Investigation of regulatory factors in the signal transduction network of herbivore-induced terpene production in maize

A thesis submitted to attain the degree of Doctor rerum naturalium
(Dr. rer. nat.)
Submitted to the Faculty of Natural Sciences I (Biological Science) of
the Martin Luther University Halle-Wittenberg

By Franziska Irmer
born 24.02.1988 in Jena

Jena, 24.11.2018

Verteidigung:
22.05.2019

Gutachter:
Prof. Degenhardt
Prof. Scheel
Prof. Novak

Summary

Maize is one of the most important food and feed crops of the world. Its vulnerability to herbivore damage leads to high losses, and extensive use of pesticides. Many maize lines have different mechanisms of defense against herbivory. One of these mechanisms is the induced production and emission of volatile terpenes when herbivores feed on the plant. The volatiles attract natural enemies of the herbivores, which kill the herbivore and thus reduce damage to the maize plant. There are efforts to breed maize plants towards optimizing this defense mechanism. In order to facilitate the selection of plants in this breeding, the complex regulatory network of herbivore-induced terpene production needs to be understood.

Previous studies have identified several terpene-producing enzymes, the terpene synthases. Statistical analyses of different maize line phenotypes and genotypes have shown that the regulation of terpene production is not limited to the terpene synthases, but involves several levels of regulation. The aim of this work was to identify such factors and investigate their role in the regulation network.

Using statistical calculations of Nested Association Mapping, a number of genes that play different roles in the regulation of terpene production was identified. Genetic comparisons between different maize lines, and looking for homologous genes in Arabidopsis and rice, have been used to validate the function of these candidate genes. The expression of the candidate genes has been compared in control plants and plants treated with indanone, which mimics the elicitor effects of herbivores feeding on the plant.

This work concentrated on two genetic regions, called Quantitative Trait Loci, which show an influence on the emission of different sets of terpenes. For QTL_{215} on chromosome 2, eight candidate genes were found. For $QTL_{991\text{-}996}$ on chromosome 9, two candidate genes were identified. For each candidate gene, the sequences were compared between different lines of the NAM population. If transposon-carrying lines were available, the effects of transposon insertions in the candidate genes were investigated.

Two transcription factors of the Ereb gene family were identified outside the QTL regions. Their expression level does not change significantly after herbivory, but sequence differences between different maize lines correlate to differences in terpene emission in these lines.

Overall, several factors influencing the production of different sets of terpenes were identified. Their interaction partners in the complex regulatory network of terpene production are yet to be determined more exactly.

Zusammenfassung

Mais ist eine der wichtigsten Nahrungs- und Futtermittelpflanzen weltweit. Die Anfälligkeit von Maispflanzen für Fraßschädlinge führt zu hohen Ernteverlusten trotz des Einsatzes von Pestiziden. Viele Kulturlinien des Mais haben natürliche Verteidigungsstrategien gegen herbivore Insekten. Einer dieser Mechanismen ist die Produktion und Abgabe von flüchtigen Terpenen nach Insektenbefall. Diese Terpene locken die Feinde der Herbivore an, die die Herbivoren abtöten und dadurch weitere Fraßschäden an der Pflanze verhindern. Es gibt Bestrebungen von Pflanzenzüchtern, diese Fähigkeit zur Insektenabwehr in neuen Maislinien zu optimieren. Um das zu ermöglichen, müssen die komplexen Mechanismen der herbivor-induzierten Terpensynthese untersucht werden. Mit dem Wissen über die zugrundeliegenden Mechanismen kann bei der Zucht gezielter selektiert werden.

Frühere Studien haben mehrere terpenproduzierende Enzyme, die Terpensynthasen, identifiziert. Statistische Analysen der Phenotypen und Genotypen verschiedener Maislinien haben gezeigt, dass die Regulation der Terpensynthese nicht nur auf Ebene der Terpensynthasen stattfindet, sondern ein Netzwerk von regulierenden Faktoren vorliegt. Ziel dieser Arbeit war, solche Faktoren zu finden und ihre Rolle im regulatorischen Netzwerk zu untersuchen.

Mithilfe der statistischen Methode des Nested Association Mapping wurden verschiedene Gene identifiziert, die an der Regulation der Terpensynthese beteiligt sind. Vergleiche der Gensequenzen zwischen verschiedenen Maislinien der NAM-Population, und die Suche nach homologen Genen in Arabidopsis und Reis, wurden genutzt um diese Gene zu charakterisieren. Die Expression der Gene in Kontrollpflanzen wurde verglichen mit der Expression in mit Indanon behandelten Pflanzen. Die Behandlung mit Indanon stellt die Reaktionen auf Herbivorie nach.

Diese Arbeit konzentrierte sich auf zwei genetische Regionen, sogenannte Quantitative Trait Loci, die einen Einfluss auf die Emission verschiedener Terpene zeigen.

Sowohl für QTL_{215} auf Chromosom 2 als auch für $QTL_{991\text{-}996}$ auf Chromosom 9 wurden mehrere Kandidatengene identifiziert und deren Sequenzen zwischen verschiedenen Elternlinien der NAM-Population verglichen. Wenn Maislinien mit Transposon-Insertionen verfügbar waren, wurden die Einflüsse dieser Insertionen auf die Terpenemission untersucht.

Zusätzlich wurden zwei Transkriptionsfaktoren der Ereb-Genfamilie außerhalb der QTLs gefunden und untersucht. Ihre Expression verändert sich nicht nach Herbivorie, aber das

Auftreten von Sequenzunterschieden zwischen verschiedenen Maislinien korreliert mit Unterschieden in der Terpenemission dieser Linien.

Insgesamt wurden zwölf Kandidatengene auf zwei Chromosomen identifiziert, und ihre Sequenzen, Expressionsmuster und Einfluss auf die Terpensynthese untersucht. Ihre Interaktionspartner im regulatorischen Netzwerk sind noch zu identifizieren.

Contents

Table of abbreviations

cDNA	complementary deoxyribonucleic acid
DMAPP	dimethylallylpyrophosphate
DMNT	4,8-dimethylnona-1,3,7-triene
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
FPP	farnesyl diphosphate
GC/MS	Gas Chromatography and Mass Spectroscopy
GGPP	geranylgeranyl diphosphate
GPP	Geranyl diphospate
GWAS	Genome Wide Association Mapping
IPP	isopentenyl pyrophosphate
JA	jasmonic acid
LB	Lysogeny broth
MEP	methylerythrolphosphate
mRNA	Messenger Ribonucleic acid
MVA	mevalonate
NAM	Nested Association Mapping
OPDA	cis-(+)-12-oxophytodien acid
PCR	polymerase chain reaction
PDMS	Polydimethylsiloxane
qRT-PCR	quantitative real-time polymerase chain reaction
QTL	quantitative trait loci
RIL	recombinant inbred lines
SNP	single nucleotide polymorphisms
SPME	solid phase micro extraction
TAE	Tris acetate EDTA
TMTT	(E,E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene
TPS	Terpene synthase
Tris	tris(hydroxymethyl)aminomethane
UFMu	Uniform Mu

1. Introduction

1.1 Maize as a food and feed product

Maize is one of the most important crops worldwide. The worldwide production in 2009 was 817 million tons – more than any other grain cultivated in the world. Grown on all continents, maize is used for food, animal feed, bioethanol production, and as a raw material for chemical industry. Additionally, it is used to produce biodegradable plastics, fabrics and adhesives.

Maize originates from teosinte, and was first cultivated as a food crop around 9000 years ago in South America (Piperno, et al., 2009). Mesoamerican farmers selectively bred maize for an increase in size of kernels, number of rows and number of cobs per plant.

Figure 1: Corn cobs uncovered by archaeologists show the evolution of modern maize over thousands of years of selective breeding (Robert S. Peabody Museum).

Maize breeding throughout history aimed for traits like increased yield, drought resistance, and nutrition values. The resistance of maize plants against pests was not considered as much, since pesticides were used to fend them off.

The modern six main types are dent corn, flint corn, pod corn, popcorn, flour corn and sweet corn. There is a large variety of different maize cultivars in these types or hybrids between these types. The different lines are for example bred for specific climates (Körber-Grohne, 1995).

Maize for food and animal feed has a highly favorable composition of nutrients. It contains large amounts of starch, oils, proteins and nutritional fiber, as well as B vitamins like thiamin and folate, vitamin C and minerals.

1.2 Genetics of maize plants

Maize plants are diploid organisms with 20 chromosomes (n=10). Due to its quick growth and easy cultivation, maize is widely used as a model organism for developmental biology (Strable, et al., 2010). The Maize Stock Center of the University of Illinois (University of Illinois at Urbana/Champaign - Department of Crop Sciences, S-123 Turner Hall, 1102 South Goodwin Avenue, Urbana, IL 61801-4730, USA) curates about 80,000 different maize samples, which exhibit a wide variety of different traits and adaptations to different environments. These maize lines are a valuable resource for genetic research.

Sequencing and annotation of the maize genome is a large ongoing project. In the first published sequence, a total number of 32,540 genes was estimated (Schnable, et al., 2009). There is a large number of transposons in the genome. Transposons are genetic elements which can switch their position in the genome, and thus lead to increased genetic variability.

1.3 The maize plant is susceptible to a variety of herbivores and diseases

Crops grown in large scale, especially in monoculture, often suffer yield loss due to herbivores or infectious diseases. Resistance against such influences has not been a main focus for plant breeding for a long time. Focus on high-yield plant lines often encouraged the use of pesticides to prevent any losses.

Maize plants can suffer from a wide range of diseases as well as attack from herbivorous insects, which can lead to significant amounts of yield loss. Above ground, any pests that can damage the leaves and such reduce the photosynthetic area of the plant, lead to lower biomass production and thus a loss in yield. Below ground, parasites can attack the roots and lead to a loss of mineral and water availability for the plant, which also lead to lowered growth and yield.

Among the most significant diseases of the maize plant are fungi like Corn Smut (*Ustilago maydis*) or Southern Leaf Corn Blight (*Bipolaris maydis*), and viruses like Maize Dwarf Mosaic Virus and Maize Streak Virus.

This work focuses on herbivores that attack maize plants. Herbivorous pests of maize include different species of armyworm (*Spodoptera*), aphids, rootworms, silkflies and corn borers. Attack from these herbivores can greatly reduce the yield, especially when young plants with vulnerable leaves are attacked. However, many plants have developed a number of natural defense mechanisms, which can be classified in two classes: constitutive and induced defenses.

Constitutive defenses include plant surface structures like thorns or a waxy cuticula or preformed defense-related chemicals, which are constantly present to impede herbivore feeding (Fürstenberg-Hägg, et al., 2013).
Insect-induced direct defense mechanisms can include the production of chemicals that are toxic for the herbivore (Rosenthal, et al., 1992), or an increase of lignification of the cell wall to make herbivory more difficult (Arimura, et al., 2011) (Hernandez-Cumplido, et al., 2016).

Compared with constitutive defenses, induced reactions have both advantages and drawbacks to the plant. A certain amount of damage is caused by the herbivore during the time needed to execute the induced defense, which could have been avoided by effective constitutive defense. But the constitutive expression of a defense competes for resources that are needed for proper growth and reproduction processes. A defense mechanism that is only induced when needed saves valuable resources for the plant. Like most plants, maize has both constitutive and induced defense mechanisms. In adult maize plants, the hardness of the leaves acts as a constitutive, passive protection against herbivores feeding on the leaves. Young plants with vulnerable, soft leaves need to rely on other protection mechanisms, like the induced defenses described in this work.

