Friedrich-Schiller-Universität Jena

Biologisch-Pharmazeutische Fakultät

Max-Planck Institut für chemische Ökologie

Abteilung Biochemie

Arbeitsgruppe Biosynthese und Funktion flüchtiger Stoffe in Gehölzpflanzen und Gräsern

Cyanogenic glycosides in *Zea mays*

Bachelorarbeit zur Erlangung des Grades eines

Bachelor of Science

vorgelegt von

Elias Kalthoff

aus Köln

Jena, den 24.09.2015

Gutachter:

Dr. Tobias Köllner

Prof. Dr. Frank Große

Contents

List of tables

List of figures

List of abbreviations

Bp	Base pair(s)
cDNA	Complementary DNA
cps	Counts per second
CNglc	Cyanogenic glycoside
DIBOA	2,4-dihydroxy-1,4-benzoxazin-3-one
DIMBOA	2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleic triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
fwd	Forward
g	Standard gravitational acceleration
H_2O	Distilled water
HCN	Hydrogen cyanide
HF	High fidelity
LB	Lysogenic broth
LOD	Logarithm of odds
NAM	Nested Associated Mapping
NEB	New England Biolabs
OD	Optical density
PCR	Polymerase chain reaction
QTL	Quantitative trait locus
rev	Reverse
rpm	Revolutions per minute
RNA	Ribonucleic acid
RNase	Ribonuclease
RT	Reverse transcriptase
SC	Synthetic complete
SNP	Single nucleotide polymorphism
SOC	Super optimal broth with catabolite repression
TAE	TRIS-Acetate-EDTA
TE	TRIS-EDTA
TEG	TRIS-EDTA-glycerol
TEK	TRIS-EDTA-KCL
TES	TRIS-EDTA-sorbitol
TRIS	Tris(hydroxymethyl)-aminomethane
UGT	UDP-glycosyltransferase
YPGA	Yeast peptone glucose agar
ROS	Reactive oxygen species

1 Introduction

1.1 Cyanogenic glycosides

Cyanogenic glycosides (CNglcs) are an ancient class of defensive secondary metabolites classified as Phytoanticipins, preformed compounds that release toxic substances upon herbivory. Chemically, CNglcs are α-hydroxynitriles, derivatives of amino acids, that are stabilized by a β-1,4 linked sugar moiety.

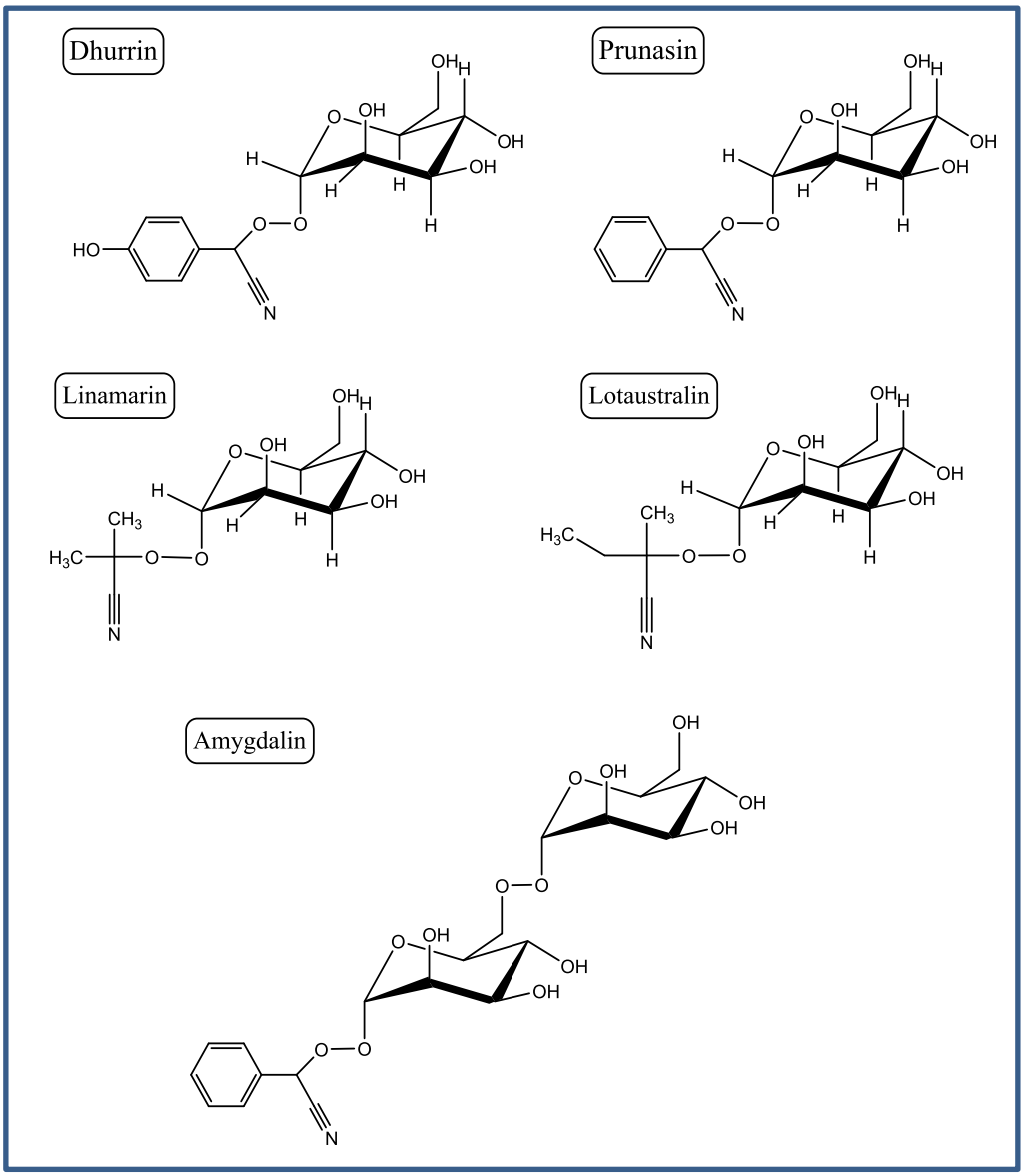

Figure 1: Structures of five common cyanogenic glycosides. Dhurrin and prunasin are derived from the aromatic parent amino acids tyrosine and phenylalanine, respectively, while linamarin and lotaustralin are normally co-occurring derivatives of leucine and isoleucine, respectively. Amygdalin is the diglycoside of prunasin and thought to be an inactive storage form. (Gleadow et al., 2014)

When the sugar moiety is hydrolyzed by the action of a β-glucosidase, the instable α-hydroxynitrile degrades into HCN and an aldehyde or keto compound. Under physiological conditions, CNglcs and β-glucosidases are separated in different compartments, the CNglcs are normally localized in the vacuole while the β-glucosidase can typically be found in the chloroplast or the apoplastic space (Zagrobelny et al., 2008), but can also be localized in a completely different tissue. Only after disruption of the plant tissue, for example by the chewing of a feeding herbivore, are the two parts of this binary system brought together.

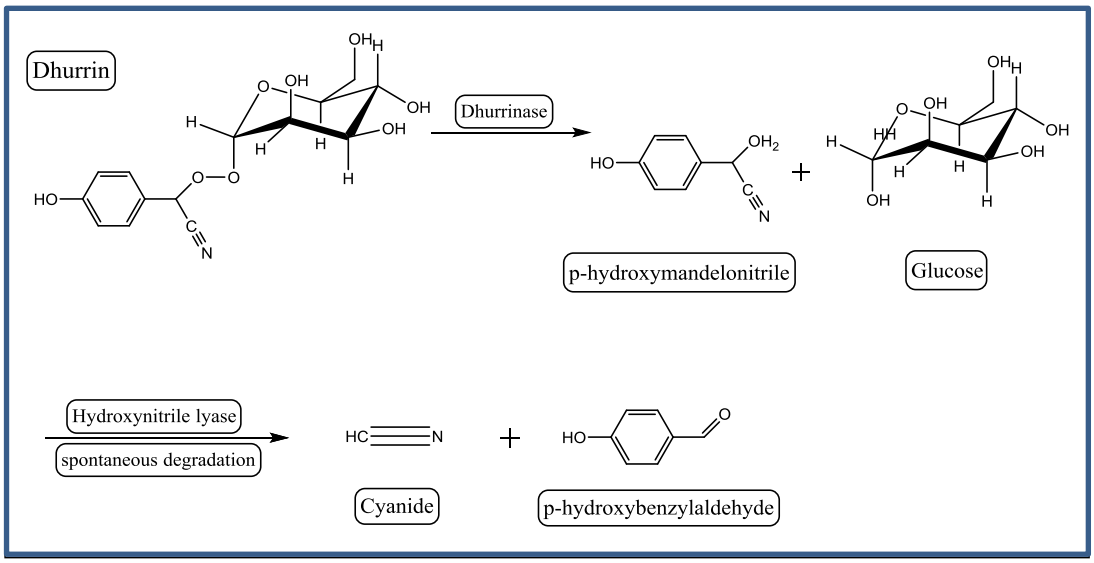

Figure 2: Activation of dhurrin. The β-1,4 glycosidic bond is hydrolyzed by the action of the β-glucosidase dhurrinase. The α-hydroxynitrile then degrades into HCN and an aldehyde. The degradation of *p*-hydroxymandelonitrile can occur spontaneously at high pH, but is also catalyzed by an *α*-hydroxynitril lyase.

Phylogenetic analysis of the enzymes involved in the production of CNglcs shows that the first step in the synthesis predates the separation of angiosperms and gymnosperms (Bak et al., 2006) which puts their age at least at 300 million years. With over 3000 known cyanogenic plants in over 130 families, CNglcs are widespread throughout the plant kingdom (Busk et al., 2002). Cyanogenic plants encompass important food crops such as Sorghum (*Sorghum bicolor*), barley (*Hordeum vulgare L.*) or cassava (*Manihot esculenta*). Even some arthropods species are able to sequester CNglcs from their food sources with the notable exception of the burnet moths (*Zygaenidae filipendulae*), which are able to *de-novo* synthesize CNglcs (Zagrobelny et al., 2008). Interestingly the low sequence identity between plant and insect genes coding for CNglcs suggests that both biosynthesis pathways

evolved independent from one another (Gleadow et al., 2014) and are not a result of horizontal gene transfer.

While the main purpose of CNglcs in plants is the deterrence of herbivores, some studies indicate that they might also be used as nitrogen and glucose storage. Experiments showed that grown Sorghum plants that were fertilized with high levels of nitrogen exhibited an increased cyanide potential (Busk and Møller, 2002). In the rubber tree (*Hevea brasiliensis)*, linamarin is transformed into the diglycoside linustatin which is then transported over a long distance to the seedlings were it is broken down by a α-diglucosidase and a β-cyanoalanine synthase to obtain aspartatic acid or asparagine (Selmar et al., 1988). Similarly bitter almonds (*Prunus dulcis)* transport the CNglc prunasin into the developing cotyledons (seed leaves) where it is further glycosylated, forming the corresponding diglycoside amygdalin. At the beginning of germination, 80% of the amygdalin is then turned over in a 3 week period (Sánchez-Pérez et al., 2008). Furthermore a possible pathway for the detoxification of CNglcs without the release of HCN has been hypothesized (Jenrich et al., 2007). These studies open up the possibility that rather than being purely defensive compounds, CNglcs are also used as a metabolic buffer tying up excess nitrogen that can be used in times of need. Additionally their role as antioxidants is being investigated. It has been shown that in the rubber tree levels of CNglcs diminish over the course of the day and are replenished in the night (Kongsawadworakul et al., 2009). Also, primary amide glycosides have been found in the rubber trees, which correspond to the CNglcs present. These amides could be products of CNglcs reacting with ROS (Reactive oxygen species), originating from exposure to high doses of sunlight (Møller, 2010). All these findings demonstrate that CNglcs are a group of secondary metabolites that have acquired a wide range of functions in the course of their evolution.

1.2 Biosynthesis of cyanogenic glycosides and the CYP79 family

Over 60 different CNglcs are known today originating from only seven different amino acids: L-valine, L-leucine, L-isoleucine, L-phenylalanine, L-tyrosine and the non proteinogenic 2-(20-cyclopentenyl)-glycine and the 2-(2'-cyclopentenyl)-glycine epimers (Gleadow et al., 2014). These amino acids are transformed into CNglcs by three different enzymes, two multifunctional cytochrome P450 monooxygenases and one glucosyltransferase. The biosynthesis pathway has first been elucidated in Sorghum (McFarlane et al., 1975; Møller and Conn, 1979; Shimada and Conn, 1977), where the parent amino acid tyrosine is transformed into dhurrin (Figure 3), the most abundant CNglc in this species. The first steps in the biosynthesis of dhurrin are two N-hydroxylations, followed by a dehydration and a decarboxylation reaction mediated by a single enzyme, the multifunctional cytochrome P450 enzyme CYP79A1 (Bak et al., 1998). The product (*E*)-*p*-hydroxyphenylacetaldoxime is then dehydrated and further hydroxylated by CYP71E1 to produce *p*-hydroxymandelonitrile. The final step in the biosynthesis of dhurrin is the glycosylation of *p*-hydroxymandelonitrile by the glucosyltransferase UGT85B1. Other CNglcs are thought to be synthesized in the same way.

Studies with radio-labelled intermediates (Møller and Conn, 1980) showed that the biosynthesis pathway is highly channeled and fluorescence tagging of UGT85B1 revealed a shift in fluorescence from cytosolic to the ER, when CYP79A1 and CYP71E1 were expressed (Kristensen et al., 2005). This shift could be the result of UGT85B1 forming non covalent bonds with the membrane-bound CYP79A1 and CYP71E1. These findings suggest that the three enzymes form a metabolon, a multi-enzyme complex that facilitates swift transportation of the instable and toxic intermediates. Such a metabolon would also help to keep undesired metabolic cross-talk at bay and possibly allow for easier regulation of the biosynthetic pathway. Analysis of the substrate specificities of the enzymes involved indicated that CYP79A1 is the most specific and rate limiting enzyme in the biosynthesis of dhurrin (Kahn et al., 1999), thereby ensuring specificity of the pathway while allowing the following enzymes to be less specific. Further investigation revealed a unique substitution in the otherwise conserved PERF motif of other P450s (Bak et al., 2006).

