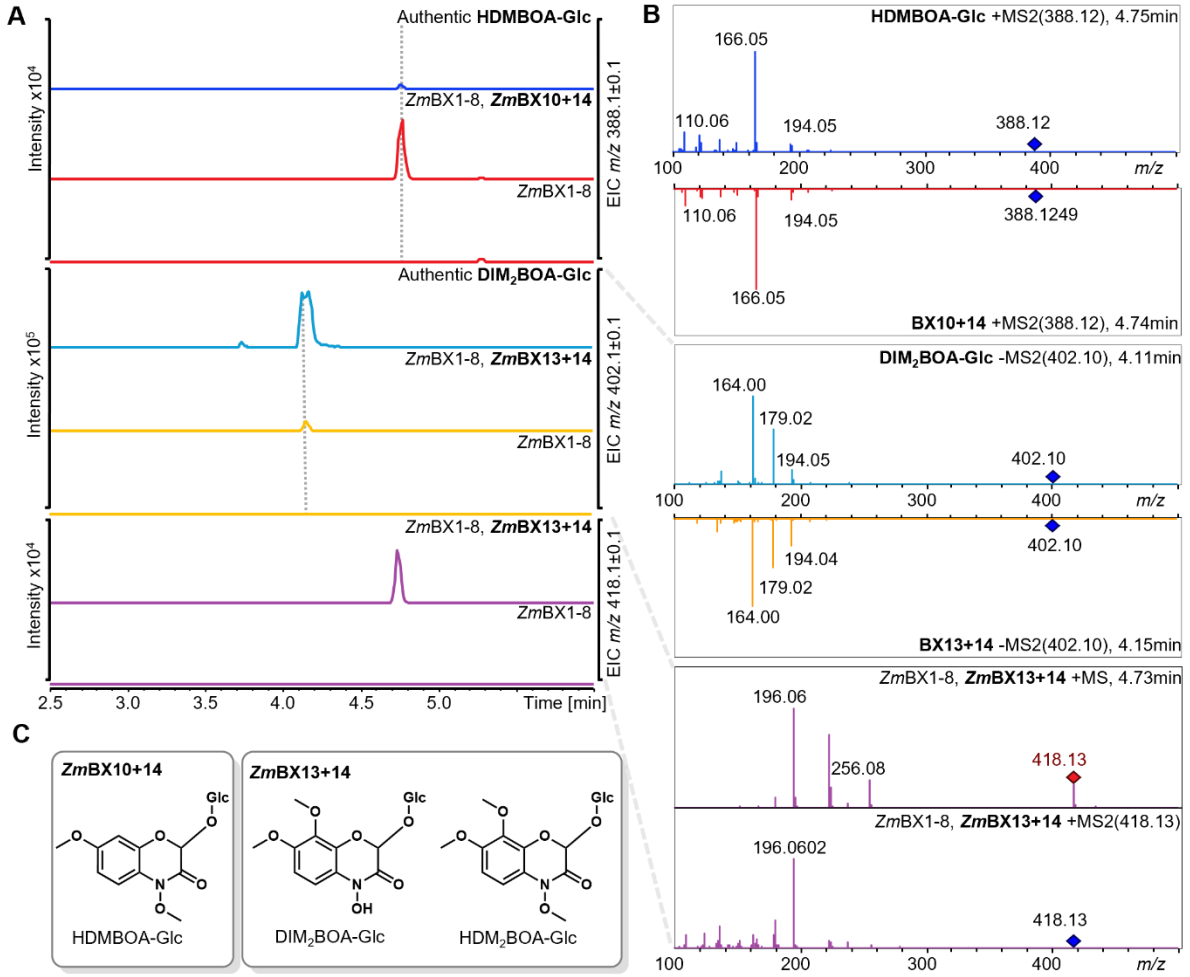

Figure 9: LC chromatograms (A) and MS/MS fragmentation (B) showing the biosynthesis of HDMBOA-Glc, DIM$_2$BOA-Glc, and HDM$_2$BOA-Glc through expression of *ZmBx1-8* and *ZmBx10+Bx14* or *ZmBx13+Bx14* in *N. benthamiana*. The compounds produced by the respective combination of enzymes are shown in **C**.

Establishment of leaf disk assays for the production of BXD in *N. benthamiana*

Transformation and substrate infiltration of whole *N. benthamiana* leaves is labor intensive and involves the use of high amounts of substrates, which can be very expensive. In alternative, it may be possible to scale this process down by assaying single infiltrated leaf disks in 48 well plates, saving time and substrate. A workflow for the biosynthesis and analysis of BXD production in leaf disk assays was established by modifying the protocol used by Hong et al [78]. Plants were transiently transformed with *ZmBx2-5*, *ZmBx8*, and *P19*. After two to three days, leaf disks were collected in a 48-well plate, followed by the addition of HEPES buffer containing the substrate solved in methanol. The substrate was infiltrated into the disks

by the application of vacuum in a desiccator. The leaf disks were extracted after approximately 18 h and analyzed by LC-qTOF (**Figure 10C**). In a first attempt expressing *ZmBx2-5*, the natural substrate indole was infiltrated. As a control, an equivalent volume of MeOH was infiltrated. Peak areas of HBOA-Glc and DIBOA-Glc were compared to peak areas obtained through the canonical method *in planta*. Each experiment was performed in triplicates.

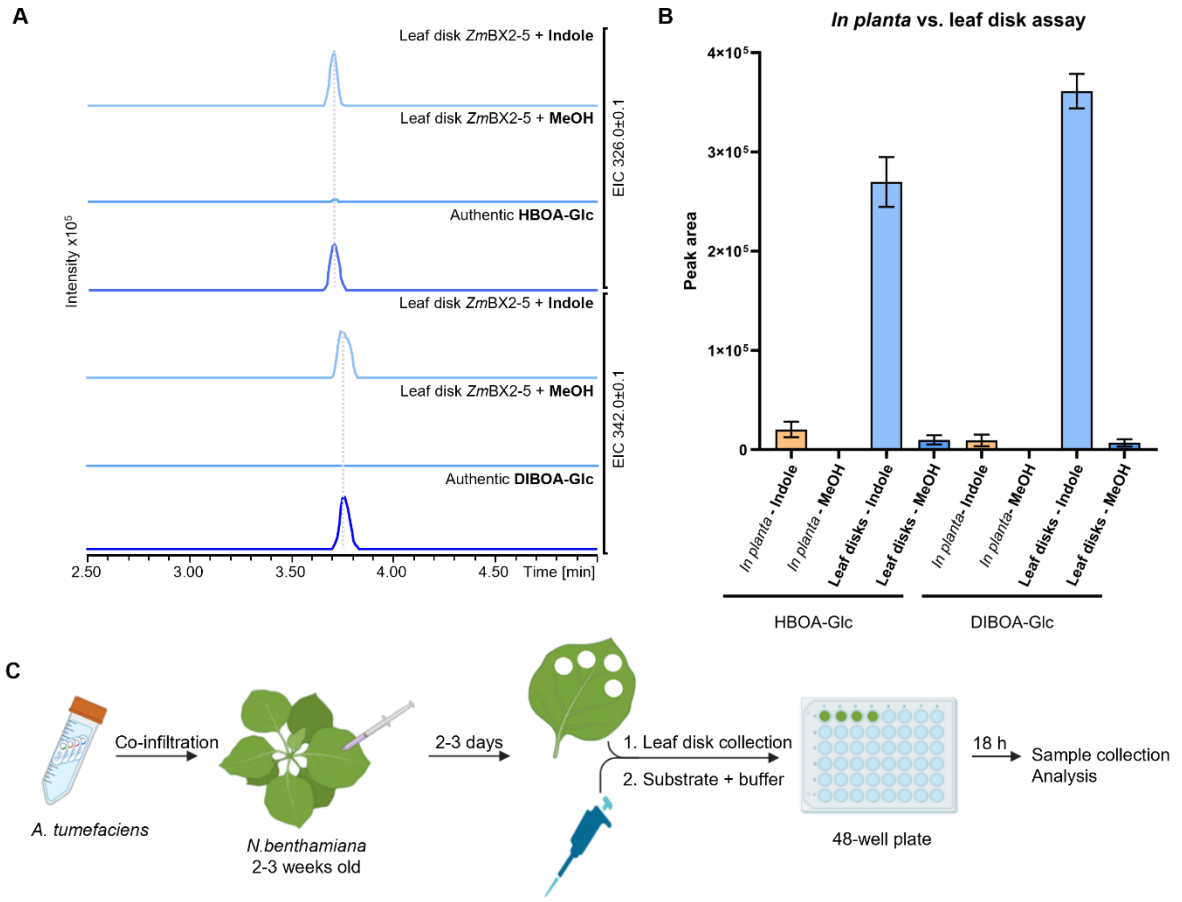

Figure 10: Workflow of preparing leaf disk assays by transiently transforming N. benthamiana leaves and incubating leaf disks in buffer containing the substrate in a 48 well plate (A). LC-chromatograms comparing the results of the production of HBOA-Glc (B) and DIBOA-Glc (C) in leaf disk assay or *in planta*. Experiments were performed in triplicates (average ± SE) and representative chromatograms are shown.

Yields observed for HBOA-Glc and DIBOA-Glc from the leaf disk assay were significantly higher than yields obtained from the substrate infiltration *in planta* (**Figure 10B**). The leaf disk method showed a higher background activity for BXDs. Peaks of low intensity of HBOA-Glc and DIBOA-Glc were detected in the control of the leaf disk assay when infiltrating only MeOH, but not with the *in planta* approach. This can be explained by the release of indole by plants upon injuring during the process of collecting the leaf disks [79]. The identity of HBOA-Glc and DIBOA-Glc was confirmed by comparison of retention time and MS/MS spectra with reference standards (**Figure 10A**).

De novo production of halogenated indole for BXD biosynthesis in *N. benthamiana*

Infiltration of halogenated indole as substrate for the production of halogenated BXDs in *N. benthamiana*, despite being a functional strategy, is a labor-intensive process and requires large amounts of expensive substrates. Therefore, attempts were made to enable the *de novo* biosynthesis of halogenated BXD either by producing the halogenated precursor indole or by halogenating the final product DIMBOA-Glc. For the first approach, several bacterial halogenases were tested to produce chlorinated or brominated indole or indole precursors. Due to the reaction mechanism of halogenases and the strong electronegativity of fluorine (**Figure 3**), no fluorinated indole could have possibly been produced with this approach. Thus, *de novo* production of fluorinated BXD was attempted by using a halide-methyltransferase to produce fluorinated S-adenosyl methionine (F-SAM) as substrate for OMTs in the BXD pathway.

De novo biosynthesis of halogenated indole

First, the direct halogenation of indole through the activity of the halogenase RebH-3LRS was explored. RebH-3LRS has been reported to halogenate the pyrrole ring of indole derivatives at position 2 or 3 (**Figure 11B**) [67]. RebH-3LRS is derived from RebH, a tryptophan halogenase from *Lechevalieria aerocolonigenes*, and was obtained by mutation of three amino acids resulting in a switch in substrate specificity from tryptophan to indole derivatives. However, RebH-3LRS would only introduce the halogenation on the pyrrole ring of indole, which would presumably prevent the ring expansion to HBOA catalyzed by BX4. RebH-3LRS was transiently expressed in *N. benthamiana* and indole, KCl, and KBr were used as substrates. The hexane-extracted samples were analyzed by GC-MS. No halogenated indole was detected (Figure S 3A).

As an alternative strategy, the biosynthesis of halogenated tryptophan was attempted, followed by conversion into halogenated indole by action of a tryptophanase. To do so, five bacterial flavin-dependent tryptophan halogenases which introduce the halogen atoms on the aromatic ring at positions 5-7 and TnaA, a tryptophanase from *E. coli*, were transiently expressed in *N. benthamiana*. Two of these halogenases, KtzQ and KtzR from *Kutzneria* species, were reported to sequentially produce dihalogenated tryptophan (**Figure 11A**) [58]. The flavin reductase *RebF* from *L. aerocolonigenes* was co-expressed with the halogenases (**Table 9**) [59].

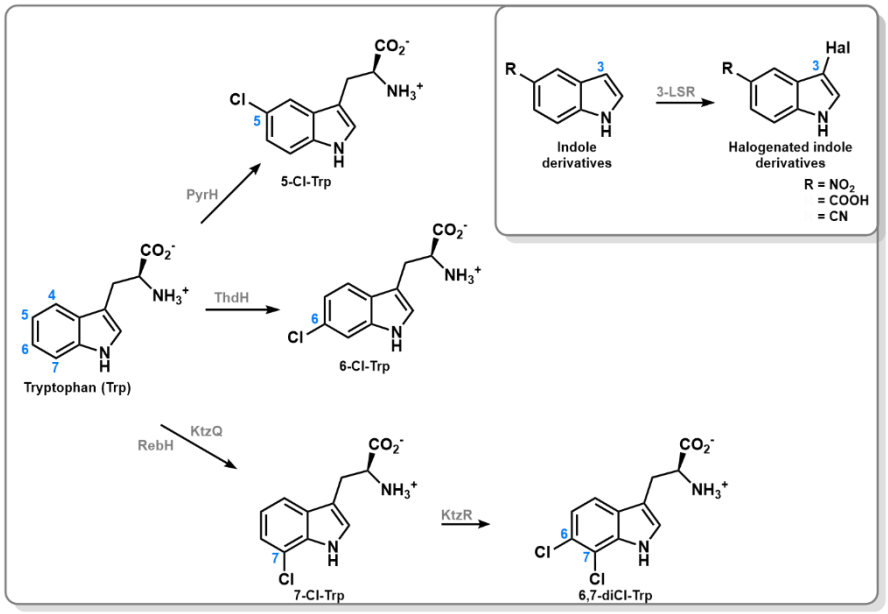

Figure 11: Halogenation of tryptophan (A) and indole derivatives (B) catalyzed by the respective enzymes (gray). (Modified from Davis et al. [59])

Table 9: Enzymes used to produce halogenated BXDs *in planta* grouped based on the substrate, indole, tryptophan (Trp), or other.

Indole halogenase	Product	Organism
RebH-3LSR	3-Hal-5-nitroindole	Modified from *L. aerocolonigenes* ATCC 39243
Trp halogenase	Product	Organism
PyrH	5-Cl-Trp	*Streptomyces rugosporus* LL-42D005
ThdH	6-Cl-Trp	*Streptomyces albogriseolus* MJ286-76F7
KtzQ	7-Cl-Trp	*Kutzneria* sp. 744
KtzR	7,6-Cl-Trp	*Kutzneria* sp. 744
RebH	7-Cl/Br-Trp	*Lechevalieria aerocolonigenes* ATCC 39243
Accessory enzyme	Function	Organism
RebF	Flavin reductase	*L. aerocolonigenes* ATCC 39243
TnaA	Tryptophanase	*E. coli*

The activity of the flavin-dependent halogenases PyrH, ThdH, KtzQ, KtzR, and RebH and the activity of the tryptophanase TnaA were tested *in planta*. Prior to combining them with the BX enzymes, all halogenases except ThdH were tested for the production of halogenated tryptophan or indole. The halogenases *KtzR* and *KtzQ* were co-expressed to produce dihalogenated tryptophan and indole. KCl and KBr were infiltrated as halide source at concentrations of 1 mM *in planta* or 0.5 and 1.0 mM in leaf disk assays. Methanolic sample

extracts were analyzed using LC-qTOF to detect halogenated tryptophan, while hexane extracts were analyzed on GC-MS to detect halogenated indole when co-expressing *TnaA*.

When expressing *PyrH* and *RebH*, a peak with the expected *m/z* of 239.0 for chlorinated tryptophan was detected in samples with and without the tryptophanase. The compounds for both halogenases eluted at 4.5 min and shared the same MS/MS fragmentation. When *KtzR* and *KtzQ* were co-expressed, small peaks eluting at the same time were present. However, the intensity of the peaks was insufficient for MS/MS fragmentation. When *PyrH* and *KtzQ/R* were co-expressed with the tryptophanase, less chlorinated tryptophan was detected (Figure S *3*E). Furthermore, leaf disk assays of PyrH, RebH, and KtzR/Q showed that less or no chlorinated tryptophan was formed when using only 0.5 mM KCl. GC-MS analysis of samples from *N. benthamiana* transiently transformed with *TnaA*, *RebF*, and *PyrH* or *RebH* confirmed the formation of chlorinated indole (**Figure 12A**). No brominated tryptophan or indole was detected for any of the flavin-dependent halogenases (Figure S 3B). In hexane extracts from PyrH and RebH infiltrated with KBr, chlorinated indole was detected, but no dihalogenated tryptophan or indole was detected when *KtzQ* and *KtzR* were co-expressed (Figure S 3C+D).

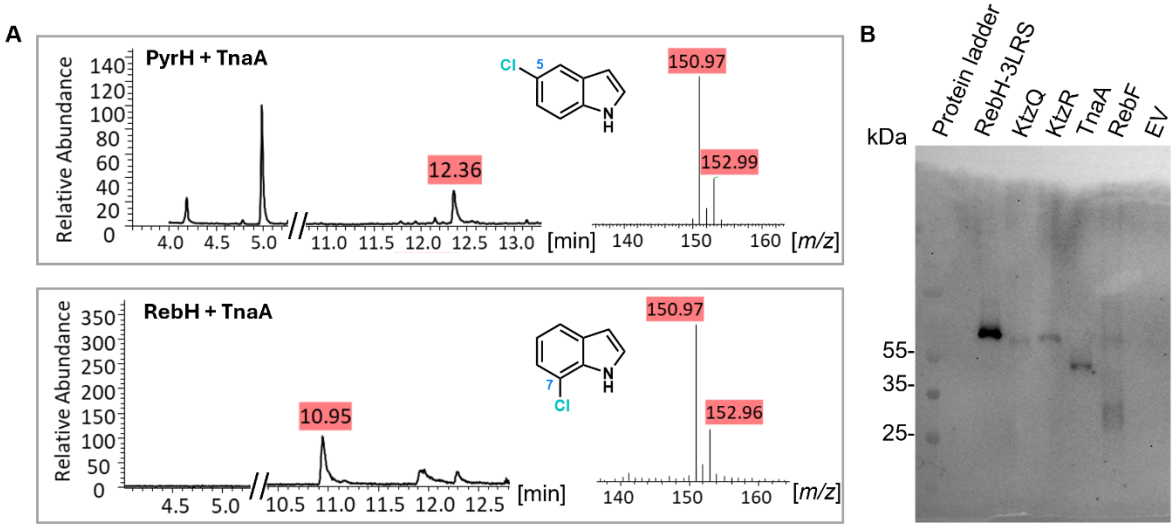

Figure 12: GC-MS analysis showing the production of 5-Cl-indole and 7-Cl-indole in planta in samples from N. benthamiana expressing the halogenases *PyrH* and *RebH*, the tryptophanase *TnaA*, and the flavin reductase (A). Western blot of His-tagged proteins after extraction from plant material and separation by SDS-PAGE (B).

To ensure the expression of bacterial genes in *N. benthamiana*, enzymes showing no or low activity were transiently expressed as His-tag fusions in *N. benthamiana,* followed by protein extraction. Proteins were separated by SDS-PAGE, blotted on a PDVF membrane and visualized using an anti-His antibody. Bands of the correct size were detected for the halogenases RebH-3LRS (58 kDa), KtzQ (58 kDa), KtzR (59 kDa), and the tryptophanase TnaA (54 kDa). Several bands of larger size were detected for the flavin reductase RebF (19 kDa). No band was present in the sample of the empty vector (EV) (**Figure 12B**).

Co-expression of flavin-dependent halogenases and BX enzymes

To produce halogenated BXDs *de novo* in *N. benthamiana*, the bacterial halogenases RebH and PyrH, which showed high yields of chlorinated indole, were co-expressed with *ZmBx2-5*. The halogenase ThdH was also tested in combination with the BX enzymes. The flavin reductase *RebF* and the tryptophanase *TnaA*, as well as *P19* and *ZmBx8*, were always co-expressed. As halide source, 1 mM KCl was infiltrated.

Chlorinated HBOA-Glc (*m/z* 360) was observed when co-expressing the BX enzymes with either *RebH* or *PyrH*. The MS/MS fragmentation and the RT (5-Cl-HBOA-Glc: 4.4 min; 7-Cl-HBOA-Glc: 4.6 min) were identical with the MS/MS fragmentation and RT observed when infiltrating 7-Cl- or 5-Cl-indole as a substrate. No chlorinated DIBOA-Glc was observed (**Figure 13**). Because 6-Cl-indole was not available as a substrate, the RT and MS/MS fragmentation of the products of ThdH could not be compared to previously produced chlorinated BXDs. No peaks with the correct MS/MS fragmentation could be detected for 6-Cl-HBOA-Glc. However, the peak observed for 6-Cl-DIBOA-Glc had a similar RT (4.8 min) as other chlorinated DIBOA-Glc derivatives and matched in MS/MS fragmentation (parent ion: *m/z* 376.04; fragment ions: *m/z* 168.00; *m/z* 195.00) (**Figure 13**).

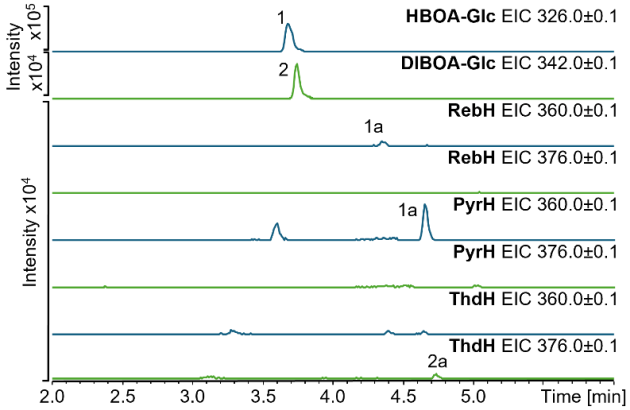

Figure 13: LC chromatograms showing the production of Cl-HBOA-Glc (1a, blue) or Cl-DIBOA-Glc (2a, green) from chlorinated indole produced *in planta* by RebH, PyrH or ThdH in combination with *Zm*BX2-5, *Zm*BX8, TnaA and RebF.

Halogenation of BXD pathway end products

Another enzyme tested to produce halogenated BXDs *in planta* was a halide-methyltransferase (HMT) from *Burkholderia xenovorans*. HMT produces fluorinated S-adenosyl methionine (F-SAM) using fluoroiodomethane (F-MeI) as a halide donor. The transfer of the halogenated methyl group onto hydroxyl groups of BXDs by the OMTs *AsBx7,* *Zm*BX7, *Zm*BX10, and *Zm*BX14 was tested. *N. benthamiana* was transiently transformed with *HMT*, *ZmBx1-8*, *P19* and *AsBx7*, *ZmBx10* and *ZmBx14*, or *ZmBx13* and *ZmBx14*. Co-expression of only *ZmBx1-8* and *P19* served as a control. Three days after transformation, F-MeI was infiltrated as a substrate. Co-expression of only *ZmBx1-8* and *P19* served as a control. Three days after transformation, F-MeI was infiltrated as substrate.

Table 10: Halide methyltransferase from *Burkholderina xenovorans* that produces F-SAM.

Enzyme	Product	Organism
HMT	F-Me-SAM	*Burkholderia xenovorans*

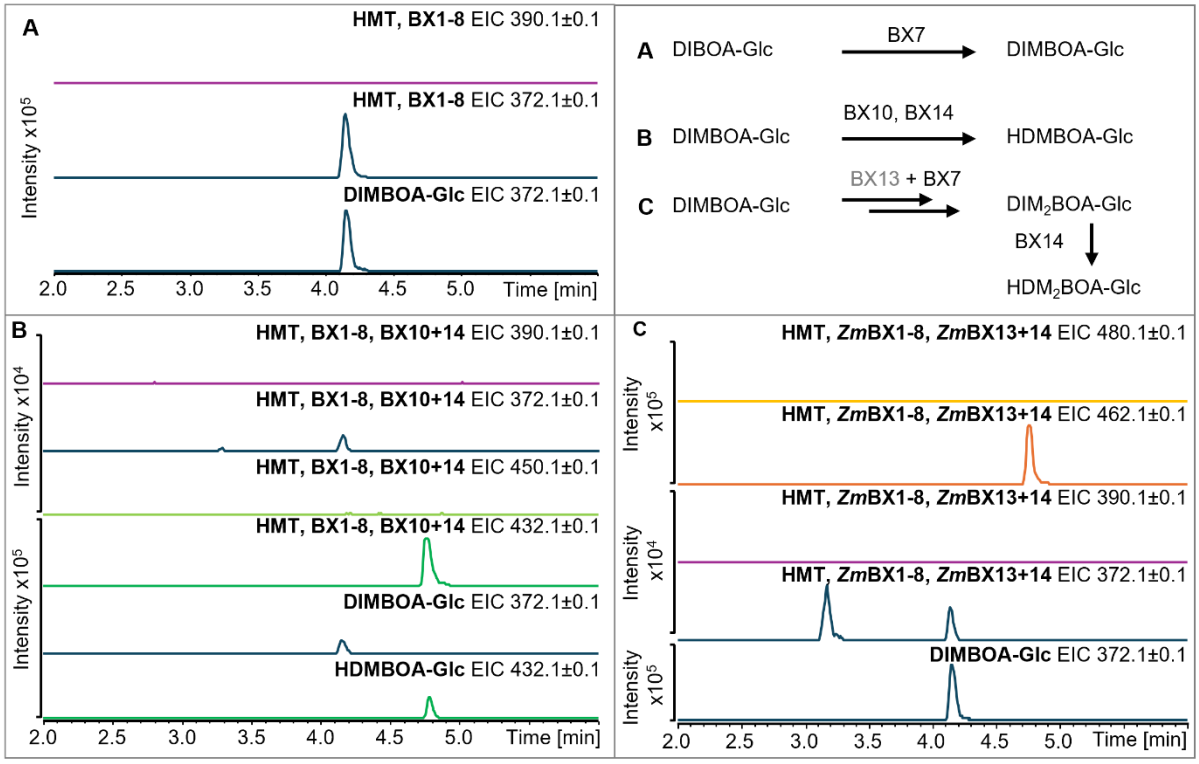

Figure 14: LC chromatograms showing the formation of non-halogenated BXDs when co-expressing the *ZmBx* enzymes with the *HMT*, but not the fluorinated derivatives (A-C). An overview of the products formed by the respective BX combination is shown on the top right. *O*-methyltransferases are shown in black (BX7, BX10, BX14). EICs of the non-halogenated BXD of authentic standards and leaf extracts, and the expected *m/z* of the fluorinated derivative are shown. DIMBOA-Glc (blue, *m/z* 372.1), F-DIMBOA-Glc (purple, *m/z* 390), HDMBOA-Glc (dark green, *m/z* 432.1), F-HDMBOA-Glc (light green, *m/z* 450), HDM2BOA-Glc (orange, *m/z* 462.1, no standard available), F-HDM2BOA-Glc (yellow, *m/z* 480.1).

When *N. benthamiana* was transiently transformed with *HMT*, *ZmBx1-8* and *AsBx7*, production of DIMBOA-Glc (*m/z* 372.0), but no fluorinated DIMBOA-Glc (*m/z* 390.1) was observed. Similarly, when testing the incorporation of a fluorinated methyl group by *ZmBx10* and ZmBx14, only non-fluorinated HDMBOA-Glc or DIMBOA-Glc were observed. Only non-fluorinated DIM_2BOA-Glc or HDM_2BOA-Glc were detected after the addition of *ZmBx13* and *ZmBx14* to the pool of co-expressed enzymes. Except for HDM_2BOA-Glc, all non-halogenated BXDs were identified by comparison of the RT and MS/MS fragmentation with reference standards (**Figure 14**).