This work focuses on maize tritrophic interaction - an induced, indirect defense mechanism of maize plants against herbivores feeding on leaves.

1.4 Tritrophic Interactions can help reduce the damage caused by herbivores to different plants

Tritrophic interactions are reactions between organisms of three trophic levels. In the case of maize herbivore defense, it includes the maize plant, an herbivore pest, and the natural enemy of the herbivore (Turlings, et al., 1995) (Tumlinson, et al., 1993) . These interactions reduce the herbivore damage to a plant by attracting natural enemies of the herbivores. In the tritrophic interaction of maize, herbivore and parasite, both the plant and the parasite that is attracted profit from the interaction – the plant suffers less damage from herbivory, and the animal finds herbivores which serve as food or host for larvae.

This interaction mechanism is known for a large number of plants, their respective herbivores and their enemies. For example, Wild Privet (*Ligustrum vulgare*) reacts upon herbivory by the caterpillar of *Pseudargyrotoza conwagana* by secreting volatiles that attract the wasp *Bracon otiosus* (Hernández, et al., 2014). The wasp oviposits in the caterpillar, which leads to death of the caterpillar. The growth of the larvae inside the host is fast enough to kill the host rapidly, thus stopping the herbivore feeding on the plant.

Similarly, the Lima Bean (*Phaseolus lunatus*) defends against the feeding of the herbivorous spider mite *Tetranychus urticae* by secreting volatiles that attract the predatory mite *Phytoseiulus persimilis* (Ozawa, et al., 2012). The predatory mite consumes the eggs of the herbivorous mite, which reduces the population of herbivores.

Another variant of tritrophic interaction was demonstrated in Arabidopsis. When Arabidopsis plants were genetically modified to produce volatile isoprenoids, they attracted carnivorous predatory mites (*Phytoseiulus persimilis*) which reduced the numbers of herbivores attacking the plants (Kappers, et al., 2005). This shows the potential to engineer terpene production in plants to reduce their susceptibility to herbivores.

Such tritrophic interactions are beneficial to the plant, as they reduce the number of herbivores, which reduces the damage to the plant (Dicke, et al., 1987). In most cases the main factors are volatile substances emitted by the plant, which draw natural enemies of the herbivore from a wide distance (Dicke, et al., 1990) (Turlings, et al., 1995).

There are local reactions where just the attacked plant part reacts, but more effective mechanisms also include a systemic reaction by the whole plant. The emission of volatiles

does not only happen above ground, but also by the root system in case of herbivory at the roots (Capra, et al., 2015).

In agriculture, tritrophic interactions are widely used to protect plants without using harmful pesticides (Sobhy, et al., 2015). Sometimes the natural enemies of the herbivores are supplied if the local ecosystem has a low amount or variety of them. This is widely favourable over the use of chemical insecticides, as it does not have a detrimental influence on soil and groundwater quality.

Breeding plants like maize with the goal to increase the ability to be part of tritrophic interactions is a current topic in maize breeding (Degenhardt, et al., 2003). Plant lines that use these reactions to defend against herbivores are less at risk for yield loss due to herbivory, and need less extensive care such as applying insecticides. In order to make breeding more effective, the mechanisms of the tritrophic interaction need to be known, so the breeder knows exactly which traits to select for. This is one of the main possible applications for the work in hand.

1.5 Tritrophic interactions in Maize increase the defense of maize plants against herbivores, thus reducing the need of pesticides

When a maize plant is attacked by caterpillars of the Egyptian cotton leaf worm (*Spodoptera littoralis)*, a variety of reactions in the plant is started. An indirect defense is the production and emission of a blend of volatile substances such as terpenes and "green leaf volatiles" including alcohols, aldehydes and esters (Pinto-Zevallos, et al., 2016). The emission of terpenes can attract insects, for example the parasitic wasp *Cotesia marginiventris* (Fontana, et al., 2011). This wasp oviposits in the caterpillars, and the growing parasitoid larvae kill their host. As a result, the number of caterpillars is significantly reduced, and the plant is subjected to less damage. Within two days after oviposition, the wasp larva grows to a size of about 5 mm (Gillett-Kaufman, 2017) – their feeding and quick growth kills the host rapidly.

After herbivory, maize plants emit a mixture of volatile terpenes. This mixture was proven to be more attractive to natural enemies of the herbivore than single terpenes (Fontana, 2011).

Among the rich variety of modern maize lines, not all plants emit the same amount and composition of volatile blend. The mixture of homo-, sesqui- and diterpenes and other volatiles in different maize breeds is highly variable. Thus, their ability to attract the different wasp species varies, and the maize lines show a different susceptibility to herbivore attacks (Degen, et al., 2012). The wasps can adapt to the different volatile blends, and learn which spectrum is linked to a herbivore feeding on the maize plant (Gandolfi, et al., 2003). There is also evidence that the parasitoid can distinguish the species and amount of herbivore infestation on the host plant (Yamauchi, et al., 2015).
A goal of plant breeding is to select for plants that produce the optimal amount and composition of volatiles that attract the wasp species present in the region where the maize is grown (Fatouros, et al., 2016). This increases the effectiveness of indirect defense against herbivory, making the maize plant more robust against the herbivores. Understanding the mechanisms of volatile production is crucial here.

1.6 Production pathways of terpenes in the maize plant

Terpenes are organic compounds derived from C5 isoprene units. They are products of the secondary metabolism of a wide variety of plant species, and are major building blocks for other products in animals, for example for steroids.
Volatile terpenes are mostly monoterpenes (C5), sesquiterpenes (C15), diterpenes (C20) and homoterpenes (C11, C16). Plants have a variety of different terpene synthase enzymes, which can produce one or more terpene products.

Terpenes are also called isoprenoids, because they are derived from isoprene units. The activated form of isoprene, isopentenyl pyrophosphate (IPP) is used as a building block for the synthesis of terpenes. IPP is created either in the cytosol in the mevalonate (MVA) pathway, which is also possible in animals; or in the plastids in the methylerythrolphosphate (MEP) pathway, which happens exclusively in plant cells, shown in Figure 2.

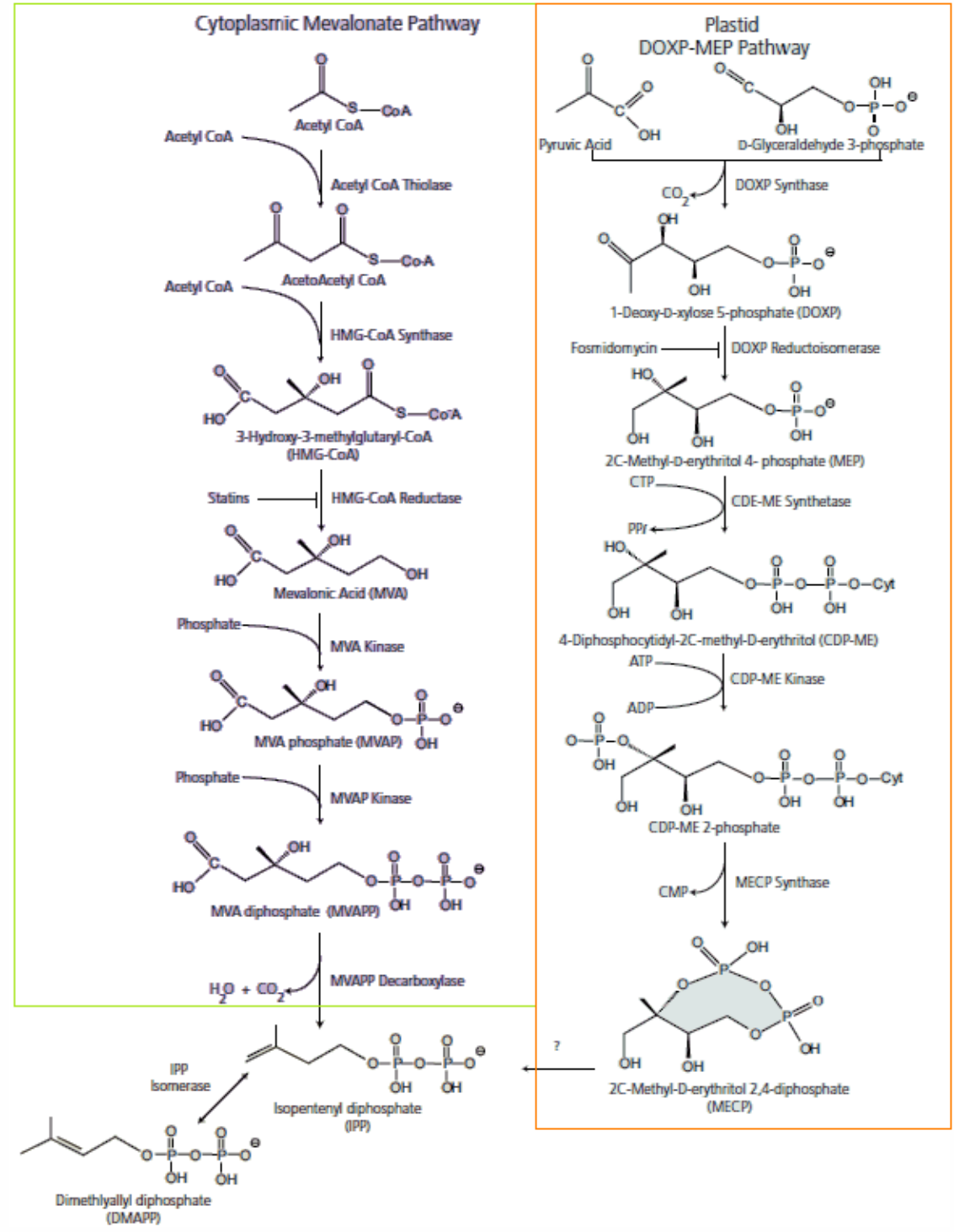

Figure 2: MVA pathway and MEP pathway leading to the synthesis of IPP and DMAPP, image source (Sigma-Aldrich, 2017)*, modified.*

The IPP and its dimer dimethylallylpyrophosphate (DMAPP) are then combined to polymers of varying sizes, which are modified to different terpenes by terpene synthases.

Both in the cytosol and in plastids, enzymes called prenyltransferases convert IPP and DMAPP into prenyl diphosphates (Dudareva, et al., 2006). Farnesyl pyrophosphate synthase

condenses two molecules of IPP and one molecule of DMAPP to the C15 molecule farnesyl diphosphate (FPP), a precursor of sesquiterpenes (McGarvey, et al., 1995). The precursor of monoterpenes, the C10 molecule geranyldiphosphate (GPP), is produced in plastids by geranyl pyrophosphate synthase, using one molecule IPP and one molecule DMAPP as substrates (Poulter, et al., 1981).
The enzyme geranylgeranyl pyrophosphate synthase condenses three molecules of IPP and one molecule of DMAPP to geranylgeranyl diphosphate (GGPP), a C20 molecule that is the precursor of diterpenes (Dudareva, et al., 2006).

1.7 A variety of Terpene Synthases in maize produce a spectrum of terpenes

Terpene production is catalyzed by terpene synthases (TPS), which are mostly named after their main products. Most terpene synthases are multi-product enzymes, allowing the plant to produce a varied spectrum of terpenes (Degenhardt, 2009).
GPP is converted into acyclic, monocyclic or dicyclic monoterpenes by monoterpene synthases (Srividya, et al., 2016). The substrate FPP is converted into a variety of sesquiterpenes by sesquiterpene synthases (Schifrin, et al., 2016). Diterpenes are produced by diterpene synthases, using GGPP as a substrate (Wang, et al., 2016). Many terpene synthases are multiproduct enzymes (Degenhardt et al., 2009). For example, TPS10 in maize produces (E)-α-bergamotene und (E)-β-farnesene. Some of these enzymes convert multiple substrates into different products. Other terpene synthases are single product enzymes, like δ-cadinene synthase in cotton (Yoshikuni, et al., 2006).