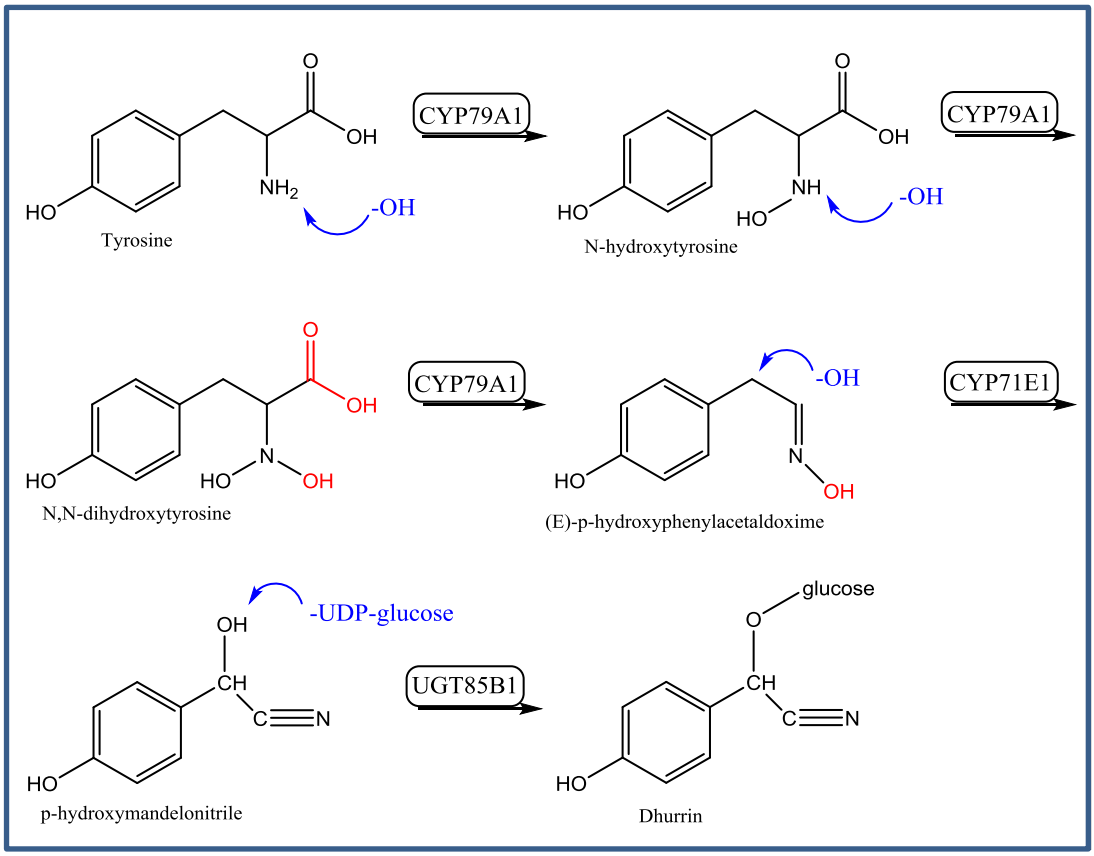

Figure 3: Biosynthesis of dhurrin. Nucleophile attack groups are shown in blue and leaving groups in red. To sequential N-hydroxylations are followed by a decarboxylation and a dehydration reaction mediated by the multifunctional CYP79A1. The resulting acetaldoxime is further dehydrated at the oxime nitrogen and C-hydroxylated by CYP71E1. The instable hydroxynitrile is subsequently converted into the CNglc by the addition of a stabilizing sugar trough the UDP-glucosyltransferase UGT85B1 (modified from Gleadow et al., 2014).

Instead of the aromatic amino acid phenylalanine, the positively charged histidine is used, giving CYP79 enzymes the family consensus PERH (Bak et al., 2006). This and another substitution in the heme-binding region are both lying within the active site, the consequences of these changes are currently not understood, but make the identification and isolation of CYP79 genes easier. The second P450 enzyme CYP71E1 belongs to the CYP71 subfamily of plant P450s (Werck-Reichert et al., 2002). It is by far the biggest subfamily and its members are involved in the production of many different secondary metabolites such as isoprenoids and terpenes as well as other plant defense products like DIMBOA and DIBOA (Gierl, 2005). The final step in dhurrin biosynthesis is the glycosylation of *p*-hydroxymandelonitrile by UGT85B1. Glycosylation is ubiquitous among secondary plant metabolites, as it promotes water solubility as well as stability, and can provide a signal for compartmentalization (Ross et al., 2001).

Most plant UGTs are localized in the cytoplasm but a hydrophobic patch has been discovered in UGT85B1 that could mediate the contact to the ER bound CYP79 and CYP71 (Thorsøe et al., 2005).

1.3 Cyanogenic Glycosides in Maize

While CNglcs are in general well studied in plants like Sorghum or cassava, their occurrence in maize is not well documented. A few older studies report certain lines of maize to be cyanogenic (Brünnich et al., 1903; Lehman et al., 1979; Erb et al., 1981a). Interestingly, like in Sorghum the greatest concentration of CNglcs was found in the developing seedlings, with concentrations diminishing over the course of plant development (Erb et al., 1981b). But while Sorghum seedlings contain up to 6 mg of Dhurrin in 100 g of dry mass, maize seedlings cyanogenic potential was reported to be much lower at 0,4 µg of released HCN per 100 g of dry mass (Erb et al., 1981b). These figures are not strictly corresponding but give a notion that concentrations of CNglcs might be significantly lower in maize than they are in Sorghum. In a newer study (Irmisch et al., 2015), an analysis of all sequenced angiosperms genomes using CYP79A1 from Sorghum as template, revealed four CYP79 genes in the genome of the maize line B73 (Irmisch et al., 2015). All these genes contain features that are typical for the CYP79 family, such as the PERH motif or the specific substitutions in the heme-binding region. One of the genes, GRMZM2G138248, showed 72% amino acid identity with CYP79A1 from Sorghum and was designated CYP79A61.While these findings revealed a candidate gene for the biosynthesis of CNglcs in maize, further experiments using the maize line Delprim showed no traceable amounts of CNglcs, after the expression of CYP79A61 was induced with caterpillar secretion (Irmisch et al. 2015).

1.4 Maize nested association mapping population

Maize is not only one of the world's most important food crops, with the International Grains Council forecasting the world production for 2015 at around 1 billion tons, but also a model organism for genetic studies. Scientific breakthroughs such as the discovery of transposable elements in the genome (McClintock, 1950) or the existence of epigenetic phenomena (Coe, 1966) were made in maize. This role as a model organism is partly due to its great genetic diversity, ease of handling as well as extensive nucleotide diversity. One field of research has been the linking of phenotypic traits to the genotype, making use of molecular markers such as SNPs (single nucleotide polymorphisms) to find QTLs (quantitative trait loci). These QTLs are regions in the genome that are correlated to a certain trait in the phenotype and therefore contain candidate genes, which are responsible for that phenotype. There are two approaches for finding QTLs, association studies and linkage mapping.

Linkage mapping is based on the assumption that the shorter the distance between genes or genetic markers on one chromosome is, the more likely they are to be inherited together in a recombination event. Therefore the segregation frequency of two genes or genetic markers during recombination is correlated to their physical distance (Griffiths et al., 1999). In most maize studies that use linkage mapping, two parent lines with relative few but well-known genetic markers are crossed to obtain inbred lines, which are then scanned for the phenotypic trait in relation to the distribution of markers. QTLs for that trait can then be determined using statistical methods, because the trait will appear most often in those inbred lines where the two genetic markers encompassing the QTL are not separated during recombination. In contrast, association studies rely on the large-scale sequencing of a genome in search of SNPs. In a case-control design, plants that exhibit the desired phenotypes and control plants are screened and the occurrence of SNPs is then correlated with the trait to ascertain one or more QTLs. To combine these two methods, the Nested Associated Maize (NAM) population has been developed (Flint-Garcia et al, 2005). The fully sequenced line B73 was crossed with 25 other parent lines, which were either densely genotyped or sequenced. These lines encompassed commercial, tropical and even a popcorn line to cover most of the genetic diversity found in different maize lines (McMullen et al., 2009). Their progeny was then self-fertilized to the F_6 generation to create 5000 RILs (Recombinant inbred lines). For these RILs

and the 26 founder lines 1106 SNPs were genotyped. These SNPs were especially chosen for their rarity and are only found on genetic regions from the B73 founder line. Finally the detailed genomic information from the 26 founder lines was overlaid onto the RILs, as the 1106 SNPs allowed the allocation of specific regions in the genome to one of the two parent lines (Yu et al., 2008). This design allows for high statistic power in a linkage analysis in search for QTLs, without the need to sequence or densely genotype the 5000 RILs. It also makes experimental procedures easier as the result of a phenotypic screening can now be correlated directly to the genomic regions using a database.

1.5 Research objectives

To establish a general understanding of the occurrence of CNglcs in maize, especially in regard to maize's great genomic diversity, it was planned to screen the NAM parental lines for CNglcs using liquid chromatography-mass spectrometry (LC-MS). Subsequently, to identify a QTL in the maize genome, the RILs of a line with a high content of CNglcs should be analyzed. Another aim was to functionally characterize the products of the four CYP79 genes discovered by Irmisch et al., utilizing molecular cloning and a yeast expression system with the goal of ascertaining which of the four CYP79-enzymes is responsible for the expected production of CNglcs in maize.

2 Material and Methods

2.1 Material

2.1.1 Plant material

The NAM parent lines used were supplied by the Northern Central Regional Plant Introduction Station (Ames, Iowa, USA) and the NAM RILS by the Maize Genetics Cooperation Stock Center (Urbana, Illinois, USA).

2.1.2 Chemicals

If not stated otherwise, all chemicals were supplied by the company Carl Roth (Karlsruhe, Germany).

2.1.3 Vectors

Table 2.1: List of vectors with their mediated resistance and their use

Name	Resistance	Use
TOPO Zero Blunt (Invitrogen, Darmstadt, Germany)	kanamycin	Routine cloning and transformation.
pESC-Leu (Agilent Technologies, Waldbronn, Germany)	ampicillin	Yeast transformation.

2.1.4 Microbial organisms

Table 2.2: List of used microbial organisms with their genotype and resistance to antibiotics

Name	Organism	Genotype/Features	Resistance/ Selection
10-beta cells (New England BioLabs, Frankfurt am Main, Germany)	*Escherichia coli*	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZ ΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	Strep, T1-Phage (fbuA2)
W(R)	*Saccharomyces cerevisiae*	Engineered from W303–1B to carry the yeast cytochrome P450 reductase gene YRED (Pompon et al., 1996)	leucin auxotrophy
WAT11	*Saccharomyces cerevisiae*	Engineered from W303-1B to carry the Arabidopsis cytochrome P450 reductase 1 (CPR1). (Pompon et al., 1996)	leucin auxotrophy

2.1.5 Oligonucleotides

Table 2.3: List of oligonucleotides with their sequence and targets. All oligonucleotides were manufactured by Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany)

Name	Sequence (5'-3')	Target site
M13 forward **M13 reverse**	GTA AAA CGA CGG CCA GCA GGA AAC AGC TAT GAC	*LacZ* gene on the TOPO®-Zero Blunt® Vector.
138248 forward	ATG GTT TCC TC TCC GCA AGC A	*CYP79-1* (GRMZM2G138248)
138248 reverse	CTA CTT ACA AAG CAA GAT CCC G	*CYP79-1* (GRMZM2G138248)
011156 forward	ATG GCG CTA GCA CCT AGC CAT	*CYP79-2/3/4*
011156 reverse	CTA GCC AGA CGC ATA GAG ATG	*CYP79-2*
0178351 reverse	TTA ACC GGA CGC CGC ATA GAG	*CYP79-3/4*
79-1x-pESC-MCS1 forward	AAG CGG CCG CAA TGG TTT CCT CTC CGC AAG CAA AT	*CYP79-1* The primers carry recognition sequences for the restriction enzymes *Bgl*II and *Not*I at their 5'-ends (underlined).
79-1x-pESC-MCS1 reverse	TTA GAT CTC TAC TTA CAA AGC AAG ATC CCG G	
Oligo(dT)$_{20}$	TTTTTTTTTTTTTTTTTTTT	poly-A tail
Gal10 forward	GGT GGT AAT GCC ATG TAA TAT G	Gal10 gene on the pESC-Leu vector.
Gal 10 reverse	GGC AAG GTA GAC AAG CCG ACA AC	

2.1.6 Laboratory equipment

Table 2.4: Laboratory equipment

Instrument	Manufacturer	Use
NanoDrop® 2000c	Thermo Scientific, Dreieich, Germany	Concentration measurements of nucleic acids.
peqSTAR® 2X Gradient Thermocycler	PEQLAB, Erlangen, Germany	PCR and incubation of reactions.
ABI Prism®-Gen-Analysator 3130xl	Applied Biosystems, Darmstadt, Germany	Capillary electrophoresis and fluorescence detection for sequencing
Mupid®-One gel chamber	Eurogenetec, Cologne, Germany	Electrophoresis
UHPLC-System Model 1260 infinity	Agilent Technologies, Santa Clara, CA, USA	Liquid chromatography
API 5000® Triple Quadrupol MS/MS-System	AB SCIEX GmbH, Darmstadt, Germany	Mass spectrometry
XDB-RP18 Column (1.8 µm particle size, 4.6 x 50 mm)	Agilent Technologies, Santa Clara, CA, USA	Liquid chromatography; exchangeable column
B6120 Kendro	Heraeus, Hanau, Germany	Incubator
CERTOMAT® BS-1	B. Braun Biotech International, Melsungen, Germany	Incubator shaker
Typ 1002	GFL, Burgwedel, Germany	Water bath
Gene Genius	SYNGENE, Cambridge, UK	Gel documentation system

2.1.7 Composition of mediums, buffers, agar plates and agarose gels

Table 2.5: Compositions of different mixtures

Compound	Components	Use
0,5x TAE	0,5 mM EDTA, 20 mM TRIS, 10 mM glacial ethanoic acid	Gel electrophoresis buffer
TEK	50 mM Tris-HCL (pH: 7,5), 1 mM EDTA, 100 mM KCL	Isolation of microsomes
TES	50 mM Tris-HCL (pH: 7,5), 1 mM EDTA, 600 mM sorbitol	Isolation of microsomes
TEG	50 mM Tris-HCL (pH: 7,5), 1 mM EDTA, 30% glycerol	Isolation of microsomes
Sodium-Phosphate Buffer	2,34 g sodium dihydrogen phosphate dehydrate, 200 ml H_2O, pH 7,4	Enzyme assays
Agarose gel	1,5% agarose (w/v), 0,5x TAE-buffer, 0,25 µl ethidium bromide per 100 ml of gel	Separation medium for electrophoresis
Agar plates	35 g LB-agar (Lennox) (Roth) in 1 l H_2O with 50 µg/ml kanamycin or 100 µg/ml ampicillin	Growing of bacterial colonies
Liquid LB-medium	20 g LB-medium (Lennox) (Roth) in 1l H_2O with 50 µg/ml kanamycin or 100 µg/ml ampicillin	Growing of overnight cultures
YPGA (glucose/galactose)	900 ml H_2O, 10 g yeast extract, 20 g bacto peptone, 74 mg adenine-hemisulfate and 100 ml sterile glucose/galactose solution (20%)	Growing of yeast cultures
SC-minimal medium (without leucine)	900 ml H_2O, 6,7 g yeast nitrogen base, 100 mg adenine-hemisulfate, 100 mg uracil, amino acids (100 mg): R,C,K,T and W, amino acids (50 mg): D,H,I,M,F,P,S,Y and V, 100 ml sterile glucose solution and 20 g agar	Selection plates for the yeast transformation

2.2 Extraction and analysis of cyanogenic glycosides

2.2.1 Germination and harvesting of *Z. mays* seedlings

Five *Z. mays* seedlings were spread out onto a filter paper and put into a petri dish. The filter paper was then moistened with 5 ml of tap water. To allow germination, the petri dishes were incubated at 28°C for five days. After the incubation period the seedlings were broken up into roots, stem and corn and immediately frozen in liquid nitrogen.