Optimization of the biosynthesis of halogenated BXDs in *N. benthamiana*

Targeting tryptophan halogenases to the chloroplast or rerouting tryptophan biosynthesis to the cytosol to enhance de novo halogenated-BXD biosynthesis

The *de novo* biosynthesis of halogenated BXDs in *N. benthamiana* was successfully achieved using bacterial tryptophan halogenases in combination with a tryptophanase and subsequent incorporation of the so produced halogenated indole by the BX enzymes *Zm*BX2-5 and *Zm*BX8. Tryptophan halogenases and tryptophanases, being bacterial enzymes, possess no localization peptide and we expected them to be localized in the cytosol of *N. benthamiana* cells. In plants, on the other hand, tryptophan biosynthesis takes place in the chloroplast. Therefore, the cytosolic localization of tryptophan halogenases and tryptophanases could limit the availability of tryptophan for these enzymes, reducing the yields of halogenated BXDs. To overcome this limitation, two different approaches were tested: The first approach aimed at increasing tryptophan concentrations in the cytoplasm. To do so, seven enzymes involved in the tryptophan biosynthetic pathway in *A. thaliana*, starting from chorismate (anthranilate synthase subunit α and β (ASA+ASB; AT5G05730.2 + AT1G25220.2), phosphoribosylanthranilate transferase (PAT; AT5G17990.1), phosphoribosylanthranilate isomerase (PAI; AT1G07780.1), indole-3-glycerol phosphate synthase (IGPS; AT2G04400.1), tryptophan synthase subunit α and β (TSA+TSAB; AT3G54640.1 + AT5G54810.1), without the chloroplastic localization peptide were transiently expressed in *N. benthamiana* (**Figure 15**). This gene combination was co-expressed with the halogenases, tryptophanase, flavin reductase, and *ZmBx2-5*, *ZmBx8*, and *P19* to produce halogenated DIBOA-Glc. In the second approach, the halogenases, the flavin reductase, and the tryptophanase were expressed with an additional N-terminal chloroplastic localization peptide (CLP) from *Arabidopsis thaliana* tryptophan synthase subunit β. To further increase the availability of tryptophan in the chloroplast, the tryptophan synthase subunits α and β (TSA and TSB) from *A. squarrosa* were additionally co-expressed.

Cytosolic expression of halogenases led to low yields in chlorinated HBOA-Glc for all halogenases expressed. No DIBOA-Glc could be detected when RebH and PyrH were used. When *RebH*, *PyrH* and *ThdH* were co-expressed with the additional truncated genes for the cytosolic tryptophan biosynthesis, less Cl-HBOA-Glc was obtained for PyrH and ThdH, while no Cl-HBOA-Glc was obtained when co-expressing *RebH*. Upon expression of *ThdH*, low amounts of 6-Cl-DIBOA-Glc were detected both when the gene was expressed alone or in combination with the cytosolic tryptophan biosynthetic pathway. Compared to cytosolic expression, the yield of chlorinated HBOA-Glc was significantly higher for both halogenases tested when expressed in the chloroplast. Addition of TSA and TSB did not further increase the yields (**Figure 15**).

However, expression of the bacterial tryptophanase not only led to the accumulation of halogenated indole, but also of competing non-halogenated indole resulting from the cleavage of non-halogenated tryptophan. Therefore, high amounts of non-halogenated BXDs were detected in the same samples.

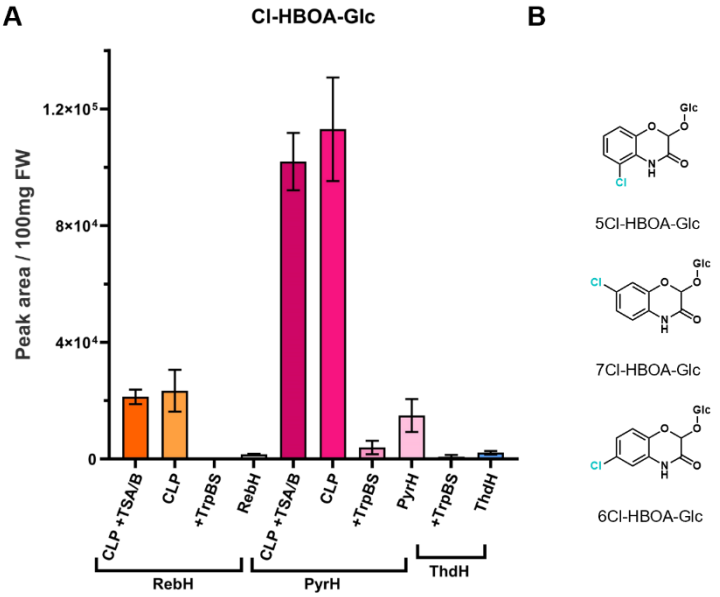

Figure 15: Comparison of yields for chlorinated HBOA-Glc depending on subcellular localization and overexpression of genes involved in Trp halogenation or Trp biosynthesis (A). CLP+TSA/B: Chloroplastic localization of the halogenases and overexpression of TSA and TSB; CLP: Chloroplastic localization of the halogenases; + TrpBS: Cytosolic overexpression of the truncated A. thaliana genes for the biosynthesis of tryptophan; RebH/PyrH/ThdH: cytosolic expression of the halogenases. The tryptophanase and flavin reductase were always co-expressed with the same subcellular localization as the halogenase in the respective experiment. Structures with the position of the added chloride atom are shown in C. RebH produces 7-Cl-Trp which yields 5-Cl-HBOA -Glc; PyrH produces 5-Cl-Trp which yields 7-Cl-HBOA-Glc; ThdH produces 6-Cl-Trp which yields 6-Cl-HBOA-Glc

In silico modeling of *Zm*BX4 and *As*BX4

To explain the observed differences in substrate specificity of BX enzymes from different species, *in vitro* modeling and docking studies should be performed. Knowledge of these mechanisms may help to tailor the substrate specificity of BX enzymes by direct mutagenesis in the future. BX4 was chosen as starting point because this enzyme showed efficient substrate conversion as well as high variability among the BX enzymes from different species. Three-dimensional models of the individual BX4s were generated using AlphaFold 2. Substrate docking was performed using Webina 1.0.5. To compare the structures, the different BX4s were aligned in PyMOL. To catalyze the ring expansion of the heterocycle of 3HI2O to form HBOA, the P450 BX4 contains a heme group as a prosthetic group.

Comparison of the active sites of *As*BX4 and *Zm*BX4 revealed that *Zm*BX4 had a larger active site cavity. While the space above the heme group of *As*BX4 is limited by a tryptophan

(Trp231), *Zm*BX4 has a serine (Ser236) at the same position (**Figure 16A+B**). Sequence alignment identified the respective amino acid residues for all BX4s. Apart from *Zm*BX4, aromatic amino acids were found at consensus position 261. *Sd*BX4, *Wr*BX4, *As*BX4, and *Lg*BX4 contained a tryptophan (Trp, W), while *Co*BX4 contained a phenylalanine (Phe, F). In contrast to these large amino acid residues, *Zm*BX4 had a serine (Ser, S). Furthermore, all BX4s except for *Zm*BX4 contained a phenylalanine at the following position 262. An alanine (Ala, A) was present at the same position in *Zm*BX4 (**Figure 16C**).

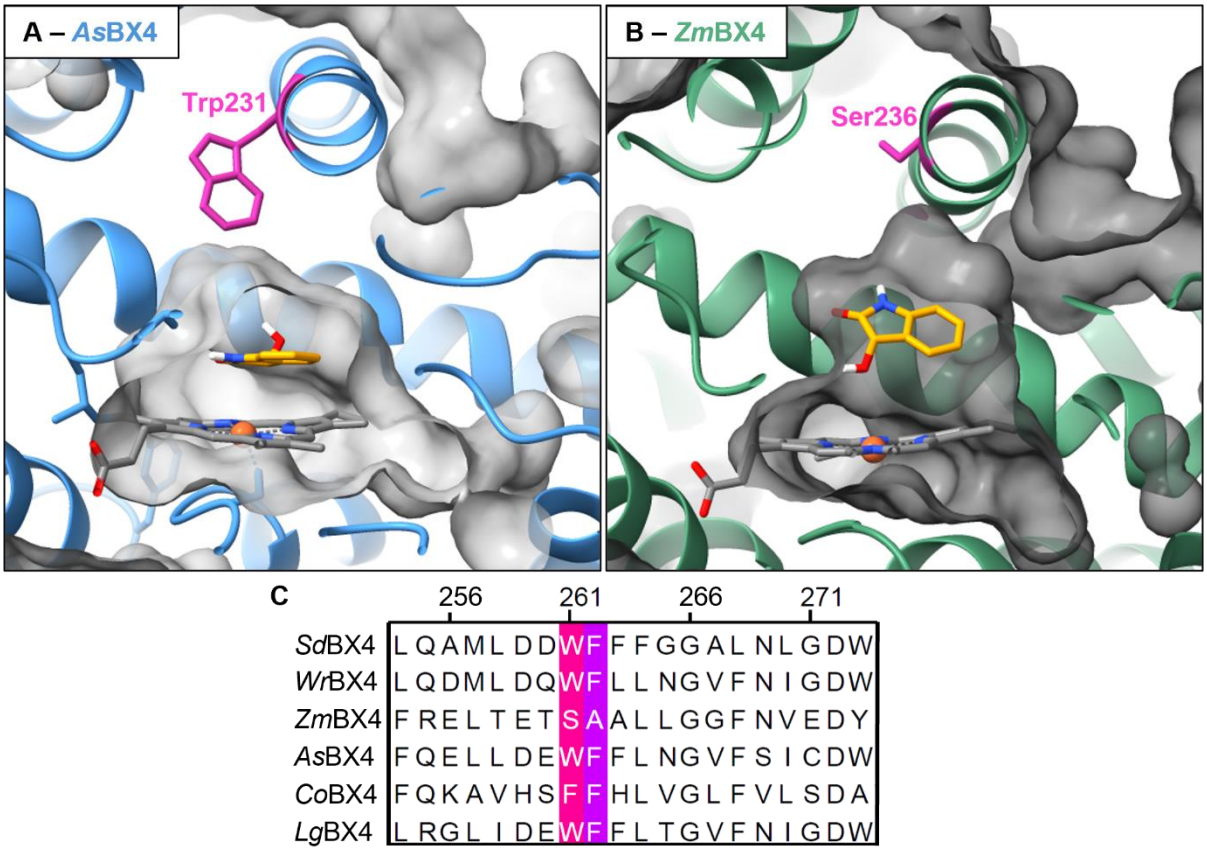

Figure 16: Active sites of *As*BX4 (A) and *Zm*BX4 (B) binding 3HI2O (orange). Amino acid residues which limit the cavity of the active site (gray) are marked in pink (Trp231 and Ser236). The heme is shown in gray. Figures were created using Chimera. **The sequence alignment of the area of interest of all BX4s is shown in C.** Trp 231 and Ser236 shown above correspond to consensus position 261 (pink). Phenylalanine and alanine of consensus position 262 are highlighted in purple. Figure is created using Jalview.

Discussion

Precursor-directed biosynthesis of halogenated BXDs in *N. benthamiana*

The biosynthesis of halogenated BXDs in transiently transformed *N. benthamiana* was successfully accomplished by precursor-directed biosynthesis. Chlorinated and fluorinated indole analogues fed to *N. benthamiana* were converted to the corresponding BXD analogues by BX enzymes from different plant species. The substrate specificity of the convergently evolved pathway enzymes was individually characterized and allowed the identification of optimal combinations of BX enzymes for the biosynthesis of halogenated BXDs *in planta*. For the seven-step pathway from indole to DIMBOA-Glc, the substrate specificity of 21 BX enzymes from six different species was characterized, and by using twelve of them, a total of 23 fluorinated and chlorinated BXD derivatives were produced starting from eight halogenated indole derivatives.

In all experiments, the yields for halogenated BXDs were normalized to those obtained with the BX enzymes of *Z. mays*, since BXD biosynthesis is best known in this species and its BX enzymes showed efficient incorporation for most substrates tested. However, the yields obtained with *Z. mays* BX enzymes were not the highest for all substrates tested. Therefore, the enzymes selected as optimal candidates belonged to different plant species and different enzyme families. This demonstrates the importance of investigating convergent evolution and its relevance as a valuable tool for the biotechnological production of natural products.

Overall, fluorinated indole showed better incorporation through the BX enzymes and higher yields than the chlorinated indole derivatives. When increasing the extent of the pathway, decreasing yields were observed for halogenated BXDs, but also for the positive control starting from indole. Higher yields were observed for DIMBOA-Glc and some halogenated analogues than for the precursor DIBOA-Glc and its analogues. This could be due to biological variance of the plants. Since experiments were conducted within different weeks, different batches of *N. benthamiana* plants were used. Even when grown in a greenhouse under stable conditions, variability in plant productivity may occur in experiments conducted in different weeks. Differences in yield were also observed depending on the position of the halogen atom. Interestingly, indole substituted at position 4 resulted in the highest yields of halogenated DIMBOA-Glc for both, chlorinated and fluorinated indole. Similarly, substitution at position 7 resulted in low yields and no incorporation further than HBOA-Glc for either derivative as well. The differences in the incorporation of halogenated indole depending on the position could be due to steric reasons, as in particular chloride is a sterically hindering substituent.

Furthermore, the high electronegativity of halogens can influence the electron distribution in the aromatic system affecting the compound reactivity.

It was previously reported that *N. benthamiana* possess endogenous BX3 oxidase activity [34]. This explains the accumulation of halogenated 3HI2O-Glc in plants only expressing *ZmBx8* and *LgBx8* (negative control). *ZmBx2* was expressed alone to test whether the endogenous BX3 activity of *N. benthamiana* would lead to similar substrate incorporation compared to the transiently expressed BX2-3 enzymes. However, co-expression of *ZmBx2-3* proved to be necessary as yields observed were higher for almost all substrates compared to expression of *ZmBx2* alone. Slightly higher yields when expressing *ZmBx2* alone were only observed for 4-Cl-3HI2O-Glc. Starting from 7-Cl-I2O as substrate for chlorinated BXDs, generally resulted in higher yields of chlorinated 3HI2O-Glc and chlorinated HBOA-Glc than starting from 7-Cl-indole. This might be due to the fact that I2O, the product of the first oxidation step catalyzed by BX2, is already a pathway intermediate. Yields of chlorinated 7-Cl-3HI2O-Glc produced by the negative control were highest, when starting from 7-Cl-I2O. This might also be due to the easier incorporation of the I2O derivative compared to indole. Unfortunately, no F-I2O was available.

All halogenated BXD analogues produced were identified by comparing the MS/MS fragmentation with authentic standards of the respective non-halogenated BXDs. Increases in *m/z* of the parent and fragment ions were in accordance with the expected + *m/z* 34 for chlorinated and + *m/z* 18 for fluorinated compounds. Halogenated BXDs showed later retention times compared to the non-halogenated BXDs in liquid chromatography. This was in accordance with later retention time also reported for halogenated derivatives of monoterpene indole alkaloids [20]. Two separate peaks with identical MS/MS fragmentation or peaks that could not be fully chromatographically resolved were observed for 3HI2O-Glc, halogenated 3HI2O-Glc analogues and some halogenated HBOA-Glc derivatives. These could result from *R*- and *S*-epimers depending on the stereochemistry at C3 (3HI2O-Glc) or C2 (HBOA-Glc) bound to the *O*-glucosyl group. Keto-enol tautomerization of 3HI2O before glucosylation could lead to the different epimers [34]. Furthermore, the peak could result from an isomer of 3HI2O-Glc, such as 5HI2O-Glc which was detected in *Arabidopsis thaliana* expressing *ZmBx1-3* [80]. Isolation of the compounds and subsequent NMR would be necessary to confirm these hypotheses.

No production of halogenated DIMBOA-Glc was expected when starting from indole halogenated at C7 as BX6-7 would catalyze the methoxylation of DIBOA-Glc at the same position. Nevertheless, a peak with the MS/MS fragmentation matching other fluorinated DIMBOA-Glc analogies was observed when *AsBx6* was among the pool of expressed genes.

Presumably, the presence of fluorine forced the methoxylation to occur at another nearby position and BX6-7 show some flexibility regarding the regioselectivity. Isolation and NMR analysis would be necessary to confirm the identity of the unknown “5-F-DIMBOA-Glc” analogue produced.

Two halogenated BXDs, 8-Cl-HBOA-Glc and 8-F-HBOA-Glc, that showed high yields as evidenced by quantification of the relative peak area were isolated from plant material and further characterized by NMR. In collaboration with Yoko Nakamura (NMR service group, MPI-CE), the identity of both compounds could be confirmed. Coupling of ^{19}F with C or H confirmed the presence and exact position at C8. The incorporation of chlorine into the BXDs could be confirmed by LC-qTOF analysis. Based on the natural isotope abundance of chlorine, the ratio of $^{35}Cl/^{37}Cl$ is about one third [81]. When analyzing the chlorinated BXD derivatives, peaks of lower intensity for parent and fragment ions in the MS/MS, which were increased by *m/z* 2, indicated the incorporation of ^{37}Cl.

Due to the large amounts of substrate needed, the analysis of different enzyme combinations and the screening of the substrate scope *in planta* can be labor and cost intensive. Furthermore, indole is a volatile and when infiltrated as a substrate can be emitted from the infiltrated leaves or transported elsewhere in the plant. Thus, the set-up of a leaf disk assay for the production of BXDs was established (modified from Hong et al. [78]) and compared to the established workflow using whole plants. The leaf disk assay led to significantly higher yields than infiltration of whole leaves. Small amounts of BXDs were produced in the control experiment where no indole was provided as a substrate. This can be explained by the release of indole caused by wounding the plants to collect leaf disks and subsequent incorporation through the BX enzymes [79]. Leaf disk assays could be a useful tool for future experiments to increase yields and throughput and to optimize parameters such as substrate concentrations or time point of substrate infiltration and sample collection. Moreover, leaf disk assays would be a good platform to further screen BX enzymes substrate scope on other halogenated indoles such as Br-indole. Since bromide is a good leaving group, brominated BXD derivatives could offer the possibility for further synthetic modification through nucleophilic aromatic substitution reactions [82,83].

For transient transformation of *N. benthamiana,* individual *A. tumefaciens* cultures, each carrying one gene of interest, were mixed prior infiltration. Successful production of halogenated BXDs was depending on a sufficient number of leaf cells being transiently transformed with all the constructs needed. To increase the rate of cells transformed with all constructs, Grzech et al. used multitranscriptional units (multi-TU) vectors, harboring up to four genes per construct, for the biosynthesis of precondylocarpine acetate in *N. benthamiana*

[84]. A similar approach could be taken, combining multiple BX enzymes of the optimal combination in one multi-TU.

De novo production of halogenated BXDs *in planta*

After the successful semi-synthetic production of halogenated BXDs *in planta*, the subsequent aim became the *de novo* production of such compounds. The first goal was the production of halogenated indole. Enzymatic modification of indole and derivatives commonly occurs at the most electron rich site C3 [61,67,85]. This position is favored since the lone pair on the nitrogen can donate electrons, displacing the resulting positive charge of C2 to the nitrogen. The indole halogenase RebH-3LSR, a mutant of the flavin-dependent tryptophan halogenase RebH from *L. aerocolonigenes*, was reported to halogenate indole derivatives with both, electron withdrawing and electron donating substituents with good yields at the C3 position *in vitro*. Thus, the enzyme was expressed *in planta*. Although the expression was confirmed by Western-blot, no halogenated indole was observed. The absence of any halogenated product could be due to non-appropriate folding or post translational modifications of the protein or the different reaction conditions in a plant system. Furthermore, Li et al. speculated that substituted aromatic substrates might be needed for interactions of the substrate with the active site of the halogenase [67].

A more successful alternative approach to produce halogenated indole in planta was the expression of bacterial flavin-dependent tryptophan halogenases. The halogenases were expressed in combination with a flavin reductase and a tryptophanase to cleave halogenated tryptophan and release the modified indole moiety. The *de novo* production of multiple chlorinated BXD derivatives was successfully achieved using three different halogenases. When first expressed in planta, the 7-Cl-tryptophan halogenase RebH from *L. aerocolonigenes* and 5-Cl-tryptophan halogenase PyrH from *S. rugosporus* produced chlorinated tryptophan with and without the addition of KCl. When expressed together with the *E. coli* tryptophanase *TnaA*, production of chlorinated indole was observed. No bromination was observed. Since more chlorinated tryptophan was detected when adding 1 mM KCl, additional salt was added in subsequent experiments. When *RebH* and *PyrH* were expressed in combination with *ZmBx2-5* and *ZmBx8*, de novo production of chlorinated HBOA-Glc was successfully achieved. Even chlorinated DIBOA-Glc was detected when co-expressing the 6-Cl-tryptophan halogenase *ThdH* from *S. rugosporus* in combination with the BX enzymes. This combination was particularly interesting since ThdH produces 6-Cl-indole, which was not at hand as a substrate when characterizing the substrate specificity of different BX enzymes.

Furthermore, it was possible to drastically increase the yields of chlorinated HBOA-Glc when adapting the subcellular localization of two halogenases and accessory enzymes to the *in*

planta production platform. Being of bacterial origin, the tryptophan halogenases were believed to be localized in the cytosol. Since tryptophan biosynthesis in plants takes place in the chloroplast, the halogenases *RebH* and *PyrH* were expressed with a chloroplastic localization peptide (CLP), in order to increase yields of chlorinated BXDs. This led to significantly increased yields of chlorinated HBOA-Glc for both halogenases when compared to the yields of the respective enzyme without CLP. To fully confirm the effective change in subcellular localization, the localization of halogenases and CLP-halogenases fused to a fluorescence protein should be assessed by confocal fluorescence microscopy. *ThdH* was not expressed with a CLP as it was included in the experiments at a later stage. Higher yields of chlorinated HBOA-Glc were observed for the co-expression of *CLP-PyrH* compared to *CLP-RebH* with the BX enzymes and could be attributed to multiple factors: one reason could be better expression and activity of PyrH *in planta* or that 5-Cl-tryptophan produced by PyrH is more easily cleaved by the tryptophanase than 7-Cl-tryptophan produced by RebH. More extended experiments would be needed to clarify which of these factors play a major role. In accordance with our results, previous studies on the activity of the flavin-dependent halogenases in *N. benthamiana* showed higher yields of chlorinated tryptophan produced by PyrH than RebH and ThdH [59]. Taking into consideration the results of the relative peak area quantification of the semi-synthetic production of halogenated BXDs, higher yields and more extended incorporation into the BXD pathway were observed starting from 5-Cl-indole than 7-Cl-indole. Thus, it is likely that better incorporation of 5-Cl-indole, produced by PyrH, through the BX enzymes additionally effects the higher yields compared to the co-expression of *RebH*.

Attempts to further increase yields by over-expression of the tryptophan synthase subunits α and β (TSA and TSB) when co-expressing *CLP-RebH* or *CLP-PyrH* did not yield further increases. Metabolic flux analysis in rice showed that the tryptophan biosynthesis is tightly regulated through feed-back inhibition of the anthranilate synthase (AS) through tryptophan. Instead of over-expression of TSA+TSB, a modified AS, which is insensitive to feed-back inhibition, might be needed, as this has shown to highly increase tryptophan concentrations in rice calli and rice plants [86].

To increase yields of *de novo* produced chlorinated BXDs by adapting the subcellular localization, an alternative approach aimed at higher cytosolic concentrations of tryptophan. Seven truncated enzymes from *A. thaliana* for tryptophan biosynthesis from chorismate were expressed without their CLP. However, less or no halogenated HBOA-Glc was obtained compared to simply co-expressing the BX enzymes and the halogenases in the cytosol. Although pathways utilizing chorismate are present in the cytosol as well, lower availability of chorismate in the cytosol compared to the plastid could be the reason for the observed results [87].

Halogenated BXDs were successfully produced *in planta* by using precursor-directed biosynthesis and entirely enzymatic approach, using bacterial halogenases. The yields of the *de novo* production were greatly increased by adjusting the subcellular localization of the bacterial genes to the plant expression system. Further modifications could be made in the future to increase and/or simplify the *de novo* production of halogenated BXDs. Davis et al. supplemented the co-infiltration mixture of *A. tumefaciens* directly with the salt solution. This leads to an immediate substrate availability of the halides, but also eliminates the additional step of substrate infiltration [59]. Watering the plants with NaCl solution might already increase the intracellular salt availability, making the additional infiltration of Cl^- also unnecessary. The salt tolerance of *N. benthamiana* was briefly tested and leaves showed no change in phenotype when watering the plants with 500 mM NaCl (Figure S 4). As a first test of the *de novo* production of BXDs *in planta*, only *ZmBx2-5* were co-expressed. Addition of *CoBx4* and *LgBx4-5* might further increase the yields, especially since *LgBx4* showed efficient incorporation of 7-Cl-indole.

A drawback of the production of halogenated indole via the halogenation of tryptophan is the production of non-halogenated indole resulting from the cleavage of non-halogenated tryptophan by the overexpressed bacterial tryptophanase. All plants produced high amounts of non-halogenated BXDs, as both the tryptophanase and the BX enzymes presumably better accepted their natural, non-halogenated substrates. One way of reducing the availability of non-halogenated indole could be the design of a tryptophanase mutant that prefers halogenated tryptophan. Favored binding of halogenated tryptophan could be promoted through enlargement of the active site or addition of amino acids that better stabilize the halogenated derivatives. Alternatively, a recycling system for non-halogenated indole could be implemented by the expression of a stand-alone TSB. After indole-3-glycerol phosphate cleavage, indole is usually channeled from TSA to TSB of the tryptophan synthase complex. TSB alone shows poor catalytic activity for the condensation of L-serine and indole into tryptophan. Using directed evolution, Buller et al. created a stand-alone TSB from *Pyrococcus furiosus,* which can utilize free indole. However, bacterial tryptophan synthase complexes are also responsible for the biosynthesis of noncanonical amino acids and are described to accept indole derivatives as well [88]. Thus, the substrate specificity of the TSB mutant should be elucidated first by comparing the catalytic activity for the utilization of indole or chlorinated indole derivatives.

Providing a modified substrate that subsequently gets incorporated through the enzymes of the pathway can be described as a bottom-up biosynthetic approach. With this bottom-up approach, losses in yields were observed with every subsequent enzymatic conversion and overall yields of halogenated derivatives were mostly lower than yields of the respective non-

halogenated BXDs when starting from the natural substrate. In a top-down approach, halogenation would be introduced at the end of the biosynthesis. Consequently, higher yields could be possible since most of the pathway enzymes would convert their natural substrates. To produce fluorinated BXDs *in planta,* a top-down approach should be developed. Fluorinated BXDs would be highly interesting, because many agrochemicals contain fluoride [47]. However, due to the high electronegativity of fluoride, enzymatic fluorination is way less common in nature than chlorination or bromination and an incorporation using the flavin-dependent halogenases as seen for chlorinated BXDs is not possible [51]. Thus, production of fluorinated BXDs was attempted by expressing a halide methyl transferase to generate F-SAM as a fluoro-methyl group that could be incorporated through the *O*-methyltransferases of the BXD pathway. Peng et al. already showed utilization of F-SAM by multiple methyl transferases *in vitro* [71]. Unfortunately, no fluorinated BXD derivatives were detected when expressing the *HMT* and *Bx* genes *in planta*. In further experiments, the ability of the BX *O*-methyltransferases should first be characterized *in vitro*. Furthermore, validation of *HMT* expression in *N. benthamiana* should be confirmed.

Rational design to improve yields of halogenated BXDs

BXD biosynthesis evolved convergently in different plant species and enzymes catalyzing the same reaction belong to different enzyme families [34]. Differences in substrate incorporation were observed among the BX enzymes catalyzing the same step of the pathway. For the rational analysis of the differences in substrate specificity, *in silico* models of all BX enzymes and the docked substrate were generated and structural characteristics were compared to experimental results. At consensus position 261 and 262 of the amino acid sequences, *Zm*BX4, the enzyme with the broadest substrate specificity for all substrates tested, carried smaller amino acid residues (Ser, Ala), while all other BX4s carried larger aromatic amino acid residues limiting the space above the catalytic heme group (*Co*BX4: Phe, Phe; others: Trp, Phe). Combining the knowledge gained from testing the substrate specificity of the BX enzymes and the rational *in silico* analysis could allow for the engineering of optimized BX enzymes.

Conclusion

Derivatization of natural products to alter their biochemical properties and biological activity is important for the development of new drugs and agrochemicals. However, total chemical synthesis or semi-synthetic derivatization of natural products can be challenging and these chemical methods are often not environmentally friendly. Given the already established herbicidal properties of halogenated BXDs, along with knowledge of the biosynthetic genes, we engineered the total biochemical production of halogenated BXDs in *N. benthamiana* in line with green chemistry principles [57,89]. The previous elucidation of the convergently evolved pathway in different plants species proved to be a valuable set of resources for the modification of BXDs. Having a pool of independently evolved enzymes to choose for each step, each one with a different substrate specificity, reduced the risk for pathway bottlenecks. If future bioassays would also confirm an enhanced antifeedant activity of halogenated BXD derivatives, genetically modified crops could be developed to directly produce halogenated BXDs. Beneficial effects of BXDs on human health are not well studied, in part due to their limited abundance. Using *N. benthamiana* as a production platform for BXDs and their derivatives could accelerate the investigation of potential medicinal application of these compounds.