Terpene synthases have been found in many different plant families. The widely used model organism *Arabidopsis thaliana* emits a relatively small variety of terpenes, and in its genome, 32 putative active TPS and 8 pseudo TPS genes have been identified (Aubourg, et al., 2002). Grape (*Vitis vinifera*) produces a larger spectrum of volatiles, and has far more TPS enzymes. So far, 69 putative active TPS and 63 pseudo TPS have been identified in *V. vinifera* (Martin, et al., 2010).

Eucalyptus species have been used for centuries for their essential oil, which contains a mixture of terpenes with antiseptic effects used in pharmacy. For the plant itself, the oil has been shown to have antifungal effects (Eyles, et al., 2003), deterring herbivores and attracting pollinators and parasitoids (Giamakis, et al., 2001). In *Eucalyptus grandis*, 113 putative functional TPS sequences have been found (Külheim, et al., 2015).
Maize produces a spectrum of volatile terpenes, varying in different maize lines. Since they are not directly used as a product of the maize plant, they have for a long time been out of focus in maize breeding. Within the last years, the role of terpenes in herbivore defense has been investigated, stressing their importance for plant health and agricultural yields.

For my work, I concentrated on a set of terpenes that are most relevant for the tritrophic interaction in maize. They are produced by the following enzymatic steps:
TPS2 converts GPP to linalool, FPP to nerolidole, and GGPP to geranyllinalool (Richter, 2014). Geranyllinalool is converted to (*E*,*E*)-4,8,12-trimethyltrideca-1,3,7,11-tetraene (TMTT) by the P450 enzyme CYP92C6. Another P450 enzyme called CYP92C5 converts nerolidole to 4,8-dimethylnona-1,3,7-triene (DMNT) (Richter, et al., 2016). TPS10 produces (E)-α-bergamotene and (E)-β-farnesene (Schnee, et al., 2006). E-β-caryophyllene is a product of TPS23 (Köllner, et al., 2008). These are just the last steps of complex secondary metabolism pathways – so not only this last step needs to be investigated, but regulation can happen at many earlier steps in the production of precursor molecules, too.

All of these terpenes are subject to a complex regulatory network, and their production is increased after herbivory (Turlings, et al., 1990). It has been shown that after herbivory, there is an upregulation of terpene synthase genes (Tamiru, et al., 2017). The reaction is specific to the kind of attack (De Vos, et al., 2005). This happens by the recognition of so-called pathogen associated molecular patterns (PAMP) (Göhre, et al., 2008) – in the case of herbivory, these are elicitor substances from the saliva of the herbivore (Lauchli, et al., 2003). This recognition starts a signal cascade, leading to the upregulation of terpene synthase expression. The steps of this signaling network are widely unknown.

Starting at receptors for elicitors in the insect saliva, the signal network leads to the upregulation of terpene production and secretion of terpenes, and the down-regulation of the whole process after the herbivore attack has been fended off. The regulation happens at several levels in the pathway, leading to a varying volatile spectrum in different maize lines.

Secondary metabolite pathways can also be regulated in a redundant manner (Chezem, et al., 2016) (Gholami, 2012). That means the inactivity of one gene does not lead to a full loss of all following steps and products, but only results in a small change in product level, because other factors partly replace the effect of the missing regulatory factor. This makes the investigation of the role of single compounds in the signal network more complicated, because classical genetic methods like knockout mutants do not show unambiguous results like a complete knock-down of the affected phenotype.

1.8 The regulation of terpene synthesis in maize after herbivory

Since the production of volatiles requires resources and energy, it would be detrimental for the plant to produce them all the time. Constitutive volatile emission compromises seed germination, growth and yield (Robert, et al., 2013). It is advantageous to produce and emit terpenes only when needed, i.e. after an herbivore attack. Thus, plants have developed a regulation network, which recognizes not only mechanical damage, which could also be caused by abiotic factors such as hail. In the latter case, an increased terpene emission would be a waste of resources for the plant, since it has no defense impact. The Theory of Optimal Defense states that each defense mechanism comes with a benefit and a cost (Zangerl, et al., 1996). The cost of terpene production is the resources that could otherwise be used for growth. It is thus evolutionary favourable to produce terpenes only when needed.

Upon herbivory, terpene emission is induced by elicitor substances in the saliva of herbivorous insects (Turlings, et al., 1998), which include fatty acid-amino acid conjugates like volicitin [N-(17-hydroxylinolenoyl)-L-glutamine] and its biochemical precursor, N-linolenoyl-L-glutamine (Yoshinaga, et al., 2010).

This induction of terpene emission has to happen fast, to stop the herbivore from further damaging the plant. Young plants benefit most from a fast reaction and protection from herbivory, as their leaves are small and vulnerable. Older plants have more robust leaves and can survive loss of a small area of leaf. Experiments have indicated that young maize plants reach the maximum of terpene emission within four hours after herbivory. The terpene production stays up for a day, then it decreases again (Schmelz, et al., 2003) (Seidl-Adams, et al., 2015).

The production of terpenes is regulated on the level of gene expression (Huffaker, et al., 2013). There is a complex chain of regulatory genes, starting at the (yet to be identified) receptors for fatty acid conjugate elicitors volicitin and N-(17-Hydroxylinolenoyl)-L-Glutamine or N-Linolenoyl-L-glutamine, which originate from the saliva of *Spodoptera littoralis*. For lab experiments, the effects of caterpillar saliva can be mimicked using a solution of indanone derivative (Schüler, et al., 2001).

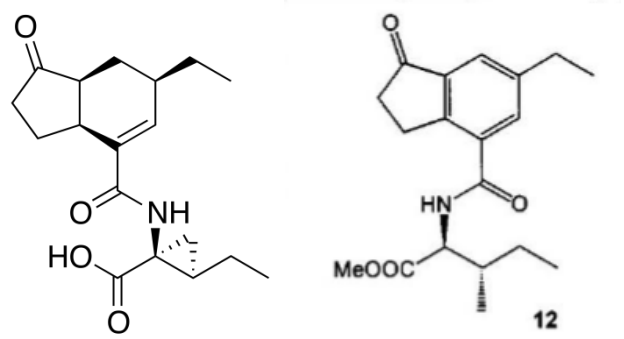

Figure 3: Coronatine (left) and 6-ethyl-indanoyl-isoleucine (right) (Schüler, et al., 2001).

It is uncertain which receptor in the leaf binds the indanone. Schüler et al (2001) suggests indanone mimics the function of coronantine and interacts with the receptors for jasmonic acid and cis-(+)-12-oxophytodien acid (OPDA). Due to its structural similarity to the keto group on 5-ring of jasmonates, indanone triggers the jasmonate response (Farmer, et al., 1992).

The production and emission of different terpenes is regulated differently, so we expect to find several regulatory factors influencing different steps of terpene synthesis. Even products of one terpene synthase do not necessarily show the same regulatory pattern.

An induced production of terpenes serves in the defense of maize plants against different kinds of attacks on the plant. Other than herbivory, factors inducing a specific upregulation of terpene synthesis can be the presence of fungi (Piesik, et al., 2011), bacteria or viruses (Becker, et al., 2014). Each of these requires a specific defense reaction – some terpenes have antifungal effects (Rao, et al., 2010), others are volatile and can attract enemies of the herbivores. Biotrophic plant pathogens are fended off by an increase in salicylic acid content, whereas necrotrophic pathogens and herbivores cause an upregulation of jasmonic acid (JA) and ethylene. This shows how a complex system of regulation can lead to a variety of useful outcomes in plant defense.

The complexity of these regulatory networks is the biggest challenge in investigating them. The signaling chain from the recognition of the elicitor in the herbivore saliva until the rise in terpene emission has many intermediate steps, several of which can be up- or downregulated.

Proving the role of one factor in this signaling chain is further complicated by genetic redundancies in the maize plant. Crossing two maize lines that differ in the functionality of one gene does not always result in a clear mendelian split of the phenotypes of their offspring. That's why more complex statistical methods are necessary to identify genes that play a role in the regulation of herbivore-induced terpene production.

1.9 Quantitative Trait Loci and Nested Association Mapping as means to elucidate terpene biosynthesis steps relevant for herbivory resistance.

Statistical analysis of genetic data allows us to find links between chromosomal regions and phenotypic traits of organisms. A Quantitative Trait Locus (QTL) is a genomic region that is linked to a specific phenotypic trait. Genetic differences in this region lead to phenotypic differences in this phenotype. For plants, there are QTLs which influence traits like flowering time, plant size, drought resistance and yield (Wang, et al., 2016).

The position of QTLs is calculated using genetic markers. Markers can be unique DNA sequence variations like single nucleotide polymorphisms (SNP) or insertions/deletions (indels), which can be identified unambiguously. They have a defined, known position in the genome. The markers need to be as constant as possible, with a low rate of random mutations, which would obscure the results. When two parent lines with a difference in markers are crossed, the resulting marker sequence in the offspring tells about the genetic region surrounding the marker. The genes in this region are usually identical to this genetic region in the parent line with the identical marker.

Depending on marker density in the region, a QTL covers a region on the chromosome that may contain several potential candidate genes. Knowledge of these QTLs can be used to make plant breeding more effective – for example, if a QTL is linked to drought resistance, one can

analyze this QTL in the plant population and predict the drought resistance of each plant, instead of having to test all plants by exposing them to different environmental conditions. This helps pick the best candidates for further breeding of a specific trait.

Nested Association Mapping (NAM) is a method to statistically correlate traits to genetic differences in plant lines. For the NAM population utilized in this study, 25 maize lines, selected for a large phenotypic and genetic variety, were crossed with the common parent line B73 (McMullen, et al., 2009). Each of the resulting F1 lines was self-crossed, until a population of 5000 homozygous recombinant inbred lines (RILs) was created (Fig. 4). The self-crossing assures that the lines are as near to homozygous as possible, because different alleles in one plant would make the analysis less conclusive.