2.2.2 Methanol extraction of cyanogenic glycosides

The harvested plant material was ground using pestle and mortar while being kept frozen with liquid nitrogen. The plant powder was then weighted and transferred into reaction tubes. Subsequently for each mg of plant powder, 3 µl of methanol were

added and the reaction tube vortexed rigorously. The resulting solution was then shaken in a paint shaker for 3 minutes and centrifuged at full speed for 10 minutes. Finally the supernatant was transferred into a glass vial for further analysis (modified from Irmisch et al., 2015).

2.2.3 Liquid Chromatography - Mass Spectrometry

For the analysis of CNglcs (dhurrin, prunasin, amygdalin, linamarin, and lotaustralin) and enzyme assays, five ml of the extracts were directly injected and analyzed by LC-MS/MS. The elution profile was: 0– 0.5 min, 5 % B in A; 0.5–6.0 min, 5–50 % B; 6.1–7.5 min 100 % B and 7.6–10.5 min 5 % B. The flow rate was set to 1.1 mL min−1. The tandem mass spectrometer was operated in negative ionization mode (ion spray voltage, −4500 eV; turbo gas temp, 700 °C; nebulizing gas, 60 psi; curtain gas, 30 psi; heating gas, 50 psi; collision gas, 6 psi). MRM was used to monitor parent ion→product ion reactions for each analyte as follows: m/z 310.0→ 179.0 for dhurrin, m/z 294.0→89.0 (CE, −22; DP, −15) for prunasin, m/z 260.0→179.0 for lotaustralin, m/z 246.0→179.0 for linamarin, and m/z 456.0→179.0 for amygdalin. If not stated above, the transition parameter settings for the CNglcs were as follows: CE, −10 V; DP, −15 V. Quantification was done by integrating the peaks for the five compounds dhurrin, linamarin, lotaustralin, prunasin and amygdalin using the Analyst 1.6 Software (AB SCIEX, Darmstadt, Germany). The analyte peak area was then translated into the concentration of analyte by using a calibration line generated with a commercial dhurrin standard (Sigma-Aldrich) that was analyzed along with the plant extracts. As no standards for the other CNglcs were available, the dhurrin calibration line was used for all of the five analytes.

Aldoximes were measured from methanol extracts using the same LC-MS/MS system. Formic acid (0.2%) in water and acetonitrile were employed as mobile phases A and B respectively, on a Zorbax Eclipse XDB-C18 column (50 × 4.6 mm, 1.8 µm; Agilent Technologies). The elution profile (gradient 1) was 0 to 0.5 min, 30% B; 0.5 to 3 min, 30 to 66% B; 3 to 3.1 min, 66 to 100% B; 3.1 to 4 min 100% B; and 4.1 to 6 min 30% B at a flow rate of 0.8 mL min^{-1} at 25°C. For the analysis of the more polar *p*-hydroxyphenylacetaldoxime, the gradient was modified as follows (gradient 2): 0 to 4 min, 10 to 70% B; 4 to 4.1 min, 70 to 100% B; 4.1 to 5 min 100% B; and 5.1 to 7 min, 10% B at a flow rate of 1.1 mL min^{-1}. The API 5000 tandem mass spectrometer was operated in positive ionization mode (ion spray voltage, 5500

eV; turbo gas temperature, 700°C; nebulizing gas, 60 p.s.i.; curtain gas, 30 p.s.i.; heating gas, 50 p.s.i.; collision gas, 6 p.s.i.). MRM was used to monitor parent ion→product ion reactions for each analyte as follows: *m*/*z* 136.0 → 119.0 (collision energy [CE], 17 V; declustering potential [DP], 56 V) for phenylacetaldoxime; *m*/*z* 102.0 → 69.0 (CE, 13 V; DP, 31 V) for 2-methylbutyraldoxime; *m*/*z* 102.0 → 46.0 (CE, 15 V; DP, 31 V) for 3-methylbutyraldoxime; *m*/*z* 175.0 → 158.0 (CE, 17 V; DP, 56 V) for indole-3-acetaldoxime; and *m*/*z* 152.0 → 107.0 (CE, 27 V; DP, 100 V) for *p*-hydroxyphenylacetaldoxime (modified from Irmisch et al., 2015).

2.3 RNA and cDNA-based methods

2.3.1 RNA extraction

The RNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) was used to extract RNA from ground plant material. The manufacturer's protocol was followed and the RNA eluded in 40 µl of RNase free water. To inhibit any RNase activity, the plant material was kept frozen in liquid nitrogen until the addition of lysis buffer. Furthermore sterile working procedures were used.

2.3.2 cDNA synthesis

After RNA-Extraction, the extracted RNA was reverse transcribed utilizing the SuperScript® III Reverse Transcriptase Kit (Invitrogen). To avoid any contaminations of genomic DNA, one µg of RNA was incubated with DNase I and DNase I buffer (Fermentas, St.Leon-Rot, Germany) for 30 minutes at 37°C to digest any DNA present. The reaction was stopped by incubating at 65°C for 10 minutes. The reverse transcription was then set up by adding deoxynucleotides and an oligonucleotide primer. After a five minute incubation period at 65°C, buffer, DTT and RNase Out (Invitrogen) were added and the following thermal cycling routine was carried out:

Table 2.6: Pipetting pattern for cDNA snythesis

Substance	Volume (µl)
RNA (1 µg in H_2O)	11
dNTPs	1
Oligonucleotide primer	1
Buffer (First Strand 5x)	4
DTT (0,1 M)	1
RNase Out	1
Reverse Transcriptase (Super Script III)	1

Table 2.7: Thermocycling program for cDNA synthesis

Step	Temperatur (°C)	Time (s)
Initial attachment	25	300
Annealing	50	3600
Reverse transcription	70	900

2.4 PCR-based techniques

2.4.1 PCR

Polymerase chain reaction (PCR) is a technique that makes use of a thermostable polymerase to achieve exponential amplification of DNA. Repeating cycles of varying temperatures are used to facilitate denaturing of the DNA strands, annealing of the primers and synthesis of a new strand. PCR was generally carried out using Phusion® High Fidelity Polymerase (Thermo Scientific), a proof-reading Phu Polymerase, as well as site specific primers. If not outlined otherwise, the following program and reaction mixture was used for routine amplification of DNA:

Table 2.8: Reaction mixture for the Phusion-PCR reaction

Substance	Volume (µl)
Buffer (Phusion HF 5x)	10
H_2O	32,5
dNTPs	1,0
Primer fwd.	2,5
Primer rev.	2,5
cDNA	1 ,0
Polymerase (Phusion)	0,5

Table 2.9: Thermocycling program for Phusion-PCR

Step		Temperatur (C°)	Time (s)
Initial denaturing		98	30
Denaturing		98	10
Annealing	**35x**	55	30
Elongation		72	45
Final elongation		72	600

2.4.2 Colony-PCR

To screen bacterial colonies for transformants with the desired insert, Colony-PCR was carried out to amplify the insert and then screen the products for the right size using gel electrophoresis. For cost saving purposes the GoTaq® -Polymerase (Promega, Mannheim, Germany), with no proof-reading capabilities, was used.

Table 2.10: Reaction mixture for Colony-PCR

Substance	Volume (µl)
Buffer (GoTaq Buffer)	5
H_2O	18,375
dNTPs	0,5
Primer fwd.	0,5
Primer rev.	0,5
GoTaq	0,125

Table 2.11: Thermocycling program for Colony-PCR

Step		Temperatur (°C)	Time (s)
Initial denaturing		95	600
Denaturing		95	30
Annealing	**35x**	55	30
Elongation		72	90
Final elongation		72	300

2.4.3 Sequencing

For sequencing with the Sanger method (Sanger et al, 1977) a special PCR routine was adopted. As the method of chain termination by dideoxynucleotides is only accurate for constructs up to 800 base pairs and the gene of interest encompassed roughly 1500 base pairs, two reactions had to be used, one with the forward and the other with the reverse primer. To label and prepare the reactions, the BigDye® Terminator Cycle Sequencing v2. Kit (PE Biosystems, Foster City, USA) was used

and the reactions prepared according to their protocol. The cycling program applied comprised the following steps:

Table 2.12: Reaction mixture for a sequencing reaction

Substance	Forward Reaction	Reverse Reaction
Big Dye Puffer	2 µl	2 µl
Big Dye	2 µl	2 µl
Plasmid	2 µl	2 µl
H_2O	3 µl	3 µl
Primer fwd.	1 µl	/
Primer rev.	/	1 µl

Table 2.13: Thermocycling program for a sequencing reaction

Step		Temperatur (°C)	Time (s)
Initial denaturing		95	600
Annealing	**35x**	50	20
Denaturing		95	10
Elongation		60	240

The DNA was then purified using the DyeEx® Spin Column Kit (Qiagen). Finally the sequencing was performed using capillary electrophoresis and fluorescence detection in the ABI Prism®-Gen-Analysator.

2.5 Microbial techniques

2.5.1 Overnight Cultures

For the purpose of multiplying transformed beta-10 cells, 3,5 ml of LB broth containing a kanamycin concentration of 50 µg/µl were inoculated with microbes using a toothpick. The microbes were then allowed to grow overnight in a shaking incubator.

2.5.2 Yeast cultivation

The two yeast strains WAT11 and W(R) were inoculated into sterile Erlenmeyer flasks containing 30-50 ml YPGA (glucose) medium and then allowed to grow in a shaking incubator at 180 rpm and 28°C until the desired OD was obtained.

2.5.3 Yeast transformation

The EasyComp™ Transformation Kit (Invitrogen) was used for the preparation of competent yeast cells and the subsequent transformation. After the preparation of competent WAT11 and W(R) cells, aliquots were created and frozen, while the remaining cells were immediately used for the transformation.

2.5.4 Microsome extraction

The two yeast strains WAT11 and W(R) were inoculated into sterile Erlenmeyer flasks containing 30 SC- minimal medium and then allowed to grow in a shaking incubator at 180 rpm and 28°C overnight. The following day, 100 ml of YPGA (glucose) were inoculated with one unit $OD_{600,}$ equaling around $2*10^7$ cells, and grown overnight at 180 rpm and 28°C. Subsequently the cultures were centrifuged at 5000 g and 16°C for 5 minutes, the supernatant was discarded and the pellet resuspended in 100 ml YPGA (galactose) medium to induce the expression. The culture was grown overnight at 160 rpm and 25°C. For the microsome extraction, the cultures were first centrifuged at 7500 g and 4°C for 10 minutes, the pellet was thereafter resuspended in 30 ml TEK buffer, centrifuged again at 7500 g for 10 minutes and then resuspended in 2 ml of TES buffer. To disrupt the cells integrity, glass spheres were added and the solution manually shaken for 1 minute and then put on ice for 1 minute. This process was repeated 5 times. The ruptured cells were then washed out with 20 ml TES buffer, collected and then centrifuged for 10 minutes at 7500 g and 4°C. The supernatant was subsequently spun in an ultracentrifuge at 100.000 g for 90 minutes. The remaining supernatant was collected and the contained microsomes homogenized in a potter and stored at -20°C in aliquots.

2.6 Molecular cloning

2.6.1 Agarose gel analysis and extraction

To confirm the right size of PCR-products and digested plasmids, agarose gel electrophoresis was performed. The gels were cast with a 1,5 % agarose content using TAE buffer as solvent. Furthermore, Midori Green Advanced DNA Stain (NIPPON Genetics, Düren, Germany) was added at a ratio of 2,5 µl/100ml for the detection of DNA with UV light. The gel was loaded with a 1kb ladder (Thermo Scientific Gene Ruler) and the samples mixed with DNA loading dye (6x Thermo

Scientific). The electrophoresis was performed at a voltage of 135V for 20 minutes. Imaging was done with a gel documenting system. If a preparative gel electrophoresis was performed, the right sized bands would be cut out using a scalpel.

2.6.2 TOPO-cloning

For the ligation and subsequent transformation of PCR-products into bacterial cells the TOPO®-Zero Blunt® Vector (Invitrogen) was used. By utilizing a topoisomerase fused to both ends of the linear vector, the need for a digestion step prior to the cloning was bypassed as the topoisomerase would ligate the blunt PCR product within 30 minutes at room temperature.

Table 2.14: Reaction mixture for TOPO-cloning

Substance	Volume (µl)
PCR-Product	1
Salt Solution	0,5
Vector	0,5
H_2O	1

2.6.3 Transformation

For the transformation of chemically competent 10-β *E. coli* cells (NEB) a simple heat shock regimen was adopted. The cells were thawed on ice and then incubated with the plasmids for 20 minutes on ice. Subsequently a heat shock at 42°C for 45 seconds was applied. Afterwards the cells were allowed to recover on ice for 1 minute, then 100 µl of SOC-medium (Invitrogen) was added, followed by a growing period of 45 minutes in a shaking incubator. Finally the transformed cells were plated out onto agar plates containing 50 mg/ml of kanamycin and incubated at 37°C overnight.

2.6.4 Plasmid purification

The Nucleo-Spin® Plasmid DNA Purification Kit (Macherey-Nagel, Dueren, Germany) was used to harvest plasmids from overnight cultures. The manufacturer's protocol was followed and the plasmid DNA eluted in 50 µl of water.

2.6.5 Enzymatic restriction

Plasmids from different colonies were screened for their incorporation of the right PCR-product by selective digestion with *Eco*RI or *Bgl*II and *Not*I (NEB). The

reactions were prepared and then incubated at 37°C for one to two hours. To confirm the right size of the incorporated PCR-product the reactions were analyzed using an agarose gel.

Table 2.15: Reaction mixture for a restriction with *Eco*RI

Substance	Volume (µl)
H_2O	5,7
Buffer (NEB 3)	1
Plasmid	3
***Eco*RI**	0,3

Table 2.16: Reaction mixture for a restriction with *Bgl*II and *Not*I

Substance	Volume (µl)
H_2O	2,4
Buffer (NEB 3)	1
Plasmid	5
***Bgl*II**	0,3
***Not*I**	0,3
BSA (10x)	1

2.6.6 Cloning into the pESC-Leu Vector

The pESC-Leu vector is a vector specifically designed to express eukaryotic genes in *Saccharomyces cerevisiae*. It carries an auxotrophic selection marker gene for leucine enabling the selective screening for the incorporation of the vector. To clone the genes of interest into the vector, recognition sites for the restriction enzymes *Bgl*II and *Not*I (NEB) had to be attached to the genes. This was accomplished using a Phusion-PCR with primers encompassing the necessary recognition sites (Table 2.3). The PCR-products were then purified using an agarose gel, cloned into a TOPO Zero Blunt Vector and transformed into 10-beta cells. The transformed colonies were then inoculated in an overnight culture and their plasmids subsequently isolated using the Nucleo-Spin® Plasmid DNA Purification Kit (Invitrogen). At that point the plasmids were digested using *Bgl*II and *Not*I, the digested reactions separated on an agarose gel and the PCR-product extracted. Finally the PCR-product was ligated into the pre-restricted pESC-Leu vector using an overnight incubation at 16°C:

Table 2.17: Reaction mixture for a ligation reaction

Substance	Volume (µl)
T4 Ligase Puffer (5x)	2
T4 Ligase	1
DNA	2 (1 µg)
H_2O	15

The ligated plasmids were then transformed into 10-beta cells and overnight cultures inoculated from the newly formed colonies. The pESC-Leu plasmids were finally isolated using the Nucleo-Spin® Kit (Invitrogen).