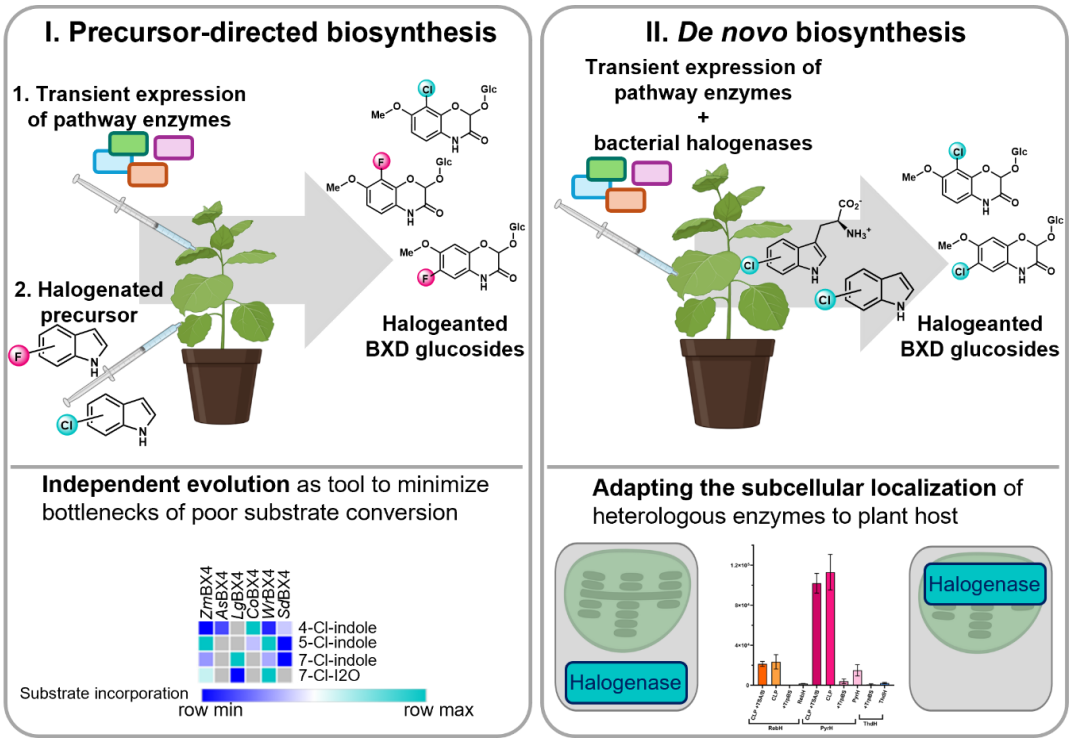

Figure 17: Overview of the two approaches, precursor directed biosynthesis and *de novo* biosynthesis, used to produce halogenated BXDs *N. benthamiana*. In order to achieve the highest yields possible through precursor directed biosynthesis starting from halogenated indole, the substrate specificity of independently evolved enzymes was characterized. Yields of *de novo* produced halogenated BXDs expressing bacterial halogenases were significantly increased by changing the subcellular localization of the bacterial halogenases from the cytoplasm to the chloroplast.

References

[1] Loiseleur, O., *Chimia*, (2017) **71**, 810.
[2] Weng, J.-K.; Lynch, J. H.; Matos, J. O.; Dudareva, N., *Nature chemical biology*, (2021) **17**, 1037.
[3] Mithöfer, A.; Boland, W., *Annual review of plant biology*, (2012) **63**, 431.
[4] Erb, M.; Veyrat, N.; Robert, C. A. M.; Xu, H.; Frey, M.; Ton, J.; Turlings, T. C. J., *Nature communications*, (2015) **6**, 6273.
[5] Kong, C.-H.; Xuan, T. D.; Khanh, T. D.; Tran, H.-D.; Trung, N. T., *Molecules (Basel, Switzerland)*, (2019) **24**.
[6] Chopra, B.; Dhingra, A. K., *Phytotherapy research : PTR*, (2021) **35**, 4660.
[7] Sparks, T. C.; Duke, S. O., *Journal of agricultural and food chemistry*, (2021) **69**, 8324.
[8] Caputi, L.; Franke, J.; Farrow, S. C.; Chung, K.; Payne, R. M. E.; Nguyen, T.-D.; Dang, T.-T. T.; Soares Teto Carqueijeiro, I.; Koudounas, K.; Dugé de Bernonville, T.; Ameyaw, B.; Jones, D. M.; Vieira, I. J. C.; Courdavault, V.; O'Connor, S. E., *Science (New York, N.Y.)*, (2018) **360**, 1235.
[9] Vollmann, D. J.; Winand, L.; Nett, M., *Current opinion in biotechnology*, (2022) **77**, 102761.
[10] Chaachouay, N.; Zidane, L., *DDC*, (2024) **3**, 184.
[11] Newman, D. J.; Cragg, G. M., *Journal of natural products*, (2007) **70**, 461.
[12] Sadre, R., *Current opinion in biotechnology*, (2024) **87**, 103142.
[13] Chauhan, D.; Hideshima, T.; Anderson, K. C., *British journal of cancer*, (2006) **95**, 961.
[14] Neumann, C. S.; Fujimori, D. G.; Walsh, C. T., *Chemistry & biology*, (2008) **15**, 99.
[15] Feling, R. H.; Buchanan, G. O.; Mincer, T. J.; Kauffman, C. A.; Jensen, P. R.; Fenical, W., *Angewandte Chemie (International ed. in English)*, (2003) **42**, 355.
[16] Lin, D.; Jiang, S.; Zhang, A.; Wu, T.; Qian, Y.; Shao, Q., *Natural products and bioprospecting*, (2022) **12**, 8.
[17] Xiong, X.; Yeung, Y.-Y., *ACS Catal.*, (2018) **8**, 4033.
[18] Boddy, C. N.; Hotta, K.; Tse, M. L.; Watts, R. E.; Khosla, C., *Journal of the American Chemical Society*, (2004) **126**, 7436.
[19] Grüschow, S.; Rackham, E. J.; Elkins, B.; Newill, P. L. A.; Hill, L. M.; Goss, R. J. M., *Chembiochem : a European journal of chemical biology*, (2009) **10**, 355.
[20] Boccia, M.; Grzech, D.; Lopes, A. A.; O'Connor, S. E.; Caputi, L., *Frontiers in plant science*, (2022) **13**, 919443.
[21] Runguphan, W.; Qu, X.; O'Connor, S. E., *Nature*, (2010) **468**, 461.
[22] Deb Roy, A.; Grüschow, S.; Cairns, N.; Goss, R. J. M., *Journal of the American Chemical Society*, (2010) **132**, 12243.
[23] Bruijn, W. J. C. de; Gruppen, H.; Vincken, J.-P., *Phytochemistry*, (2018) **155**, 233.
[24] Wouters, F. C.; Gershenzon, J.; Vassão, D. G., *Journal of the Brazilian Chemical Society*, (2016).
[25] Zhou, S.; Richter, A.; Jander, G., *Plant & cell physiology*, (2018) **59**, 1528.
[26] Frey, M.; Chomet, P.; Glawischnig, E.; Stettner, C.; Grün, S.; Winklmair, A.; Eisenreich, W.; Bacher, A.; Meeley, R. B.; Briggs, S. P.; Simcox, K.; Gierl, A., *Science (New York, N.Y.)*, (1997) **277**, 696.
[27] Richter, A.; Powell, A. F.; Mirzaei, M.; Wang, L. J.; Movahed, N.; Miller, J. K.; Piñeros, M. A.; Jander, G., *The Plant journal : for cell and molecular biology*, (2021) **106**, 245.
[28] Schullehner, K.; Dick, R.; Vitzthum, F.; Schwab, W.; Brandt, W.; Frey, M.; Gierl, A., *Phytochemistry*, (2008) **69**, 2668.
[29] Schulz, M.; Marocco, A.; Tabaglio, V.; Macias, F. A.; Molinillo, J. M. G., *Journal of chemical ecology*, (2013) **39**, 154.
[30] Rad, U. von; Hüttl, R.; Lottspeich, F.; Gierl, A.; Frey, M., *The Plant journal : for cell and molecular biology*, (2001) **28**, 633.
[31] Frey, M.; Schullehner, K.; Dick, R.; Fiesselmann, A.; Gierl, A., *Phytochemistry*, (2009) **70**, 1645.
[32] Oikawa, A.; Ishihara, A.; Iwamura, H., *Phytochemistry*, (2002) **61**, 331.
[33] Handrick, V.; Robert, C. A. M.; Ahern, K. R.; Zhou, S.; Machado, R. A. R.; Maag, D.; Glauser, G.; Fernandez-Penny, F. E.; Chandran, J. N.; Rodgers-Melnik, E.; Schneider, B.; Buckler, E. S.; Boland, W.; Gershenzon, J.; Jander, G.; Erb, M.; Köllner, T. G., *The Plant cell*, (2016) **28**, 1682.

[34] Florean, M.; Luck, K.; Hong, B.; Nakamura, Y.; O'Connor, S. E.; Köllner, T. G., *Proceedings of the National Academy of Sciences of the United States of America*, (2023) **120**, e2307981120.
[35] Dutartre, L.; Hilliou, F.; Feyereisen, R., *BMC evolutionary biology*, (2012) **12**, 64.
[36] Dick, R.; Rattei, T.; Haslbeck, M.; Schwab, W.; Gierl, A.; Frey, M., *The Plant cell*, (2012) **24**, 915.
[37] Wouters, F. C.; Blanchette, B.; Gershenzon, J.; Vassão, D. G., *Phytochemistry reviews : proceedings of the Phytochemical Society of Europe*, (2016) **15**, 1127.
[38] Niemeyer, H. M., *Journal of agricultural and food chemistry*, (2009) **57**, 1677.
[39] Poschenrieder, C.; Tolrà, R. P.; Barceló, J., *Journal of inorganic biochemistry*, (2005) **99**, 1830.
[40] Bravo, H. R.; Lazo, W., *Phytochemistry*, (1993) **33**, 569.
[41] Gleńsk, M.; Gajda, B.; Franiczek, R.; Krzyżanowska, B.; Biskup, I.; Włodarczyk, M., *Natural product research*, (2016) **30**, 1305.
[42] Macías, F. A.; Chinchilla, N.; Varela, R. M.; Molinillo, J. M.; Marín, D.; Siqueira, J. M. de, *Pest management science*, (2009) **65**, 1104.
[43] Śmist, M.; Kwiecień, H.; Krawczyk, M., *Journal of environmental science and health. Part. B, Pesticides, food contaminants, and agricultural wastes*, (2016) **51**, 393.
[44] Le-Van, N.; Wratten, S. J., *Tetrahedron Letters*, (1984) **25**, 145.
[45] Anai, T.; Aizawa, H.; Ohtake, N.; Kosemura, S.; Yamamura, S.; Hasegawa, K., *Phytochemistry*, (1996) **42**, 273.
[46] Kanchanapoom, T.; Kamel, M. S.; Kasai, R.; Picheansoonthon, C.; Hiraga, Y.; Yamasaki, K., *Phytochemistry*, (2001) **58**, 637.
[47] Jeschke, P., *Pest management science*, (2010) **66**, 10.
[48] Smith, K.; El-Hiti, G., *COC*, (2006) **10**, 1603.
[49] Braddock, D. C., *Organic letters*, (2006) **8**, 6055.
[50] Vincent, M. A.; Hillier, I. H., *Chemical communications (Cambridge, England)*, (2005), 5902.
[51] Agarwal, V.; Miles, Z. D.; Winter, J. M.; Eustáquio, A. S.; El Gamal, A. A.; Moore, B. S., *Chemical reviews*, (2017) **117**, 5619.
[52] Butler, A.; Carter-Franklin, J. N., *Natural product reports*, (2004) **21**, 180.
[53] Winter, J. M.; Moffitt, M. C.; Zazopoulos, E.; McAlpine, J. B.; Dorrestein, P. C.; Moore, B. S., *The Journal of biological chemistry*, (2007) **282**, 16362.
[54] Keller, S.; Wage, T.; Hohaus, K.; Hölzer, M.; Eichhorn, E.; van Pée, K.-H., *Angew. Chem. Int. Ed.*, (2000) **39**, 2300.
[55] Zehner, S.; Kotzsch, A.; Bister, B.; Süssmuth, R. D.; Méndez, C.; Salas, J. A.; van Pée, K.-H., *Chemistry & biology*, (2005) **12**, 445.
[56] Milbredt, D.; Patallo, E. P.; van Pée, K.-H., *Chembiochem : a European journal of chemical biology*, (2014) **15**, 1011.
[57] Schnepel, C.; Sewald, N., *Chemistry (Weinheim an der Bergstrasse, Germany)*, (2017) **23**, 12064.
[58] Heemstra, J. R.; Walsh, C. T., *Journal of the American Chemical Society*, (2008) **130**, 14024.
[59] Davis, K.; Gkotsi, D. S.; Smith, D. R. M.; Goss, R. J. M.; Caputi, L.; O'Connor, S. E., *Frontiers in plant science*, (2020) **11**, 581675.
[60] Hölzer, M.; Burd, W.; Reißig, H.-U.; van Pée, K.-H., *Adv. Synth. Catal.*, (2001) **343**, 591.
[61] Payne, J. T.; Andorfer, M. C.; Lewis, J. C., *Angewandte Chemie (International ed. in English)*, (2013) **52**, 5271.
[62] Frese, M.; Guzowska, P. H.; Voß, H.; Sewald, N., *ChemCatChem*, (2014) **6**, 1270.
[63] Glenn, W. S.; Nims, E.; O'Connor, S. E., *Journal of the American Chemical Society*, (2011) **133**, 19346.
[64] Shepherd, S. A.; Karthikeyan, C.; Latham, J.; Struck, A.-W.; Thompson, M. L.; Menon, B. R. K.; Styles, M. Q.; Levy, C.; Leys, D.; Micklefield, J., *Chemical science*, (2015) **6**, 3454.
[65] Schnepel, C.; Minges, H.; Frese, M.; Sewald, N., *Angewandte Chemie (International ed. in English)*, (2016) **55**, 14159.
[66] van Pée, K.-H.; Milbredt, D.; Patallo, E. P.; Weichold, V.; Gajewi, M., *Methods in enzymology*, (2016) **575**, 65.

[67] Li, E. H. Y.; Sana, B.; Ho, T.; Ke, D.; Ghadessy, F. J.; Duong, H. A.; Seayad, J., *Frontiers in bioengineering and biotechnology*, (2022) **10**, 1032707.
[68] Müller, K.; Faeh, C.; Diederich, F., *Science (New York, N.Y.)*, (2007) **317**, 1881.
[69] O'Hagan, D.; Deng, H., *Chemical reviews*, (2015) **115**, 634.
[70] Harper, D. B., *Natural product reports*, (2000) **17**, 337.
[71] Peng, J.; Liao, C.; Bauer, C.; Seebeck, F. P., *Angewandte Chemie (International ed. in English)*, (2021) **60**, 27178.
[72] Huo, L.; Hug, J. J.; Fu, C.; Bian, X.; Zhang, Y.; Müller, R., *Natural product reports*, (2019) **36**, 1412.
[73] Markina, N. M.; Kotlobay, A. A.; Tsarkova, A. S., (2020) **12**, 28.
[74] Chase, M. W.; Cauz-Santos, L. A.; Dodsworth, S.; Christenhusz, M. J. M., *Aust. Syst. Bot.*, (2022) **35**, 345.
[75] Bally, J.; Jung, H.; Mortimer, C.; Naim, F.; Philips, J. G.; Hellens, R.; Bombarely, A.; Goodin, M. M.; Waterhouse, P. M., *Annual review of phytopathology*, (2018) **56**, 405.
[76] Ruocco, V.; Strasser, R., *Plants (Basel, Switzerland)*, (2022) **11**.
[77] Sarrion-Perdigones, A.; Vazquez-Vilar, M.; Palací, J.; Castelijns, B.; Forment, J.; Ziarsolo, P.; Blanca, J.; Granell, A.; Orzaez, D., *Plant physiology*, (2013) **162**, 1618.
[78] Hong, B.; Grzech, D.; Caputi, L.; Sonawane, P.; López, C. E. R.; Kamileen, M. O.; Hernández Lozada, N. J.; Grabe, V.; O'Connor, S. E., *Nature*, (2022) **607**, 617.
[79] Gierl, A.; Frey, M., *Planta*, (2001) **213**, 493.
[80] Abramov, A.; Hoffmann, T.; Stark, T. D.; Zheng, L.; Lenk, S.; Hammerl, R.; Lanzl, T.; Dawid, C.; Schön, C.-C.; Schwab, W.; Gierl, A.; Frey, M., *Phytochemistry*, (2021) **192**, 112947.
[81] Kaufmann, R.; Long, A.; Bentley, H., *Nature*, (1984), 338-340.
[82] Senger, N. A.; Bo, B.; Cheng, Q.; Keeffe, J. R.; Gronert, S.; Wu, W., *The Journal of organic chemistry*, (2012) **77**, 9535.
[83] Gronert, S.; Garver, J. M.; Nichols, C. M.; Worker, B. B.; Bierbaum, V. M., *The Journal of organic chemistry*, (2014) **79**, 11020.
[84] Grzech, D.; Hong, B.; Caputi, L.; Sonawane, P. D.; O'Connor, S. E., *ACS synthetic biology*, (2023) **12**, 27.
[85] Kugel, S.; Baunach, M.; Baer, P.; Ishida-Ito, M.; Sundaram, S.; Xu, Z.; Groll, M.; Hertweck, C., *Nature communications*, (2017) **8**, 15804.
[86] Matsuda, F.; Wakasa, K.; Miyagawa, H., *Phytochemistry*, (2007) **68**, 2290.
[87] Lynch, J. H., *Horticulture research*, (2022) **9**, uhac052.
[88] Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H., *Proceedings of the National Academy of Sciences of the United States of America*, (2015) **112**, 14599.
[89] Fejzagić, A. V.; Gebauer, J.; Huwa, N.; Classen, T., *Molecules (Basel, Switzerland)*, (2019) **24**.

Appendix

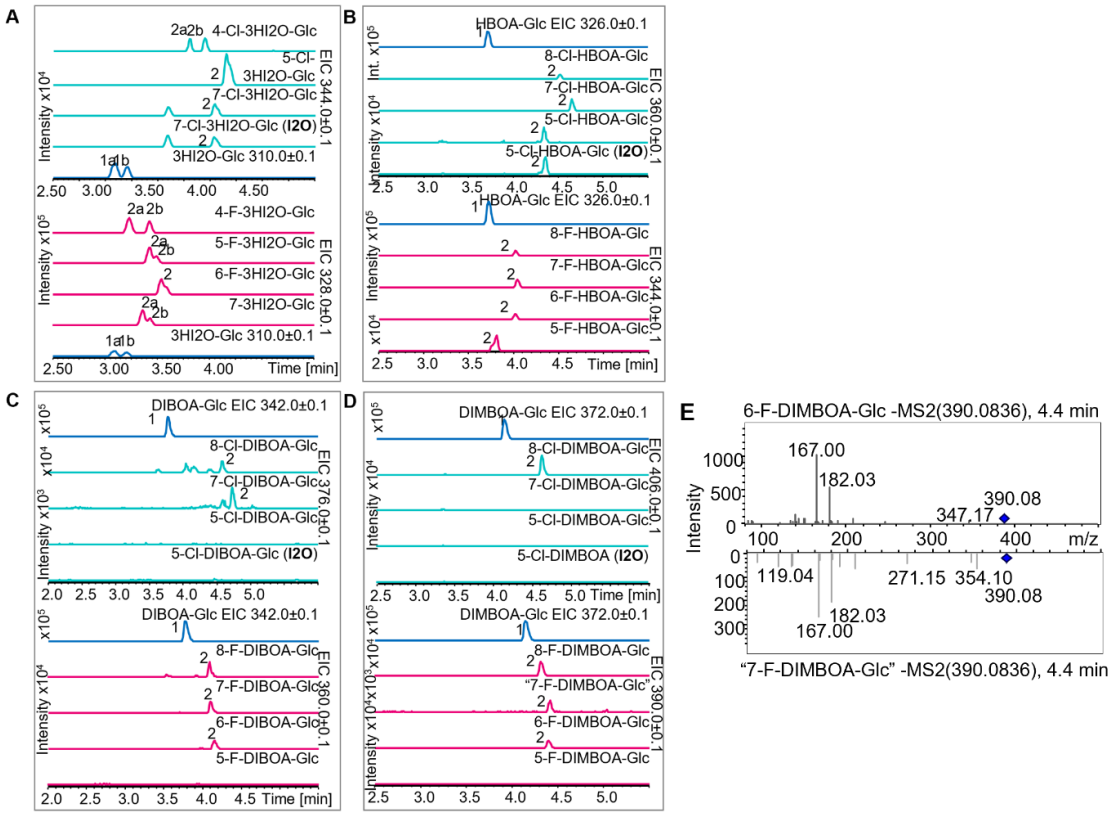

Figure S 2: LC spectra of 3HI2O-Glc (**A**), HBOA-Glc (**B**), DIBOA-Glc (**C**), and DIMBOA-Glc (**D**). Chlorinated BXDs (top half), fluorinated BXDs (bottom half). 1 – non-halogenated BXD (Authentic standards from HBOA-Glc onwards; 3HI2O-Glc produced starting from indole; 2 – halogenated BXD; a+b – putative R and S enantiomers. (**E**) MS/MS fragmentation of 6-F-DIMBOA-Glc (m/z 390) starting from 6-F-indole and the putatively fluorinated DIMBOA-Glc derivative starting from 5-F-indole expressing ZmBx2-8, CoBx4, WrBx4, LgBx5, AsBx6, LgBx8.

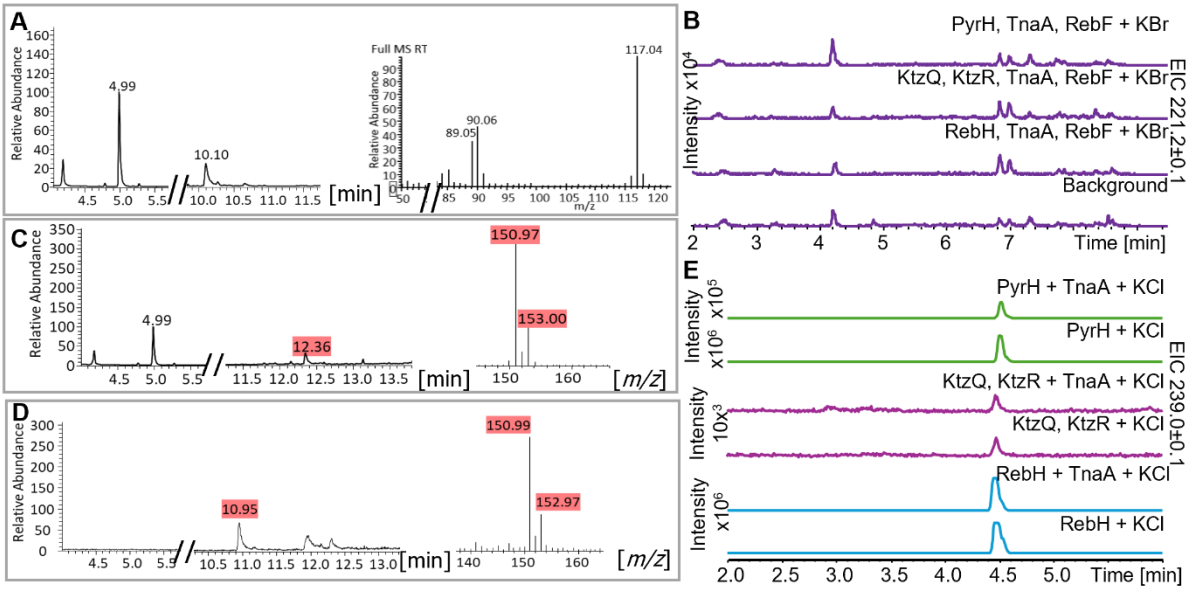

Figure S 3: **A. GC-MS analysis of plants transiently transformed with RebH-3LRS and using indole as a substrate. No halogenated indole is detected.** In samples prepared as leaf disk assays, the susbtrate is the only detected compound. **B. No detection of brominated tryptophan, when infiltrating the halogenases PyrH, KtzR/Q, RebH, with the tryptophanase (TnaA) and flavin reductase (RebF). Chlorinated indole was detected in samples from PyrH (C) and RebH (D), infiltrating KBr. Production of Cl-tryptophan *in planta* (E).**

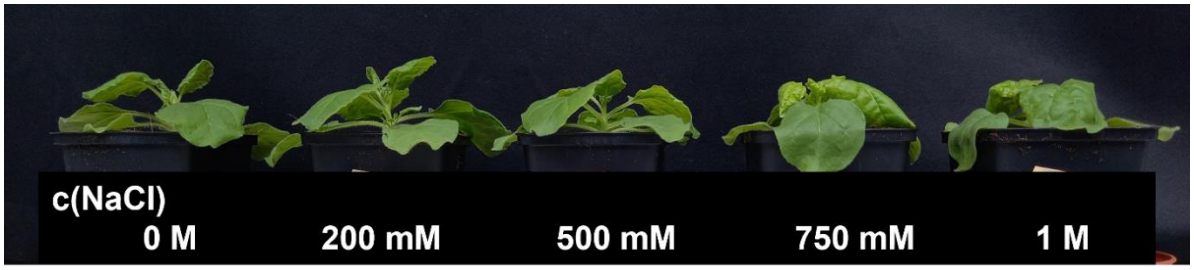

Figure S 4: Fig. Sx: watering N. benthamiana with different concentrations of NaCl.

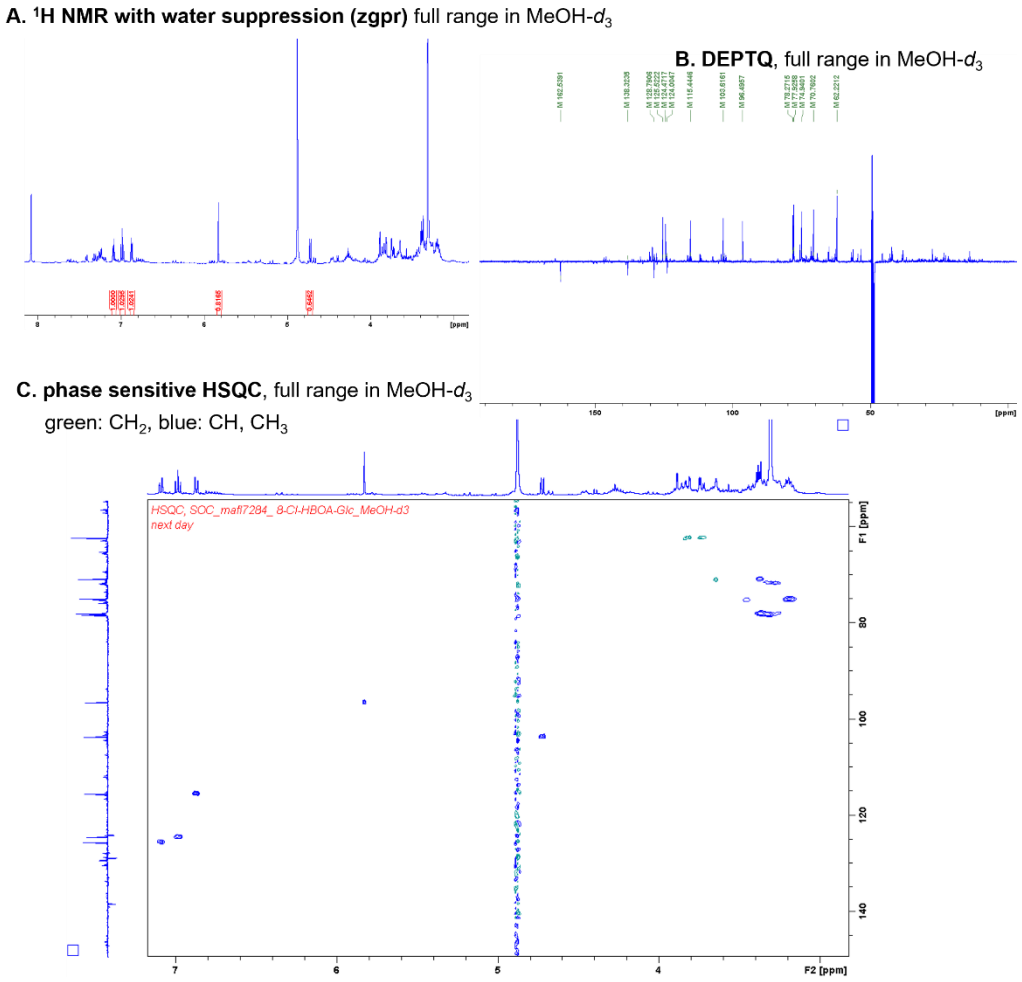

Figure S 5**:** NMR data for 8-Cl-HBOA-Glc from ¹H NMR, DEPTQ and phase sensitive HSQC measurements. NMR analysis was kindly performed by Yoko Nakamura.