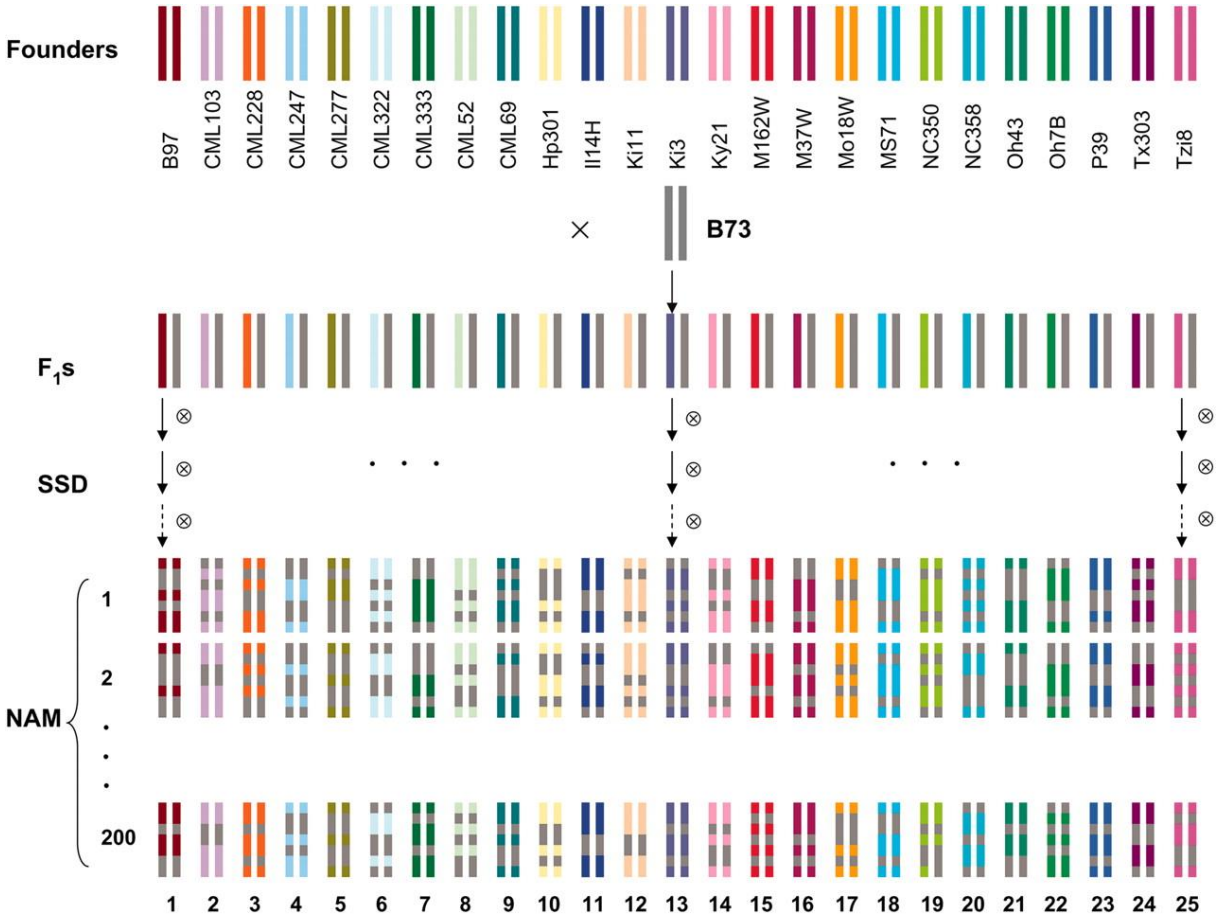

Figure 4: Structure of the NAM population; (Yu, et al., 2008).

These RILs were genotyped with 1106 SNP markers, each marker being specific for either B73 or the respective parent line. These markers were then statistically correlated with the

traits (like the amount of a certain volatile emitted by the plant). This indicates that the region around the SNP contains a gene which is relevant for the trait.

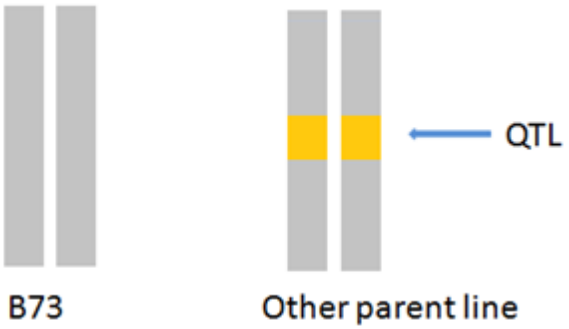

Figure 5: Using markers to visualize genetic differences between B73 and other parent lines, and thus calculate QTLs. The difference in the marker sequence shows that one genetic region is different between the two maize lines.

QTLs were calculated for the emission of each terpene as a unique trait. There are QTL which are only relevant for one terpene, while others are valid for several terpenes. Figure 6 displays the position of the highest significance QTL for each terpene.

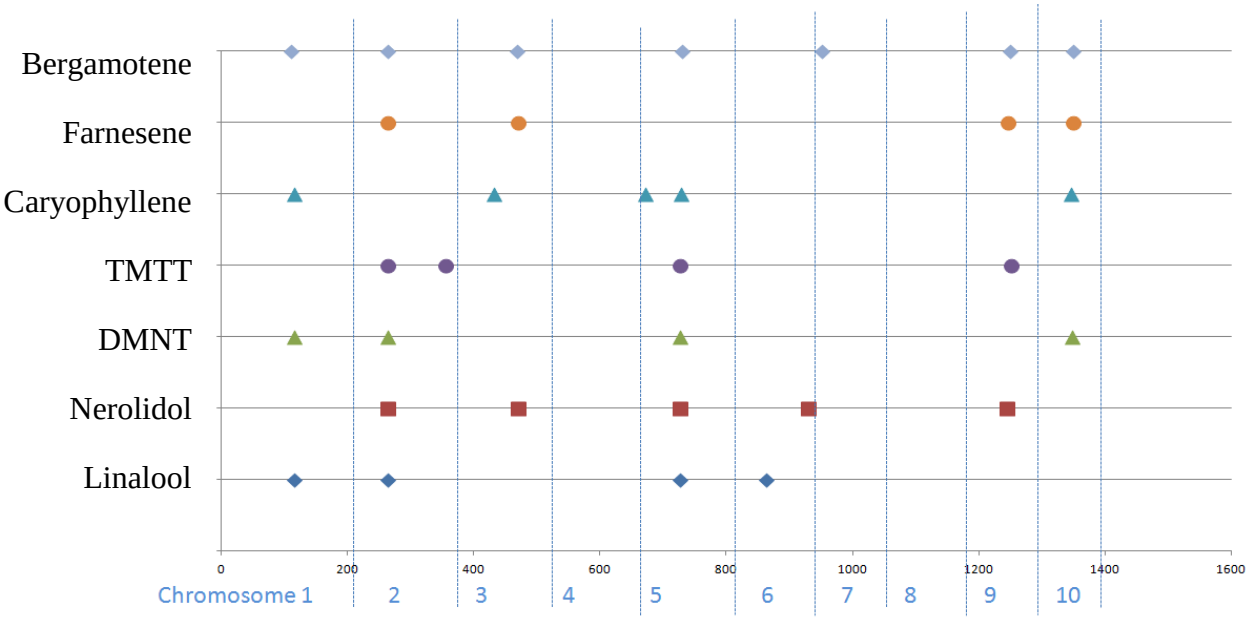

Figure 6: QTL for the emission of several terpenes, listed by cumulative chromosome position, (figure provided by Annett Richter).

One interesting result of this calculation is that several terpenes which are produced by different terpene synthases, can still share a QTL for regulation. This suggests that the regulation takes place at an earlier level than the later synthesis steps of the terpene

biosynthetic pathway, or one factor can regulate several enzymes of terpene production. Redundancies of enzymes in the secondary metabolism can lead to several factors having an influence on the regulation of one target enzyme.

In the simplest case, a QTL would point out an active terpene synthase gene. If some maize lines carry mutations which lead to an inactivity of this gene, the region shows up as a highly significant QTL for the products of this particular enzyme. Such mutations can be in the enzyme itself, leading to a loss of its function. Or they can affect the expression of the respective terpene synthase genes. For example, the QTL might be associated with nucleotide alterations situated in the promotor region of the terpene synthase gene and affect the binding of transcription factors, thereby preventing gene transcription. When a terpene synthase is not produced, this results in the absence of its products in the volatile spectrum of that plant line.

2. Objective of this work

When maize is attacked by herbivore, receptors in the maize leaf recognize substances in the caterpillar saliva and initiate a signal cascade that leads to the emission of a blend of volatile terpenes. These terpenes attract wasps that kill the caterpillars.

The emission of terpenes in maize is an inducible reaction and is specifically upregulated after herbivory. While most terpene synthases have been identified, the regulatory genes are widely unknown.

The aim of this work is to use quantitative genetic methods like NAM to identify genes that are part of the regulation process for the emission of different terpenes. Statistical correlations between genotypes and phenotypes are used to find chromosomal regions that are involved in terpene regulation. Specifically, my work focusses on two QTL loci. QTL_{215} on chromosome 2 is associated with the emission of bergamotene, farnesene, TMTT, DMNT, linalool and nerolidol. The second locus, $QTL_{991\text{-}996}$ on chromosome 9 is associated with the regulation of the herbivore-induced emission of bergamotene, farnesene, TMTT and nerolidol.

Within these regions, candidate genes are identified and their potential roles in the regulatory cascade investigated. To this aim, expression level and sequences of genes in different maize lines are analyzed.

3. Methods

3.1 Maize lines

Seed material of the common parent line B73, supplied by KWS (Einbeck) was used for most experiments, since its genome is almost fully sequenced and available to the public.

For comparison of candidate gene sequences, parent lines from the NAM population were used. Seeds of these parent lines were acquired from Maize Genetics Cooperation Stock Center operated by USDA/ARS, located at the University of Illinois, Urbana/Champaign.

UfMu transposon insertion lines were also obtained from the Maize Genetics Cooperation Stock Center.

maize lines	**properties**	**supplier**
B73	common parent line	KWS
Tzi8	NAM parent lines	Maize Genetics Cooperation Stock
Oh7b		Center
CML333		
Hp301		
W22	background line for UFMu lines	Maize Genetics Cooperation Stock Center
UFMu-00647	Mu-Transposon in	Maize Genetics Cooperation Stock
UFMu-05301	pentatricopeptide gene	Center
UFMu-03724	Mu-Transposon in kinase gene	Maize Genetics Cooperation Stock Center

Table 1: Maize lines used for this work, their properties, and suppliers of seeds.

3.2 Cultivation of maize plants

Maize seeds were incubated in petri dishes with tap water at 37°C overnight in the dark, and another 24 hours at room temperature in the dark. Then, they were planted in soil (Tonsubstrat by Klasmann, Gross-Hesepe, Germany) and further grown in a plant chamber by Snijders

(Tilburg, Netherlands) under the following conditions: Light 1 $mmolm^{-2}$ s^{-1} for 16 hours a day at 22 °C, then darkness for 6 hours at 18 °C. Air moisture was constant at 65 %.
If seed batches revealed fungal contamination that hindered plant growth, the seeds were disinfected by submerging them in 70% ethanol for 30 seconds, and then washing them with distilled water before starting the standard procedure of overnight incubation in water at 37 °C.

3.3 Induction of maize plants with indanone solution

Induction and harvesting of young plants was performed after approximately three weeks of growth. The third leaf of each plant was used.
A solution of Indanone derivative was used. 15,2 mg indanone were dissolved in 1 ml dimethyl sulfoxide (DMSO). For induction, 5µl of this solution were diluted in 1 ml of tap water. The maize leaf was cut off and placed in a tube with this induction solution. The leaf was then kept in the growth chamber under the same conditions. All samples were preserved by freezing in liquid nitrogen and storing at -80 °C. For further processing, the plant material was ground with mortar and pestle in liquid nitrogen without thawing.

3.4 Collection and analysis of volatiles

Frozen plant material was used for volatile analysis. The material was ground to a fine powder in liquid nitrogen. The sample remained frozen to prevent evaporation of volatile compounds. The plant material was placed in a glass vial and incubated with a solid phase micro extraction (SPME) fiber (Gerstel, Agilent, Mülheim a.d. Ruhr) at 42°C for 20 min in the air space above the plant material. At this temperature, the sample thawed rapidly, and the terpenes evaporated into the gaseous phase. The surface of the SPME fiber is covered with Polydimethylsiloxane (PDMS) that binds the volatiles in the headspace. The adsorption of volatiles is enriched linear to their concentration in the sample. Thus, the SPME sample represents the amount and proportions of the volatile blend in the headspace of the sample (Romeo, 2009).