2.7 Enzyme Assays

For the enzyme assays, reaction mixtures were prepared using the microsomes extracted from the pESC-Leu transformed yeast strains WAT11 and W(R) (Table 2.2). The amino acids used were L-valine, L-leucine, L-isoleucine, L-methionine, L-phenylalanine, L-tyrosine and L-tryptophan. The reaction mixtures were incubated at 25°C for 2 hours in a shaking incubator. Afterwards 300 µl of methanol were added to stop the reaction and the samples cooled on ice for 1 hour. The samples were finally centrifuged for 5 minutes at 7500 g, the supernatant collected and then analyzed with the LC-MS.

Table 2.18:Reaction mixture for enzyme assays with the CYP79 substrates

Substance	Volume (µl)
H_2O	22
Microsomes	20
Buffer (sodium-phosphat)	225
NADPH (100 mM)	3
Amino acid (10 mM)	30

2.8 Statistical analysis

The linkage mapping of the B73xCML69 RILs was done by Prof. Dr. Matthias Erb (University of Bernd).

2.9 Computer based analysis

Using the flanking markers PHM1184.26 and PHM2438.28, the Maize Genetics and Genomics Database Genome Browser was used to find the corresponding genes in between the markers. They were then manually examined for UGT, CYP71 and CYP79 genes.

2.10 Sequence alignment

Using the GRMZM2G138248 nucleotide sequence found in the NCBI databank, other plant UGTs characterized in literature as well as UGTs that are involved in cyanogenesis, a sequence alignment was performed utilizing MEGA5. The alignment was followed by a phylogenetic analysis using the Maximum Likelihood method. The analysis involved 36 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions with less than 80% site coverage were eliminated. There were a total of 1290 positions in the final dataset. Additionally a phylogenetic tree was drawn. The tree was inferred using the maximum likelihood method and n = 1000 replicates for bootstrapping. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site.

3 Results

3.1 Numerous maize lines produce Dhurrin and Lotaustralin

To obtain an overview over the distribution and concentration levels of five of the most abundant CNglcs (dhurrin, prunasin, linamarin, lotaustralin and amygdalin) in maize, seedlings of the 26 NAM parent lines were cultivated and methanol extracts prepared. Care was taken to separate roots and shoots, to allow quantification of different levels of CNglcs in these two tissues. While the cultivation of seedlings succeeded in general, no roots could be harvested from seedlings of the lines B97, CML52, NC350, NC358, OH43, and Tzi8. The collected plant material was then ground, the CNglcs extracted and finally analyzed using LC-MS. To be able to compare the concentrations of dhurrin with those found in Sorghum, extracts from Sorghum seedlings were also analyzed.

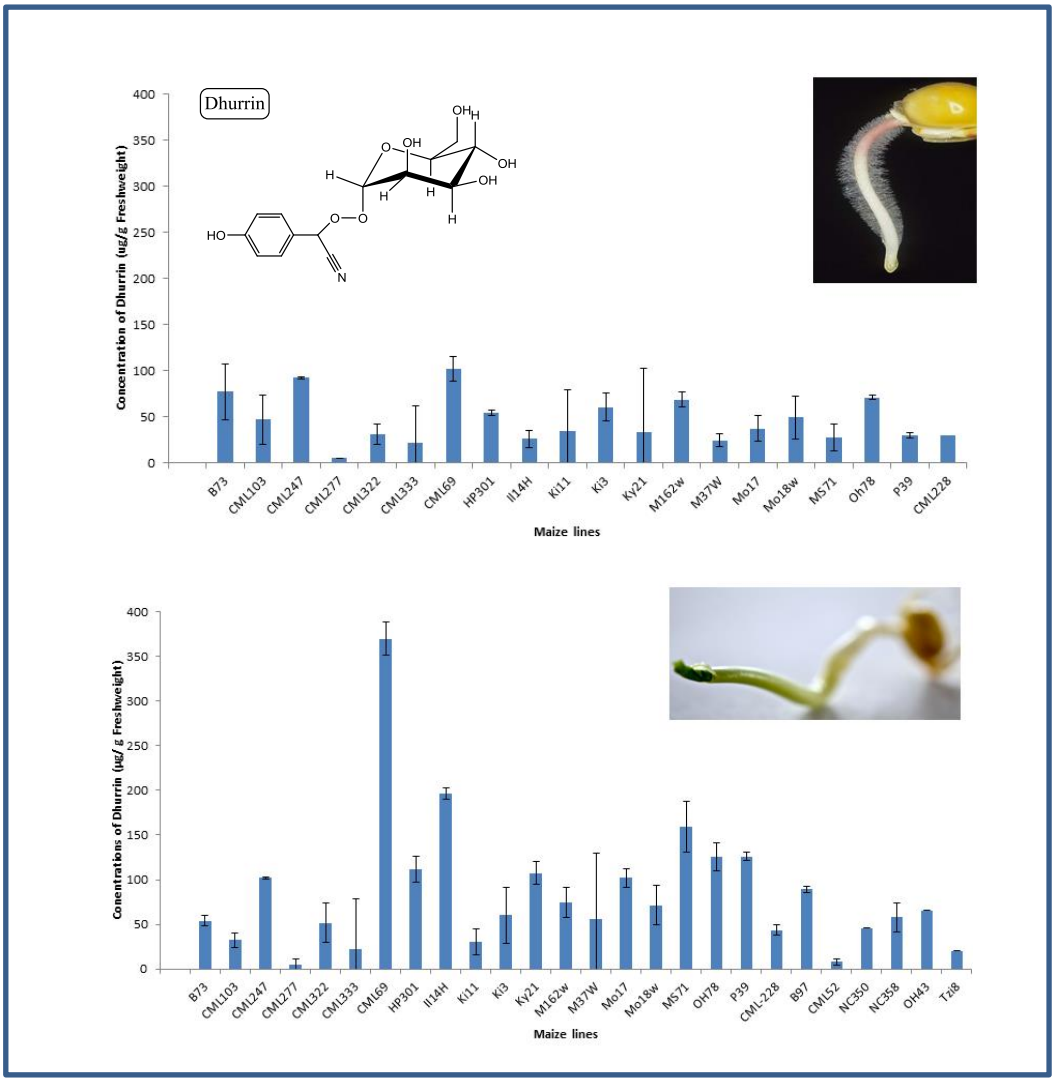

Figure 4: Concentrations of dhurrin in the roots and shoots. Quantification of the LC-MS data was done using a calibration line generated with a dhurrin standard.

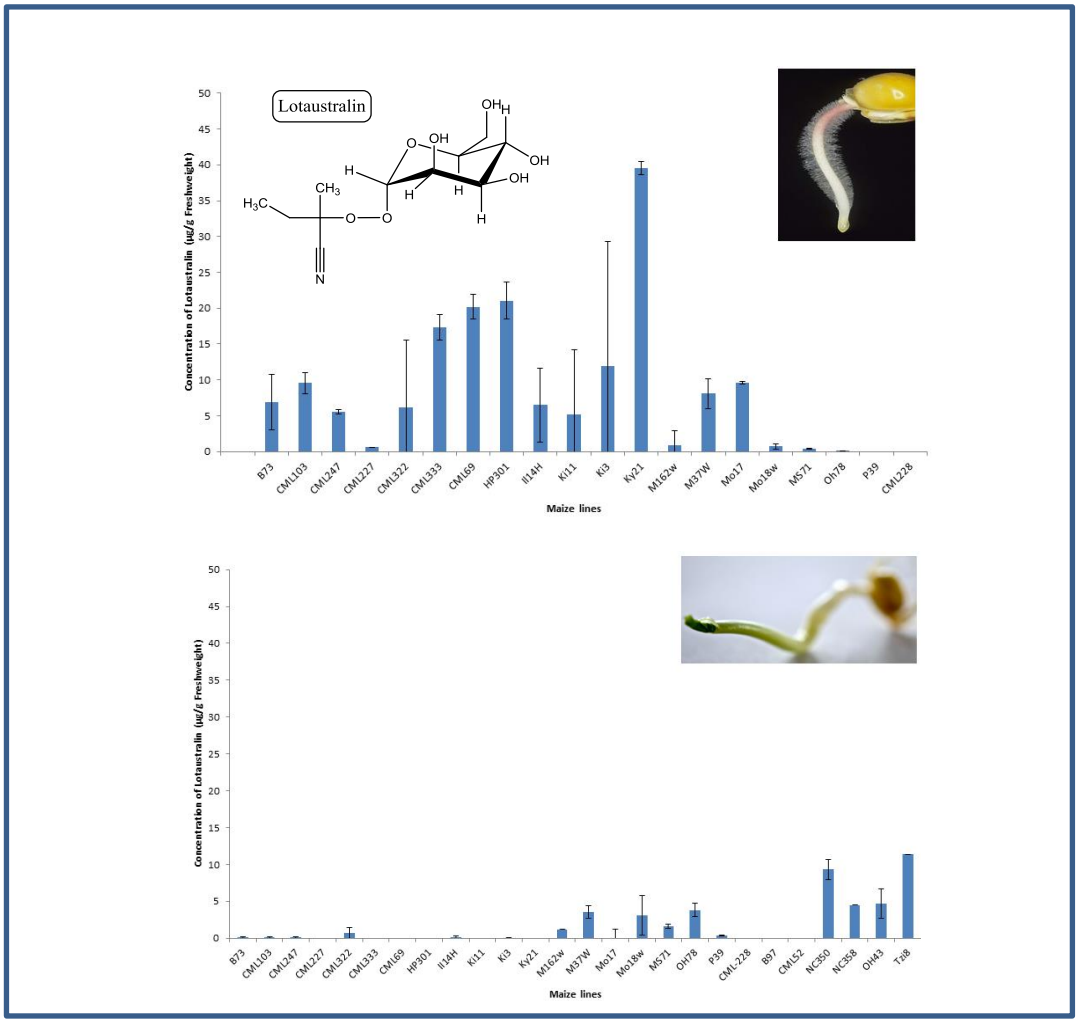

Figure 5: Concentrations of lotaustralin in the roots and shoots. Quantification of the LC-MS data was done using a calibration line generated with a dhurrin standard.

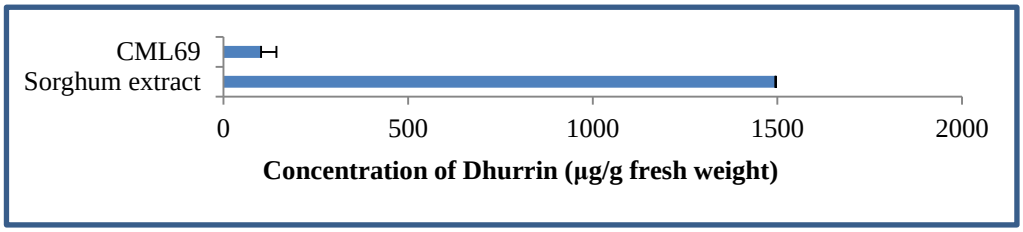

Figure 6: Comparison of the NAM line CML69 and Sorghum extract in terms of dhurrin concentration.

The results showed that all of the 26 examined maize lines produced at least small amounts of dhurrin (Figure 4 and 5), with the measured concentrations ranging from around 5 to 370 µg/g fresh weight. No clear difference could be detected regarding the concentration levels in the different tissues. In 15 maize lines, more dhurrin was detected in the shoots than in the roots, most notably CML69 with a difference of around 260 µg/g fresh weight. In 5 other lines the concentrations were higher in the roots than in the shoots. As the highest concentrations of dhurrin in both roots and shoots were found to exist in the line CML69, it was chosen for the nested association mapping. Furthermore a tenfold higher concentration of dhurrin was found in Sorghum comparing CML69 and Sorghum extracts (Figure 6).

While dhurrin was abundant in the NAM lines, no or only neglectable amounts of prunasin, linamarin or amygdalin could be detected. Lotaustralin (Figure 5) was found to be also widespread throughout the NAM population, but in up to tenfold smaller amounts than dhurrin. The highest concentration was detected in Ky21 in the roots with up to 40 µg/g fresh weight. There was also found to be a striking discrepancy in the distribution of lotaustralin regarding the two different tissues, as the shoots exhibited significantly lower levels of lotaustralin than the roots.

Statistical testing for significant differences in the concentrations of CNglcs between the lines proved unfeasible as the variances were too high for an ANOVA test and the sample pool too small for a Kruskal-Wallis or other test for samples with a high variance.

3.2 Nested associated mapping with B73xCML69

In search for genes corresponding to the higher concentrations of dhurrin in CML69 compared to B73, extracts of the B73xCML69 RILs were screened for the existence of dhurrin using LC-MS. The results were then statistically evaluated.

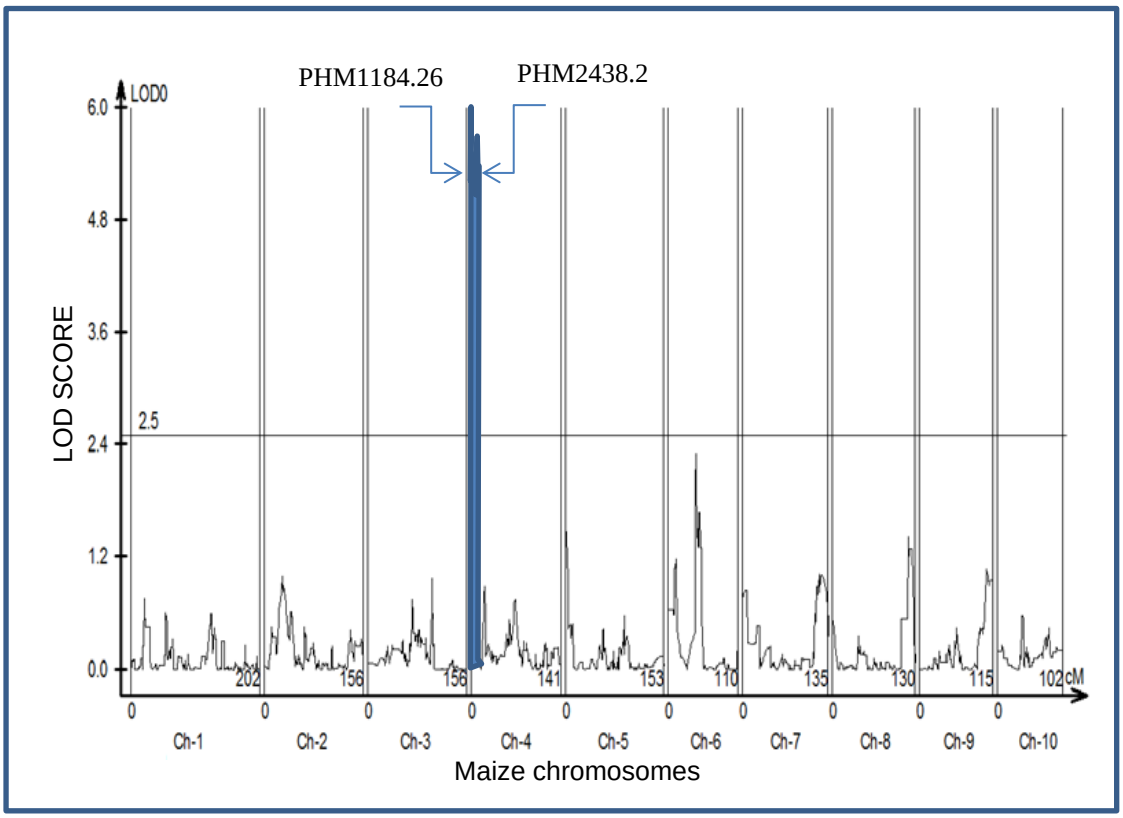

Figure 7: Nested associated mapping of the B73xCML69 RILs in search for QTLs correlating with the production of CNglcs. The LOD score calculates the probability of an actual correlation of a marker and a phenotype versus a random occurrence. Significant QTLs (LOD>2,5) in blue with surrounding markers annotated.