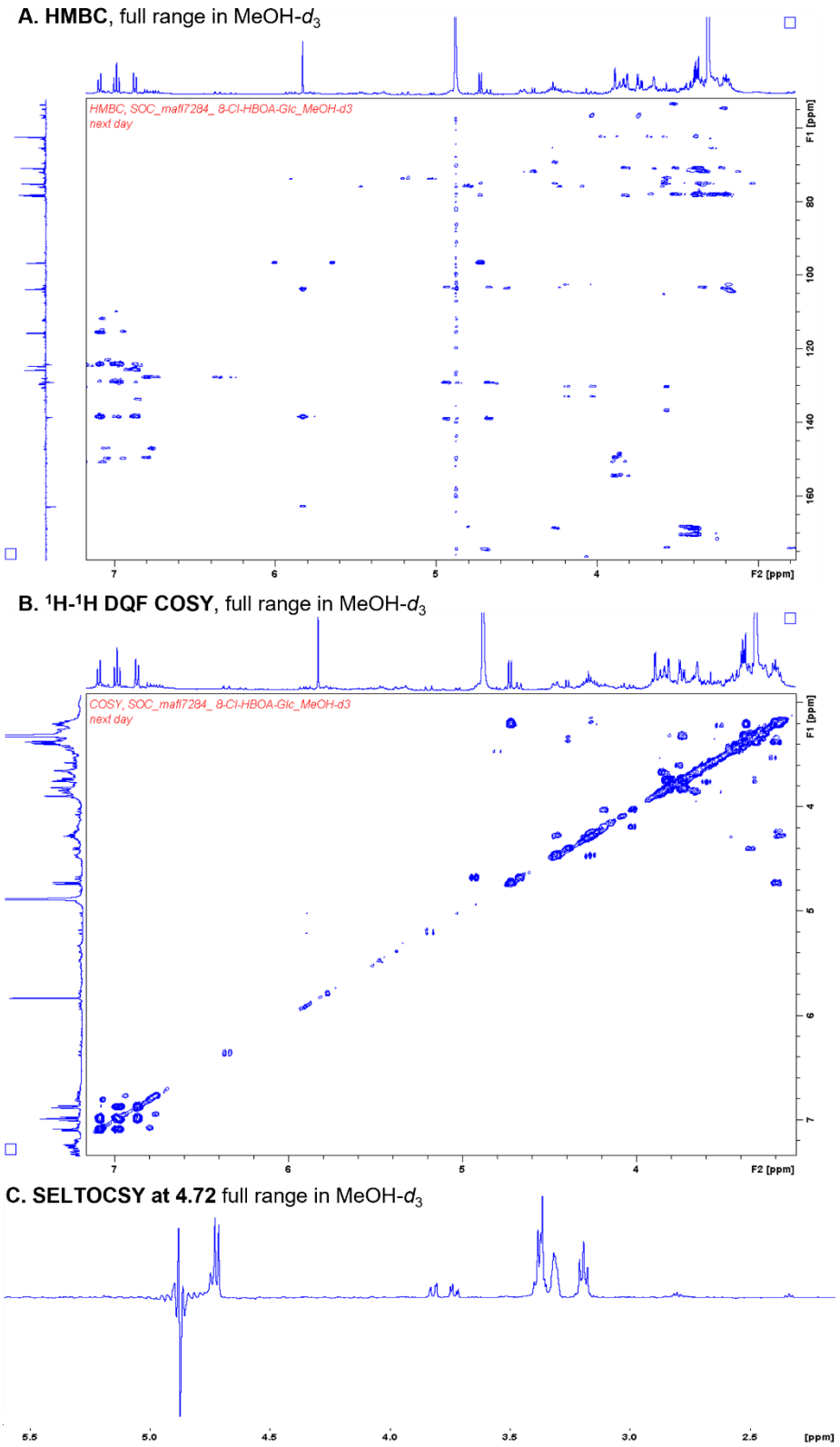

Figure S 6: NMR data for 8-Cl-HBOA-Glc from HMBC, ^{1}H-^{1}H DQF COSY and seltocsy measurements. NMR analysis was kindly performed by Yoko Nakamura.

A. ^{1}H NMR with water suppression (zgpr) full range in MeOH-d_3

B. DEPTQ, full range in MeOH-d_3

C. Phase sensitive HSQC, full range in MeOH-d_3

green: CH_2, blue: CH, CH_3

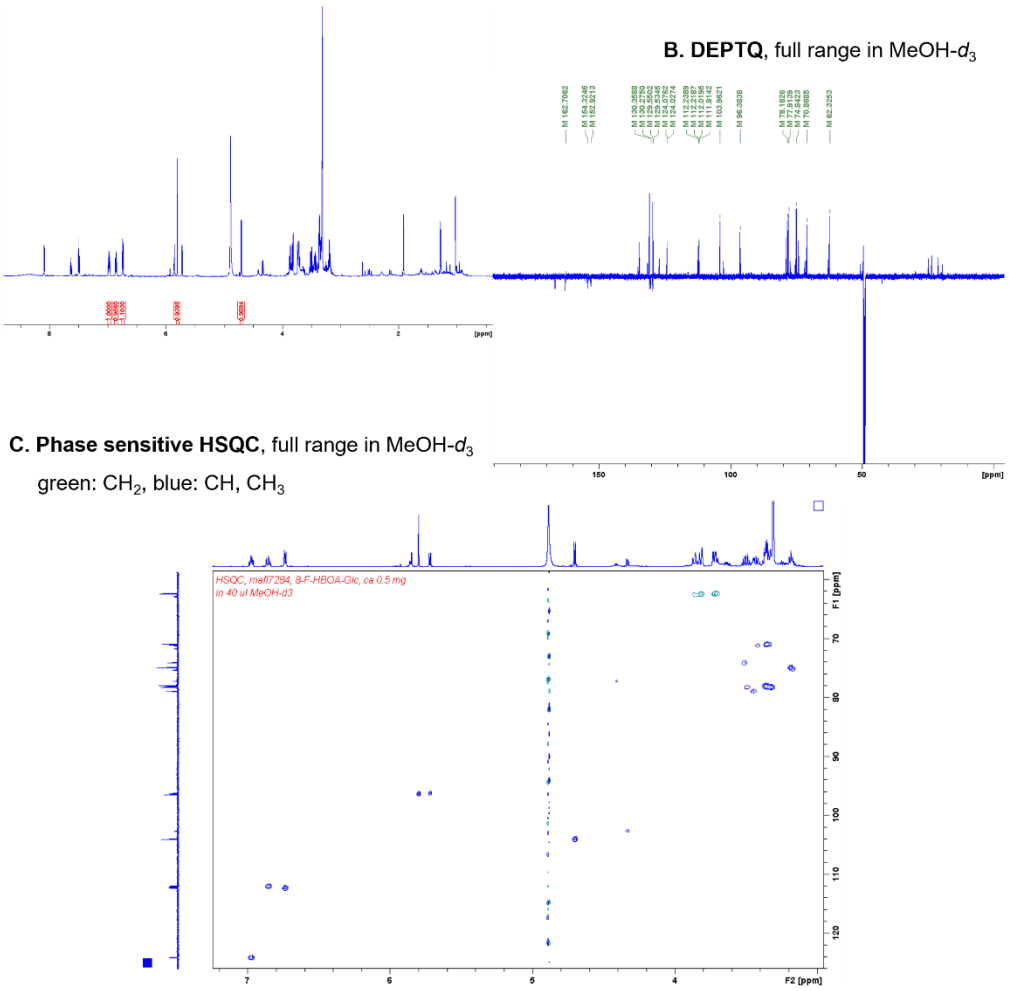

Figure S 7: NMR data for 8-F-HBOA-Glc from ^{1}H NMR, DEPTQ and phase sensitive HSQC measurements. NMR analysis was kindly performed by Yoko Nakamura.

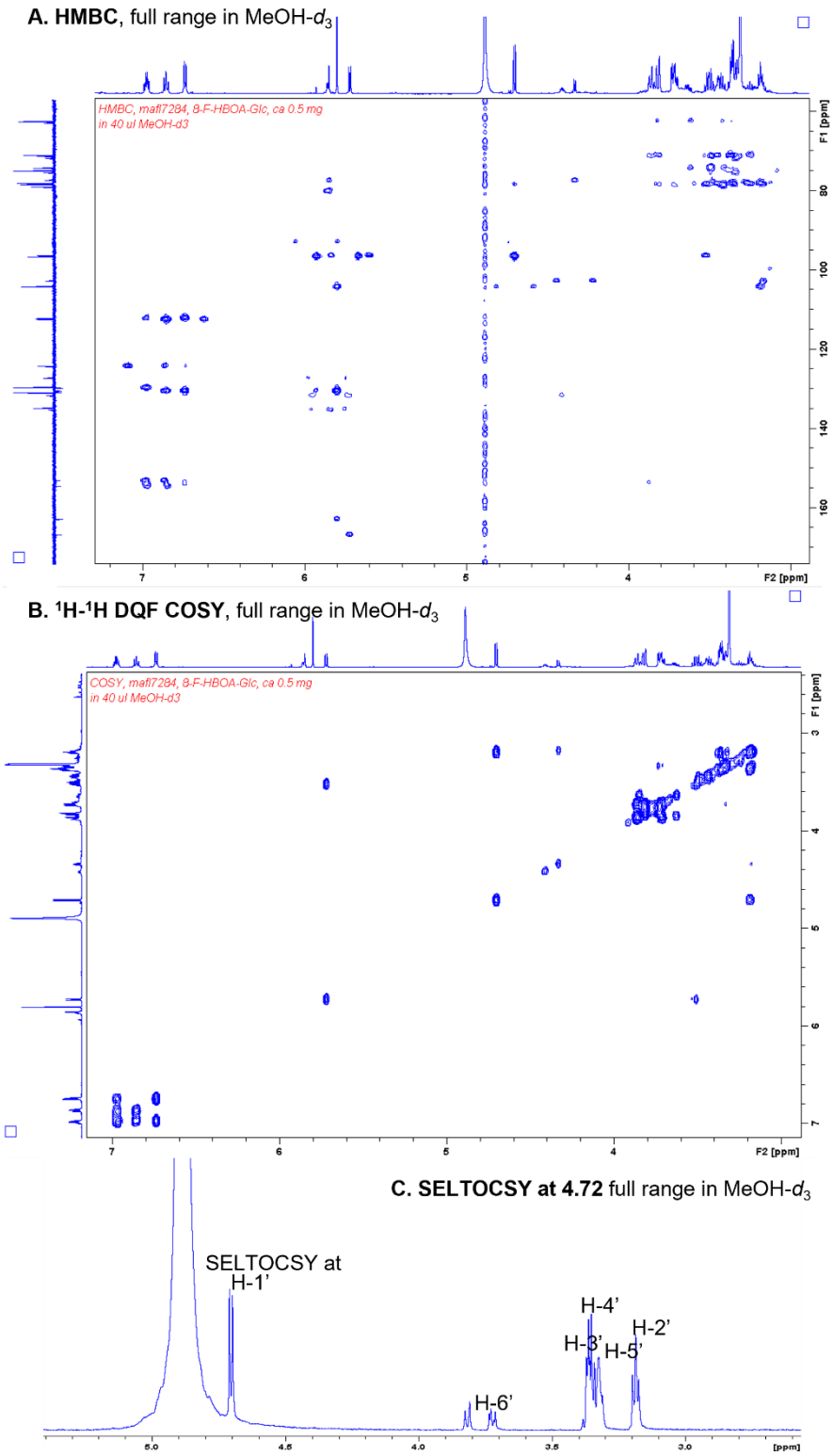

Figure S 8**:** NMR data for 8-F-HBOA-Glc from HMBC, 1H-1H DQF COSY and seltocsy measurements. NMR analysis was kindly performed by Yoko Nakamura.

Table S 1: Sequences already cloned into the respective vectors provided by M. Florean and L. Caputi.

Provided by Matilde Florean
Zm_Bx2:Ω ATGGCTGCTCAACTGCATCACGCACTATACGAGCTGCTGCACGAAGCAGCAGCAGCGCAACGAGCTCTGCTGCTCGCGATCCCCTTCTCC CTGCTGCTCCTGCCGCTCCTGCTCCGCTACCTCGCAGCGTCGGCGTCGGCGTCGGCGACGAAGAACGACGGAGCGGCGCCGGCCAGC GACCCCGACAAGCTCTTGAGCCTGCTGCCGTCGCCGCCCATGAAGCTCCCCATCATCGGGCACCTGCACCTGATGGGGGACATCCCCTA CGTGTCCCTCGCCGCGCTCGCCACCAGGTACGGCCCGGACCTGATGCTGCTCCGCCTCGGCGCCGTGCCGACCGTGGTGGTCTCGTC GCCGCGCGTCGCCGAGGCCGTGCTCCGCACCTACGACCACGTCTTCTCCTCCCGCCCGCGCTCCCTGGTGTCCGACATCATCATGTACG GCGCCACCGACTCGTGCTTCGCGCCCTACGGCGACCACTTCCGCAAGGCGCGGAAGCTGGTCACCGTGCACCTGCTCAACGCCAGCAA GGTGCGGTCGCAGCGGCCCGCGCGGGAGGAGGAGGTGCGCGGCGCGCTCGACAGGGTGCGCCGCCGCCGCGCGGGAGCCC CTGGACATGAGCGAGCTCCTGCACTCCTTCGTCAACAACCTCGTCTGCCGCGCCGTCTCGGGCAAGTTCTCCATGGAGGAAGGCCGCAA CAGGCTGTTCAGGGAGCTCACCGACATCAACGCGGGACTCCTCGGAGGGTTCCACATCCAGGACTACTTCCCCAGCCTGGGCCGGATCG AGCTCGTCAGGAAGGTGGCGTGCGCCAAGACCAGGCGCGTCAGGAAGCGCTGGGACGACCTCCTCGACAAGCTCATCGACGACCATGC GGCCAGGATGGCAACCCATCAGGACGAGGACGACGATAAGGACTTCATATACGTATTGCTGTCTCTTCAGAAGGAGTATGGCCTCACCAGA GACCATATCAAGGCTATCTTGATAGACATGTTCGAGGCCGGAACCGACACCTCGTACATGACGTTGGAGTTCGCCATGACGGAGCTCATAC GGAAGCCACACCTGATGAAGAAGCTGCAGGAAGAAGTGCGGCGGAACGTGCCGGCGGGGCAGGAGATGGTCACCGAGGACAACCTCC CCGGCATGACCTACCTCAAGGCCGTCATCAAGGAGACGCTCCGCCTGCACCCGCCGGTCCCGCTCCTCCTCCCGCACTACTCCCTGGAC GCCTGCGAGGTCGCCGGCTACACCATCCCGGCCAACACCCGCGTCGTCGTCAACGCCTGGGCGCTCGGCCGCCACAGCGGCTACTGG GAGCGCGAGAACGAGTTCGTCCCGGAGAGGTTCCTCAGCGGGGACGTGGCCGGCGGCGTCGACCTCAAGCCCAACGAGTTCCAGTTC CTGGCGTTCGGGTCCGGCCGGAGGATGTGCCCCGGGGTCCACTCGGCGTCCGCGACCATCGAGGCGATGCTGTCCAACCTCATGTACC GGTTCGACTGGCAGCTGCCGGCGGGGATGAAGGCGGAGGACGTGGACATGACCGAGGTGTTTGGGATCACGGTCTCGAGGAAGGAGA AGCTGCTCCTAGTCCCACAGGCTGCGTGA Zm_Bx3:Ω ATGGCCCTTGGAGCTGCGTACCATCACTACCTGCAGCTCGCCGGCGACCATGGCACCGCCACACACGCACTGCTCCTCGGCGTACTTATC TTCCTCGTCATCCGCCTAGTATCCGCAAGGCGAACCGGCACAACCTCAGCCAACAAACGTAAGCAGCAACAGCGGCTTCCGCTCCCGCCT TCGCCCCCGGGCAAGCTGCCCATCATCGGCCACCTCCACCTCATCGGCGCCGAGACCCACATCTCGATCCGGGACCTGGACGCCAAGC ACGGGCGCAACGGCCTGCTGCTCCTCCGCATCGGCGCCGTGCCCACCCTGTTCGTGTCCTCGCCCAGCGCCGCCGACGCCGTCCTGC GCACCCAGGACCACATCTTCGCGTCGCGGCCGCCGTCCATGGCCGCCGAGATCATCCGCTACGGGCCCTCGGACGTCGCGTTCGTGCC CTACGGCGAGTACTGGCGGCAGGGCAGGAAACTCCTCACCACGCACATGCTCAGCACCAAGAAGGTGCAGTCCTTCCGCCATGGCCGA CAGGAAGAGGTGCGGCTCGTCATGGACAAGATCCGCGCGGCTGCCACCGCCGCGCCGCCGGCGGCGGTGGACCTGAGCGATCTCCTG TCCGGGTACACCAACGACGTGGTGTGCCGCGCGGTGCTGGGCGCGTCGCACCGGAACCAAGGCCGGAACAGGCTGTTCAGCGAGCTG ACAGAGATCAACGTGTCCCTCCTCGGCGGGTTCAACCTCGAGGACTACTTCCCCCCGAATATGGCCATGGCGGACGTGCTCCTCCGGCT GGTGTCCGTCAAGGCCCGGCGGCTCAACCAGAGGTGGAACGACGTGTTCGACGAGCTCATCCAGGAACACGTGCAGAGCAGGCCCTCC GGCGAGAGCGAGGAGAGCGAGGCGGACTTCATACACGTTCTGCTCTCCATACAGCAAGAGTACGGCCTCACTACGGACAATCTCAAGGC CATCTTGGTGGACATGTTCGAGGCAGGTATAGAGACATCTTACCTGACGCTAGAGTACGGCATGGCCGAGCTCATCAACAACAGGCACGTC ATGGAGAAGCTACAGACGGAGGTGAGGACGACGATGCGCTCGCCGGACGGCAAGAAGCTGGACATGCTAGCGGAGGAGGACCTCGGCA GCATGCCCTACCTCAAGGCGACCATCAAGGAGACGCTGCGCCTGCACCCGCCGGCGCCGTTCCTCCTCCCGCACTACTCCACCGCCGA CTCCGAGATCGACGGCTACTTCGTCCCAGCCGGGACGCGCGTCCTCGTCAACGCCTGGGCGCTCGGGAGGGACCGCACGACGTGGGA GAAGCCGGAGGAGTTCATGCCGGAGCGGTTCGTGCAGGAACCGGGCGCCGTGGACGTCCACATGAAGGGCAAGGACCTCAGGTTCATC CCGTTCGGCTCCGGCCGGAGGATCTGCCCCGGGATGAACTTCGGATTCGCGACCATGGAGATCATGCTGGCCAATCTCATGTACCATTTT GATTGGGAGGTTCCGGGCTCTGCAGCAGGGGTGTCCATGGACGAGTCCTTTGGGCTGACGCTTCGCCGGAAGGAGAAGCTTCTCCTTGT TCCAAGGATTGCTTCCTGA Zm_Bx4:Ω ATGGCTCTCGAAGCAGCGTACGACTACCTGCACGTCGCCGTCGTCCAGTGCACGCCCACACAAGCAGCAGCAGTCCTCGGCGTCCTCCT CCTCCTCGCCATCCGGCTAGCCGCGGCGGCGAGGAGCTCGTCCGCGACGTCCCCCAAATGGAAGCAGCACAGGCTGCCGCCGACGCC TCCGGGGAAGCTCCCCATCATCGGGCACCTCCACCTCATCGGCTCCCACCCGCACGTCTCCTTCCGCGACCTCCACGCCAAGTACGGCC ACAACGGGCTGATGCTGGTGCAGGTCGGCGCCGTGCCCACCATCGTGGTGTCCACGCCGCAGGCCGCCGAGGCCGTGCTCCGCACGC ACGACCACGTGCTGGCGTCCCGCCCGCGGAACCCCGTCGCCGACATCATCCGCTACAACTCCACGGACGTGGCCTTCGCGCCGTACGG CGAGTACTGGCGCACGGCGCGGAAGGTGGTCAACACGCACCTGCTCAGCGCCAAGATGGTCTTCTCCAAGCGGCGGGAGCGGGAGGA GGAGGTGCGCCTGGTGGTGGCCCGGATCCGCGACGCCGCCGAGGCGTCCCCGGGGACGGCGCTCGACATGACGGAGCTGCTGGGCG GCTACGCCAGCGACTTCGTGTGCCGCGCCGTGCTCGGCGAGTCCCACCGGAAGCAGGGCCGGAACAAGCTGTTCCGGGAGCTGACGG AGACCAGCGCCGCCCTCCTGGGCGGGTTCAACGTGGAGGACTACTTCCCGAAGCTGGCGGACGTGGACCTGTTCCTCCGGATCATCTG CGCCAAGGCCAAGGGCGTCAGCAAGCGGTGGGACAGCCTCTTCAACGAGCTGCTGTCCGAGTACGCGCTCTCCGGCGGCAAGCAGGG CGACCACAACTCTGAAGACTTCGTGCACCTCTTGCTGTCGCTGCAGAAAGACTACGGTCTCACCACGGATAACATCAAGGGCATACTTGTG AACATGTTCGAAGCAGCCATCGAGACATCATTCCTGGTGCTGGAGTACTCCATGAGCGAGCTGATGAACAACCGCCACGTCCTGGCCAAG CTCCAGAAGGAGGTGAGGACGGCGACGCCCGACGGCCGCATGGTGATGGAGGAGGACCTGAGCCGGATGCCATACCTGAAGGCCACCA TCAAGGAGTCCATGCGCATCCACCCGCCGGCGCCGTTCCTCCTGCCGCACTTCTCCACCCACGACTGCGAGATCAACGGCTACACCATC CCCGCCGGCACGCGCGTCATCGTCAACGCGTGGGCGCTCGCCAGGGACCCGACGTGCTGGGACAAGGCCGAGGAGTTCTTCCCGGAG CGCTTCCTGGAACAAGGCCGCGACGCCGAGGTCGACATGTACGGCAAGGACATCCGGTTCGTGCCGTTCGGGGCGGGGCGCAGGATCT GCGCGGGGGCCACGTTCGCCATCGCCACCGTCGAGATCATGCTCGCGAACCTCATCTACCATTTCGACTGGGAGATGCCCGCCGAGATG GAGCGCACCGGGGCCAAGGTGGACATGTCCGATCAGTTCGGGATGACGCTTCGCCGGACGCAGAAACTTTACCTTGTTCCTAGAATTCCC AAATGA Zm_Bx5:Ω ATGGCACTCCAGGCAGCCTACGAGTACCTGCAGCAGGCCGTCGGCCATGGCGCGTGGTCGTCCACGCAGACGCTGACGCTGCTGCTCA TCGCCGTACCCACCGTACTACTGCTGCTAGCGTCCGTCGCCAAGAGCACGTCGTCGTCCGGTAAGGGCAAGCCGCCGCTCCCTCCCTCG CCGCCGGGCACCCTCCCCATCGTGGGGCACCTACACCACATCGGGCCCCAGACCCACATCTCGCTCCAGGAGCTGGTGGCCAAGTACG GGCACAACGGGTTCCTGTTCCTCCGCGCCGGCGCCGTGCCCACCCTGATCGTGTCGTCGCCCAGCGCCGCCGAGGCTGTGATGCGCAC CCACGACCACATCTTCGCGTCCCGGCCGTGGTCCATGGCCTCCCACATCCTCCGCTACAACACCTGCGACGTGGCCTTCTCGCCGCTCG GCGAATACTGGCAGCAGACCAGGAAGCTGATGAACACGCACCTGCTCAGCAACAAGAAGGTCTACTCCTTCCGCCATGGCCGCGAGGAA GAGGTGTGCCTCGTCGTGGACAACCTCCGCGAGGCGGCCGCCAAGTCGCCGTCGACGGCCGTGGACATGAGCGAGGTGCTGGCGGCG TACACCAACGACGTGGTGAGCCGGTCGGTGCTGGGCTCGACGCACCGGAAGAAAGGCCGGAACACGCTGTTCAGGGAGATGACCATGA CCAACGTGGACCTCCTGGTGGGGTTCAACCTGGAGTACTACATCCCGCGGTGGCCGCTCACGGACCTGCTCTTCAGGCTCGTGTGCTGG AAGGTCACGCGCCACCTCAAGCGCTGGGACTCCCTGCTGGAGGAGGTGATCCACGAGCACGTGGAGATGAGGAAGCTGTCCGGCGACA AGGAGAAGGAGTCGGACGACTTCATCGACATCTTCCTCTCCAGATACGAGGAGTACGGCTTCACCATGGATAACGTCAAGTCCCTCCTCAT GAACGTGTTCGAGGCAGCCATCGAGACCTCATATCTGGTGCTGGAGTCCGCCATGGCCGAGCTCATGAACCACAGGCGCGTCATGAAGA AGCTGCAAGCGGAGGTACGGGCGTACGGAGCGGAGAAGAAGCTGGACATGATCAGGGAGGACGACCTGAGCAGCCTGCCGTACCTAAA GGCGTCCATGAAGGAGGCGCTGCGGCTGCACCCACCGGGGCCCCTGCTGCTGCCGCACTACTCCACCGCCGACTGCCAGATCGACGG GTACCACATCCCCGCCAACACGCGCGTCCTCGTGAACGGCTGGGCCATCGGCAGAGACCCGGCGGTCTGGGAGAAGCCCGAGGAGTTC ATGCCGGAGAGGTTCATGCGCGACGGCTGGGACAAGTCCAACAGCTACAGCGGCCAGGACTTCAGGTACCTGCCGTTCGGGTCTGGGC