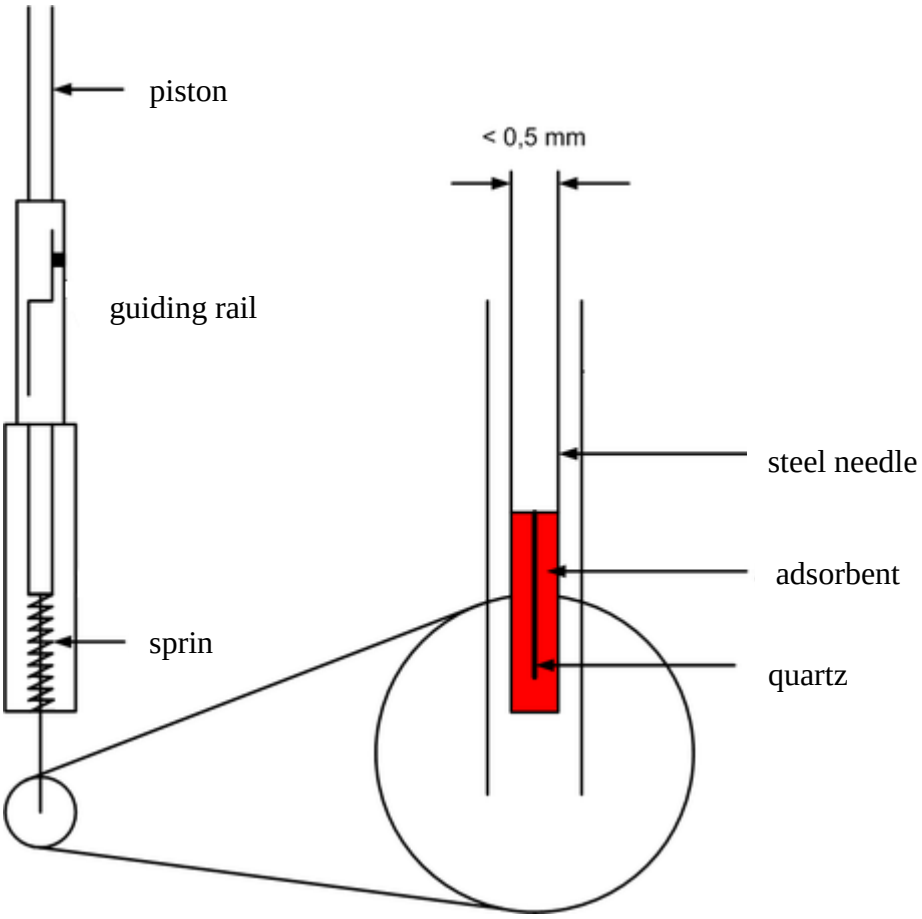

Figure 11: SPME sampling device; modified image based on (Mattern, 2017).

3.5 Gas Chromatography and Mass Spectroscopy

The SPME fiber was then inserted into a GC/MS (GC-2010 and GCMS-QP 2010 Plus, Shimadzu, Duisburg). With hydrogen as a carrier gas and a flow rate of 1 ml/min, the following program was run: 80 °C for 3 min, then raising the temperature by 7 °C/min to 200 °C, in the next step increasing the temperature to 300 °C within one minute, and holding this temperature for 2 min. The rising temperatures cause a thermic desorption of the volatile compounds from the SPME fiber surface. The volatiles were then separated by their binding to the gas chromatography column. Attached to the GC column, there was a mass spectrometer. The results of the gas chromatography were analyzed with the help of Shimadzu GCMS Postrun software. For the mass spectra, the libraries Shim2205 by Shimadzu and Wiley 8 by Hewlett Packard were used.

3.6 Extraction of genomic DNA

Total genomic DNA was extracted from the frozen plant material according to the protocol of (Brandstädter, et al., 1994).
500 mg to 1 g of raw material was frozen in liquid nitrogen, ground to a fine powder, and resuspended in 1ml of extraction buffer (100 mM Tris-HCl, pH 8; 50 mM EDTA; 500 mM NaCl; 10 mM mercaptoethanole; 1.5% SDS). The sample was incubated at 65 °C for 10 min. Then, 300 µl of 3M potassium acetate with 2M acetic acid were added. After 10 min incubation on ice, the samples were centrifuged for 5 min at 20.000 g and 5 °C. The supernatant was transferred into a new tube and 500 µl of Phenol-Chloroform-isopropylalcohol (25:24:1) were added. The sample was again centrifuged at 6.000 g for 6 min. The aqueous upper phase was precipitated with 500 µl isopropanol at -20 °C for 10 min and centrifuged at 20.000 g and 4 °C for 10 min. The resulting pellet was washed with 700 µl of 70 % ethanol, centrifuged for 2 min at full speed and then dried. The clean pellet was redissolved in 50 µl of water.

3.7 Concentration and purity of DNA samples

The concentration and purity of DNA was determined by InFinite 200 PRO NanoQuant by Tecan (Männerdorf, Switzerland) which measures the absorption at a wavelength of 260 and 280 nm. A DNA sample with a ratio of absorption at 260 nm to absorption at 280 nm of 1.8 was considered to be of good purity (ThermoScientific, 2008).

3.8 Extraction of RNA

The frozen, ground plant material was also used for the extraction of RNA. The RNeasy® Plant kit by Qiagen (Hilden, Germany) was used. A maximum of 100 mg frozen plant material was put into room temperature buffer RLT containing β-mercaptoethanol to avoid RNA degradation. Then the sample was filtered through a shredder column. The supernatant of the flow-through was mixed with 0.5 volume ethanol and applied onto a filter column. The

column was washed twice with wash buffer. Afterwards, the RNA was eluted in 25 µl RNAse-free water. During all steps, it is important to avoid RNAse contamination from surfaces and skin.

3.9 Concentration and purity of DNA samples for cDNA synthesis

The concentration of RNA was determined using the Nanoquant (Tecan, Männedorf, Switzerland). A ratio of absorptions at 260 nm to 280 nm of around 2 is considered to be a good purity for RNA samples. Lower ratios indicate contaminations with proteins or solvents from the purification process (ThermoScientific, 2008). It is important to use the same amount of RNA for cDNA synthesis in all samples, so the cDNA can be used for quantitative experiments.

3.10 removal of genomic DNA from RNA samples

Prior to cDNA synthesis, all genomic DNA had to be removed from the RNA sample. A mixture of genomic DNA and cDNA would cause wrong results in the quantitative analysis of cDNA samples. There would be false positives of genes that are present in the genomic DNA of a maize line, but not expressed.
The genomic DNA was removed by DNA digestion using the RQ1 DNAse kit by Promega (Fitchburg, USA). 3 µg of RNA were incubated with 1µl RQ1-DNAse buffer and 1.5 µl RQ1-DNAse, in a total volume of 10 µl. After incubation at 37 °C for 30 min, 1 µl of RQ1 DNAse stop solution was added. A 10 min incubation at 65 °C stopped the DNAse activity to avoid loss of cDNA due to remaining DNAse activity later on.

3.11 cDNA synthesis

1.2 µg RNA were used for cDNA synthesis. The synthesis was done with the First Strand cDNA Synthesis Kit by Thermo Scientific (Waltham, USA). With 6 µl RNAse-free water, 1 µl random primer and 1 µl Oligo-DT, the RNA sample was incubated for 5 min at 65 °C, then cooled down on ice. To this sample, 4 µl of 5x concentrated buffer, 1 µl RNAseOut, 1 µl of reverse transcriptase and 2µl dNTP were added. The cDNA synthesis is done in a thermocycler with the following programme:

5 min 25 °C
1 h 45 °C
5 min 70 °C

The random primers and oligoDT mix ensured that all mRNA sequences were transcribed to cDNA at an amount linear to their presence in the RNA sample. RNAse-free ingredients and RNAseOut prevented the degradation of RNA during the cDNA synthesis process.

3.12 DNA amplification by Polymerase Chain Reaction

DNA fragments were amplified using the polymerase chain reaction (PCR) with gene-specific primers with a length of 18-22 base pairs. In the PCR, the double stranded DNA was denatured into single stranded DNA. The primers bind to a specific site in the single-strand DNA, and are a starting point for the polymerase to copy the DNA. Primer construction has to ensure they are specific to one gene, and do not form hairpin loops within the primer. Primers binding to themselves would make them ineffective for DNA amplification. The GC content of a primer defines the strength of its binding to the target, and thus the annealing temperature. Higher GC contents mean a stronger bond and a higher annealing temperature. Choosing the correct annealing temperature can help avoid unspecific binding.

Different polymerases can be used, depending on the properties of the sample. The different polymerases also have different proofreading efficiencies, allowing for a different rate of random errors.
For most samples, the goTaq polymerase (Promega, Madison, USA) was used.

goTaq

2.5 µl goTaq buffer
0.25 µl fwd primer
0.25 µl rev primer
0.25 µl dNTP mix
0.0625 µl goTaq polymerase
6.1875 µl H_2O
2.5 µl cDNA

Advantage PCR

Advantage PCR was used for fragments that could not successfully be amplified in goTaq PCR. It has a higher efficiency and improved proof-reading. The following ingredients were used:

0.5 µl Advantage Taq polymerase [ClonTech]
36.5 µl H_2O
5 µl Advantage buffer
2 µl dNTP mix
2 µl fwd primer
2 µl rev primer
2 µl cDNA

All PCRs were done in peqSTARn 96 UNIVERSAL thermocyclers by peqlab (Erlangen).
The following programme was used:

initial denaturation of DNA	94°C for 2-5 min	
denaturation of DNA double strand	94°C for 30 s	repeat 39x
annealing of primers	58-62°C for 30s	
extension of complementary strand	72°C	
completion of DNA synthesis	72 °C for 5 min	
storage	4°C	

Table 2: PCR steps.

The annealing temperature depends on the GC content of the primers. The higher the GC content of the primers, the higher the temperature needs to be. Primer design and determination of annealing temperatures were done with the help of PrimerFox (Fuchs).
The length of the extension step depends on the length of the target sequence. For goTaq polymerase, a velocity of 1000 base pairs per minute is estimated.

3.13 Quantitative Real Time PCR

Quantitative Real Time PCR (qRT-PCR) is a method to determine the amount of a certain DNA fragment in a sample. When cDNA is used, it demonstrates the expression level of the fragment in the plant sample. The master mix for qRT-PCR contains a fluorescent dye, which intercalates in the product, and alters the fluorescence of the sample at a specific wave length. After each amplification step, the fluorescence was measured. This measurement represents the amount of product present in the solution at this specific timepoint. Since the amount of product directly depends on the number of templates for this sequence in the cDNA sample, this method is suitable to measure the relative amount of RNA that was present in the cell when the leaf was harvested and frozen. All expression levels are compared to a so-called housekeeping gene, which is expressed at a constant level in all cells (Lin, et al., 2014).
In this method, it is especially important that the primers are very specific. Primers binding to a wrong target would cause a wrong positive for the expression of the target gene. The typical target size is about 200 base pairs. Longer target sequences are not recommended.

For qRT-PCR, SYBR Green master mix by BioRad (iTaq universal SYBR green One-Step Kit, BIO-RAD, Munich) was used.

10 µl SYBR Green
5 µl forward primer
5 µl reverse primer
1 µl template cDNA
8 µl water

The qRT-PCR was done in a CFX96 Real Time thermocycler by BIO-RAD (Munich).

initial denaturation	95°C for 10 min	
Denaturation	95°C for 30 s	repeat 40x
Annealing	56-68 °C for 30s	
Extension	72 °C for 30 s – 1 min	
melting curve	56-95 °C for 30s each	

Table 3: qRT-PCR steps.