The nested associated mapping revealed a significant QTL in the forward part of chromosome four in the B73 genome (Figure 7). Using the surrounding markers PHM1184 and PHM2438, the Maize Genetics and Genomics Database Genome Browser was used to examine the genes in between the markers. The investigation of that genomic region showed no CYP71 or CYP79 genes, but revealed an UDP-glucosyltransferase (GRMZM2G085854) as well as multiple transcription factors and protein kinases. A phylogenetic tree was constructed, using the sequences for GRMZM2G085854, UGT51B1 from Sorghum and other UGTs reported to be involved in the biosynthesis of CNglcs, as well as unrelated UDP-glucosyltransferases (Figure 8).

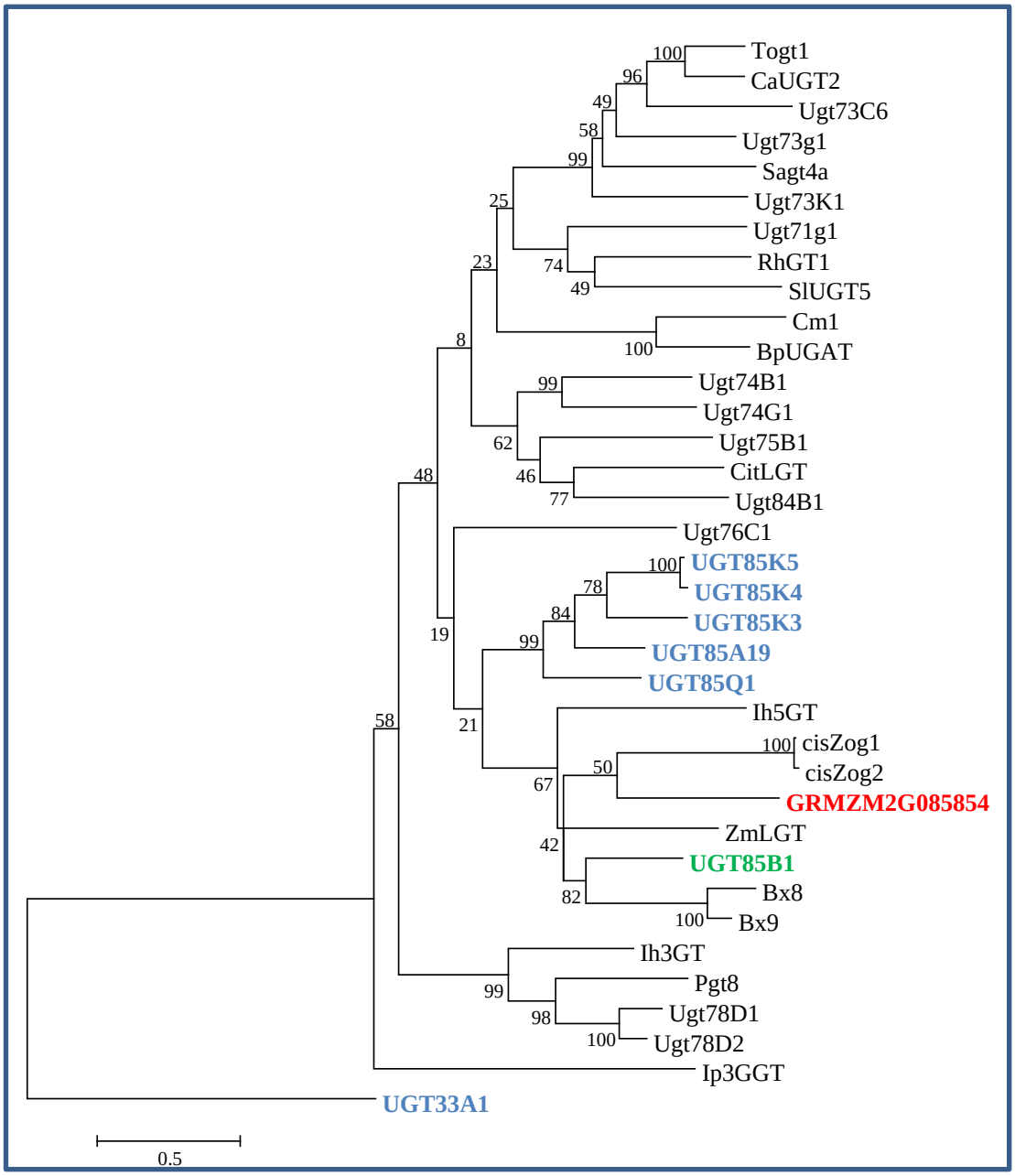

Figure 8: Sequence alignment and phylogenetic analysis of GRMZM2G085854 and other UGTs (Gleadow et al., 2014 and Bowles et al., 2005). The tree was inferred using the maximum likelihood method and n = 1000 replicates for bootstrapping. Bootstrap values are shown next to each node. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The glucosyltransferase involved in the biosynthesis of CNglcs in the insect species *Zygaena filipendulae* was used as an outgroup. Accession numbers for the cyanogenesis related UGTs are listed in the appendix. Accession numbers for the other UGTs are listed in Bowles et al., 2015. UGTs connected to CNglc biosynthesis are shown in blue, GRMZM2G085854 in red and UGT85B1 from the dhurrin biosynthesis pathway in Sorghum in green.

The phylogenetic analysis revealed that all of the plant UGTs that have been linked to cyanogenesis in literature are grouped into one branch. While these UGTs seem to be related, the bootstrap values inside the branch were not high enough to deduce a common lineage. As predicted, the outgroup UGT33A1 from the burnet moth (*Zygaena filipendulae*) showed no relation to the plant UGTs (Zagrobelny et al., 2008).

3.3 Cloning of maize CYP79s

Aiming to identify the CYP79 gene responsible for the production of dhurrin and lotaustralin in the NAM lines, multiple PCR based attempts were made to clone these genes, using cDNA generated from plant extracts as templates. The NAM lines Mo17 and Mo18w showed PCR products during preliminary testing while CML69, CML277 and Ky21 were included into the testing after the results of the screening for CNglcs were evaluated (Figure 4 and 5). CML69 was chosen for its high content of dhurrin, while CML277 was included for the low content. It was hoped that thereby different alleles of the same genes could be isolated and analyzed, to reveal different substrate specificities. To also cover the production of lotaustralin, Ky21 was chosen, as it held the highest concentration of this compound in the screening.

While PCR products appeared several times for all three primer pairs (Table 2.3) and all maize lines, cloning the products proved difficult. Many PCR products could not be inserted into the TOPO cloning vector or did not give rise to bacterial colonies after plating out the transformation mixtures. Furthermore, upon sequencing, some products that were successfully cloned were found to be made up of unrelated DNA. In the end only the CYP79A1 gene, already characterized in literature (Irmisch et al., 2015) was successfully cloned from CML69 and Mo17 cDNA.

3.4 Yeast expression and enzyme assays

The cloned alleles of CML69 and Mo17 were chosen to be expressed in both the WAT11 and the W(R) *S. cerevisiae* strains, purposely designed to carry cytochrome P450 reductase genes for the testing of P450 enzymes. The enzyme assays were aimed at revealing possible differences in the substrate specificities for the two

CYP79A1 alleles. The yeast strains were transformed, cultivated and microsomes containing overexpressed protein were extracted. Using these microsomes, different enzyme assays were prepared with possible substrate amino acids. Detection of the aldoxime products, generated by the enzymes, was then achieved with LC-MS.

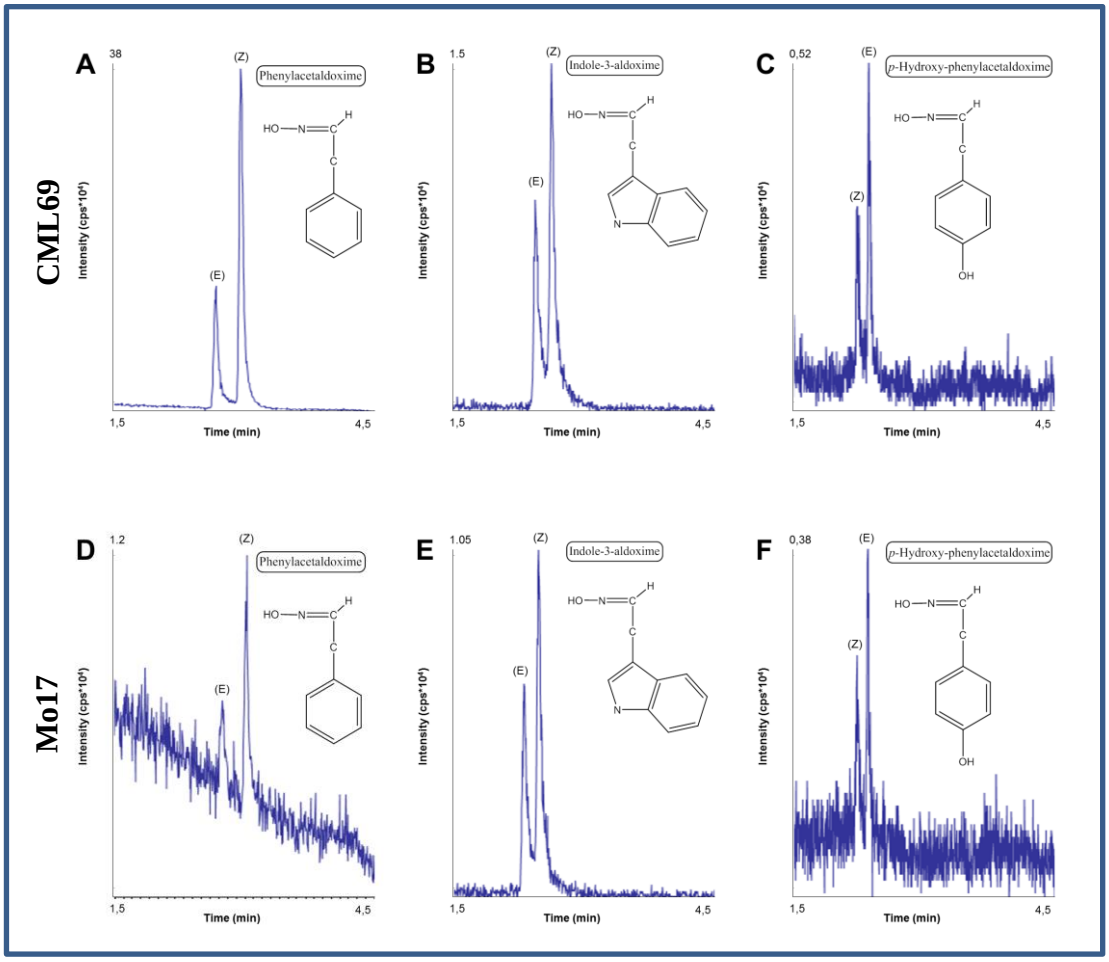

Figure 9: LC-MS analysis of enzyme assays using the microsomes harvested from the transformed W(R) cells and aromatic amino acids as substrates. Peak intensity was measured in counts per seconds (cps) over the retention time. The structures of the detected compounds are shown next to the diagrams. Diagrams [A], [B] and [C] correspond to the CML69 assays and [D], [E] and [F] to the Mo17 assays.

LC-MS analysis of enzyme assays revealed that the activities of the W(R)-expressed enzymes were significantly higher than the activities of enzymes expressed in WAT11 (data not shown). However, the relative activities for W(R)-expressed and WAT11-expressed enzymes were similar. Also for both alleles no products could be detected for non-aromatic parent amino acids.

Comparing the two alleles (Figure 9) it seemed that the CML69 allele produced more than thirty times more phenylacetaldoxime from phenylalanine than its Mo17 counterpart. The other aromatic amino acids, tryptophan and tyrosine, were also accepted as substrates although with significantly less product being catalyzed. The

intensities of indole-3-aldoxime were doubled in CML69 compared to Mo17. *p*-Hydroxyacetaldoxime, the intermediate in the biosynthesis of dhurrin from tryrosine, was found to be in similar ranges in the CML69 assay compared to the Mo17 assay. Overall, *p*-hydroxyacetaldoxime had the lowest intensities compared to the other products.

To confirm the difference in the production of phenylacetaldoxime between the two alleles, three technical replicates of the enzyme assays, using only aromatic amino acids as substrates, were performed (Figure 10).

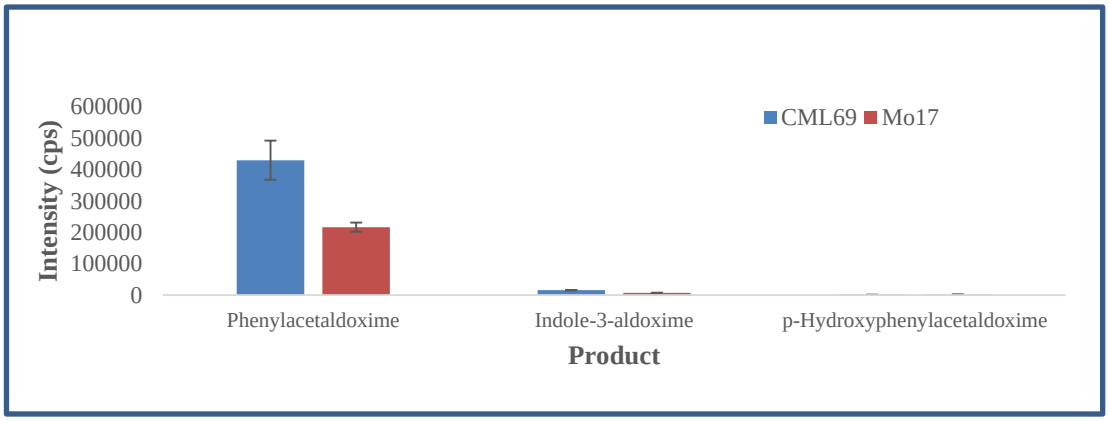

Figure 10: Results of the enzyme assay replicates. Intensity was measured in highest counts per second (cps). The mean intensity of the three replicates is shown.

While in the first enzyme assay (Figure 9) the production of phenylacetaldoxime was thirty times higher in CML69 than in the Mo17 allele, the three other replicates (Figure 10) only showed a twofold amount of phenylacetaldoxime for the CML69 allele compared to the Mo17 allele. The technical replicates of the enzyme assays thereby indicate that, while the absolute intensities between the two expressed alleles where higher in CML69 than in Mo17 for all of the substrates, the relative intensities for the different substrates were alike. The intensities for phenylacetaldoxime are in the magnitude of 10^5 for both alleles, for indole-3-aldoxime in the magnitude of 10^4 and for *p*-hydroxyphenylacetaldoxime at around 10^3. This suggests that more CML69 enzyme was expressed during the yeast cultivation, explaining the higher concentrations of product in the enzyme assays.