GCCGGATCTGCCCTGGGGCCAACTTCGCGCTCGCGACCATGGAGATCATGCTCGCCAACCTCATGTACCATTTCGACTGGGAGGTCCCC
AATGAGAAGGAAGACGGTGGCGGGAAGGTGAGCATGGACGAGACGTTCGGGCTGATGCTTCGCAGGAACGAGCCGCTCTACCTTGTTC
CTAGGGCCGTCTAG
Zm_Bx6:Ω
ATGGCTCCAACGACCGCCACCAAGGACGACAGTGGCTACGGCGACGAGCGCCGGCGCGAGCTGCAGGCGTTCGACGACACGAAGCTG
GGGGTGAAAGGTCTGGTGGACTCCGGCGTCAAGTCCATCCCGTCCATCTTCCACCACCCGCCGGAAGCCCTGAGCGACATAATTTCTCC
GGCGCCTCTGCCGTCGTCGCCACCCTCCGGCGCCGCCATCCCGGTGGTGGACCTCTCGGTGACGCGGCGCGAGGACCTAGTCGAGCA
GGTGAGGCACGCGGCCGGGACGGTGGGCTTCTTCTGGCTGGTGAACCACGGCGTGGCGGAGGAGCTGATGGGCGGCATGCTCCGCGG
GGTGCGCCAGTTCAACGAGGGCCCCGTGGAGGCGAAGCAGGCGCTCTACTCCAGGGACCTGGCTCGCAACCTGCGCTTCGCCTCCAAC
TTCGACCTCTTCAAGGCCGCTGCGGCCGACTGGCGCGACACGCTCTTCTGCGAGGTGGCTCCGAACCCGCCGCCGCGGGAGGAGCTG
CCGGAGCCACTGAGGAACGTGATGCTGGAGTACGGCGCGGCGGTGACGAAGCTGGCGCGGTTCGTGTTCGAGCTGCTGTCCGAGTCCC
TCGGCATGCCCAGCGACCACCTGTACGAGATGGAGTGCATGCAGAACCTGAACGTGGTGTGCCAGTACTACCCGCCGTGCCCGGAGCCG
CACCGCACCGTGGGCGTCAAGAGGCACACGGACCCTGGCTTCTTCACCATCCTCCTGCAGGACGGCATGGGCGGCCTGCAGGTGCGCC
TCGGCAACAACGGGCAGAGCGGCGGCTGCTGGGTGGACATCGCTCCCCGGCCTGGGGCGCTCATGGTCAACATCGGTGATCTTCTGCA
GCTCGTGACCAATGACCGGTTCAGGAGCGTGGAGCACCGGGTGCCAGCCAACAAGAGCAGCGACACGGCGAGGGTCTCCGTGGCGTC
CTTCTTCAACACGGACGTGAGGAGGTCCGAGAGGATGTACGGCCCCATTCCAGACCCCAGCAAGCCCCCGCTGTACAGGAGCGTCAGG
GCCCGAGACTTCATCGCAAAGTTCAACACCATAGGACTCGATGGCCGCGCGCTCGACCACTTCAGGCTCTAG
Zm_Bx7:Ω
ATGGGGCACCAGGCGCAGCACGGCACGGACGACACCGAGGAGCTGCTTGCAGCCCACCGGCAGCTGTGGTGCCATGCCCTGGGCTAC
GTCAAGTCCATGGCGCTCAAGTGCGCCCTGGACCTGCGCATCCCCGACACCATCGACCGCTGCGGCGGGAGCGCCACCCTGGGCGAG
CTGCTCGCCGCAAGCGAGATCCCGGCGTCCAACCACGACTACCTCCGACGGGTCATGCGCACGCTGACAGCCATGCGCATCTTCGCGGT
CAGCCACGACGACCCCGCCAAGGCCGACGACGCGGCCGCCATCTCTTACCAGCTCACCCCGGCGTCCCGGCTGCTCGTCAGCAGCAGC
AGCAGCAGCGTCGACGCCGCCGCCGCCGGCGCCGGCGCCTCGAAGGAGAACACTACTACCCCGAGCATCCTCCCCAACATCGCCCACC
TGGTCCGCCCCAACACCATCTCCCTGCTGTTCAGCATGGGCGAGTGGATGAAGGACGAGAGCGCGGCGTCCGTGTCGCTGTACGAGAC
GGTGCACCGGCAGGGGATGTGGGCGTGCGTGGAGGACGACGCGGCCAACAGGGCCAGCTTCTACGAGAGCATGGACGCCGACACCCG
GCTCGTGATGCAGGCCGTGGTCCGCCGGTGCCCGCACGTGTTCGACGGCATCAAGTCGCTCGTCGACGTCGGCGGCGGCCGCGGCAC
CGCCGCCGCCGCCGTCGTCGCCGCCTTCCCGCACATCCAGAGGTGCACCGTCATGGACCTCCCCCACGTCGTCGCCGAGGCTCCCGCC
GGCACCGCCGGCCTGTCCTTCCACGGCGGCGACATGTTCGAGCACATCCCGTCGGCTGACGCACTCATGCTCAAGTGGATCCTGCACGA
CTGGGATGAGGACAAGTGCATCAAGATAATGGAGCGCTGCAAGGAAGCGATCGGTGGCAAGGAAGCAGGGGGGGAAGGTGATAATAATAG
ACACGGTGCTCGGATCTCGGGCAGACGACGACGACGACGACAAGACGTGCAGGGAGACGTATGTGTTGGATCTCCACATACTGAGCTTC
GTGAACGGGGCGGAGCGGGAGGAGCACGAGTGGAGGAGGATCTTCCTGGCGGCTGGGTTCCGTGACTACAAGATCACGCACACCCGA
GGCATCCCGTCCATCATCGAGGTCTTCCCG
Zm_Bx8:pCambia
ATGGCAGCATCGTGCGGCGGCCGCGTGGTGGTCTTCCCGTTCCCGTTCCAGGGCCACTTCAACCCGGTGATGCGCCTGGCCCGCGCGC
TGCACGCCCGGGGCGTCGGGATCACCGTGTTCCACACCGCCGGCGCGCGGGCGCCGGACCCGGCCGACTACCCCGCCGACTACCGCT
TCGTGCCCGTGCCCGTGGAGGTGGCCCCGGAGCTGATGGCGTCCGAGGACATCGCCGCCATCGTCACGGCGCTCAACGCCGCCTGCG
AGGCGCCCTTCCGGGACCGCCTCTCGGCGCTGCTCTCCGCCGCCGACGGCGAGGCGGGGGAGGCGGGGGGCCGCGTCCGCTGCGTC
CTCACCGACGTCAGCTGGGACGCCGTGCTGTCGGCGGCCCGCGGCCTCGGCGTGCCCGCGCTCGGCGTCATGACGGCCAGCGCCGC
CACGTTCCGCGTCTACATGGCGTACCGCACCTTGGTCGACAAGGGATACCTGCCGGTGAGAGAGGAGCGCAAGGACGACGCGGTCGCC
GAGCTACCCCCGTACCGCGTGAAGGACCTGCTGCGGCACGAGACGTGCGACCTGGAGGAGTTCGCGGACCTGCTGGGCCGCGTGATC
GCGGCGGCGCGGCTGTCCTCGGGGCTCATCTTCCACACGTTCCCCTTCATCGAGGCCGGCACGCTGGGCGAGATCCGGGACGACATGT
CGGTGCCGGTGTACGCCGTGGCGCCGCTCAACAAGCTGGTGCCGGCGGCCACGGCCAGCCTGCACGGCGAGGTGCAGGCGGACCGG
GGCTGCCTGCGCTGGCTGGACGCGCAGCGGGCGCGCTCCGTGCTGTACGTGAGCTTCGGGAGCATGGCGGCCATGGACCCGCACGAG
TTCGTGGAGCTGGCGTGGGGGCTGGCCGACGCCGGCCGCCCCTTCGTGTGGGTGGTCCGGCCCAACCTCATCCGCGGCTTCGAGTCG
GGCGCGCTGCCCGACGGCGTGGAGGACCGGGTGCGCGGCCGCGGCGTCGTCGTCAGCTGGGCGCCGCAGGAGGAGGTGCTCGCGC
ACCCGGCCGTGGGCGGCTTCTTCACCCACTGCGGCTGGAACTCCACCGTGGAGGCCGTGTCGGAGGGCGTGCCCATGATCTGCCACCC
GCGCCACGGGGACCAGTACGGCAACGCGAGGTACGTGTGCCACGTCTGGAAGGTGGGCACGGAGGTCGCCGGGGACCAGCTGGAGAG
AGGGGAGATCAAGGCCGCCATCGACAGGCTCATGGGCGGCAGCGAGGAAGGGGAGGGCATCAGGAAGAGGATGAACGAGCTCAAGATC
GCTGCGGACAAGGGCATCGATGAATCTGCTGGGTCGGATTTAACTAATTTGGTTCATCTCATAAACTCCTACTGA
As_Bx2/3:Ω
ATGCAGATGAAGAAAACGGTGGCTATCGTTGGTGCTGGAATTAGTGGCCTCCTTGCCTGCAAATATGCCCTCTCCATCGGCCTCAATCCCG
TCGTCTTCGAGGCTGAAAGCTGCGTGGGCGGCCTCTGGAACCACACCATTGAATGCACCAGGTTGCAGAGCCTTAAGGACTATTTCCAGT
TCTCCGATTTCCCATGGCCGTCTACGATCAAGACCATGTTCCCTAAGAACGAGCAGGTGCTCGAGTACGTCAACTCCTACGCCGATCATTT
CCAGTTGCGGCAGTACGTAAGGTTCAATTCCCAAGTCATGAGCATAGGCTACGAGGGGGAGTCCAAGGAAGAAATGGGGAGCTGGAAATC
GTGGGGCGGGACTGGGATGGCTTTCGGCTCAAAAGGGAAGTGGAGGCTACAAGTTGAAGCTGGCGGTGGCATGCAAGAGTACCATGTC
GCCGACTTTGTGATTGTCTGCATTGGAAAGTTCAGCGGCTTGCCAGCAGTCCCGGACTTCCCTCCGGGGAGTGGTCCGGAGGTCTTCTC
CGGCAAGGTCCTCCACTCCATGGAGTACTCCGACATGGACAACGCCAGCGCCGCCCAGTTCATAAAAGGGAAAAGAATCGCCGTCATCG
GCTCGGGCAAGTCGGCCATTGATATAGCATTCGAGTGCGGACGTGCCAACGGTCGCGACACGCCTTGCACAGTCATCCATAGAAATGCTC
ATTGGATGTTGCCTTACGAGAAGCCGTGGCTGCTCGCTTTCGCCCTCTTATGCTTCACCCGCTTCTCCGAGCTATTGGTCCATAAGCCCGG
TGAAGGCTTCTTTGCTAGCGTATTGGCCACCTTGCTCTCGCCTTTGAGATGGCTGGTGTCAAAGTTTCTGGAGATCAGTTTGCGATGGAGA
CTTCCCTTAAAAAAGTACGACATGATTCCATCGGAGAACTTCCTCGAAGAAGCCAATTCTTGTCAAGTAATCATGTTGCAGGACAATTTCTTC
GACAAAGTGGTCGATGGGAGTATTGTTTTTAAGAAATCAGAGCACTTTAGCTTCTGCAAAGAAGGCTTGATCGTTGAGGGCGAAGATAGCC
CTGTAGAGGTTGACATAGTCATCTTCGCCACCGGATATAAAGGCGATGAGAAGCTCAAGAATATCTTTGCATCTCCGGCTTTCCAGAACCAC
ATATTTGGATCTCGTAACTCCACCGTCCCTCTTTTCAGGCAAATGATCCACCCACGAATTCCACAGCTCGCCGTGGTAGGCTACTCTGAGTC
CCTCTCAAATTTGTACACATTCGAAATGCGATGCAGATGGCTAGCATTCTTTCTGGACCGAGCCTTTGCTTTGCCGAGCATAAAGGAGATGG
AGGACGACATCAAGATGTGGGAGAGGTACATGAAGAAGTATGCTGGGAACGGGGAGTTCAAGCGTGGCTGCATTGCCGGTATCCACATCT
GGTACAACGATCAACTCTGCAAAGACATTGGGTGCAGTCCAAGGAGAAAGAAGGGGATCTCGTCGGAATTGTTCCAGCCATACGGACTGG
AGGACTACGATGGAATCACTCCGGCTCAGATCAACCCCAGTTATTTTGCCAAGTGA
As_Bx4:Ω
ATGGCCAAAGTCTGTAACCAAATTCTCCACACAGGCTCGTCATTCTCAATGGGCATGCAGATGGAGGGGTTGTTGTGGTGGGTACTCGCC
GTGGCTTTGGCATACATAGTAGCATCCTTCTTCTTCAACAATGGGAAGAAACCAAAACTTCCCCCAGGCCCTAAACCATGGCCGATTATTGG
CAATTTGAACCTCATTGGACCCATCCCGCATCAATCCCTTCACGATCTGTTCCAAAAATACGGCGAGTTGATGCACCTGAAATTCGGCCATT
TCCCGGTGGTGGTCGCGTCCACGCCTGAGATGGCCAAGCAGTTCCTGAAGACGCACGACACCGCCTTCGCCTCGAGGCCGGCGCTCGA
CGCGGGGCGGTTAACGAGCTACAACTACACGGACTTGACTTGGGCACCGTACGGGCCGTACTGGCGTCTGGCGCGCAAGATCTACGTCG
TGGACGTGTTCAGCGCCAAGAAGCTGGACTCGGACGCCTTCCACCAGATCCGCGTGGAGGAGAGGCAGACCTTCTTCTCCCTCATGCAC
TCGCTTGCGGGGAAGCCCGTGGTGCTCCGAGGCCACCTGGGTCGCGCCGCCGTGATCACCATAAGCCGGATGGTGATGGGGAAGGCTT
ACTTTGGTGAGGACGAATCGTGCCTTATAGGGACCGATGGGTTTCAGGAGCTTCTGGATGAGTGGTTCTTCCTCAACGGGGTGTTCAGCA
TCTGTGATTGGATCCCGTGGCTCCGTTTCATCGACCCTCAGGGATACGTGAGGAGGATGAAGAACCTGCACAAGGAACTGGACCGGATTT
TCGACTTCGTGGTCGCTGATCACGAAGCCACCAGGATGAAAGACACTGATTCATGCTCCGAAAAAGATGTCGTCTACAAGATGCTGGACAT
GACTGAGGAGCCCACTCTTGGAGTCGCACTTTCAAGGGAACATGTCAAGGCATTATTGCAGAACCTAATACTCGGAGGCACGGATACGAC
ACCAACCTCGGTGGAATGGATCCTGTGCGAATTAATGAAGAATCCTCCCATAGTTCGAAAGGCAAAAGAAGAGCTGAATAGGGCAGTTGGA

AGGGACAGATGGGTAGAAGAAAGTGACTTCCCTGAACTCCCATACTTAGATGCCATCATCAAGGAGACGATGAGATTGCATCCGATCGCCA
CGTTTCTAGCGCCTCACTGCGCCATGGAGGACTGCAAGGTCGCCGGCTACGACATACCCAAAGGCGCGACAGTCTTCATCAACTCATGGA
GCCTGGGGAGGGACCCTGATTCATGGGACAGGCCCTTGGAGTTCTTGCCAGAGAGGTTCATGGAGAAGGAGATTGATATAATGGGAAGTA
ACTTTGCGATGCTGCCATTCGGGGCGGGCCGGAGAAGATGCCCTGGATACGGTCATGCGCTGAAGAACATCAGGGCAATCGTGGCTAAC
TTGCTGCATGGATTCGAGTTGAAGTTGCCGAATGGGATGAAGCCTTCAGATGTGAATATGGAAGAAGAGTACGGACTCACCACACATCCTA
AGGAACCCCTTGCACTCATCATGGAACCTACCCTCCCACCCCACCTTTACTAA
As_Bx5:Ω
ATGGAAGGCAGTCTGCTGTACACAATTATTACAGCCGCTCTGGTCATCATCTTTGTCCTTTTCCTGAGTCACCGGCGGAGATCACGTGGCT
ATCAGCTGCCTCCGAGCCCGGCGGCGGCCCTCCCGATCATCGGTCACCTCCACCTGCTGAAGGAGCCACTCCACCAGACGCTCCAGCG
TTTCTCCGAAGCCTGCGGGCCGATCTTCTTGCTCAGGCTAGGCGTACGCAAGGTCGTCGTCGTGTCGTCGCCGGAACTGCTGGAGGAAT
GCTACACAGCGTACGACGTCGCCCTGTCCAACCGGCCGCGCATCATGTCGGACAAGTACATCGGCTACGACCACACCACCATGGCCGGC
GCGCCTTATGGCGACAGGTGGCGCAGCCTCCGCCGCCTGGCAGCGCAGGAAGTTCTCTCCACCGCGAGGATCAACTCGTTCACGGAGA
TCAGGCACGACGAAGCTAAGAGGACGCTGCAAAGCATGGTGAGCCGCGTGGGGGAAAGCGGGAAGCTACAGCTGAGCCTGAGACCCGA
GCTTTACGACCTGATCTTCAACGTAATAATGAGAATGTTGACAGGGAAGAGATACATAAAAGGTGAGTACCTGTCCCATGACCGACTGGCCG
ACGGATACAGGGAGATCATGAGTGAGTTTTTTGAGAATGCCCAGACGTCGAATCCTGAGGATTTCTTGCCTTTCCTGAAATGGGTCGATTAC
CGTGGTCTGAAGAAGAAGATGGCTTCTCTGGGTGGTCGGCTGGACAACTTCTACCAGAGCTTATTGCAGGAGCATCGACAGGAGAAGAG
GAATACCATCATCGGACATTTACTATCTCTGCAAGAATCCGATCCTCAATTCTACGACGACCAACTTGTCAAAGGCTTTATGACGAACATGAT
CGTTGGCAGCCTGGACACATCCGTGGTGACCATAGAATGGGCTATGTGCCTTTTGCTCAACAATCCGAACATCTTCAAGACAGCTAGACAG
GAGCTCGACTCGCAAATAGGCCTTGATCGTCTTGTTGAGGAACAAGACCTATCCAACTTGCCTTACCTTCATGGCATCATCATGGAGACACT
ACGGATGTTCCCACCTGGGCCGCTGATGTTGCCTCGTGAGTCGTCTACCGACTGTGTGATTGGGGGATATGATGTCCCAAATGGAACCAT
CGTATTGGTCAATGCTTATGCTATACAACGAGACCCTAAGTCTTGGGACGACCCCGCGAGCTTCAAGCCTGAGAGATTCATAGGGAAGGAT
GTTGAGTTGCGGAAGATGTTGCCCTTTGGCGTGGGTAGAAGGTCGTGCCCCGGTTCAGTTTTGGCCCATCGGATGTTGGGACTCATTTTG
GGATCGTTGATTCAATGCTTTGATTGGGAGAGGATGAACTCAGAAAAAGTTGATTTAACTGAGGGGGCTGGACTAACGCTGCCCAAATTGC
AACCACTAGAGGCCATATGTAAACCAAGAGTGAGCATGGTTAAGGTTATTCAAGAATCAGATAATGTGGATTAG
As_Bx6:Ω
ATGGCGGCAGCAAACCTCGTAAGTGGAACCTCTGTGACAACCATCCCAAGCAAATACATTCTGTCTGAATTGGACAGGCCGAAGCTGTCTC
AAGTTATCGAATCCAATGACAACATTCCTGTCATCGACTTGGCTTTGGGAGACGATGCCCTAATCACGAAACAAATGGGGGAGGCTGCTCG
GGAGTATGGCTTCTTCCAAATCATAAATCACGATGTACCAAAGGAAGTTGTGGAGGGAATCATAGGAGTGGCTCGTGAATTCTTCAGTTTAC
CTGTGGAAGAGAAGATGAAGTTGTGCTCAGATGACACTTCAAAACCTACAAGATTATCCACAAGTTTTAATTTGAACAATGAGAGTGCTCAG
AGCTGGAGGGATTACCTGAGGATTCAGTGTTATCCCTTGGAGAAGAACGTGGATGAGTGGCCATCAGTCCCCTCCTCTTTCAAGGATATTG
TTAGCAATTACTACAAGGAAGTGAGACAAGTTGCCTTCAAACTAGAGGAATACTTGTCAAAGAGCTTGGGATTAGAGAAAGACAGCCTCAAA
ACTCTCCTAGGTGAAGATGAACAAGGACAACACATGACCATCAACTATTATCCACCATGCCCGCAGCCCGACCTAACCTTGGGTATCCCAC
CACATACAGATCCTAATACCCTCACCATTCTTCTCCCCGACACCGATGTTTCTGGCCTTCAAGTCCTCAAGGATGGCAAGTGGATGGCAGTA
AAACCCTGCCCTGGTGCCTTTATTATCAACATCGGTGACCAACTCCAAGCAATGAGCAACGGGCAGTACAAGAGCGTTTGGCATCGAGCTA
TTGTGAATATAGATAAGCCAAGGATTTCAATACCATCCTTCATATGCCCAAGCAACACAGCAAAAATAAGAGCTCCAGATGATCTTATCAGTAA
GGGCCAAACTGCACAATATACAGAGTTCACCTTTGCTGAGTATTATGACAAGTTTTGGAAGCAGAGAAAGGAACATCAACATGGCTTAGACC
TCTTCAAGAATTAA
As_Bx7:Ω
ATGATGGATCATCATGAGTCCGATGAGCAAGAAGCTGCTTTCTCATTTGCTTGGCATCTATCCCATGGTGCTGTGCTTCCAATGGCAGTCAA
ATGCGCCATAGAGCTCGATCTCTTCCAACTCATTAAGAAAGCTGGCCCTGATGCTGCTGTTTCTGCTTCTCAACTTGCAGCTCAGATTCCAA
ATACCAGAAACCCTGATGTCGTCGCCATGATCGAAAAACTTCTCCGCTTCCTGGCTGCAAACTCCGTTCTCAGCTGCCAACTCACACCCAT
GCCGGACCTCACGCTTCAGCGGACCTACTCCCTCACGCCTGTTTGCAACTATTTGATTGCAGATGAGGATGGGGCTTCCCTGGCTCCGGC
GTGTCTCATTAATCACGACAAGGTTTCGATGGAGGCTTGGTACCATTTGAAAGATGCAATTGTTGAGGGAGGGATTCCATACAATAGAGCAT
ACGGAATGAGCAGCTATGAACATCTGGGCAAGAACCCAAGAACGAACCAGCTCTTCAACCAATCTATGTATGGCGCTTCGACCATCACTAT
GAAGAGAATTCTTGAAAACTATCCAGGGTTTGAGGGCGTGAAATCCCTGGTGGATGTTGGCGGTGGAACTGGAGCTTCACTCAAGTTGAT
CCTTGCCAAACACCCATCCATTAAGGGCATTAACTTTGACTTGCCCCACGTTGTCCAAGATGCTCCATCTCATCCAGGGATTGAGCATGTTG
GCGGGGACATGTTTATGAGCGTGCCACAAGGGGACGCAATCTTTCTGAAGTGGATCTGCCACAATTGGAGCGATTCACAGTGTGAGAAAC
TATTGAAGAACTGCTATGAAGCACTCCCAGAAAATGGAAAAGTGATCATTGTTGAAACCATTCTTCCCAATTCTGACCCAAAAAAGAGTGCA
CCAGAAGTTGAGAAGTTGGCTTATCAAGCAGCAGCAGGGTATGATCTGATTATGTTTGCTTTTAATCCAGGTCGAGAGAGGTCAGAAGAGG
AGTTAGAGGATTTAGGCAAACATGCTGGCTTTAAACATTTCCACAAAGTTTGCCATGCATTTGGATACTGGATCATGGAATTTCTTAAATGA
Lg_Bx2/3:Ω
ATGGAGATCATTTCTTGGTTTCTATTGGCCTTACTATATGTTGCTTCCCTAGCCCTAATCTCTAAAATTCGACGCCCTCCGATTCAAAATCTCC
CACCAGGCCCAAAACCATGGCCTATAATTGGTAACCTAAACCTTATTGGCTCAATCCCACATCAATCCCTTCACTATCTATCTCAAAAATATGG
CGAAATAATGCTACTAAAATTCGGTAAATTTCCCGTCATAGTAGCATCATCTCCTAACATGGCCAAACTATTCTTAAAGGTGCATGATGCAGAT
TTTGCCTCAAGGCCTTTACTTGCCGCTGGAAAATACGCCAGCTACAACTACTCCGACTTGACGTGGGCCCCTTACGGCCCGTATTGGAGAC
AAGCGCGTAAAATCTACCTATCGGAGGTGTTMAAMAMAAAAAAAWTTGAATTTTTCGAGCAAATTCGTGTTCAGGAGGCGAGTAATTTGTT
GTCCCGTCTACGACAGCTGTCAGGAAAGCCGGTCGTGCTAAGAGACGAGTTGCAACGGTACACGCTATCGTYTATAAGTAGGAYGTTATCA
AGAGACGTATGCTTTACTTCAGAACAGGATGACTTACGAGGGTTGATAGATGAGTGGTTTTTTCTGACTGGGGTGTTCAATATTGGGGATTG
GATACCGTGGTTGAGTTTTCTTGATTTGCAGGGATATGTTAAGAGAATGAAGGTATTGAACAAGAAACTTGATAGGGTTTTTGATTATGTGATT
GATGAACACTTAGCTAAAAGGGCTAAAGATAAAGATTTCGTCGCAAAAGATATTGTAGATGTGTTCTTGCAGATCTCTGAGGATCCTAATCTT
CATGTTGAACTCACAAGATATCATGTCAAGGCATTGTTACAGAACCTAATGGTCGGGATCACAGATACCTCGGCTACAACAATCGAGTGGGC
TTTGAACGAAGTTCTCAGGCATCCTCTCATTATGAAAAAGGCGAAAGAAGAGATCGACGGGGTAATTGGAAGAAACAGATGGGTGCAAGAG
ACCGACKTGTCTCAATTGCCTTACATAGAATCYWTYATCATGGAAACATTCAGGTTGCATCCACTTGCAACATTCTTAGCCCCTCATTGCGCT
ATCAATGACTGCAATGTCGCAGGATACGACATAGCTAAAGGGACGATGGTCCTCATCAACACATGGAGCATAGGGAGAGACTCTAACTCAT
GGGATGCACCAGAAAAATTCCTGCCAGAGAGGTTTATGAGAGAAAAAATCGATGCACTGGGTAATAATTTTTCCTTTTTGCCATTTGGGTCG
GGTAGGAGAAGATGTCCCGGGTACAAACTTGCACTTAAGATGATTCCAACAACGTTTGCCAATTTGTTGCATGGATTTAACTTCAAATTGGTA
GAGGGGATGAAGCTTCAAGATATATGCATGGAGGAAGAGTATGGACTTACTACACAYCCTAAAGTGCCTCTTTCCATCATCATGGAACCTAC
ACTTCCACCCCACCTTTATTAA
Lg_Bx4:Ω
ATGATGTTAATGGAGAAAAGAGTAGCCATTATTGGAGCAGGAATTTCTGGGCTTCTTGCCTGTAAATACACTGCTTCTATAGGCTTAAAACCC
ATCGTTTTCGAGGAACAAGATAATGTGGGCGGGTTATGGAATCATACAATCGAGATCACAAGACTTCAACAGGTTAAACAATCTTTCCAGTTT
TCTGATTTTCCATGGCCTTCCTCTGCTGATGACAACCCAAAATCTGCCCAGCTTTTGGAATATCTCCAATCTTATGCCCACAATTTTGACCTT
TTGCCCTTCATAAAGTTCAATTCCAAAGTGATGAACATAGATTATGTGGGAGAATCCAAAGAAGAGATGCAGAGTTGGGATTTGTGGGGTGG
ATCTGGAAAAGCTTTTGGGTCAAAAGGGAGGTGGATTCTCAAAGTTCTTCACACCAAAGACGACTCTATTAAGGTGTATGAAAGTGAATTTG
TAGTTGTCTGCACTGGTAGGTTTAGTGGGCTACCAAATATCCCGGAGTTTGCTCCGGGATATGGTCCTGAAATATTTTCAGGGCAAGCGTTG
CATTCAATGGACTACTCTAATATGGACAATGCAAGTGCTGCACAGTTTATTAAAGGAAAAAGAATTGCCATCATTGGTTCAGGAAAATCAGCG
TTCGACATTGCATTTGAATGTGCAAATGCCAACGGTAATGATAATCCTTGCACACTGGTCCAAAGAACGATTCATTGGACGTTTCCTAGTACA
AAAACATGGGGCTTGAATTTCGCTTTCTTGATCTTCACTCGGTTTTCTGAGCTAATGGTTCGTAAACCGGGCCAAGGGTTCGTATCTAGTGC
GTTGACGACATTACTTACACCTCTGCGATGGGGGATGTCCAAGTACGTGGAGAGCTATCTAAGGTGGAAACTTCCGTTGAAAAAGTACGGA
ATGGTTCCGAAAGAAAGCTTTGCTGCAGAATCCTCTGCTTGCCAGATATTCTACCTCCCTCGTGGAAACTTCTATGGCAAAGTTGAAGATGG
AAGCATTGTTTTAAAGAAATCACAAGAATTTAGATTCTGCAAACATGGCTTGATCATTGATGGAGATGATAAGAATCCCATAAACGCAGATATC