For all samples, the housekeeping gene HG5/HG6 (putative APT1A) was used as reference. Gene expression levels were calculated relative to this gene, using the CFX manager software. For each gene, at least three biological replicates were used.

3.14 Electrophoresis

Both RNA and DNA samples were analyzed by gel electrophoresis. Agarose gels are used to analyze PCR fragments by size. The sample buffer gives all fragments an equal charge, which makes them move in an electric field in the gel chamber. The smaller a fragment is, the faster it moves through the gel. A visible dye is added to see the progress of the run. The gel contains ethidium bromide, which intercalates in the DNA fragments, and makes them visible under UV light. Next to the samples, one lane is loaded with a size marker. This marker is a mixture of fragments of known sizes. The distance these fragments have moved is compared with the position of the sample bands, to determine the size of the gene fragments in the samples.

All samples were run in 1.5 % agarose gels in TRIS-Acetate-EDTA (TAE) buffer (40mM TRIS-acetate pH 7.8, 2 mM EDTA) with 0.5 µg/l ethidium bromide. Samples were mixed with sample buffer (20 mM Tris-HCl, pH 8.0; 120 mM EDTA; 50 % glycerin; 0.1 % [w/v] brome phenole blue) before applying them to the gel pockets. As size marker the 1 kb DNA ladder (Invitrogen, Carlsbad, USA) was used. After running the gel at 100V in a i-Mupid mini electrophoresis chamber (Helixx, Ontario, Canada) the DNA bands were viewed under UV light and documented with the Syngene G:Box (Syngene, Cambridge, Great Britain).

3.15 Gel elution

Visible bands under the UV light can be cut out from the gel. The DNA is purified using the NucleoSpin® Gel and PCR Clean-up kit by Macherey-Nagel (Macherey-Nagel, 2014).
The agarose gel piece is put into binding buffer and dissolved by heat. The buffer contains chaotropic salts, which lead to a binding of the DNA to a filter column. Contaminations are removed by an ethanolic wash buffer. The DNA is eluted under low salt conditions.

3.16 Cloning vectors for DNA amplification

In order to amplify a PCR product, it was cloned into a vector. The choice of vector depended on the size and properties of the PCR product.

3.16.1: TOPO vector

The most used vector was TOPO (Invitrogen, Carlsbad, USA), which includes a resistance against Ampicillin and Kanamycin for selection of clones. The cloning site is inside a *lacZα* site, allowing for selection of colonies by blue-white-screening. Blue-white screening is based on the function of β-galactosidase in the lacZ operon. Cells with the basic vector contain a full β-galactosidase gene, and can express the enzyme which converts X-gal to 5-bromo-4-chloro-indoxyl, which spontaneously convert 5,5'-dibromo-4,4'-dichloro-indigo, which gives the

colony a blue colour. If an insert was cloned into the vector, it interrupts the lacZ operon. No β-galactosidase can be formed, the cells do not convert X-gal, and the colonies stay white.

An EcoRI cleaving site and other endonuclease cleaving sites allow for enzymatic excision of the cloned sequence.

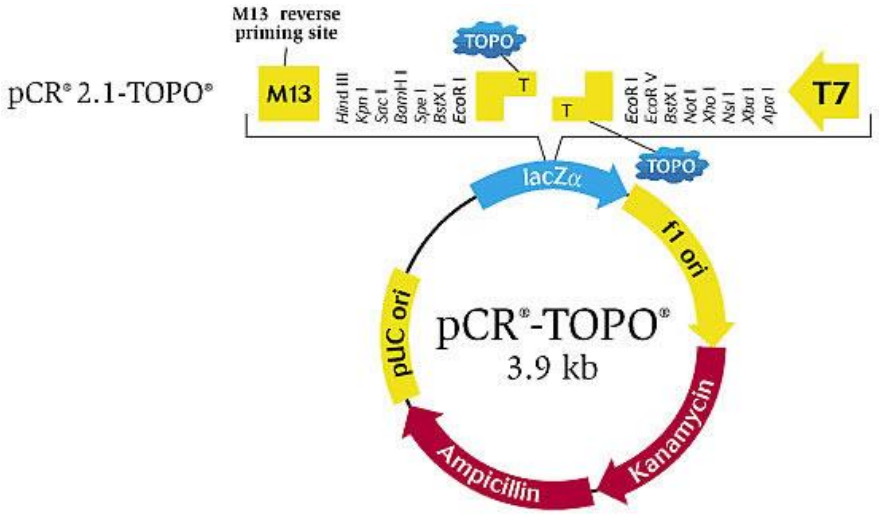

Figure 12: map of TOPO cloning vector (ThermoFisher).

A mixture of 0.5 µl salt solution, 0.5 µl vector and 2 µl gel extract were incubated at room temperature for 30 min. Then, it was added to 50 µl competent TOP10 cells on ice and incubated another 30 min on ice. A heat shock of 45 s at 42°C gets the vector into the cells. After adding 200 µl SOC medium the cells were incubated in a shaker at 37°C for 3 h, then plated on LB agar with Kanamycin and incubated at 37°C overnight.

3.16.2: pJet vector

If TOPO cloning was not successful, a blunt end ligation in pJet (Thermo Scientific) was another option.

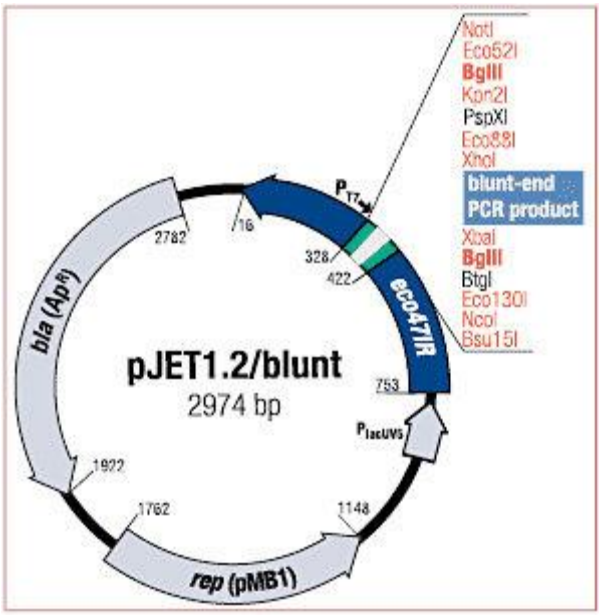

Figure 13: map of PJet vector (ThermoFisher).

To create blunt ends in the PCR product, the following steps apply:

5µl 2x reaction buffer

1µl PCR product from gel elution

0.5µl blunting enzyme

7.5µl H_20

This mixture was incubated at 70 °C for five minutes. Then, each 0.5 µl of vector and T4 ligase were added, and incubated at room temperature for 5-10 minutes. The further steps, starting with addition of 50 µl TOP10 cells, were the same as for TOPO cloning.

3.16.3: pDrive vector

Another alternative was the Qiagen PCR cloning kit, using the pDrive vector.

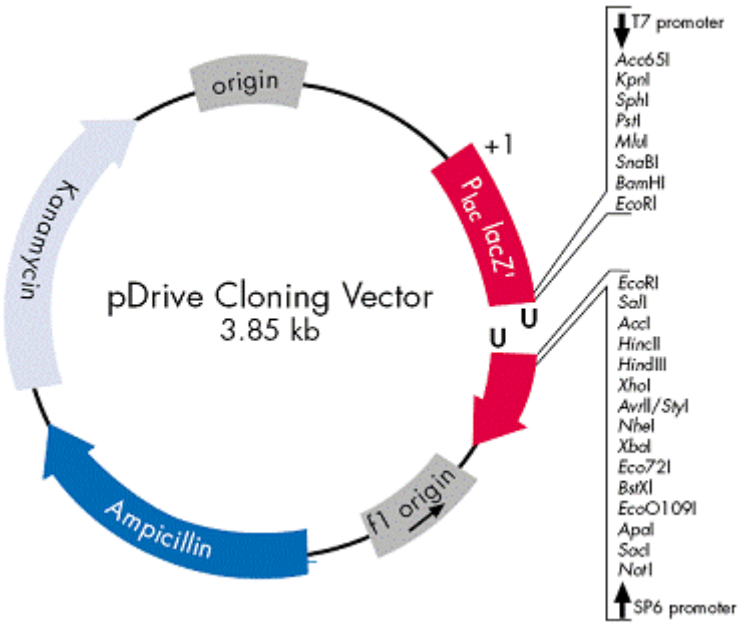

Figure 14: map of pDrive vector; (Qiagen, 2017).

0.5µl pDrive vector
2.5 µl 2x Master Mix
2 µl PCR product from gel elution
were incubated for 20 min to 2h at 4-16 °C. The following steps, starting with TOP10 cells, were the same as for other vectors.
After adding the DNA and vector mix to the frozen cells and incubating on ice for 30 min, the cells were shocked at 42 °C in a water bath for 45 s. Then they were placed on ice again, and 200 µl of SOC medium were added. After 3 h of incubation at 37 °C, the cells were plated on LB-Agar with Ampicillin or Kanamycin.

3.17 Cultivation media

media:	
SOC medium:	2% tryptone (Sigma-Aldrich, St. Louis, USA)
	0.5% yeast extract (Sigma-Aldrich, Munich, Germany)
	0.5% glucose
	10 mM NaCl
	2.55 mM KCl
	21.6 mM $MgCl_2$
	20 mM Mg_2SO_4
LB agar:	32g LB medium (Applichem, Darmstadt, Germany) in 1 l *a.dest.*
LB culture medium:	25 g LB medium (Applichem, Darmstadt, Germany) in 1 l *a.dest.*

Table 4:Composition of cultivation media.

To the LB medium, antibiotics were added for selection of successful transformation. The TOPO vector carries a resistance to Ampicillin (100 mg/l in LB medium) and Kanamycin (50 mg/l in LB medium).

3.18 Microorganisms for cloning and sequencing

For cloning experiments, the *E.coli* strain TOP10 (Invitrogen, Karlsruhe) was used. It has the genotype F− mrcA Δ(mrr − hsdRMS − mcrBC), Φ80lacZΔM15 ΔlacX74 deoR recA1, araD139 Δ (ara − leu)7697 galU galK, rpsL (StrR) endA1nupG.

After transformation, *E.coli* strains were grown on LB agar plates containing Ampicillin or Kanamycin for selection. These plated cultures were grown at 37 °C overnight. Positive clones were selected by colony PCR, and cultivated in liquid LB medium at 37 °C and shaking at 220 rpm overnight.

3.19 Colony PCR

To check if the resulting colonies on the agar plates included the correct DNA fragment, a colony PCR was performed with the respective primers for the cloned fragment.

5 µl goTaq buffer
0.5 µl forward primer
0.5 µl reverse primer
0.5 µl dNTP mix
0.125 µl goTaq
17.375 µl H_20

As template, a small amount from the colony was picked with a sterile toothpick.
The colony PCR uses the same program as the PCR that generated the fragment. The resulting product size was checked by agarose gel electrophoresis. A wrong product size usually meant the gene fragment was incomplete, or primers bound to a random sequence of the genome.