4 Discussion

4.1 The maize NAM lines produce dhurrin and might also contain lotaustralin

Screening of the NAM lines revealed that all of them contained dhurrin, many also showing small amounts of lotaustralin. As the NAM lines encompass a wide array of genetic variability, these results hint at maize being a cyanogenic plant in general, with the exception of the line Delprim and possibly more being non-cyanogenic. CML69 was found to be the line with the highest amount of dhurrin with concentrations of up to 370 µg/g fresh weight. This supported information in literature (Dunstan et al., 1902), stating the concentrations in maize to be significantly lower than in Sorghum (Figure 6). Using the CML69 x B73 RILs for a nested associated mapping, the difference in the concentrations of dhurrin between CML69 and B73 could be traced to a genetic region on chromosome four of the maize genome. This region contained multiple genes with transcriptional significance such as transcription factors and protein kinases as well as the glucosyltransferase GRMZM2G085854 which could potentially be the cause for the increased production of dhurrin in CML69. Phylogenetic analysis of different UGTs (Figure 8) grouped GRMZM2G085854 in close range to other glucosyltransferases related to cyanogenesis in plants, but bootstrap values were not high enough to deduce a common lineage. This corresponds to information in literature describing multiple independent neofunctionalizations of UGTs, resulting in different lineages, as the most likely scenario for the recruitment of plant UGTs into the CNglc biosynthesis pathway (Gleadow et al., 2014).

While the screening for CNglcs showed small amounts of lotaustralin (Figure 5), the significance of these findings remains unclear. Due to the lack of authentic standards for the other CNglcs, the dhurrin calibration line was used for all CNglcs, potentially distorting the results. Additionally, literature (Gleadow et al., 2014) reports the co-occurrence of lotaustralin and linamarin due to their structural similarities, whereas in this screening only lotaustralin could be detected. In light of these facts the screening results for lotaustralin cannot safely be considered accurate.

4.2 GRMZM2G138248 might be the CYP79 enzyme involved in the biosynthesis of dhurrin in maize

Enzyme assays with CYP79A1 (Figure 9) and different amino acid substrates revealed that the gene products accepted tyrosine, the parent amino acid for the production of dhurrin, as well as the other aromatic amino acids as substrates. Furthermore, it is the only CYP79 enzyme that could be cloned from seedlings of the dhurrin-producing NAM lines. Additionally, it shares 72 % amino acid identity with CYP79A1 (Irmisch et al., 2015), the enzyme responsible for the production of dhurrin in Sorghum. While these findings suggest that GRMZM2G138248 plays a role in the production of dhurrin in maize, other conclusions can be drawn. The enzyme assays also showed that the production of phenylacetaldoxime was more than 100 times higher than the production of *p*-hydroxyacetaldoxime, suggesting that phenylalanine is the primary substrate for the enzyme. While these results are adverse to the conclusion that GRMZM2G138248 is the gene accountable for the biosynthesis of dhurrin, it cannot be excluded as these findings could be mitigated by locally high concentrations of the substrates or a low K_M for tyrosine *in vivo*.

4.3 Three other CYP79 enzymes are not expressed in the NAM line seedlings

The three other CYP79 enzymes could not successfully be cloned from cDNAs which were obtained from different NAM line seedlings. This is likely because they are not expressed in those seedlings. But with the description of the production and turnover of amygdalin and prunasin in bitter almonds (Sánchez-Pérez et al., 2008), an example for how these genes could still be responsible for the production of dhurrin is available. In the bitter almond, prunasin and amygdalin are stored in the seedlings and, upon germination, turned over, acting presumably as a storage compound in addition to their role in plant defense. Similarly, dhurrin or the corresponding diglycoside proteacin could be produced by one or more of the three CYP79s in the maize seeds to act as a defense measure for the developing seedlings. Like bitter almonds, developing maize seedlings have been shown to have diminishing cyanogenic potentials (Erb et al., 1981a) over the course of their development, indicating a similar storage function as CNglcs are presumed to have in other plants. Furthermore, as Irmisch et al. (2015) report, no traces of CNglcs could

be found in the CYP79A1-expressing maize line Delprim. This further strengthens the hypothesis that dhurrin production in maize takes places before the germination and is mediated by one of the three other CYP79s.

4.4 Future prospects

To identify the CYP79 responsible for the production of dhurrin in the NAM lines, further studies seem to be necessary. Using additional enzyme assays, the K_M of CYP79A1 for tyrosine could be studied, and it could be determined if the low production of *p*-hydroxyphenylacetaldoxime found in the enzyme assays carried out in this study could be compensated for, by a high enzyme specificity towards tyrosine. Furthermore, by using qRT-PCR to monitor the expression levels of the four CYP79s in different developmental stages of maize seeds and developing plants, evidence could be given for the time point when dhurrin is produced and if it is indeed stored in the seeds as a preformed defense and nitrogen storage compound. A linkage mapping with CML69 and a maize line without CNglcs, such as Delprim, could also be effective in revealing the CYP79 mediating the biosynthesis of dhurrin, but such a mapping would be expensive and time consuming as only the NAM RILS are readily available and already genotyped. Finally, using primers specific for the glycosyltransferase found in the nested associated mapping, cDNA clones of GRMZM2G085854 could be recovered and characterized for their potential activity. These clones could be subjected to enzyme assays containing *p*-hydroxymandelonitrile as substrate to determine if the enzymes are able to transfer UDP-glucose onto the nitrile and thus might be responsible for the increased levels of dhurrin found in CML69 compared to B73.

5 Summary

5.1 English

In the present study, evidence for the occurrence of the CNglc dhurrin in maize is given and the possible CYP79 candidate genes, coding for the first enzyme of the biosynthesis pathway, are explored. Liquid Chromatography-Mass Spectrometry was used to determine the concentration levels of dhurrin in the NAM lines and B73xCML69 RILs. The LC-MS data from the RILs was used for a nested associated mapping and with the UDP-glucosyltransferase GRMZM2G085854, a candidate gene for the increased concentration of CML69 compared to the reference line B73 was found. Additionally, in several cloning experiments two different alleles for CYP79A1 one of the four CYP79 genes found in maize could be amplified. These alleles were subsequently transformed in a yeast expression vector, expressed and the gene products purified. The purified enzymes were then tested for their amino acid specificities. The enzyme assays showed that CYP79A1 accepted phenylalanine instead of tyrosine, the parent amino acid for dhurrin, as its primary substrate. Furthermore no altered specificities between the two alleles could be detected in the assays. Therefore it remains unclear which CYP79-enzyme is involved in the biosynthesis of Dhurrin in maize.

5.2 Deutsch

In der vorliegenden Arbeit wurde das Vorkommen des cyanogenen Glykosides Dhurrin in Mais nachgewiesen und vier verschiedene CYP79 Gene, die möglicherweise das erste Enzym des Biosynthesewegs codieren, untersucht. Mit Hilfe der Flüssigchromatographie-Massenspektrometrie wurden die Konzentrationen an Dhurrin in sämtlichen NAM Linien und den B73xCML69 RILs untersucht. Die Ergebnisse der Konzentrationsbestimmung für die B73xCML69 RILs wurden anschließend verwendet, um ein Nested Associated Mapping durchzuführen. Mit der UDP-Glycosyltransferase GRMZM2G085854, wurde ein Kandidatengen für die unterschiedlichen Konzentrationen an Dhurrin in CML69 verglichen mit B73 gefunden. In zusätzlichen Klonierungsexperimenten konnten zwei verschiedene Allele von CYP79A1, eines der vier CYP79-Enzyme die in Mais vorkommen,

amplifiziert werden. Diese Allele wurden anschließend in einen Hefeexpressionsvektor kloniert, exprimiert und die Genprodukte aufgereinigt. Die aufgereinigten Proteine wurden dann auf ihre Spezifitäten gegenüber verschiedenen Aminosäuren getestet. Die Tests zeigten, dass CYP79A1 primär Phenylalanin und nicht Tyrosin, die Aminosäure aus der Dhurrin synthetisiert wird, als Substrat akzeptiert. Außerdem konnten in diesen Testreihen keinerlei Unterschiede in den Substratspezifitäten zwischen den beiden Allelen gezeigt werden. Es ist daher weiterhin unklar welches der vier CYP79-Enzyme in die Biosynthese von Dhurrin involviert ist.

6 Bibliography

Bak, S., Kahn, R.A., Nielsen, H.L., Møller, B.L., Halkier, B.A., 1998. Cloning of three A type cytochromes P450, CYP71E1, CYP98, and CYP99 from Sorghum bicolor (L.) Moench by a PCR approach and identification by expression in Escherichia coli of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant Mol. Biol. 36, 393–405.

Bak S., Paquette S.M., Morant M., Rasmussen A.V., Saito S., Bjarnholt N., Zagrobelny M., Jørgensen K., Hamann T., Osmani S., Simonsen H.T., Pérez R.S., van Hesswijck T.B., Jørgensen B., Møller B.L., 2006. Cyanogenic glycosides; a case study for evolution and application of cytochromes P450. Phytochem. Rev. 5, 309–329.

Bowles D., Isayenkova J., Lim E., Poppenberger B., 2005. Glycosyltransferases: managers of small molecules. Current Opinion in Plant Biology vol. 8 no. 3 254–263.

Brunnich JC., 1903. Hydrocyanic acid in fodder-plants. J Chem Soc. 83:788–96.

Busk P.K., Møller B.L., 2002. Dhurrin synthesis in Sorghum is regulated at the transcriptional level and induced by nitrogen fertilization in older plants. Plant Physiol. 129, 1222–1231.

Coe E. H. Jr., 1966. The properties, origin, and mechanism of conversion-type inheritance at the B locus in maize. Genetics, vol. 53, no. 6, pp. 1035–1063.

Deepak G., Shiv K., Asha D., Kumari A., 2010. Advances in cyanogenic glycosides biosynthesis and analyses in plants: A review. Acta Biologica Szegediensis, 1-14.

Erb N., Zinsmeister H., Lehmann G., Michely D., 1981a. Der Blausäuregehalt von Getreidearten gemässigter Klimazonen. Z Lebensm Unters Forsch.,173(3):176–9.

Erb N., Zinsmeister H., Nahrstedt, A., 1981b. Die cyanogenen Glykoside von Triticum, Secale und Sorghum. Planta Medica 01, 84-89.

Flint-Garcia S., Thuillet A., Yu J., Pressoir G., Romero S., Mitchell S., Doebley J., Kresovich S., Goodman M., Buckler E., 2005. Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant Journal 2005, 1054-1064.

Gierl A., 2005. The Cytochrome P450 Superfamily of Monooxygenases. Handbook of Maize, 731-739.

Gleadow R.M., Møller B.L., 2014. Cyanogenic Glucosides: Synthesis, Physiology, and Plant Plasticity. Annual Review of Plant Biology 65(1).

Griffiths AJF, Gelbart WM, Miller JH, et al., 1999. Modern Genetic Analysis.

Irmisch S., **Zeltner P.**, **Handrick V.**, **Gershenzon J.**, **Köllner T.G., 2015**. The maize cytochrome P450 CYP79A61 produces phenylacetaldoxime and indole-3-acetaldoxime in heterologous systems and might contribute to plant defense and auxin formation. BMC Plant Biology 2015, 15:128.

Jenrich, R., Trompetter, I., Bak, S., Olsen, C.E., Møller, B.L., Piotrowski, M., 2007. Evolution of heteromeric nitrilase complexes in Poaceae with new functions in nitrile metabolism. Proc. Natl. Acad. Sci. USA 104, 18848–18853.

Kahn, R.A., Fahrendorf, T., Halkier, B.A., Møller, B.L., 1999. Substrate specificity of the cytochrome P450 enzymes CYP79A1 and CYP71E1 involved in the biosynthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor (L.) Moench. Arch. Biochem. Biophys. 363, 9–18.

Kongsawadworakul P., Viboonjun U., Romruensukharom P., Chantuma P., Ruderman S., Chrestin H., 2009: The leaf, inner bark and latex cyanide potential of Hevea brasiliensis: evidence for involvement of cyanogenic glucosides in rubber yield. Phytochemistry, 70:730-739.

Kristensen C., Morant M., Olsen CE., Ekstrom CT., Gal- braith DW., Møller BL., Bak S., 2005. Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. Proc Natl Acad Sci USA 102:1779–1784.

Lehmann G., Zinsmeister HD, Erb N., Neunhoeffer O., 1979. Content of hydrocyanic acid in corn and cereal products. Z Ernahrungswiss.,18(1):16–22.

McClintock B., 1950. The origin and behavior of mutable loci in maize.
Proceedings of the National Academy of Sciences of the United States of America, vol. 36, no. 6, pp. 344–355.

McFarlane IJ., Lees EM., Conn EE., 1975. The in vitro biosynthesis of dhurrin, the cyanogenic glycoside of Sorghum bicolor. J. Biol. Chem. 250:4708–13.

McMullen MD. et al., 2009. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Science 325: 737-740.

Møller BL., Conn EE., 1979. The biosynthesis of cyanogenic glucosides in higher plants: N-hydroxytyrosine as an intermediate in the biosynthesis of dhurrin by Sorghum bicolor (L.) Moench. J. Biol. Chem. 254:8575–83.

Møller BL., Conn EE., 1980. The biosynthesis of cyanogenic glucosides in higher plants. Channeling of intermediates in dhurrin biosynthesis by a microsomal system from Sorghum bicolor (linn) Moench. J Biol Chem.

Møller BL., 2010. Functional diversifications of cyanogenic glucosides.Current Opinion in Plant Biology. Current Opinion in Plant Biology vol. 13 no. 3 338–347.

Pompon D., Louerat B., Bronine A., Urban P., 1996. Yeast expression of animal and plant P450s in optimized redox environments. Methods Enzymol. 272:51–64.

Ross J., Li Y., Lim E. and Bowles D.J., 2001. Higher plant glycosyltransferases. Genome Biology. 2(2):reviews 3004.1–3004.6.

Sánchez-Pérez R., Jørgensen K., Olsen C.E., Dicenta F., Møller B.L., 2008. Bitterness in almonds. Plant Physiol. 146, 1040–1052.

Sanger F., Nicklen S. and Coulson A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America, 74, 5463-5467.

Selmar D., Lieberei R., Biehl B., 1988: Mobilization and utilization of cyanogenic glycosides. Plant Physiol. 86:711-716.

Shimada M., Conn EE., 1977. The enzymatic conversion of p-hydroxyphenylacetaldoxime to p-hydroxymandelonitrile. Arch. Biochem. Biophys. 180:199–207.

Thorsøe K., Bak S., Olsen C, Imberty A, Breton C., Møller B., 2005. Determination of Catalytic Key Amino Acids and UDP Sugar Donor Specificity of the Cyanohydrin Glycosyltransferase UGT85B1 from *Sorghum bicolor*. Molecular Modeling Substantiated by Site-Specific Mutagenesis and Biochemical Analyses. Plant Physiology vol. 139 no. 2 664-673.

Werck-Reichhart D., Bak S., Paquette S., 2002. Cytochromes P450. Arabidopsis Book. 1: e0028.

Yu J., Holland J. B., McMullen M. D., Buckler E. S., 2008. Genetic design and statistical power of nested association mapping in Maize. Genetics 178, 539.