GTCATCTTCGCCACTGGATACAAAGGCGATGAGAAGCTCAAGAGCATATTTGCATCTCCAACCTTTCAAAACTATATTGTTGGATCTCCAACC
TCCACAGTTCCTCTCTATAGACAGATGATCCAACCAAGAATTCCACAGTTAGCAGTAATAGGGTACTCGGAGTCATTTTCGAATCTTTTCACG
TTCGAAATGAGGTGCAAATGGTTGGCCTTCTTTCTCGACCAGGCCTTCGAGTTACCGAGTATAAGAGCTATGAAGAAGGAGGTGGAAGCGT
GGGAGACGTACATGAAGAGGTACGCGGGAAACAATAAGTTCAGACGAGCTTGCATCGGTCGTGTACACGTGTGGTACAATGATCAACTATG
CAGAGATATTGGATGCAATCCAAGAAGAAAGAAGGGATTTTTATCTGAGTTGTTTGAGCCTTATGGATCGGCTGACTATGTTGGACTATAG
Lg_Bx5:Ω
ATGGCATGGAGGTGATAGACTTCACCACCCTTCTTGCCTTTTCATCTCTCATCTTCTTACTAATCAAGGCATGGCGAAACCCAAACCCACAA
AATTTACCTCCAAGCCCTCCGAGCCTTCCGGTGATCGGCCACCTCCACCTTCTCGGCGGAGGGGCAATCCAGGCATTGGCAAGAATAAGG
AAGAAATACGGCCCCATCGTGTCTCTCAAGCTCGGAGAAGTTACTACGGTGGTGATTTCGTCGCGAGAAGCGACTAAAGAGGCGCTGAAA
ATCCACGACCCCGCCTGCGCCGACAGGCCCGACAGCACAGCTGTGAAGATAATGTTCTACGGCTGCGAAGACATCTCCTTCGGTCCCTAC
AACGAGAATTGGAGGCAATTGCGTAAAATATGCATTATGGAGATGCTTAGTCCTAAGAATGTCAAATCTTTCGGGTACATTAGGGAAGAGGA
GATCGATTCTCTCCTAGAATCTCTGCGATTATCGTCTGGAGAAGCTGTCAATCTCACGGAAAAGATATCAACCTTAACTTTTGCGGTGACTTG
CAGAGCGTCGTTTGGTAGAGTGTTGAAGGGTCATGGTACGGTTGTTGCCTTGTTGAAGGAGCTATCGCCGATGTCTACGGGATTTCAGGT
GGCTGATTTGTTTCCGTCTATGAAGTTTTTGCATCCCTTCAGTTTGAATAAATACAGGGCGCAACAAATGCGACGTGAGATGGACGACATTC
TTGATACCGCCGTGGAGGAACAAAAATTGAAGAAGAGTGGCGAGTTTGAAGCCGAAAATTTTGTGGATGTTCTTCTCAGGGTGCAGAAAAA
TAAAGAGCTCAAATTCCCCATCACCACCGACAACATCAAATCGGTCATTTTGGACATATTCGCCGGTGGAATTGAGACAACGGCTACCACTA
ATGACTGGGCAATGGCGGAGCTGATGAGAAATCCGCGTGTGATGGCGAAAGTACAGTCGGAAATAAGGGAGGCTTTGAAGGGAAAGACG
ACGGTGGAAGAAAAGGATATACAGAGCCTAAAATACCTGAAATTGGTTGTCAAAGAAAGTCTAAGGCTTCATCCTCCGGTCCCTCTTCTCCC
TAGAAAATGCAGGCAAGAATGCAGAGTGGGTGGGTTTACAATTCCCAACAAAGCTAAAGTTTTGATCAACGTCCACTCCCTGGGTCGGGAT
CCCCTGTACTGGAATGACCCGGAAACCTTTCTACCCGAAAGATTCGAAAACAGTTCACTAGATTTCCTGGGGAATGACTGCGAATTCCTTC
CATTCGGGGCGGGTAAAAGGAATTGCCCGGGCCTGAATTTCGGGATAGCAAACGTCGAGTTTCCGTTGGCTCAACTTCTTTACCACTTCGA
TTGGAAGTTGCCGCAAGGAATGAGTCATGCTGATGTGGACATGACCGAGGCTGAAGGAATGGCTCTCTTAAGAAAGAGTCCTCTTATGGTG
ATTCCTACACCATATAAGCAGCCTCTTTGA
Lg_Bx8:Ω
ATGAGTTCATTTTCAAAAGCAGAAAGTAGTGAACAGAAACCCCATGCGGTATTCATACCTGCACCAGGACAGGGACATATCAACCCCATGCT
CATGTTAGCAAAGCTCCTCCACTATAACGGCTTCCACGTCACCTTTGTCAACACCGAGTTCATTCACCGCCGCCTCCTTAAGTCCTGCGGC
CCTTCCGCCCTCCGCGGCCTGCCGGGCTTCCGCTTCGCCACCATTCCCGACGGTCTCCCTGAATCCTCCACTCGAGACGTGGCTGCCCT
TTGTGAGTCCTTCACCACCACCTGCTTGGAGCCCTTCTGTGACCTGCTTTTGGAGATCAAAAGCTCCGGCGGGGGCGTGCCGCCGGTGA
GCTGCATAGTCGCCGATGGGGAGATGAGCTTCACGCTGAAGGCGGCGGAGAGATTTGGGCTGCCGGAGGTGCTGTTCTGGACCACAAG
TGCTTGTGGACTCTTGGGGTATACGCAGTATCGTAGACTTGTTGAGAAAGGCTACGTCCCTCTTCAAGACATGAGCCAGGGGTATTTGGAC
ACAACAGTTGATTGGGTTCCAGGAATCAAGAATATCCAACTCAGAGATTTCCCAGCTTTCATCAGAACTACTGATCCTAAAGACTTCATGCTT
AACTTTAAAACACAAGAGGTTGAGGCAATTCCCAGAGCGAAAGCTCTGATCCTCAACACCTTCGACGCATTGGAGCAAGATGTTTTGAATG
CTCTCTCAGACATGTTTCCAACAATTTACACCGTCGGCCCTCTTCAACTCATGATGAATCATGTCCACGACCAAACACTGGATTCCATCACC
TCTAGCCTCTGGAAAGAAGACTTCCAGTGCATCGAATGGCTCGATTCTAAAGAGCCTGGATCTGTGGTCTACGTGAATTTTGGCAGCATCA
CTGTGGTTACTGCTCAGCAGCTAACCGAATTCGCGTGGGGCCTAGCTAATAGCAACAAACAATTCTTATGGGTTATCAGGCCTGATATAGTG
GCAGGGGAATCGGCCCTGGTGCCGCCTGAATTTGTGGAAAAGACTAAAAATAGAAGCCTGCTGATCAGCTGGTGTCCGCAGGAGCAGGT
CCTGAAACACCCTGCGATCGGAGGGTTCTTGACGCACAGTGGATGGAATTCGACTCTCGAAAGCGTTATCAATGGGGTGCCCATGATTTG
CTGGCCTTTCTTCGCAGAGCAGCAGACGAATTGTCGGTTCAGCTGCGTGGAGTGGGAGATGGGAATGGAGATAGACAACGATGTGAAAA
GAGACGAGGTGGAGTTTCTTGTGAGGGAGCTTATGGATGGAGAAAAGGGCAAGAAAATGAAGGAGAAGGCTATGGAGTGGAAGAACAAA
GCAGAAGCAGCAACTGTGCCTGGTGGTTCTTCATACTTGAATTTCGAGAAATTGATCAAGGAGCTTTTGCCATAA
Co_Bx2/3:Ω
ATGGCATGGAGGGGCGAGTGGGGATCATAGGAGCTGGACTCAGTGGCCTTCTTGCCTGCAAATATGTCTCACAGATGGGCTTCACTCCGG
TCGTCTTTGAAGCCAAAGCTGAGATCGGCGGGGTGTGGAATCAAACTATAGAGACTACTAAGCTTCAGACTGTCAAGCAGTACTACCAGTT
TATGGACTTCCCATGGCCTAGTTCTGTGAAGGAAGACTTCCCAGATCACAACCAAGTTCTAGATTATATTGAGTCTTACGCTCGCCACTTTG
GTTTGCTTCAGTACATTAAGTTCAACTCAAGAGTCACCTCCATTGACTTTGAGGGAGTATCAGATAAAGAGATGTCGTCGTGGGACTTCTGG
GGAGGTAACAACAACCCCTTTGGCTCCAAAGGGAAATGGAATATCACAGTCGAGGATACGCTGACACAATCCACTGAGGTTTATCAAGTAA
AATTTGTGATACTATGCATCGGAAGGTTTAGTGATGTTCCAAAGATTCCGGATTTCCCACCCAAAAGAGGTCCAGAAGCCTTTGATGGGAAG
GTAATTCATGCGATGGAGTATGCTGCACGTGGTGCTGAATGCATCAGAGGGAAACGGGTCGCAATTGTGGGAATACAGAAATCAGCACTGG
ACATTGCCGTCGAATGTGCGGCGGAAAATGGGTTAGAGCATCCTTGTACTATGATATACCGAAGGGCACATTGGAATTTTCCAGATTACTATC
CATGGGGAGTTCCGCTCGCTTTGCTCTATTTCAATCGATTTGCTGAGTTATTGATTCACAAACCCGGTGAAGGCTACCTGCTCAGTCTCTTG
GCTACGTTACTTTCGCCCCTGAAATGGTCACGGACCAAATTCGTCGAGAGCTATCTCAAGTGGAAGCTTCCCATGAAGAAGTACGACATGC
TACCTGAACATAGTTTCGATAGAGATTTATCTGCTTCTACAATTTCAATCACACCAGAAACCTTTTACGAAAAGGTGAAAAAAGGAAGTATTGT
TTTGAAGAAATCAAGTGGCTTTAGCTTCTGTAAGAAAGGTTTGATGCTAGACGGTGTGGCTGATCCTTTGGAAGTAGATGTTGTGATCCTTG
CCACTGGGTACAAAAGCGACGAGAAACTCAAGGACATTTTTGCATCATCCACATTTAAAGATCACATTATGTCATCCTCAATTCAATTCTACA
GGGAATGTGTACACCCACGAATTCCTCAACTAGCAGTCATCGGGTATTCAGAGAGCTTATTCAATCTGTATCCATTCGAGATGCGGTGCAGA
TGGCTGGGACATCTTCTAAAAGGTATGTTTAAGTTGCCTAGGATTAAAGAGATGGAAATGGATGTCAACAGGTGGGAACAGTATATGAAGAA
ATACACCGGTGAGTACCACCGGAGATCTAGCTCGAGTTGGATACAAACATGGTGTAATGATCAGCTGTGCAAAGACATGAAATGCAACCCG
AGAAGGAAAGATGGAGTTTTGGCAGAACTATTTGAGCCATACGGACCCATGGATTATGCTGATCTCTCTTAA
Co_Bx4:Ω
ATGGCATGTATTTTCATGTCACTCACTCTCACTCAGTGTTCACTCCCTCACTCAGTGTTCACTCCTCAACAATGGAATCACTTAACCAATCTT
ATCCAGTGATTATTTGTTTCTGGAGCTTCGCATTGTTATTATTACTGTACACCATGCCCATATCATGGAAGAGAAGAAGGGATATGGCACCCG
AACCGTCGGGGCGCATGGCCTCTGATAGGCCACCTGCCTCTGCTGATGGGTCACACCTTACCACAGAAAACATTGGGGGCACTGGCCGATA
AGTATGGCCCCGTTTACACCATTCGTCTTGGTTTGCGTAAAGTATTGGTGATCAATAGCTGGGAGGCTGCCAAGGAATGCTTCACCACCAAT
GACAAGGCTTTTGCACACCGCCCCAGCTCGGTTGCCTTGAGGATCATGGGCTATAACAATGCCATACTCTTCTATGCATCCTATGGCCCCTA
CTGGCGCGAGGTTCGCAGGATAGCCATATTGGAACTTCTCTCAAATCGCCGGATTGAATTGATAAAGCACATCCAAGTCTCAGAAGTTGAC
GTGTCAATGAGAGAACTGTACAAAATGTGGCAAGCAGGTTGCACTGATGGCGGGTATTCGCTCGTGGAGATGAAACAGTGGTTCATTGACA
TAAATCTGAATATGATTGTGAGGATGGTTGTTGGTGAAAGATACTTCGGGACTAGGATGGACGATGACGGAATACGATTCCAGAAGGCCGTT
CATTCCTTCTTTCATTTGGTTGGGCTTTTTGTCTTGTCTGACGCGGTTCCGTTTCTAGGGTGGTTGGATTTTAGGAAGAAAGAGATGAAAAG
AACTGCAAAAGAATTGGACAGGATCATGCAAAAATGGCTTGACGAACATAAACAGAAAAAACTAGCTGGTGGTGGGAAAGGCGAGCAAGAT
TTCATGGACATGATGCTGTCCGTGCTTGAAGATAAAAAGCTCCCTGATTTCGACGCCGATACCATCAACAAGGCTGTCAGCATGAATATGAT
CTTTGGCGGCGTCGATACTATAACTGTGGGTCTAACCTGGACTTTTTCGTTGCTGATAAACAATCCACGAGTTCTAAAGAAAGCCCAAGATG
AAATTGAGGCAACAGTAGGAAAACATCGGCCGGTGGGGGCGTTAGATGTGGAGAAGCTAGTGTATCTCCAAGCAATAGTAAAAGAGTCATT
ACGTCTGTACCCACCCAGCCCTATGCCGCCACAACATCTCACCGTGGAAGATTGTACTGTGGCTGGCTACCACGTCCCAGCTGGCACATG
CTTGGTGGCAAACATTTGGAAGATCCAACGAGATCCGCGCGTGTGGTCCGATCCTTGTGAGTTCCAGCCTGAGAGGTTCCTTGGGGACG
GAGCAAATGTTGACCTTAGAGGGAAGAATTTTGAGTTTATTCCTTTCGGAGCTGGACGACGTCAATGCCCAGGAATTTCATATGCACTTCAA
GTACTCAACTTAGTATTAGCGCGTCTGTTACAGGGTTTCGATTTCAAGACCAAATCAAATCAAGCCCTGGACATGACTGAAAGCCCGGGACT
GACTAATATGAAAGCTACTCCTATCCATGTATTCATCACTCCTCGTCTCCCTTCGGATCTTTATTGA
Co_Bx5:Ω
ATGGCATGAAGCTATTTAAGGTGAAACAACTACCAGTTTGCTTCAAGACCTTTCTCCACAAACTTAATAGCATGGCTCCTTCCTCCTTCTCTT
GGGCATGGCCTTGCGCTGTGTTTATGCTCGTTGCATATTTTGTATACCTCCTGAGGAAGAAAAGTAGCCTTTGTTTGCCTCCAGGTCCTAAA
GGACTACCCATCGTCGGAAACCTTTTCATGCTAGGTGCACTTCCCCACCGTGTTCTCCATCGTTTTGCCCAGCAATATGGCCCGATAATGCA

CCTACAACTCGGCTTCATCCATACCATTGTCGTTTCCTCCCCTCAGGCTGCTGAATTATTCCTCAGAACACATGACCGCATTTTCTCTGGTAG
GCCTTACAACCAAGCAGCCGGGTACATTTCTTATGATACGAAAGGATTAACTTTCTCTAGTTATGGTAGCTACTGGCGAAACACACGTAGGG
TTTGTGCTCTAGAGTTACTTAACCACCTCAAAATTGATGCCTTCAAACACGTGAGGACAGAAGAGATGAATTTGTTCATTGAGTCCCTCAAA
GAGGCTTCAACTGCTAAGGTTGCGGTGGATTTAAGCACCAAACTGAAGGGTCTTAACACTGATATTTCCTTCCGTATGGTTTTGGGGAAGAA
ATTAGTACACAGGGCTGATCAAGAAAAGTTCAGTGCCGTAGTTCATGAAGGAATGCATTTGGCTGCATTCTTTAACATTGCTGATTATTTTCC
CTACGTAGGTGCACTTAACCTTCAAGGTAAGATGAAAAGAATTAAGGCGGTACACAGGTTCTATGATGATATTCTTGACAAGATCATTGAAGA
GCACGTCCGAGTTGGGCCTGGCGGGGGTACTCCACCGTTTATTGATATGATGATGTCCTACATGGAATCCAAAGAAAATGACTTTCTCAAG
GACCGATCCAATGTAAAAGCCATTGTACTGGACATGCTTTTGGGTGCCATGGATACAGCATCAACTGCGATTGAATGGGTCATGGCAGAGC
TCATTAAGCATCCAAGTCTAATGAAGAAAGTTCGACAAGAGGTAGACGGAGTGGTGGGTTTGGACAGAATGTTAGACGAGTCTGACCTTGT
GCATCTGGAGTACCTAGACATGGTCATCAAGGAGACCTTCCGAGTCCACCCAGTGGCACCGTTACTAATACCTCATAGTAGCATGGAAGATT
GCACTGTTAACGGATACCAAATACCTCGAAATTCTCGCATCATTATCAACACATTTGCCATAGGCCATGACCCCGAAGCGTGGCCTGATCCA
GATGAGTTCAAACCAGAGAGATTTGTTGGTAAAAATATGGATTACAAAGGAAGTGCCTTCCAGTTTCTTCCATTTGGATCAGGGAGAAGGGG
TTGTCCTGGGATAACTATGGGCGTCAGAGTGGTAAGATTGGTTGTAGCACAACTGGTGCATTGCTTTGATTGGGAACTTCCAAATGGTATGG
TATGCACTGAACTGGACATGGATGAAGATTTTGGACTTACAGTAGCTAGAGCCAATCCTCTATTAGTTATTCCCACTTATCGTCTTCGCATATG
A

Sd_Bx4:Ω

ATGGAGAGCTCCTCGGCGCTGTTGCTCTTAGCATTCATAGCTGCAGTCGCCTTTATATCAAGAATAATCAATAGAAAACCCCCCACAAAACTT
AAACATCCTCCAGGTCCAAAACCATGGCCGATTATCGGCAACATGCACCTTATTGGTTCCATCCCACATCAATCCCTTCACTATTTATCCCAA
AAATACGGAGAGCTCATGTTGTTGAAGTTTGGAAAATTTCCAGTTGTGGTTGCAGCATCCCCAGAGATGGCCAAACAGTTTCTAAAGATATA
TGACACGTCCTTCGCCTCGAGGCCTGCCCTTGCCGCTGGGAAATACACAAGCTTTAACTACTCAGACTTGACATGGGCTCCTTACGGCCC
ATATTGGAGACAGGCACGTAAGATCTACCTCACTGAGGTTCTCAACGCAAAAGCAATCAAAACGTACGAGCATGTTCGCATGGAGGAGAGG
CGTACTTTCTTCTCCCGTCTCCATTCCCTTTCGGGGAAGTCAATTGTGCTACGGGAACAATTATGTCGTTACACCCTCTCTACCATGAGCAA
CGTGATATTCGGTCGATTAGTCTTCGACAGTCATGAAGCTAAAGGCGGCAATGCTGTGAAGCTGGATGAGTTGCAAGCGATGCTGGACGA
CTGGTTCTTTTTCGGTGGTGCGCTCAATCTCGGGGATTGGATTCCATGGCTGAATTACATCGACCCGCAGGGGTATGTTAAGAGGATGAAA
GCCCTGAACAGGAAGTTTGACAGATTTTACACTTCCGTGTTTTCTCATCATCTTGCCAATAGGGACCCGTCGGCCCCAAAAATTCCACTAGA
CTCATTGTTGGGGATGCCTGGGGATCCGAATCTTGATGTTAAATTGACTGAGGAAAACGTCAAGGGATTGCTCATGGATTTGCTGGCTGGT
GGCACTGACACATCAGCCACCACTGTTGAATGGGCAGTCCACGAACTTTTGAGGCACCCACACGTGTACAGCAAGGTGAAAGAAGAGATA
GACAGAGTTGTTGGAAGAAATAGATGGGTGGAAGAAGAAGATTTCCCTCAATTAACTTACTTAGATGCCGTCATCAAGGAGTCGATGAGATT
ACATCCACTTGCTACATTTTTAGCCCCGCACTGTGCGATGGAAGACATTAACATAGCAGGCTACGACATATCAAAGGGGACCATGATATTAGT
GAACACTTGGAGCTTAGGAAGGGACCCTAAGGTATGGGATAACCCTGAAAAATTCTCGCCTGAGAGGTTTTTGGTGGAAGATATCGACATAT
TGGGATCGAATTTCGCATTGTTGCCGTTTGGCTCTGGCCGGAGGAGATGCCCTGGCTATAAACTTGGACTGAAACTTGTGAGATCTACGTT
GTCTAATTTGGTGCATGGATTTAACTGGAGATTGCCCGAGGGAATGACTGTCAAAGATGTGTGTATGGAGGAACTGTATGGACTCACCACA
CGTCCCAAAATCTCTCTTGAGATTATTTTGGAGCCCTCACTTCCACTCCACCTTTACTAG

Wr_Bx4:Ω

ATGGAAGTTTTAAATAGCACCAGCCTTGCTCATCGTTTTAGCATCAAGATGGAAAATTCTTGGCTCGTCTTAGCCTTTTCATTGCTCGCAGTG
CTCGTTTTCCTCTCAAAATTAACCAATCATAGCAGCCGCCAGAAGAAAAAACTTCCTCCAGGTCCAAAACCATGGCCTATTATTGGCAACTT
GAACCTTCTGGGATCTATCCCGCATCAATCTTTACACTTGTTAGCCGAGAAATATGGAGATTTAATGTGCCTGAAATTCGGTTCCAAACCTGT
TTTAGTAGCTTCATCTCCAGAAATGGCTAAACTATTCTTGAGAACCCATGACAACGTCTTTGCATCTCGCCCACCTACAGCTAGTGGCAAGTA
CACTGCTTATGATTACACCGATGTTACATGGGCGCCTTACGGTCCATACTGGCGACATGCTCGCAGAATTTACCTGAATCATATATTCAACCC
AAAAAGTCTCGACTCCTTTGAATATCTTCGTGTTCAGGAAAGAAAGGCTTTGATTTCTGGCCTTCATGTGCTCTCAGGAAAGCCTTTTGTGT
TAAGAGAACACCTATCTCGTTACAGTGTCTCTAGCTTAAGCAGAATGGTTATGGGCAACAAGTATTTCTCGGAAACAAAAGTAATCCGTGATG
AATCCATTTGCACCCTGGAAAAGTTGCAGGACATGCTAGACCAGTGGTTTTTGCTTAATGGTGTGTTCAATATTGGGGATTGGATACCATCT
CTTAGCAGATTTGATTTGCAGGGCTATGTGAAGCAAATGAAGGAGTTGTATAAGAATTTTGAAATATTCAATAACCACGTGCTTGATGATCATC
TGACCAAGAGAAATGCAGACAAAGATAATTTTATCCCCAAAGACATGCTTGACATCATATTGAAGATTGCTGAGGACCCTAATCTTGAGATTA
AGCTTACAAGAGATAATGTGAAGGCGTTATTACAGGAGCTATTAGTGGGTGGTTCAGATACCTCGGTTACAACAGTAGAGTGGGCAATGAAT
GAACTTCTTAGACATCCACACCTCATTAAAAAGGCTACCGAAGAGCTTGATAGGGTTATAGGGAGAGATAGATGGGTAGAAGAGGGTGATTA
TTCAAAACTACCTTTCATTGAGGCCATCATAAAGGAAACCATGAGGATCCATCCGCTGGCTACATTGCTTGCACCTCATTATGCCATTGAAGA
CTGCAATGTTGCTGGTTATGATGTTTCTGAAGGAACCACAGTCCTGATAAATATATGGAGTATGGGAAGGAATCCAAAATACTGGGATTCACC
AGAAGAATTCTTGCCAGAAAGGTTTCTTGAAAAGGATATTGATGTGAAGGGACAAAACAACTTTGTCCTAATTCCATTCGGATCTGGCCGGA
GAAGGTGCCCAGGCTACACACTAGGGCTCAGAATTGTTAGAGCAACTTTGGCAAACTTGTTGCATGGCTTCAACTGGAAATTACCTCCTGG
AATGAAGCCAGAAGATGTCTGCGTGGAAGAACTCTATGGACTCACTACTCGACCGAAGATTTCATTGTCGGTCATTGTTGAACCTCGATTAC
CTCTCCATCTCTATTAG

P19:Ω

ATGGAACGAGCTATACAAGGAAACGACGCTAGGGAACAAGCTAACAGTGAACGTTGGGATGGAGGATCAGGAGGTACCACTTCTCCCTTC
AAACTTCCTGACGAAAGTCCGAGTTGGACTGAGTGGCGGCTACATAACGATGAGACGAATTCGAATCAAGATAATCCCCTTGGTTTCAAGG
AAAGCTGGGGTTTCGGGAAAGTTGTATTTAAGAGATATCTCAGATACGACAGGACGGAAGCTTCACTGCACAGAGTCCTTGGATCTTGGAC
GGGAGATTCGGTTAACTATGCAGCATCTCGATTTTTCGGTTTCGACCAGATCGGATGTACCTATAGTATTCGGTTTCGAGGAGTTAGTATCAC
CGTTTCTGGAGGGTCGCGAACTCTTCAGCATCTCTGTGAGATGGCAATTCGGTCTAAGCAAGAACTGCTACAGCTTGCCCCAATCGAAGT
GGAAAGTAATGTATCAAGAGGATGCCCTGAAGGTACTGAGACCTTCGAAAAAGAAAGCGAGTAA

Provided by Lorenzo Caputi

ThdH:pTURBO

ATGGATAACAGGATCAAGACCGTGGTGATCCTCGGAGGTGGAACTGCTGGATGGATGACTGCTGCTTATCTTGGAAAGGCTCTCCAGAAC
ACCGTGAAGATCGTTGTTCTTGAGGCTCCTACCATCCCTAGAATCGGAGTTGGAGAAGCTACTGTGCCTAACCTCCAGAGGGCTTTCTTCG
ATTACCTTGGAATCCCTGAAGAGGAATGGATGAGAGAGTGCAACGCTTCTTACAAGATGGCTGTGAAGTTTATCAATTGGAGGACCCCTGG
TGAGGGATCTCCTGATCCTAGAACTCTCGATGATGGACACACCGATACTTTCCACCACCCTTTCGGACTTCTCCCATCTGCTGATCAGATCC
CTCTCTCTCATTACTGGGCTGCTAAGAGACTTCAGGGTGAGACTGATGAGAACTTCGATGAGGCTTGCTTCGCTGATACCGCTATCATGAA
CGCTAAGAAGGCTCCTAGATTCCTCGATATGAGAAGGGCTACCAACTACGCTTGGCACTTCGATGCTTCTAAGGTGGCAGCTTTCCTCAGA
AACTTCGCTGTGACTAAGCAGGCTGTTGAGCACGTTGAGGATGAGATGACTGAGGTGCTCACTGATGAGAGGGGATTCATCACTGCTCTC
AGGACCAAGTCTGGAAGAATCCTCCAGGGTGATCTCTTCGTGGATTGCTCTGGATTCAGAGGACTCCTCATCAACAAGGCTATGGAAGAAC
CTTTCATCGATATGTCTGATCACCTCCTCTGCAACTCTGCTGTTGCTACTGCTGTGCCTCACGATGATGAGAAGAATGGTGTTGAGCCTTAC
ACCTCTTCTATCGCTATGGAAGCTGGATGGACCTGGAAGATCCCTATGCTTGGAAGATTCGGATCTGGACACGTGTACTCTGATCACTTCGC
TACTCAGGATGAGGCTACCCTCGCTTTCTCTAAGCTTTGGGGACTCGATCCTGATAACACCGAGTTCAACCACGTGAGATTCAGGGTGGGA
AGAAACAGAAGGGCTTGGGTTAGGAACTGTGTGTCTGTTGGACTCGCTTCTTGCTTCGTTGAGCCTCTTGAGTCATCTGGAATCTACTTCA
TCTACGCTGCTATCCATATGCTCGCTAAGCACTTCCCTGATAAGACCTTCGATAAGGTGCTCGTGGATAGGTTCAACAGGGAAATCGAAGAG
ATGTTCGATGATACCAGGGATTTCCTCCAGGCTCACTACTACTTCAGTCCTAGAGTGGATACCCCATTTTGGAGGGCTAACAAAGAGTTGAA
GCTCGCTGATTCTATCAAGGATAAGGTGGAAACCTACAGGGCTGGACTCCCTGTTAACCTCCCTGTTACTGATGAGGGAACCTACTACGGA
AACTTCGAGGCTGAGTTCAGAAACTTCTGGACCAACGGATCTTACTACTGCATCTTCGCTGGACTCGGACTTATGCCTAGAAACCCTCTTC

CTGCTCTCGCTTACAAGCCTCAATCTATTGCTGAGGCTGAGCTTTTGTTCGCTGATGTGAAGAGGAAGGGTGATACCCTCGTTGAGTCTCT CCCTTCTACTTACGATCTCCTCAGACAGCTCCACGGTGCTTCATGA

Table S 2: Gene sequences that were cloned in this study. For addition of the CLP or 6His-tag, the respective primers listed below were used.