3.20 Overnight cultures and purification of DNA for sequencing

If the cloned fragment had the correct size, the colony was used for an overnight culture. Again, a sample was taken with a toothpick, and placed into a tube with 3 µl liquid LB medium with Kanamycin. After an overnight incubation in a shaker at 37 °C, the DNA was extracted from this culture.
For DNA extraction, the NucleoSpin® Plasmid kit by Macherey-Nagel was used. The overnight cultures are harvested by centrifugation, and resolved in buffer from the kit. A second buffer was added to start alkaline lysis of the cell structures. Afterwards, an acidic buffer stopped the reaction, and allows for precipitation of cell wall debris, proteins and liquids. The precipitate was removed by centrifugation. The sample was applied to a filter column that binds DNA. Ethanolic washing buffer removed remaining salts and metabolites. The DNA was eluted using a low ionic strength Tris-HCl buffer. (Macherey-Nagel, 2015)

3.21 EcoRI digestion

Before being sequenced, the extracted DNA was checked for the correct insert again. The EcoRI cleaving site in the vectors was used to excise the fragment.

7.5 µl H_20
1 µl EcoRI buffer
1 µl DNA

This mixture was incubated at 37 °C for 1 h. Then it was analyzed using agarose gel electrophoresis. The correct result presented two bands: the vector and the insert. If this was the case, the sample was sent to Eurofins DNA (Eurofins) for sequencing.

4. Results

4.1 Mapping of QTL_{215} and $QTL_{991\text{-}996}$

The foundation of my work is the QTL map of major maize terpene biosynthesis loci (as determined by Dr Annette Richter, see Fig. 7). For my work, QTL_{215} on chromosome 2, and $QTL_{991\text{-}996}$ on chromosome 9 were chosen. QTL_{215} is associated with the regulation of bergamotene, farnesene, TMTT, DMNT, nerolidol and linalool emission.

$QTL_{991\text{-}996}$ is associated with the emission of bergamotene, farnesene, TMTT and nerolidol. This region showed QTLs associated with seven volatile traits. Since these QTLs were situated very closely to each other, the resolution of the mapping was not sufficient for a separation of these loci. It is also very likely that they correspond to one single locus. For the purposes of this thesis, this area was considered as one QTL.

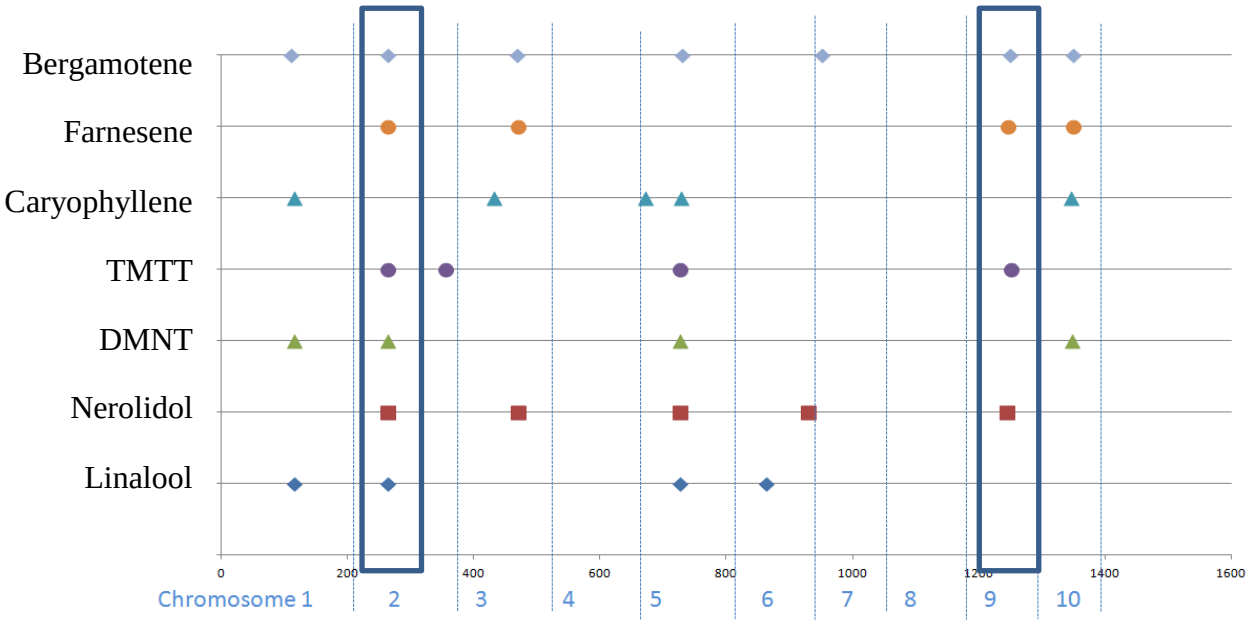

Figure 7: Position of QTLs chosen for this work, listed by cumulative chromosome position. Each QTL is involved in the regulation of several terpenes. QTL_{215} on chromosome 2 is involved in the regulation of bergamotene, farnesene, TMTT, DMNT, nerolidol and linalool. $QTL_{991\text{-}996}$ on chromosome 9 is relevant for the regulation of bergamotene, farnesene, TMTT and nerolidol. Figure by Annett Richter (modified).

Both of these QTL do not contain any known terpene synthase genes, so we assume there are regulatory factors in these regions. These can be regulatory genes working on the level of terpene synthase expression, or on any step in the production of precursors for terpenes. The aim of this work is to identify such regulatory factors.

For each QTL, the statistical calculation of QTL position was run with several sets of lines from the NAM population. Each of these sets included or excluded different parent lines. The goal was to find populations that showed this QTL as statistically relevant, and populations that didn't. If taking a parent line out of the population significantly lowers the statistical relevance of a QTL, we assume this line is highly relevant for the QTL. Often, it was one or few lines that determined the statistical value of the QTL, and taking them out of the population made the QTL disappear completely.
When a parent line is relevant for a QTL, that means there is a genetic difference between this line and the common parent B73, which is related to the phenotypic difference. This could be a gene that is inactive in one of the lines, or has an altered functionality.

4.2 Genome wide association mapping of QTL_{215} as a method to narrow down the QTL region

Genome Wide Association Mapping (GWAS) is an additional calculation of the QTL position for each trait with a set of higher density markers. There were 1.6 million markers with known position on the 26 parent lines of NAM generated (Yu, et al., 2008). These markers are projected onto the population of 5000 recombinant inbred lines, using the known genome parts of the RILs.

The NAM calculations indicate a genetic region for each QTL. In order to narrow this region down, and thus simplify the identification of candidate gene, the GWAS method was used.

Each SNP was calculated 100 times, and the resulting number of iterations each SNP came up states its significance for the trait. The following image illustrates the SNPs of QTL_{215} on chromosome 2:

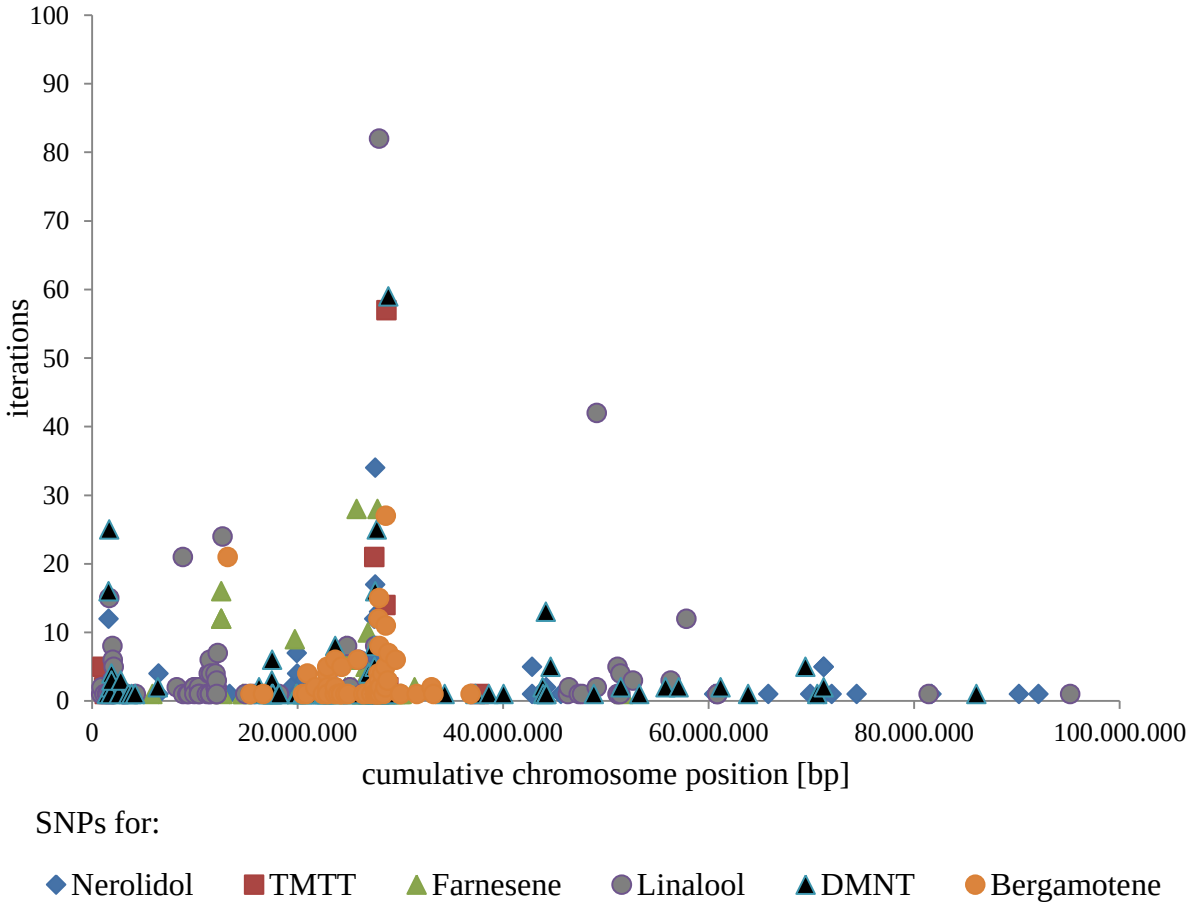

Figure 8: GWAS of QTL_{215}, QTL position calculated for each single terpene, and shown by number of hits for each position. X-Axis depicts the cumulative chromosome position in basepairs, Y-Axis shows the number of iterations calculated for each position and the terpene it relates to, indicated by colour of the spot. A higher number of iterations hints at a higher influence of this SNP position on the phenotypic trait.

QTL_{215} is associated with the regulation of several terpene traits. These traits result in significant associations in the SNP iteration counts, the highest one being linalool with 82 iterations. Other SNPs with high significance are associated with the emission of DMNT and TMTT. The position of these SNPs with high iteration numbers was used to locate the QTL more precisely, thus narrowing down the number of potential candidate genes within the genetic region.

GWAS calculations and graphs were done in cooperation with Annett Richter.

4.3 Three most relevant parent lines for QTL_{215}

Each QTL calculation was performed with different sets of the parent lines included and excluded from the database. This way, the relevance of a line for a QTL can be determined. A QTL is associated to one or more genes that influence a phenotype. If a line shares the same allele for this gene with B73, with an identical gene sequence and function of the gene product, leaving this line out of the dataset for this QTL does not cause a different result of the statistical calculations. A line with a different allele shows different sequences in the gene(s). This leads to a difference in regulatory activity, and a different phenotype. In this case, the maize line does have an influence on the calculations - if this line is removed from the dataset, the variety of alleles in the dataset is reduced and the QTL loses part of its statistical significance.