Zagrobelny M., Bak S., Moller BL., 2008. Cyanogenesis in plants and arthropods. Phytochemistry. 69:1457-1468.

7 Appendix

7.1 cDNA sequences

Table 7.1: cDNA sequence of the transformed and expressed CML69 allele

ATGGTTTCCTCTCCGCAAGCAAATAAATTTCCTCTCCGCAAGCAAATTTCCTCGCCCTCCTCTTCGTGCATGACTC
CTCATGTGCTATTAGTCCTAGTGATTCTCCTGTACCTCGTCAGAACCCTCAGGCCGTGGCGGAGCCGGAGGAACA
GCACATGCTGCACCCCTTCGCTCCCGCCGGGGCCTGTGCCATGGCCCGTCGTTGGCAACCTGCCGGAGATGATGC
TGAGCGACAAGCCGGCGTTCCACTGGATCCACCATATCATGAAGGAGGCGGGAACCGACATCGCCTGCATCAAA
CTGGGCGGTGTCCACGTCATCCCCATCAGCTGCCCCAAGATCGCGCTCGAGGTGCTAAGCAACCAGGATGCCAAC
TTCGCTTCTCGCCCGCTCACCTTCGCCTCCAAGACCTTCAGCAGAGGCTACAGGGACGCTGCAATGTCCCACTGC
GGCGACCACTGGAAGAAGATGCGCCGCGTGCTGGCCTCCGATATCGTCTGCCCCTCCCGTCACAGGTGGCTCCAC
GACAAGCGTGCCGACGAGGCAGACAACCTCACCCGGTACGTCTACAACCTCGCGACGAAGGCGGGGTCGTCCGG
CGGCGCCGTTGACGTTAGGCATATTGCGCGCCACTACTGCGGGAATGTCGTCCGCCGCCTCATGTTCAACACGCG
CTACTTTGGCAAGCCACAGCCTGACAGTGGGCCTGGTCCACTGGAGGTACAGCATGTGGACGCCGTCTTTACCTC
CCTCGGCCTTCTCTACTCCTTCTGTGTcTCCGACTATCTcCCGTGGCTGCTGGGCCTCGACCTAGACGGCCACGAGA
AGATGGtCAAGGAGGCCaATGAGAGGGTGACCAGGCTGCACGACGCGGTCATCGACGAGCGGTGGAGGCTGTGG
AAGAGCGGCGAGAGGCGGGAGCTCGAGGACTTCCTAGACGTGCTCATCATGCTCAAGgACGCCGAGGCCGGCAG
ACCGGTTCTCAGCATCGAGGAGGTCAAAGCTCTGCTAATGGATATCACGTTCGCGTCCATGGaCAACCCGTCGAA
TGCAGTGGAGTGGGCTCTGGCCGAGATGGTGAACAACCCGGAGATGCTTAAGAAGGCAGTGGACGAGATCGACA
GCGTCGTTGGCCAGGAACGGCTAGTGCAGGAGCACGACATCCCACGGCTAAACTACACCAAGGCCTGCATCCGG
AGGCGTTCCGCCTGCACCCTGTAGCGCCTTTCAACGTTCCCCACGTCGCTGTCGCGGACGCCACCGTTGCCGGCT
ACCGCATCCCCAAGGGCAGCCACGTCATCCTCAGTCGCACTGGCCTTGGACGCAACCCTGATGTCTGGGACGACC
CTCTTCGCTTCGACCCCGAACGCCACATCCCTTTGAGCCCCGAGATGGAGGTCTCACTCGTTGAGCGTGATCTAC
GGTTTATCTCCTTCAGCACTGGTCGTCGCGGCTGCATCGCTGCAGCTCTTGGCACTGATATGACCATCATGCTCTT
CGGTAGGCTCCTGCAGGGGTTCTCTTGGAGCAAGATGACCGGGGTGGCGGCCATCGACCTTAGCGAGTCCAGGC
ACAACACTTTCATGGCACGGCCACTCGTGCTACAGGCTAAGCCACGGCTTCCAGCGCACCTCTACCCCGGGATCT
TGCTTTGTAAGTAG

Table 7.2: cDNA sequence of the transformed and expressed Mo17 allele

ATGGTTTCCTCTCCGCAAGCAAATAAATTTCCTCTCCGCAAGCAAATTTCCTCGCCCTCCTCTTCGTGCATGGCTC
CTCATGTGCTATTAGTCCTAGTGATTCTCCTGTACCTCGTCAGAACCCTCAGGCCGTGGCGGAGCCGGAGGAACA
GCACATGCTGCACCCCTTCGCTCCCGCCGGGGCCTGTGCCATGGCCCGTCGTTGGCAACCTGCCGGAGATGATGC
TGAGCGACAAGCCGGCGTTCCACTGGATCCACCATATCATGAAGGAGGCGGGAACCGACATCGCCTGCATCAAA
CTGGGCGGTGTCCACGTCATCCCCATCAGCTGCCCCAAGATCGCGCTCGAGGTGCTAAGCAACCAGGATGCCAAC
TTCGCTTCTCGCCCGCTCACCTTCGCCTCCAAGACCTTCAGCAGAGGCTACAGGGACGCTGCAATGTCCCACTGC
GGCGACCACTGGAAGAAGATGCGCCGCGTGCTGGCCTCCGAGATCGTCTGCCCCTCCCGTCACAGGTGGCTCCAC
GACAAGCGTGCCGACGAGGCAGACAACCTCACCCGGTACGTCTACAACCTCGCGACGAAGGCGGGGTCGTCCGG
CGCCGTTGACGTTAGGCATATTGCGCGCCACTACTGCGGGAATGTCGTCCGCCGCCTCATGTTCAACACGCGCTA
CTTTGGCAAGCCACAGCCTGACAGTGGGCCTGGTCCACTGGAGGTACAGCATGTGGACGCCGTCTTTACCTCCCT
CGGCCTTCTCTACTCCTTCTGTGTCTcCGACTATCTCCCGTGGCTGCTGGGCCTCGACCTAGACgGCCACGAGAAG
ATGGTCAAGGAGGCcAATGAGAGGGTGACCAGGCTGCACGACGCGGTcATCGACGAGCGGTGGAGGCTGTGGAA
GAGCGGCgAGaGGCGGGaGCTCGAGGACTTCCTAGACGTGCTCATCATGctCAAGgacGCCGAGGCCGGCAGACCGG
TGCTCAGCATCGAGGAGGTCAAAGCTCTGCTAATGGATATCACGTTCGCGTCCATGGACAACCCGTCGAATGCAG
TGGAGTGGGCTCTGGCCGAGATGGTGAACAACCCGACGATGCTTAAGAAGGCAGTGGACGAGATCGACAGCGTC
GTTGGCCAGGAACGGCTAGTGCAGGAGCACGACATCCCACGGCTAAACTACACCAAGGCCTGCATCCGCGAGGC
GTTCCGCCTGCACCCTGTAGCGCCTTTCAACGTTCCCCACGTCGCTGTCGCGGACGCCACTGTTGCCGGCTACCGC
ATCCCCAAGGGCAGCCACGTCATCCTCAGTCGCACTGGCCTTGGACGCAACCCTGATGTCTGGGACGACCCTCTT
CGCTTCGACCCCGAACGCCACATCCCTTTGAGCCCCGAGATGGAGGTCTCACTCGTTGAGCGTGATCTACGGTTT
ATCTCCTTCAGCACTGGTCGTCGCGGCTGCATCGCTGCAGCTCTTGGCACTGATATGACCATGATGCTCTTCGCTA
GGCTCCTGCAGGGGTTCTCTTGGAGCAAGATGACCGGGGTGGCGGCCATCGACCTTAGCGAGTCCAGGCACAAA
CTTTCATGGCACGGCCACTCGTGCTACAGGCTAAGCCACGGCTTCCAGCGCACCTCTACCCCGGGATCTTGCTTTG
TAAGTAG

7.2 LC-MS data

7.2.1 NAM lines screening

Table 7.3: LC-MS data of analyzed root samples.

Sample Names	Analyte Peak Area (counts)				
	Dhurrin	**Prunasin**	**Linamarin**	**Lotaustralin**	**Amygdalin**
B73-1	1.19E+05	0.00E+00	1.85E+03	1.49E+04	2.31E+02
B73-2	4.44E+05	0.00E+00	1.62E+03	2.42E+04	3.08E+01
B73-3	5.93E+04	0.00E+00	2.00E+02	1.75E+04	1.38E+02
CML103-1	4.46E+04	0.00E+00	5.39E+02	8.07E+03	7.70E+01
CML103-2	2.89E+05	0.00E+00	2.32E+03	4.39E+04	7.15E+02
CML103-3	4.55E+04	0.00E+00	9.22E+01	2.60E+04	2.77E+02
CML247-1	1.77E+05	0.00E+00	4.62E+02	1.13E+04	9.22E+01
CML247-3	3.19E+05	0.00E+00	6.16E+02	1.92E+04	2.31E+01
CML227-1	9.03E+03	0.00E+00	3.08E+02	2.37E+03	9.24E+01
CML227-2	1.90E+04	0.00E+00	5.54E+02	2.95E+03	9.23E+01
CML227-3	1.66E+04	0.00E+00	1.08E+02	5.38E+02	7.70E+01
CML322-1	8.39E+04	-5.55E+02	3.23E+02	1.68E+04	1.23E+02
CML333-1	7.81E+04	0.00E+00	9.69E+02	5.26E+04	1.54E+01
CML333-2	9.57E+04	0.00E+00	5.40E+03	8.71E+04	1.23E+02
CML333-3	4.16E+01	1.84E+02	7.70E+01	3.61E+02	9.22E+01
CML69-1	4.86E+05	0.00E+00	2.15E+02	6.40E+04	1.54E+01
CML69-2	2.11E+05	0.00E+00	2.79E+04	5.07E+04	1.85E+02
CML69-3	1.28E+05	0.00E+00	3.38E+02	4.83E+04	2.31E+02
HP301-1	1.23E+05	0.00E+00	1.15E+04	4.75E+04	6.15E+01
HP301-2	9.95E+04	1.55E+03	2.81E+04	6.09E+04	5.39E+01
HP301-3	2.17E+05	4.28E+03	2.53E+04	6.18E+04	9.23E+01
IL14H-1	1.89E+05	0.00E+00	2.18E+03	1.22E+04	1.38E+02
Dhurrin std 0.005 mg mL	8.02E+04	9.23E+01	9.23E+01	6.16E+01	4.62E+01
Dhurrin std 0.01 mg mL	1.26E+05	1.54E+01	9.23E+01	5.85E+02	4.62E+01
Dhurrin std 0.02 mg mL	8.37E+04	7.68E+01	1.54E+01	7.70E+01	1.38E+02
Dhurrin std 0.04 mg mL	1.91E+05	1.54E+01	9.23E+01	0.00E+00	4.62E+01
Dhurrin std 0.05 mg mL	1.69E+05	1.54E+01	1.54E+01	4.85E+02	4.61E+01
5HIAA std 0.005 mg mL	4.02E+01	9.23E+01	1.54E+01	3.84E+02	0.00E+00
5HIAA std 0.01 mg mL	8.74E+01	9.22E+01	9.23E+01	2.46E+02	4.62E+01
5HIAA std 0.02 mg mL	3.07E+01	2.77E+02	0.00E+00	5.38E+02	9.23E+01
5HIAA std 0.04 mg mL	1.38E+02	1.54E+01	1.54E+01	1.46E+02	4.62E+01
5HIAA std 0.05 mg mL	1.38E+02	8.47E+01	9.23E+01	2.31E+02	0.00E+00
IL14H-2	2.09E+04	0.00E+00	1.23E+02	1.22E+04	1.85E+02
IL14H-3	2.19E+03	0.00E+00	6.53E+02	2.89E+04	2.31E+02
Ki11-2	1.28E+05	0.00E+00	6.46E+02	4.18E+04	6.92E+01
Ki3-1	4.29E+04	0.00E+00	6.15E+02	8.02E+03	1.54E+02
Ki3-2	2.82E+05	0.00E+00	9.69E+02	5.63E+04	2.15E+03
Ky21-1	1.20E+05	0.00E+00	4.77E+02	1.33E+05	1.23E+02
Ky21-2	1.06E+04	0.00E+00	1.75E+03	1.50E+04	9.23E+01
Ky21-3	1.43E+05	0.00E+00	2.77E+03	1.70E+05	1.39E+02
M162w-1	6.37E+01	9.23E+01	1.54E+01	3.83E+02	4.62E+01
M162w-3	3.70E+05	0.00E+00	1.13E+04	5.31E+03	2.31E+01
M37w-1	1.01E+05	0.00E+00	1.00E+02	1.99E+04	1.08E+02
M37w-2	7.30E+04	0.00E+00	6.31E+02	1.42E+04	9.23E+01
M37w-3	2.61E+04	0.00E+00	9.24E+01	3.18E+04	9.25E+01
Mo17-1	7.72E+04	0.00E+00	3.15E+03	3.73E+04	8.47E+01
Mo17-2	1.37E+05	0.00E+00	7.15E+02	2.05E+04	9.23E+01
Mo17-3	8.82E+04	0.00E+00	2.97E+03	2.01E+04	1.69E+02
Mo18w-1	6.21E+04	1.66E+03	3.38E+02	1.52E+03	9.23E+01
Mo18w-2	1.88E+05	5.36E+03	3.85E+02	2.08E+03	1.00E+02
Mo18w-3	1.48E+05	3.14E+03	3.54E+02	3.29E+03	2.23E+02
MS71-1	1.99E+05	0.00E+00	8.31E+02	3.23E+03	5.69E+02
MS71-2	2.36E+04	0.00E+00	3.07E+01	1.35E+03	1.54E+01
MS71-3	4.62E+02	1.80E+03	1.54E+01	6.83E+01	9.22E+01
Oh78-1	2.69E+05	0.00E+00	5.58E+03	5.41E+02	1.15E+02
Oh78-2	1.36E+05	2.56E+03	5.93E+02	6.00E+02	6.92E+01
Oh78-3	1.67E+05	3.46E+03	2.61E+02	4.61E+01	9.23E+01
P39-1	7.46E+04	0.00E+00	2.00E+02	2.69E+02	1.23E+02
P39-2	8.71E+04	0.00E+00	2.23E+02	3.08E+01	2.77E+02
CML228-1	2.84E+04	0.00E+00	3.07E+01	9.14E+01	6.92E+02

CML228-2	8.49E+03	0.00E+00	6.23E+02	4.46E+02	4.85E+02
CML228-3	9.69E+02	6.93E+01	2.92E+02	6.93E+01	2.00E+02
Sorghum Std.	6.45E+05	0.00E+00	2.08E+02	5.23E+02	1.38E+02