BX10, BX13 and BX14 from *Z. mays* for the expression of the extended BXD pathway
*Zm*_BX10:Ω ATGGCCCTCATGCAGGAGAGTAGTAGCCAGGATTTGCTCCAAGCTCACGACGAGCTCTTGCACCATTCCCTGTGCTTCGCCAAATCGCTC GCGCTCGCCGTGGCGCTAGACCTCCGCATCCCCGACGCGATCCACCACCACGGGGCCGGCGGCGCCACCCTTCTCCAGATCCTCGCCG AGACTGCGCTCCACCCAAGCAAGCTTCGCGCCCTTCGCCGCCTCATGCGCGTGCTCACCGTCACGGGCATCTTCAGCGTCGTCGAGCAA CCACCAGCAGGTGGTGGTGATGATTCAACCGTCCACACGTCGGACGACGAAGCTGTCGTCGTCTACAGGTTGACGGCAGCCTCCCGCTT CCTCGTCAGCGACGACGTGAGCACGGCGACCTTGGCTCCCTTTGTGAGTCTGGCGCTCCAGCCTATCGCTGCCTGTCCGCACGCCCTG GGTATCTCCGCGTGGTTCCGGCAGGAGCAGCACGAGCCGTCCCCGTATGGCCTGGCGTTCCGCCAGACCCCAACGATCTGGGAACATG CTGACGACGTAAACGCCTTGCTGAACAAAGGCATGGCCGCGGACAGCCGCTTCCTCATGCCAATTGTGCTGAGGGAGTGCGGCGAGACG TTTCGTGGGATCGACTCGTTGGTTGACGTCGGTGGTGGCCATGGTGGCGCCGCCGCCACCATCGCCGCCGCCTTCCCCCACCTCAAGT GCAGCGTGCTTGACCTCCCGCACGTTGTCGCCGGTGCTCCGTCTGATGGCAACGTGCAGTTCGTCGCAGGCAATATGTTTGAGAGTATTC CACCTGCAACCGCTGTTTTCCTCAAGAAAACTCTACATGACTGGGGTGACGATGAGTGTGTCAAGATATTGAAGAATTGCAAGCAAGCCAT ATCTCCACGGGATGCAGGTGGGAAGGTAATAATCTTGGATGTGGTAGTTGGATATAAACAGTCAAACATAAAGCATCAAGAGACACAAGTTA TGTTTGATTTGTATATGATGGCGGTTAACGGAGTTGAGCGTGACGAGCAAGAGTGGAAGAAGATCTTCACTGAAGCTGGATTCAAAGACTA CAAAATTCTACCCGTCATTGG-TGATGTATCGGTCATCATCGAGGTCTATCCTTGA *Zm*_BX13:Ω ATGGCTCCGACCGCCGCCAAGGACAGTGGCTCCGCCGACCGCTGGCGCGAGGTGCAGGCGTTCGAGGACAGCAAGCTGGGTGTGAAG GGCCTGGTCGACTCCGGCGTCAAGTCCATCCCGGCCATGTTCCACCACCCGCCGGAGTCCCTCAAGGATGTCATTTCTCCACCAGCTCT GCCCTCCTCCGACGATGCGCCGGCCATCCCAGTGGTGGACCTGTCGGTCGCGCGGCGCGAGGATCTGGTGGCCCAGGTGAAGCACGC GGCCGGGACGGTGGGCTTCTTCTGGGTGGTGAACCACGGCGTGCCGGAGGAGCTGATGGCCAGCATGCTCAGCGGGGTGCGCCAGTT CAACGAGGGCTCTCTGGAGGCCAAGCAGGCGCTCTACTCACGGGACCCCGCTCGCAACGTGCGCTTCGCTTCCAACTTCGACCTCTTCG AGTCCGCGGCGGCCGACTGGCGCGACACGCTCTACTGCAAGATTGCGCCGGATCCGGCGCCGCGGGAGCTCGTGCCGGAGCCTCTCC GGAACGTGATGATGGAATACGGCGAGGAACTGACGAAGCTGGCGCGGTCCATGTTCGAGCTGCTGTCGGAGTCCCTCGGCATGCCCAGC GACCACCTGCACAAGATGGAGTGCATGCAGCAGCTCCACATCGTGTGCCAGTACTACCCGCCATGCCCGGAGCCGCACCTCACCATGGG CGTCAGGAAGCACTCGGACACTGGTTTCTTCACCATCCTCCTGCAGGACGGCATGGGCGGCCTGCAGGTGCTAGTCGATCGCGGAGGC GGCCGCCAGACGTGGGTGGACGTCACTCCCCGACCTGGGGCACTCATGGTCAACATGGGCAGCTTTCTTCAGCTTGTGACGAATGACCG GTACAAGAGCGTGGATCACCGGGTGCCGGCTAACAAGAGCAGTGACACGGCGAGGGTCTCGGTAGCTGCCTTCTTCAACCCCGACGAG AAGAGAACCGAGAGGCTGTATGGCCCAATTCCAGACCCCAGCAAGCCTCCGCTGTACAGGAGCGTCACGTTCCCGGACTTCATCGCCAAT TTCAACAGCATCGGACTGGAACGGTCGCGTGCTCGACCACTTCCGGTTGGAAGAANAACGGCCC *Zm*_BX14:Ω ATGGCACTCATGCAGGAGAGCAGCCAGGACTTGCTCGAAGCGCACGACGAGCTCTTCCACCACTGCCTGTGCTTCGCCAAATCGCTCGC GCTCGCCGTGGCGCAGGACCTCCGCATCCCCGACGCGATCCACCACCACGGAGGCGGCGCCACCCTCCACCAGATCCTCGCCGAGGC CGCGCTCCACCCAAGCAAGCTTCGCGCCCTACGCCGCCTGATGCGCGTGCTCACCGTCTCGGGCGTCTTCACCGTCCAGTATTCTTCAA CCGTCGACGCGTCGGACGGAGCTGATGTCGTCTACAGGCTGACGGCAGCCTCCCGCTTCCTCGTCAGCGATAGCGACGAGGCGGGCAC GGCGTCCTTGGCTCCCTTTGCGAACCTGGCGCTCCACCCTATCGCCATCTCCCCGCACGCCGTGGGCATCTGCGCGTGGTTCCGGCAG GAGCAGCACGACCCGTCCCCGTACGGCCTGGCGTTCCGCCAGATCCCGACCATCTGGGAGCATGCTGACAACGTAAACGCCCTACTGAA CAAAGGCTTGCTCGCGGAAAGCCGCTTCTTGATGCCAATCGTACTCAGGGAGTGCGGAGACGAGGTGTTCCGTGGGATCGACTCGTTGG TCGACGTCGGCGGTGGGCACGGTGGCGCCGCCGCCACCATCGCCGCCGCATTCCCGCACGTCAAGTGCAGCGTGCTTGACCTCCCGC ACGTTGTCGCCGGTGCTCCATCCGATGCCTGCGTGCAGTTCGTTGCGGGCAATATGTTCCACAGTATTCCACCTGCAACCGCCGTTTTCTT CAAGACAACTCTATGTGACTGGGGTGACGACGAGTGCATCAAGATATTGAAGAATTGCAAGCAAGCCATATCTCCACGGGATGAGGGTGG GAAGGTGATAATCATGGACGTGGTAGTCGGGTATGGGCAGTCAAACATGAAGCGCCTAGAGACACAAGTTATGTTTGATTTGGTTATGATGG CGGTCAATGGAGTCGAGCGCGACGAGCAAGAGTGGAAGGAGATGTTCATTGAAGCTGGATTCAAAGACTACAAAATCCGACCAGTAGCTG GCCTCATGTCGGTCATCGAGGTCTATCCATGAATGGCACTCATGCAGGAGAGCAGCCAGGACTTGCTCGAAGCGCACGACGAGCTCTTCC ACCACTGCCTGTGCTTCGCCAAATCGCTCGCGCTCGCCGTGGCGCAGGACCTCCGCATCCCCGACGCGATCCACCACCACGGAGGCGG CGCCACCCTCCACCAGATCCTCGCCGAGGCCGCGCTCCACCCAAGCAAGCTTCGCGCCCTACGCCGCCTGATGCGCGTGCTCACCGTC TCGGGCGTCTTCACCGTCCAGTATTCTTCAACCGTCGACGCGTCGGACGGAGCTGATGTCGTCTACAGGCTGACGGCAGCCTCCCGCTT CCTCGTCAGCGATAGCGACGAGGCGGGCACGGCGTCCTTGGCTCCCTTTGCGAACCTGGCGCTCCACCCTATCGCCATCTCCCCGCACG CCGTGGGCATCTGCGCGTGGTTCCGGCAGGAGCAGCACGACCCGTCCCCGTACGGCCTGGCGTTCCGCCAGATCCCGACCATCTGGGA GCATGCTGACAACGTAAACGCCCTACTGAACAAAGGCTTGCTCGCGGAAAGCCGCTTCTTGATGCCAATCGTACTCAGGGAGTGCGGAGA CGAGGTGTTCCGTGGGATCGACTCGTTGGTCGACGTCGGCGGTGGGCACGGTGGCGCCGCCGCCACCATCGCCGCCGCATTCCCGCA CGTCAAGTGCAGCGTGCTTGACCTCCCGCACGTTGTCGCCGGTGCTCCATCCGATGCCTGCGTGCAGTTCGTTGCGGGCAATATGTTCCA CAGTATTCCACCTGCAACCGCCGTTTTCTTCAAGACAACTCTATGTGACTGGGGTGACGACGAGTGCATCAAGATATTGAAGAATTGCAAG CAAGCCATATCTCCACGGGATGAGGGTGGGAAGGTGATAATCATGGACGTGGTAATTCGGGTATGGGCAGTCAAACATGAAGCGCCTAGA GACACAAGTTATGTTTGATTTGGTTATGATGGCGGTCAATGGAGTCGAGCGCGACGAGCAAGAATGGAAAGGAAATGTTCATTGAAGCTGG ATTCAAAGACTACAAAATCCGACCAGTAGCTGGCCTCATGTCGGTCATCGGGTCTATCCATGAGCNTTTCTGCGGGGAGGCCCCCAGAATT TCTTAAAGGCCAAGAAGGGTCTTTTCCACCAATGGCCTTGGTTTTGCCAAATCGTTTGCGTTTCCCGTGGCGCAGGACCTTCGGNTTCCC CGAAGGGAATCCACCACCACGGAAGGGGCGCCACCCTCCACCCAGATCCTCGCCGAGGCCGCGTTCCACCCAAGCAAGCTTCGCGCCC TACGCCGCCTGATGCGCGTGCTCACCGTCTCGGGCGTCTTCACCGTCCAGTATTCTTCAACCGTCGACGCGTCGGACGGAGCTGATGTC GTCTACAGGCTGACGGCAGCCTCCCGCTTCCTCGTCAGCGATAGCGACGAGGCGGGCACGGCGTCCTTGGCTCCCTTTGCGAACCTGG CGCTCCACCCTATCGCCATCTCCCCGCACGCCGTGGGCATCTGCGCGTGGTTCCGGCAGGAGCAGCACGACCCGTCCCCGTACGGCCT GGCGTTCCGCCAGATCCCGACCATCTGGGAGCATGCTGACAACGTAAACGCCCTACTGAACAAAGGCTTGCTCGCGGAAAGCCGCTTCTT GATGCCAATCGTACTCAGGGAGTGCGGAGACGAGGTGTTCCGTGGGATCGACTCGTTGGTCGACGTCGGCGGTGGGCACGGTGGCGCC GCCGCCACCATCGCCGCCGCATTCCCGCACGTCAAGTGCAGCGTGCTTGACCTCCCGCACGTTGTCGCCGGTGCTCCATCCGATGCCTG CGTGCAGTTCGTTGCGGGCAATATGTTCCACAGTATTCCACCTGCAACCGCCGTTTTCTTCAAGACAACTCTATGTGACTGGGGTGACGAC GAGTGCATCAAGATATTGAAGAATTGCAAGCAAGCCATATCTCCACGGGATGAGGGTGGGAAGGTGATAATCATGGACGTGGTAGTCGGGT ATGGGCAGTCAAACATGAAGCGCCTAGAGACACAAGTTATGTTTGATTTGGTTATGATGGCGGTCAATGGAGTCGAGCGCGACGAGCAAG AGTGGAAGGAGATGTTCATTGAAGCTGGATTCAAAGACTACAAAATCCGACCAGTAGCTGGCCTCATGTCGGTCATCGGGTCTATCCATGA
Chloroplastic localization peptide (CLP)
Chlroplastic localization peptide (CLP) ATGGCAGCCTCAGGCACCTCTGCTACTTTCAGAGCCTCCGTTTCTTCAGCTCCTTCCTCTTCTTCCCAATTGACCCATTTGAAATCACCCTT CAAAGCTGTCAAATATACGCCTCTGCCATCGTCTCGCTCCAAGTCATCATCCTTCTCCGTCTCCTGCACCATCGCCAAGGACCCGCCTGTT CTC

Bacterial halogenases

HMT:Ω

ATGGGTAGCGATCCAACCCAACCTGCCGTACCAGACTTTGAGACACGAGATCCGAACTCACCCGCCTTCTGGGATGAACGTTTCGAGCGG
AGATTCACTCCATGGGATCAGGCAGGAGTCCCAGCCGCATTCCAGAGTTTTGCGGCCAGGCACTCTGGTGCAGCAGTTCTCATTCCTGGT
TGCGGCAGCGCTTACGAGGCAGTCTGGTTGGCTGGCCAAGGAAATCCAGTTAGGGCGATAGATTTCAGTCCTGCCGCTGTCGCAGCAGC
CCACGAGCAATTAGGCGCCCAACATGCACAGTTAGTTGAACAGGCCGACTTCTTCACATACGAACCACCTTTCACACCTGCCTGGATTTAT
GAGCGCGCTTTTCTATGTGCATTACCTTTGGCTAGACGAGCTGACTACGCGCATAGGATGGCAGACCTTTTGCCTGGCGGCGCTTTGTTAG
CAGGATTCTTCTTCTTAGGTGCAACTCCAAAGGGCCCACCATTCGGGATAGAGAGGGCAGAATTGGATGCCTTGCTCACACCCTATTTCGA
CCTTATTGAAGACGAGGCTGTACATGACAGCATTGCTGTCTTCGCAGGTAGAGAACGTTGGCTTACCTGGAGAAGACGTGCG

RebF:Ω

ATGACGATTGAATTCGACAGACCCGGCGCACACGTCACGGCAGCAGACCATAGGGCACTTATGAGTCTGTTTCCCACGGGTGTCGCAGTG
ATAACTGCTATCGACGAGGCCGGGACCCCTCACGGCATGACCTGTACCTCTCTCACCTCAGTGACTCTGGATCCACCGACTCTTTTGGTTT
GCTTAAACCGAGCTTCTGGCACTTTGCACGCTGTCCGAGGTGGACGTTTTGGGGTAAATCTCCTTCATGCACGTGGACGACGCGCCGCC
GAGGTGTTCAGCACTGCCGTTCAAGATCGGTTCGGGGAAGTGAGGTGGGAACACTCTGATGTGACGGGTATGCCATGGTTGGCGGAAGA
CGCTCATGCGTTCGCCGGGTGCGTTGTTCGTAAGTCTACCGTTGTCGGAGACCATGAAATCGTCCTCGGGGAAGTGCACGAGGTGGTCC
GAGAGCACGACTTGCCACTACTTTATGGAATGCGAGAGTTTGCTGTGTGGACCCCGGAAGGATAA

RebH-3LRS:Ω

ATGTCAGGTAAAATCGACAAAATCCTTATAGTGGGAGGAGGGACTGCTGGCTGGATGGCAGCTTCCTATTTGGGGAAAGCCCTTCAGGGC
ACGGCCGATATTACTCTCCTACAGGCCCCAGATATCCCAACACTCGGGGTTGGCGAAGCTACAATACCTAATTTACAGACTGCCTTCTTCGA
TTTCCTGGGTATTCCTGAAGACGAGTGGATGAGAGAATGTAATGCGTCTTATAAGGTGGCAATTAAGTTCATTAATTGGAGAACTGCAGGGG
AAGGAACCAGTGAAGCACGTGAATTAGACGGTGGACCTGACCACTTTTATCATAGTTTCGGCTTGCTTAAGTACCACGAGCAGATCCCTCT
TAGTCATTATTGGTTTGATCGGCTCTACCGGGGTAAAACTGTGGAACCATTCGACTACGCATGCTACAAAGAACCTGTGATACTTGACGCGA
ACCGGAGTCCGCGCAGACTCGACGGGAGCAAAGTGACTTCATATGCGTGGCACTTCGATGCTCACCTTGTTGCTGACTTCCTACGACGTT
TCGCTACTGAAAAGCTAGGTGTAAGACACGTCGAAGATCGAGTTGAACACGTACAAAGAGATGCGAACGGTAATATCGAGTCCGTGAGAAC
CGCAACGGGCAGAGTTTTCGACGCAGATTTGTTTGTTGACTGTTCTGGATTTCGTGGCTTGTTAATAAATAAGGCAATGGAAGAGCCTTTCT
TAGATATGAGTGACCATCTTTTAAATGACTCAGCCGTCGCGACTCAAGTTCCACACGATGATGACGCCAATGGCGTTGAACCGTTTACATCT
GCAATAGCAATGAAAAGCGGATGGACTTGGAAAATCCCGATGCTGGGCCGTTTCGGTACTGGATACGTATATTCTTCAAGGTTTGCCACTGA
GGACGAGGCCGTTAGAGAATTCTGTGAGATGTGGCACCTAGACCCCGAAACTCAACCATTAAACAGGATCCGTTTCAGGGTAGGTAGGAA
CAGAAGGGCATGGGTCGGGAACTGCGTATCAATTGGAACAAGCTCTTGTTTCGTGGAACCACTCGAATCAACAGGCATCTATTTTGTTTAC
GCTGCTCTATACCAATTGGTTAAGCACTTTCCAGACAAATCACTCAACCCTGTTCTAACTGCACGATTCAATAGGGAAATTGAGACTATGTTT
GACGATACACGAGATTTTATACAGGCACACTTCTACTTCTCCCCAAGAACCGACACGCCGTTTTGGCGGGCTAATAAGGAACTTCGACTTG
CCGACGGGATGCAGGAGAAAATCGATATGTACCGGGCCGGAATGGCGATAAACGCCCCCGCGTCCGACGATGCTCAGCTTTATTACGGGA
ATTTCGAGGAAGAGTTCCGTAACTTCTGGAACAATTCCAATTACTATTGTGTACTGGCTGGGTTAGGCCTAGTCCCTGACGCCCCATCCCCT
AGATTAGCACACATGCCTCGAGCCACCGAATCCGTCGACGAGGTTTTCGGCGCCGTAAAGGATAGACAGAGAAATTTGCTCGAAACTTTGC
CTTCCTTACATGAGTTCCTTAGGCAGCAGCACGGTCGCTGA

KtzQ:Ω

ATGGATGATAATCGCATACGATCAATTCTTGTCCTTGGAGGAGGTACCGCCGGTTGGATGTCCGCGTGCTACCTCAGTAAGGCCCTCGGCC
CAGGAGTCGAGGTGACCGTCTTAGAGGCACCCTCTATCAGCAGAATTAGGGTTGGGGAAGCTACTATCCCCAATCTCCACAAAGTGTTCTT
CGATTTCCTGGGCATTGCTGAGGACGAATGGATGAGGGAGTGTAACGCGTCCTACAAAGCTGCTGTCCGTTTTGTCAATTGGCGCACCCC
AGGAGACGGACAGGCCACGCCCCGTCGTAGGCCAGATGGGCGCCCTGATCATTTCGATCACTTGTTCGGACAGCTCCCTGAGCACGAGA
ACCTACCATTGTCACAATATTGGGCCCACAGACGCCTGAATGGGCTCACTGACGAGCCTTTCGATCGGAGTTGCTACGTGCAACCTGAGCT
TTTGGACAGGAAACTCTCCCCACGTTTGATGGATGGGACTAAACTGGCTTCATATGCATGGCATTTTGACGCTGATTTGGTTGCGGACTTCT
TGTGCCGATTTGCAGTTCAAAAGCTTAACGTAACACACGTTCAAGACGTCTTTACTCACGCCGATTTAGACCAGCGTGGCCACATCACTGC
CGTAAATACGGAATCTGGCCGTACACTAGCCGCGGACCTATTCATAGACTGTTCAGGTTTCCGGAGCGTGCTAATGGGCAAGGTAATGCAA
GAGCCTTTTCTGGACATGAGCAAGCACCTTTTGAATGACAGAGCAGTTGCTCTGATGTTGCCACACGACGATGAGAAAGTTGGGATAGAG
CCATACACAAGTTCTCTAGCCATGAGAAGCGGTTGGAGCTGGAAGATCCCATTACTCGGTAGATTTGGTAGCGGGTACGTTTACAGTTCCC
AGTTTACCTCACAAGATGAAGCAGCAGAGGAGTTGTGCCGCATGTGGGACGTCGACCCAGCCGAACAGACATTCAACAACGTCCGGTTTC
GTGTGGGTCGATCACGACGGGCATGGGTGAGAAATTGCGTGGCCATAGGGGTTAGCGCTATGTTTGTGGAGCCTCTTGAAAGTACGGGGT
TATATTTTAGTTATGCCTCTCTTTACCAGCTGGTCAAGCACTTCCCTGACAAACGTTTTAGGCCTATTCTTGCTGACCGCTTCAACCGTGAGG
TAGCAACCATGTACGACGACACACGAGACTTCCTTCAAGCCCACTTCTCTCTATCTCCTAGGGACGACAGTGAATTCTGGCGCGCGTGCAA
GGAGCTCCCCTTCGCGGACGGTTTCGCAGAGAAGGTGGAAATGTACAGAGCTGGTCTCCCTGTGGAACTACCGGTTACAATTGACGATGG
CCACTATTATGGTAACTTCGAGGCTGAGTTCAGAAATTTCTGGACTAATAGTAACTACTATTGCATATTTGCGGGATTAGGTTTCCTTCCAGAA
CACCCGTTACCCGTCCTTGAATTTAGGCCTGAAGCTGTCGACAGGGCCGAGCCTGTGTTCGCTGCCGTACGAAGACGAACTGAGGAACT
CGTCGCTACTGCTCCGACCATGCAGGCATATCTTAGAAGGCTGCACCAAGGGACTTAA

KtzR:Ω

ATGGTGATAGTAGGAGGAGGCACCGCCGGATGGATGACAGCAGCGTACCTTAAGACAGCCTTCGGGGACCGCCTCTCAATTACAGTGGT
AGAGTCATCTAGGATTGGAACAATTGGCGTCGGTGAAGCTACTTTCTCTGACATTCAACACTTCTTCCAGTTTCTCAATCTTCGTGAGCAG
GACTGGATGCCTGCATGCAATGCAACTTACAAACTTGGCATCAGATTCGAGAACTGGCGTCACGTCGGACATCATTTCTATCAACCATTTG
AACAAATTCGTCCAGTATACGGCTTTCCGCTTACTGACTGGTGGCTCCACGACGCTCCCACAGACAGGTTTGACACTGACTGTTTCGTTAT
GCCTAACCTATGCGAGGCAGGTAGATCTCCTCGGCATCTGGATGGTACGTTGGCCGACGAAGACTTTGTTGAGGAAGGAGATGAGTTAG
CTAACAGAACAATGAGCGAGCATCAAGGGAAGAGTCAGTTCCCATACGCCTACCATTTTGAGGCTGCACTTCTTGCCAAATTCTTGACAG
GCTACGCGGTGGACCGGGGAGTAGAACACGTAGTCGACGACGTCCTGGATGTTAGACTTGACCAACGGGGCTGGATCGAACATGTGGT
CACGGCAGAGCACGGAGAGATACATGGAGACCTTTTCGTCGATTGTACTGGATTCAGAGGCCTGCTCCTGAATAAAGCACTTGGAGTACC
TTTCGTGAGTTATCAAGACACCCTACCAAACGACAGCGCAGTGGCGCTGCAAGTGCCCTTGGACATGCAACGTAGGGGTATTGTTCCAAA
CACCACTGCTACAGCCCGTGAGGCAGGTTGGATCTGGACTATACCCCTTTTCGGACGGGTTGGAACGGGTTATGTGTACGCTAAAGATTA
CCTCAGTCCAGAAGAGGCCGAGCGTACTTTGAGGGAATTCGTAGGTCCTGCAGCAGCTGACGTCGAAGCGAACCATATCCGCATGAGGA
TCGGACGAAGTCAAGAGTCTTGGCGAAATAATTGTGTAGCCATTGGACTCAGCAGTGGATTCGTTGAACCACTCGAGTCAACAGGGATTT
TCTTCATCCACCACGCGATTGAACAGCTAGTTAAGCACTTTCCTGCAGCTGACTGGAACCCAAAATCTCGTGACATGTACAACTCCGCTGT
TGCGCATGTAATGGACGGCATTAGGGAATTCTTAGTCATCCATTACAGGGGTGCCGCACGAGCCGATAACCAATATTGGCGAGACACTAA
AACACGTCCTTTGCCAGATGGTCTGGCTGAACGAATCGAATGCTGGCAGACTCAGCTACCTGACACCGAGACTATCTACCCATACTATCA
CGGGCTGCCACCGTACAGCTACATGTGCATTTTGATGGGTGGTGGAGCTATTAGAACCCCAGCTAGCGCCGCATTAGCCTTGACAGACC
AGGGCGCTGCCCAGAAGGAGTTCGCAGCTGTACGCGACAGGGCAGCACAACTTAGGGATACACTTCCCTCACACTACGAGTACTTGGCA
AGGATGCGTGGGTTAGACGTTTAA

PyrH:Ω

ATGATTAGGAGTGTCGTTATAGTAGGTGGTGGAACGGCTGGCTGGATGACAGCATCTTATCTAAAGGCAGCATTCGACGACCGAATTGACG
TTACCCTAGTTGAGAGTGGAAATGTGCGACGGATCGGGGTTGGTGAAGCCACATTTTCAACCGTTAGGCATTTCTTCGACTATCTGGGTCT
CGACGAAAGAGAATGGCTTCCAAGGTGCGCTGGTGGCTACAAATTAGGCATCCGATTCGAGAATTGGTCCGAGCCAGGCGAATACTTCTA
CCACCCGTTCGAGCGGTTAAGAGTGGTTGACGGCTTCAACATGGCCGAGTGGTGGTTAGCGGTAGGTGACAGGCGTACGTCTTTTAGTGA
AGCATGCTACTTGACTCACAGGCTCTGTGAAGCTAAGCGCGCACCAAGGATGCTAGACGGGAGCTTGTTTGCTTCACAGGTTGATGAATC
CTTAGGCAGGAGCACTTTAGCTGAGCAGAGGGCCCAGTTTCCTTACGCATATCACTTTGACGCCGACGAAGTTGCGAGGTATTTGTCCGA
GTATGCTATCGCTCGAGGAGTACGCCATGTTGTGGATGACGTGCAGCATGTCGGACAAGACGAGCGTGGATGGATCTCAGGAGTACACAC
AAAGCAACACGGCGAAATATCTGGAGACCTCTTTGTGGACTGTACTGGTTTCAGGGGTTTATTGATCAACCAAACGCTGGGTGGAAGGTTC