Table 7.4: LC-MS data of analyzed shoot samples

Sample Names	Analyte Peak Area (counts)				
	Dhurrin	**Prunasin**	**Linamarin**	**Lotaustralin**	**Amygdalin**
B73-1	1.54E+05	0.00E+00	6.46E+02	1.23E+02	1.54E+01
B73-2	1.76E+05	0.00E+00	4.91E+04	7.38E+02	1.54E+01
B73-3	1.05E+05	0.00E+00	3.18E+04	1.19E+03	1.54E+01
B97-1	2.09E+05	0.00E+00	5.42E+04	1.77E+02	1.54E+01
B97-1	2.66E+05	0.00E+00	1.36E+05	2.62E+02	1.85E+02
B97-1	2.41E+05	0.00E+00	1.30E+05	1.23E+02	4.15E+02
CML103-1	5.80E+04	0.00E+00	3.63E+04	4.02E+02	9.23E+01
CML103-2	9.88E+04	0.00E+00	7.88E+04	1.00E+03	1.38E+02
CML103-3	1.03E+05	0.00E+00	4.61E+04	4.61E+01	9.23E+01
CML247-1	2.72E+05	0.00E+00	2.46E+03	0.00E+00	1.38E+02
CML247-2	3.24E+05	0.00E+00	4.65E+03	0.00E+00	1.85E+02
CML247-3	2.56E+05	0.00E+00	3.10E+03	8.61E+02	2.46E+02
CML227-2	9.52E+03	0.00E+00	7.12E+03	2.30E+02	9.23E+01
CML227-3	1.68E+04	0.00E+00	9.76E+03	1.34E+02	9.23E+01
CML322-1	1.30E+05	2.96E+04	1.31E+04	1.19E+02	1.38E+02
CML322-2	1.72E+05	0.00E+00	2.31E+04	3.07E+01	9.23E+01
CML322-3	1.13E+05	0.00E+00	1.49E+04	6.42E+03	9.23E+01
CML333-1	3.09E+03	0.00E+00	3.08E+01	0.00E+00	9.23E+01
CML333-2	1.78E+05	0.00E+00	2.47E+04	1.68E+02	2.92E+02
CML333-3	1.51E+03	1.31E+02	3.07E+01	0.00E+00	4.61E+01
CML52-1	1.40E+04	0.00E+00	1.44E+04	3.08E+01	1.23E+02
CML52-2	9.29E+03	0.00E+00	7.76E+03	3.91E+02	1.38E+02
CML52-3	4.27E+04	0.00E+00	1.53E+04	1.15E+02	7.69E+01
Dhurrin std 0.1 mg mL	8.03E+05	1.69E+02	2.31E+01	0.00E+00	1.39E+02
Dhurrin std 0.2 mg mL	4.31E+05	1.38E+02	1.23E+02	0.00E+00	1.54E+01
Dhurrin std 0.4 mg mL	1.14E+06	6.54E+02	2.31E+01	0.00E+00	2.08E+02
Dhurrin std 0.8 mg mL	1.58E+06	1.32E+03	1.38E+02	0.00E+00	1.08E+02
Dhurrin std 1 mg mL	1.55E+06	1.29E+03	1.23E+02	0.00E+00	3.07E+01
CML69-1	1.26E+06	0.00E+00	2.37E+04	-1.91E+02	2.54E+02
CML69-2	9.65E+05	0.00E+00	2.07E+04	1.74E+01	8.46E+01
CML69-3	7.45E+05	0.00E+00	1.66E+04	4.85E+01	1.54E+01
HP301-1	2.15E+05	1.46E+03	6.10E+04	2.00E+02	1.54E+01
HP301-2	3.85E+05	4.00E+02	1.28E+05	1.08E+02	9.23E+01
HP301-3	2.97E+05	2.62E+02	9.04E+04	8.43E+01	6.93E+01
II14H-1	4.55E+05	0.00E+00	3.45E+04	1.65E+03	1.00E+02
II14H-2	5.33E+05	0.00E+00	2.58E+04	3.71E+02	2.31E+02
II14H-3	5.89E+05	0.00E+00	5.61E+04	3.54E+02	9.23E+01
Ki11-1	1.16E+05	0.00E+00	1.70E+05	1.01E+02	5.38E+01
Ki11-2	6.91E+04	0.00E+00	1.03E+05	4.52E+00	1.54E+01
Ki11-3	6.00E+04	0.00E+00	1.13E+05	1.26E+02	9.23E+01
Ki3-1	1.40E+05	0.00E+00	1.34E+05	1.24E+02	1.85E+02
Ki3-2	2.26E+05	0.00E+00	5.97E+04	1.17E+02	9.23E+01
Ki3-3	1.17E+05	0.00E+00	6.97E+04	1.69E+02	9.23E+01
Ky21-1	2.27E+05	0.00E+00	5.58E+04	7.32E+01	9.22E+01
Ky21-2	4.53E+05	0.00E+00	7.60E+04	1.10E+02	1.23E+02
Ky21-3	1.84E+05	0.00E+00	3.31E+04	1.49E+02	5.38E+01
M162w-1	2.10E+05	0.00E+00	4.54E+02	2.69E+02	1.31E+02
M162w-2	2.98E+05	0.00E+00	1.18E+03	4.12E+01	9.23E+01
M162w-3	1.86E+05	0.00E+00	3.46E+02	5.40E+01	4.62E+01
M37W-1	1.95E+05	0.00E+00	1.88E+04	7.65E+03	1.54E+01
M37W-3	1.05E+05	0.00E+00	3.34E+04	1.22E+04	9.23E+01
Mo17-1	2.42E+05	5.29E+03	7.69E+04	-8.62E+01	8.46E+01
Mo17-2	3.14E+05	0.00E+00	5.77E+04	1.39E+02	9.24E+01
Mo17-3	2.65E+05	1.71E+03	1.54E+01	1.83E+02	1.38E+02
Mo18w-1	1.39E+05	3.57E+03	2.31E+02	5.38E+01	4.61E+01
Mo18w-2	2.35E+05	5.37E+03	5.33E+04	2.28E+04	9.23E+01
Mo18w-3	2.00E+05	1.45E+03	3.61E+02	2.82E+03	3.00E+02
MS71-1	4.85E+05	0.00E+00	1.69E+02	4.42E+03	1.38E+02
MS71-2	3.07E+05	0.00E+00	4.46E+02	3.48E+03	2.77E+02

MS71-3	4.85E+05	0.00E+00	6.14E+01	6.23E+03	3.15E+02
NC350-1	1.13E+05	0.00E+00	1.12E+04	2.16E+04	8.46E+01
NC350-3	1.32E+05	0.00E+00	9.52E+04	2.90E+04	1.23E+02
NC358-2	1.55E+05	0.00E+00	5.35E+04	1.23E+04	9.23E+01
Oh43-2	2.12E+05	0.00E+00	1.08E+02	1.32E+04	2.31E+02
Oh43-3	2.04E+05	0.00E+00	5.38E+01	2.25E+04	1.38E+02
Oh43-4	1.59E+05	0.00E+00	7.69E+02	7.83E+03	6.14E+01
Oh43-5	5.18E+04	0.00E+00	7.69E+01	3.24E+02	9.23E+01
Oh43-6	2.52E+05	0.00E+00	4.62E+01	2.10E+04	2.31E+01
Oh78-1	2.43E+05	0.00E+00	7.39E+02	5.91E+03	6.92E+01
Oh78-2	4.85E+05	0.00E+00	1.22E+03	1.30E+04	1.38E+02
Oh78-3	2.78E+05	0.00E+00	3.54E+02	1.28E+04	8.46E+01
P39-1	2.57E+05	0.00E+00	2.31E+02	1.03E+03	1.08E+02
P39-2	3.57E+05	0.00E+00	4.15E+02	1.60E+03	4.61E+01
P39-3	3.96E+05	0.00E+00	1.69E+02	1.58E+03	3.69E+02
Tzi8-1	5.46E+04	0.00E+00	1.01E+03	3.08E+04	1.85E+02
CML228-1	1.41E+05	0.00E+00	4.61E+01	3.31E+02	2.62E+02
CML228-2	1.00E+05	0.00E+00	1.54E+01	2.46E+02	7.69E+01
CML228-3	1.11E+05	0.00E+00	3.23E+02	2.92E+02	1.85E+02

7.2.2 Nested associated mapping with B73xCML69 RILs

Table 7.5: LC-MS data of the analyzed NAM RILs.

Sample	Analyte Peak Area (counts)	Sample	Analyte Peak Area (counts)	Sample	Analyte Peak Area (counts)	Sample	Analyte Peak Area (counts)
Z009E001	747000	**Z009E054**	200000	**Z009E105**	81100	**Z009E156**	2130000
Z009E002	1120000	**Z009E055**	164000	**Z009E106**	1080000	**Z009E157**	76300
Z009E003	98000	**Z009E056**	427000	**Z009E107**	595000	**Z009E158**	56100
Z009E004	146000	**Z009E057**	1200000	**Z009E108**	70400	**Z009E159**	233000
Z009E005	1170000	**Z009E058**	307000	**Z009E109**	478000	**Z009E160**	414000
Z009E007	818000	**Z009E059**	370000	**Z009E110**	606000	**Z009E161**	405000
Z009E008	284000	**Z009E060**	590000	**Z009E111**	1100000	**Z009E162**	162000
Z009E009	1060000	**Z009E061**	92200	**Z009E112**	13800	**Z009E163**	796000
Z009E010	/	**Z009E062**	241000	**Z009E113**	134000	**Z009E164**	1960000
Z009E012	175000	**Z009E063**	32200	**Z009E114**	73100	**Z009E165**	1330000
Z009E013	145000	**Z009E064**	18200	**Z009E115**	871000	**Z009E166**	416
Z009E014	1680000	**Z009E065**	65500	**Z009E116**	964000	**Z009E167**	127000
Z009E015	41300	**Z009E066**	884000	**Z009E117**	59200	**Z009E168**	80500
Z009E016	162000	**Z009E067**	101000	**Z009E118**	162000	**Z009E169**	105000
Z009E017	113000	**Z009E068**	62200	**Z009E119**	2940000	**Z009E170**	1130000
Z009E018	85200	**Z009E069**	2250000	**Z009E120**	1150000	**Z009E171**	1130000
Z009E019	1120	**Z009E070**	341000	**Z009E121**	5160	**Z009E173**	887000
Z009E020	89000	**Z009E071**	46500	**Z009E122**	1160000	**Z009E174**	91500
Z009E021	2020000	**Z009E072**	67600	**Z009E123**	908000	**Z009E175**	6320
Z009E022	36100	**Z009E073**	106000	**Z009E124**	1300000	**Z009E176**	87500
Z009E023	15900	**Z009E074**	633000	**Z009E125**	395000	**Z009E177**	630000
Z009E024	/	**Z009E075**	49000	**Z009E126**	23400	**Z009E178**	48700
Z009E025	271000	**Z009E076**	487000	**Z009E127**	6130	**Z009E179**	72000
Z009E026	56200	**Z009E077**	1000000	**Z009E128**	64300	**Z009E180**	887000
Z009E027	56900	**Z009E078**	3970	**Z009E129**	60500	**Z009E181**	971000
Z009E028	15800	**Z009E079**	210000	**Z009E130**	337000	**Z009E182**	537000
Z009E029	59	**Z009E080**	949000	**Z009E131**	33400	**Z009E183**	96000
Z009E030	/	**Z009E082**	203000	**Z009E132**	673000	**Z009E184**	1190000
Z009E031	37900	**Z009E083**	359000	**Z009E133**	149000	**Z009E185**	137000
Z009E032	94200	**Z009E084**	18300	**Z009E134**	130000	**Z009E186**	2700
Z009E033	496000	**Z009E085**	158000	**Z009E135**	29000	**Z009E187**	150
Z009E034	702000	**Z009E086**	87300	**Z009E136**	337000	**Z009E188**	846000
Z009E035	438000	**Z009E087**	159000	**Z009E137**	120000	**Z009E190**	97800
Z009E036	54200	**Z009E088**	125000	**Z009E138**	2080000	**Z009E191**	1000000
Z009E037	17800	**Z009E089**	/	**Z009E139**	56600	**Z009E192**	559000
Z009E038	42	**Z009E090**	25200	**Z009E140**	1080000	**Z009E193**	759000
Z009E039	1150000	**Z009E091**	288	**Z009E141**	415000	**Z009E194**	781000
Z009E040	170000	**Z009E092**	36300	**Z009E142**	15200	**Z009E196**	49900
Z009E041	142000	**Z009E093**	146000	**Z009E143**	310	**Z009E197**	1430000
Z009E042	948000	**Z009E094**	37100	**Z009E144**	896000	**Z009E198**	94000
Z009E043	633000	**Z009E095**	1430000	**Z009E145**	76700	**Z009E199**	768000

Z009E044	157000	**Z009E096**	113000	**Z009E146**	1620000	**Z009E200**	76900
Z009E045	1130000	**Z009E097**	1610000	**Z009E147**	172000	**Z009E214**	618000
Z009E047	102000	**Z009E098**	129000	**Z009E149**	76300	**Z009E219**	60900
Z009E048	66800	**Z009E099**	29800	**Z009E150**	882000	**Z009E223**	2050000
Z009E049	330000	**Z009E100**	41800	**Z009E151**	56600	**Z009E243**	41800
Z009E050	3630	**Z009E101**	700000	**Z009E152**	1000000	**Z009E257**	83200
Z009E051	73400	**Z009E102**	557000	**Z009E153**	73100	**Z009E259**	78600
Z009E052	33000	**Z009E103**	663000	**Z009E154**	152000	**Z009E266**	89200
Z009E053	3070000	**Z009E104**	1400000	**Z009E155**	80600	**Z009E284**	504000

7.3 Accession Numbers

Table 7.6: Accession numbers of the UGTs used in the phylogenetic analysis.

Name	Origin	Accession number/Chromosomal location
UGT85K5	*Manihot esculenta*	G3FIN9
UGT85K4	*Manihot esculenta*	G3FIN8
UGT33A1	*Zygaena filipendulae*	D2JLK8
UGT85Q1	*Linum usitatissimum*	I2BHA2
UGT85A19	*Prunus dulcis*	B2XBQ5
UGT85K3	*Lotus japonicus*	chr3.CM0241.610.r2.m

7.4 Enzyme assays

Table 7.7: Results of the enzyme assay replicates. The highest peaks (counts) are listed.

Allel and Substrate	CML-Tyr	CML-Trp	CML-Phe	Mo-Tyr	Mo-Trp	Mo-Phe
1.Replicate	1100	14000	550000	1300	6000	220000
2. Replicate	1600	16000	400000	1150	7000	190000
3. Replicate	1900	17000	340000	2100	8000	240000
Mean	1533	15666	430000	1516	7000	216666

8 Acknowledgement

First and foremost I would like to thank Dr. Tobias Köllner for providing me with the opportunity to write my Bachelor thesis in his group and for the assistance and help he has given me and the patience he has invested in me. Besides of my supervisor I would also like to thank Dr. Sandra Irmisch for her guidance during my cloning experiments and Vinzenz Handrick for his help with the LC-MS and data examination.

I thank our technicians Bettina Raguschke and Katrin Luck for sequencing my many cloning attempts and helping me with the yeast expression. I would also like to take this opportunity to thank my lab partners Andrew Farthing, Nathalie Lackus and especially Jan Günther for the many small things he has helped me with and questions that he has answered. Further thanks go to the cake group for the many delicious cakes and conversations.

Finally my sincere thank goes to Prof. Dr. Gershenzon and the MPI for Chemical Ecology for allowing me to write my Bachelor thesis there and to Prof. Dr. Große for his willingness to review it.

9 Statement of authorship

I hereby declare, that I have written this Bachelor thesis autonomously and without any other help or sources than those denoted.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Jena, den 24.09.2015

_____

(Elias Kalthoff)