CAAAGTTTTAGTGATGTTCTCCCCAACAATAGAGCTGTGGCCCTTCGCGTCCCCCGTGAAAATGACGAGGATATGAGACCATATACTACCGC
AACAGCTATGTCTGCTGGTTGGATGTGGACTATTCCGTTGTTTAAAAGAGATGGAAACGGCTATGTTTATTCTGACGAATTCATAAGTCCCGA
GGAAGCCGAGCGGGAACTTAGATCCACAGTGGCGCCAGGAAGAGACGACCTTGAAGCTAATCACATACAAATGAGAATAGGTAGGAACGA
GAGAACTTGGATAAATAACTGCGTGGCGGTCGGTCTCTCTGCTGCTTTCGTCGAACCATTAGAAAGTACTGGAATTTTCTTTATTCAACATG
CAATTGAACAACTAGTTAAGCACTTTCCAGGAGAGCGGTGGGATCCTGTGTTGATCTCAGCATATAACGAACGCATGGCCCATATGGTCGAT
GGAGTCAAAGAGTTCCTGGTGTTACATTATAAGGGTGCTCAGAGAGAAGATACACCTTACTGGAAGGCAGCTAAGACAAGAGCAATGCCAG
ATGGATTGGCAAGGAAATTAGAGCTGAGCGCCAGCCACTTACTCGACGAACAAACAATATACCCGTACTATCACGGCTTTGAGACATATTCT
TGGATAACCATGAATTTGGGACTAGGTATAGTGCCTGAGAGGCCACGTCCTGCATTGTTACACATGGACCCGGCACCCGCATTGGCAGAGT
TCGAGCGATTGCGCCGAGAGGGAGACGAGCTGATAGCTGCTTTACCTAGTTGCTATGAATACCTCGCATCAATACAGTGA
RebH:Ω
ATGTCAGGTAAAATCGACAAAATCTTAATAGTTGGTGGCGGGACGGCTGGCTGGATGGCAGCATCTTACTTAGGGAAAGCATTGCAAGGGA
CAGCTGACATTACCCTGCTTCAGGCTCCCGACATACCTACACTGGGTGTCGGAGAGGCTACAATTCCTAATTTGCAGACAGCTTTCTTCGA
CTTTCTTGGTATTCCGGAAGATGAGTGGATGCGTGAGTGCAATGCGTCATATAAAGTTGCCATAAAGTTTATTAACTGGAGAACCGCAGGTG
AGGGGACTTCCGAGGCTAGAGAGTTGGACGGTGGACCTGACCACTTTTACCACTCTTTCGGATTGCTAAAGTACCACGAGCAAATCCCCC
TAAGTCACTACTGGTTTGACCGTAGTTACAGAGGAAAGACCGTAGAGCCATTCGACTATGCGTGTTATAAGGAACCAGTTATTCTCGACGCC
AACAGGTCTCCTAGACGACTTGACGGGTCAAAAGTGACTAACTACGCTTGGCATTTCGATGCTCATCTTGTTGCCGACTTCCTACGGCGGT
TTGCCACCGAGAAATTAGGAGTTCGACACGTCGAGGATCGTGTTGAGCATGTACAGAGAGACGCTAATGGAAACATTGAATCTGTGAGAAC
AGCCACAGGGCGAGTTTTCGACGCTGACCTGTTTGTGGACTGTAGTGGCTTTCGGGGTCTTTTGATCAACAAAGCCATGGAAGAGCCCTT
CTTGGATATGAGCGATCACCTTTTAAATGACTCCGCTGTTGCAACTCAAGTTCCCCACGACGACGACGCAAATGGTGTGGAGCCATTCACG
TCAGCTATTGCTATGAAGTCAGGTTGGACATGGAAGATTCCCATGCTAGGAAGGTTTGGCACTGGCTATGTCTACTCTTCACGCTTCGCCAC
CGAAGATGAAGCTGTAAGGGAATTTTGTGAGATGTGGCATCTGGATCCGGAAACCCAGCCACTTAACCGTATCCGTTTCCGTGTAGGCAGA
AACCGTAGAGCGTGGGTAGGTAACTGTGTGAGTATAGGCACTAGTTCATGCTTCGTGGAGCCTCTTGAATCAACAGGAATATATTTCGTCTA
CGCTGCGCTCTATCAATTAGTGAAGCACTTCCCTGACAAGAGTCTCAACCCAGTATTGACTGCAAGGTTCAATCGGGAGATAGAAACAATGT
TTGATGATACTAGAGACTTCATACAAGCACACTTCTACTTTAGTCCCAGAACGGATACACCATTTTGGCGAGCTAACAAAGAACTAAGACTTG
CCGACGGAATGCAGGAAAAGATAGACATGTACCGAGCTGGTATGGCGATAAACGCACCAGCTAGTGATGACGCCCAGCTTTACTATGGCAA
TTTTGAGGAAGAGTTCAGGAACTTTTGGAATAACTCTAACTACTACTGTGTGTTAGCAGGTCTCGGGCTAGTACCTGACGCACCAAGTCCTA
GGCTTGCACACATGCCTCAAGCAACTGAGAGCGTCGACGAAGTTTTCGGGGCCGTGAAAGATCGTCAAAGAAATTTACTAGAGACATTGC
CTAGTTTGCATGAGTTCTTACGGCAGCAACACGGTCGTTAA

Truncated genes for the tryptophan biosynthesis in *A. thalliana*

ASA:Ω
ATGTCTGTTTCTCCGGAAGCTTCAATAGTAAGTGATACAAAGAAGTTGGCAGATGCTTCTAAGAGTACAAACCTTATACCAATTTACCGCTGT
ATATTCTCTGATCAGCTTACTCCTGTTCTTGCTTACCGTTGTTTGGTCAAAGAAGATGACCGTGAAGCTCCTAGCTTTCTTTTCGAGTCCGTT
GAGCCTGGTTCTCAGATGTCTAGCGTTGGTCGTTATAGCGTTGTTGGGGCTCAGCCTGCGATGGAGATCGTGGCAAAGGAGAATAAAGTTA
TTGTAATGGATCACAACAATGAAACCATGACTGAGGAATTCGTCGAAGATCCAATGGAGATCCCAAGAAAAATCTCTGAGAAATGGAACCCT
GATCCTCAACTAGTTCAGGACCTTCCAGATGCGTTTTGTGGTGGGTGGGTTGGTTTTTTCTCGTACGACACTGTTCGTTATGTTGAGAAGA
GGAAATTGCCATTTTCAAAGGCCCCTGAGGATGATAGGAACTTGCCAGACATGCATCTTGGTCTGTACGACGATGTAGTTGTATTTGATCAC
GTGGAAAAGAAAGCATATGTCATTCACTGGATTAGACTAGATGGGAGCCTTCCTTACGAAAAGGCATACAGTAATGGAATGCAACATTTGGA
GAACTTGGTGGCCAAGTTACATGATATTGAGCCGCCAAAACTGGCTGCAGGTAACGTGAATCTTCAGACACGACAATTTGGGCCATCTTTG
GATAATTCAAACGTGACATGCGAAGAGTACAAGGAGGCTGTGGTCAAGGCCAAAGAACATATACTTGCAGGAGACATATTTCAGATCGTGCT
GAGTCAACGTTTTGAGCGGCGAACATTTGCAGACCCCTTTGAAGTTTATAGAGCACTAAGAGTTGTGAATCCAAGTCCGTATATGGGTTATT
TGCAGGTTGGTGATGGAGAGAACATTTCTCAACAACTTTGCATTTTGTACTTGTGGATAAACCATCAATGCTGTTTTCGTTCTGTTATTCAGG
CTAGAGGATGCATTTTGGTAGCATCAAGTCCAGAAATTCTCACCAAAGTAAAGCAGAACAAGATAGTGAATCGGCCATTGGCAGGAACCAG
CAAGAGAGGGAAGAATGAAGTTGAGGATAAGAGATTAGAAAAGGAACTGCTAGAGAATGAAAAGCAATGTGCTGAGCACATCATGTTGGTT
GATCTCGGTCGCAACGATGTTGGAAAGGTTACGAAATACGGATCAGTGAAAGTAGAGAAGCTTATGAACATCGAACGTTATTCCCATGTTAT
GCATATAAGCTCCACGGTGACAGGAGAATTACAAGATGGTTTGACTTGCTGGGACGTACTACGTGCGGCTTTACCAGTGGGAACAGTTAGT
GGTGCACCAAAGGTCAAAGCTATGGAACTAATCGATGAGCTAGAGCCAACGAGGCGTGGACCATACAGTGGCGGTTTTGGTGGAGTCTCC
TTCACTGGTGACATGGACATTGCTTTATCCCTTAGGACAATCGTTTTTCCGACAGCATGTCAATACAATACAATGTACTCTTACAAGGATGCTA
ACAAACGGCGTGAGTGGGTGGCTTATCTTCAAGCTGGAGCTGGTGTAGTAGCTGATAGTGACCCGCAAGACGAACACTGTGAGTGCCAGA
ACAAAGCCGCTGGTCTTGCTCGAGCCATCGACTTGGCTGAATCTGCATTTGTGAAAAAATGA
ASB:Ω
ATGGCGGAATCGAATTCGATACCTTCCGTTGTTGTCAATTCCTCTAAGCAGCATGGTCCAATCATCGTGATTGATAATTACGACAGCTTCACA
TACAATCTCTGCCAGTATAAACAAAATTTTGAGAATTGTTATCTATTTCTACAGTATATGGGAGAGCTAGGATGCCATTTTGAAGTTTACCGCAA
TGATGAACTTACAGTAGAAGAGCTGAAAAAAAAAAATCCAAGAGGGGTGTTGATTTCTCCAGGGCCTGGTACCCCTCAAGACTCTGGGATT
TCCTTGCAAACTGTTTTGGAACTCGGACCACTTGTTCCTTTATTTGGAGTATGTATGGGTTTGCAGTGTATAGGAGAAGCATTTGGAGGAAA
GATTGTGCGGTCACCATTTGGTGTTATGCATGGGAAAAGCTCAATGGTTCACTATGATGAGAAAGGAGAAGAAGGCTTGTTCTCTGGATTAT
CAAACCCTTTCATTGTAGGTAGATATCACAGTCTCGTGATCGAAAAAGATACATTTCCTAGTGATGAACTCGAGGTTACAGCATGGACAGAAG
ATGGTCTGGTAATGGCTGCCCGTCACAGAAAGTACAAGCATATACAGGGAGTCCAATTTCATCCGGAGAGTATTATAACAACTGAGGGCAAG
ACAATTGTCCGCAATTTCATCAAAATAGTAGAGAAAAAGGAGTCCGAGAAGCTGACATAG
IPGS:Ω
ATGGCTCAACAGTCGGATTTGAAGGAAAGCTTAGCTGTGTCTTCTTCATCAGTAGAGGATAAAGGAAATGTTCTTAGAATCAAGGAATGGGA
AGTAGAGATGTATCAGGAGGAATTAGCTATTAGTCAAGGTATTAGGATAAGGAGAAAACCACCAAGTAAGGCTCCTTTGGGATACTCTGGAC
CATTTGAGTTGAGATTGCATAACAATGATGCTGATTCTCCTCGTAATATCTTGGAGGAGATCACATGGTACAAAGACGTAGAAGTTTCCCGGA
TGAAGGAGCTAAATCCGCTTGACGTGCTGAAGAAAGCTGTAGAGGATGCTCCTCCTACTAGGGATTTTGTTGGGGCTCTTAGGATGGCTCA
TAAAAGAACTGGCTTCCCTGGCTTGATAGCTGAGGTTAAGAAGGCTTCTCCAAGTAGAGGAATCTTAAAAGAGAATTTTGACCCGGTCGAG
ATTGCTCAAGCTTATGAAAAAGGCGGAGCAGCATGCCTCAGCGTTTTGACAGACCAGAAGTATTTCCAGGGAGGCTTTGAAAACTTGGAAG
CAATAAGGAGCGCTGGTGTGAAGTGTCCACTATTATGCAAAGAGTTTGTTGTAGATCCATGGCAGATCTACTATGCTCGGACTAAAGGCGCA
GATGCAGTACTGCTTATTGCTGCTGTATTGGCTGACCTAGAAATAACCTTCTTGCTTAAGATTTGCAAGAAGCTTAGCTTGGCTGCCCTTGTT
GAGGTACATGATGAGAGAGAGATGGGTCGTGTACTTGGAATAGAAGGAATCGAGCTTGTTGGCATCAATAACCGAAGTCTTGAAACATTTG
AAGTGGACATCAGTAATACAAAGAAGCTTCTTGAAGGAGAGCATGGTAGACAAATCCGTGAGAGAGACATGATTGTGGTTGGCGAATCTGG
TCTGTTTACACCAGATGACATTGCCTATGTACAAGCAGCCGGAGTCAAAGCAGTTTTGGTTGGAGAATCCATTGTGAAGCAGAACGACCCT
GAGAAAGGAATAGCTGGACTTTTTGGCAGGAACATTTCTCATACTTAG
PAI:Ω
ATGTGTGGCATCACATCAGCCAGAGATGCAGCTATGGCAGTTGAGGCTGGTGCCGATTTTATTGGGATGATTATTTGGCCACATTCAAAACG
ATCTATTTCTCTTTCTGTCGCAAAGGATATCTCCAAAGTGGCCAGGGAAGGTGGGGCCAAACCTGTTGGTGTTTTTGTTGAAGATGATGAGA
ACACGATACTGAGAGCAGCTGATTCCTCTGACCTTGAGCTTGTGCAGCTTCATGGTAATGGTTCGCGTGCTGCTTTCTCGAGGTTAGTACG
CAAGAGGAGAGTCATATATGTCCTTAATGCAAATCAGGATGGAAAGCTCTTGAATGAAGTTCCTGAAGAGGATTGTCATCTTGCTGATTGGAT
TCTTGTGGATAGTGCAACGGGTGGGAGTGGACACGGATTTAACTGGGCTCAATTCAAGTTGCCTTCCGTCAGAAGCAGAAATGGGTGGCT
CTTAGCTGGGGGAATCAATCCAACAAATGTTTCAGAAGCTCTTTCTATCCTTCAACCTGATGGAATTGATGTTAGTAGCGGTATTTGCGGTAC
AGACGGTATCCAGAAGGATAAGTCTAAGATAAGCTCCTTTATAACTGCAGTTCGCTCTGTACACTACTAA

PAT:Ω
ATGAGCGGTGGATTTAGTGCAGCTGAAGCTCTGCCACCTGCTTGTGCAAATGCTAGCCCTTCTTCCATTAAATCTTTCAACCAGTTGATTGA
AACCTTGATTGATCGGGTTGATCTATCTGAAACTGAGGCTGAATCATCTCTTGAGTTTTTGCTGAATGAGGCAAACGAGGCGCTGATCAGTG
CCTTTCTAGTTCTTCTGAGAGCTAAAGGAGAGACATACGAAGAGATTGTGGGGTTAGCTAGGGCAATGATGAAGCATGCCAGGAAAGTGGA
AGGATTAGTTGATGCTGTGGACATAGTTGGAACAGGTGGTGATGGAGCAAACACAGTCAATATATCAACTGGATCTTCAATACTTGCTGCAG
CTTGTGGTGCAAAAGTAGCAAAGCAAGGGAATCGTTCAAGTTCTTCTGCTTGTGGAAGTGCTGATGTACTAGAGGCACTAGGAGTGGTTCT
GGACTTGGGACCAGAGGGCATTAAAAGATGTGTTGAAGAAGGGGGGATCGGTTTTATGATGTCACCAATGTATCATCCAGCTATGAAGATC
GTTGGTCCTGTTCGGAAAAAGCTTAAAATTAAAACTGTTTTTAACATATTGGGACCTATGCTTAATCCTGCTAGAGTTTCTTATGCTGTCGTTG
GTGTATACCACAAAGATCTGGTTGTTAAGATGGCAAAGGCACTGCAACGATTTGGAATGAAAAGAGCACTGGTGGTTCATTCATGTGGTCTA
GACGAAATGAGTCCGTTAGGAGGAGGACTAGTTTATGATGTAACTCCGGAAAAGATCGAAGAATTCTCATTCGACCCATTGGACTTTGGTAT
TCCTCGTTGTACTCTTGAGGATTTGCGAGGTGGAGGTCCAGACTACAATGCGGATGTGCTAAGACGTGTGCTTTCAGGGGAAAGTGGAGC
AATTGCGGATTCATTGATCCTAAACGCAGCTGCAGCTCTTCTGGTTAGCAACCGAGTTCAAACGCTAGCTGAAGGAGTAACCGTGGCACGT
GAAGTACAATCGTCTGGGAAAGCTATCAAGACGCTTGATTCTTGGATAAACATCTCAAACTTAGCTCAGAAATCTCAGTGA
*At*_TSA_truncated:Ω
ATGGCTTCTCTCTCCACCTCTTCTCCTACTCTCGGTCTCGCTGATACTTTCACACAGCTCAAAAAACAAGGCAAAGTAGCATTCATACCATAT
ATCACAGCTGGTGATCCGGATCTCTCTACTACTGCTGAAGCATTGAAAGTTCTTGATGCTTGTGGTTCTGACATAATCGAATTGGGTGTTCC
TTACTCTGACCCTTTAGCTGATGGACCTGTTATTCAGGCTGCGGCAACAAGGTCGTTGGAGAGGGGAACAAACCTCGATAGCATCCTTGAG
ATGTTGGATAAGGTTGTTCCACAAATATCTTGTCCGATTTCGTTGTTCACGTATTACAACCCGATTCTTAAACGTGGGTTGGGGAAGTTCATG
TCCAGCATCAGAGCTGTTGGTGTACAGGGACTTGTGGTTCCCGATGTTCCTCTTGAGGAAACTGAGATGCTGAGAAAAGAAGCCCTCAAC
AACGACATTGAGCTGGTCCTACTCACTACACCAACCACACCAACAGAGCGAATGAAGCTAATTGTTGATGCGTCAGAAGGATTTATTTACCT
TGTGAGCTCAATTGGAGTGACTGGTGCACGATCATCTGTAAGCGGAAAGGTTCAGTCGCTCTTGAAGGATATCAAAGAGGCGACAGACAA
GCCGGTGGCAGTCGGTTTTGGAATATCAAAGCCGGAGCATGTGAAACAGATAGCTGGATGGGGAGCTGATGGAGTGATCGTAGGCAGTG
CAATGGTAAAGCTATTGGGTGATGCAAAGTCGCCAACTGAAGGACTTAAGGAGCTTGAGAAACTAACCAAGTCTCTTAAATCTGCTCTTCTT
TGA
*At*_TAB_truncated:Ω
ATGGCCGCCGGATCTGACCCGGCCCTGTGGCAACGACCCGATTCGTTCGGTCGGTTTGGGAAGTTTGGTGGGAAGTATGTACCTGAAAC
CCTTATGCACGCTCTATCTGAGCTTGAATCCGCTTTCTATGCTCTTGCCACCGACGATGATTTCCAGAGAGAGTTGGCTGGAATCTTGAAGG
ACTATGTGGGTAGAGAAAGTCCTCTGTATTTTGCAGAGAGGCTTACGGAGCATTACAGGCGCGAGAATGGCGAAGGGCCTCTTATATACTT
GAAGAGAGAAGACTTGAATCACACAGGAGCTCACAAGATTAACAACGCTGTGGCTCAGGCTCTTCTTGCTAAGCGGTTGGGGAAGAAGAG
GATTATTGCTGAGACAGGAGCGGGTCAACATGGTGTGGCTACAGCTACTGTGTGTGCCCGTTTTGGTTTGGAGTGTATTATCTATATGGGTG
CTCAAGATATGGAGAGACAAGCACTCAATGTGTTCAGAATGCGACTTCTTGGTGCCGAGGTGAGAGGAGTCCACTCTGGAACAGCGACAT
TGAAGGATGCGACATCTGAAGCGATAAGAGATTGGGTGACAAATGTGGAGACTACCCATTACATATTGGGATCTGTGGCGGGTCCTCATCC
TTACCCCATGATGGTCAGAGACTTTCACGCTGTGATTGGTAAAGAAACAAGGAAACAAGCGTTGGAGAAATGGGGCGGGAAGCCGGATGT
CTTGGTGGCTTGTGTCGGTGGTGGTTCAAACGCTATGGGACTCTTCCATGAATTTGTGAATGACACAGAGGTCCGGATGATCGGTGTGGA
AGCAGCGGGATTCGGATTGGACAGTGGCAAACACGCTGCTACATTGACAAAGGGAGATGTTGGTGTCCTCCATGGAGCTATGAGTTACTT
GCTGCAAGATGATGATGGACAAATCATTGAACCACACTCCATCAGTGCAGGATTGGACTACCCTGGAGTCGGACCCGAGCACAGTTTCTTT
AAAGACATGGGACGGGCTGAATACTATAGCATAACCGACGAAGAAGCGTTGGAAGCGTTCAAGAGAGTGTCGCGGTTAGAGGGAATCATT
CCAGCACTGGAGACCTCACACGCACTAGCTTACCTCGAGAAGCTATGTCCCACATTATCAGACGGGACGAGAGTGGTCTTGAACTTCAGC
GGAAGAGGAGATAAAGATGTTCAGACAGTGGCTAAATATCTTGATGTTTGA

Table S 3: List of primers used in this study.

Extension of BXD variety	
ZmBX10_Omega_fw	TGAATTTTGCAGCTCGATGGCCCTCATGCAGGAGA
ZmBX10_Omega_rv	CCACAACAAGCACCGTCAAGGATAGACCTCGATGATGAC
ZmBX13_Omega_fw	TGAATTTTGCAGCTCGATGGCTCCGACCGCCG
ZmBX13_Omega_rv	CCACAACAAGCACCGCTAGACGTGGTGTGCAGGAG
ZmBX14_Omega_fw	TGAATTTTGCAGCTCGATGGCACTCATGCAGGAGAG
ZmBX14_Omega_rv	CCACAACAAGCACCGTCATGGATAGACCTCGATGACC
Bacterial genes + His-tag	
RebF_Omega_fw	TGAATTTTGCAGCTCGATGACGATTGAATTCGACAGACC
RebF_Omega_rv	CCACAACAAGCACCGTTATCCTTCCGGGGTCCACA
RebF_His_Omega_rv	CCACAACAAGCACCGTTAGTGATGGTGATGGTGATGTTTAAATCCTTCCGGGGTCCA
TnaA_Omega_fw	TGAATTTTGCAGCTCGATGGAAAACTTTAAACATCTCCCTGA
TnaA_Omega_rv	CCACAACAAGCACCGTTAAACTTCTTTAAGTTTTGCGGTGA
TnaA_His_Omega_rv	CCACAACAAGCACCGTTAGTGATGGTGATGGTGATGTTTAAAAACTTCTTTAAGTTTTGCGGTG
3-LRS_Omega_fw	TGAATTTTGCAGCTCGATGTCAGGTAAAATCGACAAAATCCT
3-LRS_Omega_rv	CCACAACAAGCACCGTCAGCGACCGTGCTGC
3-LRS_His_Omega_rv	CCACAACAAGCACCGTTAGTGATGGTGATGGTGATGTTTAAAGCGACCGTGCTGCTG
KtzQ_Omega_fw	TGAATTTTGCAGCTCGATGGATGATAATCGCATACGATCA
KtzQ_Omega_rv	CCACAACAAGCACCGTTAAGTCCCTTGGTGCAGCC
KtzQ_His_Omega_rv	CCACAACAAGCACCGTTAGTGATGGTGATGGTGATGTTTAAAAGTCCCTTGGTGCAGC
KtzR_Omega_fw	TGAATTTTGCAGCTCGATGGTGATAGTAGGAGGAGGCA
KtzR_Omega_rv	CCACAACAAGCACCGTTAAACGTCTAACCCACGCA

KtzR_His_Omega_rv	CCACAACAAGCACCGTTAGTGATGGTGATGGTGATGTTTAAAAACGTCTAACCCACGCAT
PyrH_Omega_fw	TGAATTTTGCAGCTCGATGATTAGGAGTGTCGTTATAGTAGGT
PyrH_Omega_rv	CCACAACAAGCACCGTCACTGTATTGATGCGAGGT
RebH_Omega_fw	TGAATTTTGCAGCTCGATGTCAGGTAAAATCGACAA
RebH_Omega_rv	CCACAACAAGCACCGTTAACGACCGTGTTGCTGCC
Addition of chloroplastic localization peptide	
CLP_Omega_fw	TTTATGAATTTTGCAGCTCGATGGCAGCCTCAGGCAC
PyrH_CLP_rv	TCCTAATCATGAGAACAGGCGGGTCCTTG
PyrH_CLP_fw	GCCTGTTCTCATGATTAGGAGTGTCGTTATAGTAGGT
RebH_CLP_rv	TACCTGACATGAGAACAGGCGGGTCCTTG
RebH_CLP_fw	GCCTGTTCTCATGTCAGGTAAAATCGACAA
RebF_CLP_rv	CAATCGTCATGAGAACAGGCGGGTCCTTG
RebF_CLP_fw	GCCTGTTCTCATGACGATTGAATTCGACAGACC
Golden Braid *CLP-TnaA*	
GG1_Alpha_TnaA	CACCACAGGTCTCGAATGATGGCAGCCTCAGGCAC
GG2_Alpha_TnaA	CACCACAGGTCTCGGAGAACAGGCGGGTCCTTG
GG3_Alpha_TnaA	CACCACAGGTCTCGTCTCATGAAGGATTATGTAATGGAAAACT
GG4_Alpha_TnaA	CACCACAGGTCTCGAAGCTTAAACTTCTTTAAGTTTTGCGGTGA
Sequencing primers	
Alpha_seq_fw	GGTGATTTTGTGCCGAGCTG
Alpha_seq_rv	GGACGAACGGATAAACCTTTTCAC
Omega_seq_fw	AGATCTGAGTTTCTCCGGTTGT
Omega_seq_rv	ATGCTCAGAACTGCAATCAGAA
pTURBO_seq_fw	ATGATTCGGAGGCTACTGT
pTURBO_seq_rv	AACACTATGCGTTATCGTACG
Cytosolic Trp- biosynthesis genes	
ASA_Omega_fw	TGAATTTTGCAGCTCGATGTCTGTTTCTCCGGAAGCTTCA
ASA_Omega_rv	CCACAACAAGCACCGTCATTTTTTCACAAATGCAGATTCAGC
ASB_Omega_fw	TGAATTTTGCAGCTCGATGGCGGAATCGAATTCGATAC
ASB_Omega_rv	CCACAACAAGCACCGCTATGTCAGCTTCTCGGACTC
PAI_Omega_fw	TGAATTTTGCAGCTCGATGTGTGGCATCACATCAGCC
PAI_Omega_rv	CCACAACAAGCACCGTTAGTAGTGTACAGAGCGAACTG
PAT_Omega_fw	TGAATTTTGCAGCTCGATGAGCGGTGGATTTAGTGCAGC
PAT_Omega_rv	CCACAACAAGCACCGTCACTGAGATTTCTGAGCTAAGTTT
IPGS_Omega_fw	TGAATTTTGCAGCTCGATGGCTCAACAGTCGGATTTGAAG
IPGS_Omega_rv	CCACAACAAGCACCGCTAAGTATGAGAAATGTTCCTGCC
TSA_Omega_fw	TGAATTTTGCAGCTCGATGGCTTCTCTCTCCACCTC
TSA_Omega_rv	CCACAACAAGCACCGTCAAAGAAGAGCAGATTTAAGAGAC
TSB_Omega_fw	TGAATTTTGCAGCTCGATGGCCGCCGGATCTGA
TSB_Omega_rv	CCACAACAAGCACCGTCAAACATCAAGATATTTAGCCACTGT

Eigenständigkeitserklärung

1. Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich trage die Verantwortung für die Qualität des Textes sowie die Auswahl aller Inhalte und habe sichergestellt, dass Informationen und Argumente mit geeigneten wissenschaftlichen Quellen belegt bzw. gestützt werden. Die aus fremden oder auch eigenen, älteren Quellen wörtlich oder sinngemäß übernommenen Textstellen, Gedankengänge, Konzepte, Grafiken etc. in meinen Ausführungen habe ich als solche eindeutig gekennzeichnet und mit vollständigen Verweisen auf die jeweilige Quelle versehen. Alle weiteren Inhalte dieser Arbeit ohne entsprechende Verweise stammen im urheberrechtlichen Sinn von mir.
2. Ich versichere des Weiteren, dass die vorliegende Arbeit bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt wurde oder in deutscher oder einer anderen Sprache als Veröffentlichung erschienen ist.
3. Mir ist bekannt, dass ein Verstoß gegen die vorbenannten Punkte prüfungsrechtliche Konsequenzen haben und insbesondere dazu führen kann, dass meine Prüfungsleistung als Täuschung und damit als mit „nicht bestanden“ bewertet werden kann. Bei mehrfachem oder schwerwiegendem Täuschungsversuch kann ich befristet oder sogar dauerhaft von der Erbringung weiterer Prüfungsleistungen in meinem Studiengang ausgeschlossen werden.

……..

Ort und Datum

……..

Unterschrift