For QTL_{215}, in addition to the common parent line B73, the three lines CML333, Oh7b and Tzi8 had the highest influence on the calculation of QTL_{215}. When a dataset without these two lines was used, the QTL lost most of its significance. These maize lines also showed different terpene emission patterns, as presented in the following image.

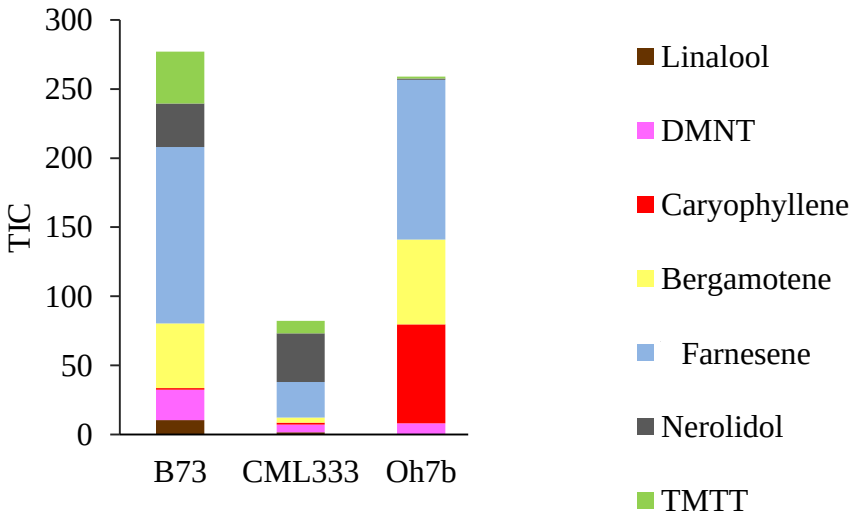

Figure 9: Terpene emission of the maize lines B73, CML333, Oh7b and Tzi8 after herbivory, measured by SPME and gas chromatography.

CML333 emitted a much lower total amount of terpenes than B73. The emission of linalool, DMNT, E-β-caryophyllene and α-bergamotene was very low. β-Farnesene and TMTT emission were also reduced in comparison to B73. Only the amount of nerolidol was slightly increased by contrast to B73.

In Oh7b, the overall amount of terpenes was similar to B73. Linalool was not produced at all in this line, DMNT amounts were lower than in B73. The E-β-caryophyllene content was highly increased in Oh7b. The emission of α-bergamotene and β-farnesene were similar to B73. Nerolidol and TMTT emission were strongly reduced.

Tzi8 showed a small reduction in overall terpene emission compared with B73. Linalool, DMNT, E-β-caryophyllene, bergamotene and β-farnesene contents in Tzi8 were similar to B73. Nerolidol and TMTT production were reduced.

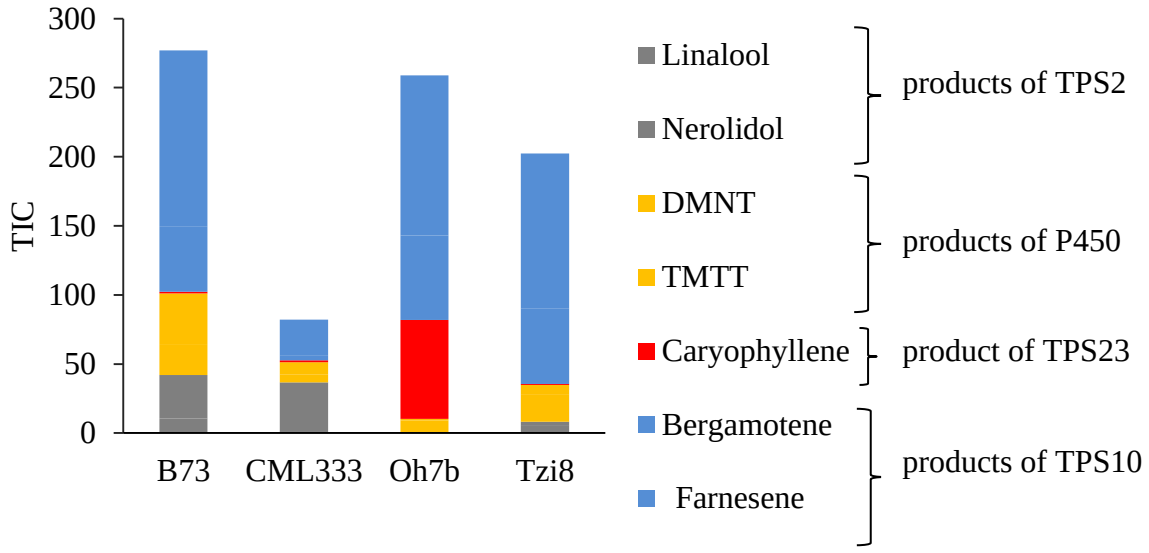

Figure 10: Terpene emission of the maize lines B73, CML333, Oh7b and Tzi8 after herbivory, sorted by producing enzymes.

The variation of volatile emission between these maize lines was not limited to the products of one terpene synthase. Products of TPS2, P450, TPS23 and TPS10 were influenced by this QTL. This means the QTL contains either one or more regulatory factors that regulate more than one enzyme.

The QTL is not defined as a point, but a statistically relevant genomic region, thus allowing for a range for potentially important candidate genes. It is even possible that there is more than one important regulatory factor in a QTL region. If genes are closer to each other than the marker distribution in this area, they appear as one QTL. For QTL with several statistical important parent lines, it is also possible that the different lines refer to different regulatory genes within the same QTL region.
Candidate genes were picked by looking for regulatory genes or genes that resemble regulatory genes from other plants, within close proximity to the QTL. Homologous genes are ones which share a common ancestor gene, thus having a related sequence, and often sharing the structure and activity of the gene product. Such related genes can for example be found in rice (*Oryza sativa*, a member of the Zea family, thus closely related to *Zea mays*), or *Arabidopsis thaliana* (a commonly used model organism which is completely sequenced) and other member of the Brassicaceae.

In order to cause a difference in regulation between B73 and the other parent lines, there is most likely a sequence difference among the alleles of the candidate gene. To investigate this, the genetic sequences were compared in the HapMap database. Since this database can have gaps or inconsistent genetic reads, the data were confirmed by own sequencing.
Genes that play a role in the regulation of terpene production, were also expected to be regulated in their expression level by herbivory. Not all steps of the biosynthetic pathway are necessarily changed in their expression level. Genes can also be constantly expressed, but the product can be activated or deactivated in response to herbivory. For example, one step in a regulatory chain can be a kinase that phosphorylates a protein which is always present - thus raising or lowering its activity. In this case, the gene might not display an altered expression level, but the signal transduction is regulated post-transcriptionally. Differences in the gene sequence can lead to differences in phosphorylation sites, or active sites. So it is necessary for to investigate for sequence differences of the gene in the parent lines, even if it does not indicate insect-induced transcription changes.

Still, a gene expression being herbivore-regulated was considered a strong hint that it is part of a signal chain. Therefore, the candidate genes were tested by qRT-PCR using cDNA from B73 control and induced plants. All samples labeled as “induced” were taken after 24 hours of incubation of a leaf taken from 2-3 week old seedlings in water with an indanone elicitor solution mimicking an herbivore feeding on the plant.

For qRT-PCR, all expression data was referenced to a housekeeping gene. Housekeeping genes are mostly genes of the basic metabolism. Their suitability as references comes from their constant expression level in all plants. All qRT-PCR experiments were done with three biological replicates.
It is possible that the induction of maize leaves or young plants is not successful, if the plants were stressed by drought, too much water, or parasites in the soil. Before cDNA was used for qRT-PCR, each batch of plant material was tested with S33 and S35 primers for *tps2*. Since *tps2* is strongly induced by herbivory or indanone treatment, its expression level demonstrates if the indanone induction was successful. Thus, false negative results for the induction of candidate genes were avoided.

4.4 Identifying candidate genes for QTL_{215}s

A QTL does not point out a specific gene, but a region on a chromosome. Depending on the local density of markers, the size of this region varies. This region may contain up to several hundred genes, for of which the function is unknown. The first step in this work was to look for genes that had sequence similarities to regulatory genes in other plants, like rice and Arabidopsis, using the sequence comparison data given in maizegdb.org database. Examples for such regulatory genes in other plants are kinases, transcription factors, myb factors, penta- and tetratricopeptides.

As a next step, the sequences of these genes between B73 and the most significant parent lines for the respective QTL were compared. For this step the HapMap database (HapMap-Project) was used. The HapMap database contains a full genome of B73, and a large number of short reads for other parent lines of the NAM population.
The HapMap database contains short reads for all NAM parent lines, which were assembled using the genomic sequence of B73. There might be gaps of missing data between reads, or several reads for one sequence with different results. Thus, all data from HapMap has to be confirmed by sequencing of the relevant genes from the maize lines. If reads for one sequence part were contradictory, they were not considered to be reliable results and were not included.

If there is a genetic difference between the NAM parent lines and B73 lines, this suggests that the gene product could be altered in its function. If a regulatory gene is inactive in one maize line and active in another, and this difference leads to a different phenotype, the region near this gene will provide a QTL for this phenotype.

If a gene plays a regulatory role, its expression is often upregulated during the reaction (Tsuda, et al., 2014) (Xiong, et al., 2001). The time point of the increased expression can give a hint about its position in the regulatory genes – early genes are upregulated fast, while genes that affect a later step in regulation may have a raised expression level after several hours.
Downregulation after herbivory is also possible as a part of the signal chain. Potentially, the gene product has a derepressing role for its target in the signal transduction chain, and its absence allows for said target factors to be activated. Alternatively, it can be a late step of the cascade, downregulating one of the earlier steps after they have fulfilled their role.
Not all parts of a signal transduction chain have to show changed expression levels after induction of the signal cascade. Part of the signal transduction can be the modification (for example methylation, phosphorylation or glycosylation) of a protein, the delivery of a protein to a different cell compartment, or the assembly of a complex (Xiong, et al., 2002).
The expression level of candidate genes was tested by quantitative real-time polymerase chain reaction (qRT-PCR), using cDNA derived from mRNA from induced leaves and non-induced control leaf samples. This cDNA directly represents the levels of mRNA, and thus the levels of transcript in the tissue.

To confirm the significance of a gene for regulatory processes, UniformMu (UFMu) maize lines were investigated. These are lines that contain short transposon inserts, called Mu transposons. Such a transposon inserted into a gene can disrupt its sequence and the structure of the gene product, thus reducing its activity or rendering it inactive. In these lines, one can measure the terpene emission, and the expression level of genes expected to be regulated by the candidate gene, such as terpene synthases.

QTL_{215} is a region around the position of 27.8 Mbp on chromosome 2. This QTL influences the herbivore-induced emission of bergamotene, farnesene, TMTT, DMNT, nerolidol and linalool. There are no known enzymes of terpene biosynthesis within this genetic region. Therefore, the QTL was expected to contain one or more regulatory factors for terpene synthesis. The QTL has an influence on the products of different enzymes – TPS10 producing

bergamotene and farnesene, P450 enzymes producing TMTT and DMNT, and TPS2 producing nerolidol and linalool. This suggests that there is more than one regulatory factor, or a general factor that influences several products.

Within this region, genes that resemble regulatory genes were searched using the maizeGDB database. Such regulatory genes can be similar to genes in rice or Arabidopsis. Closely related genes from other plants can be a hint to a similar function in maize. Relevant similar genes are for example kinases, transcription factors or proteins that are known to have the ability to bind to genetic sequences.

The following candidate genes on QTL_{215} were investigated for this project:

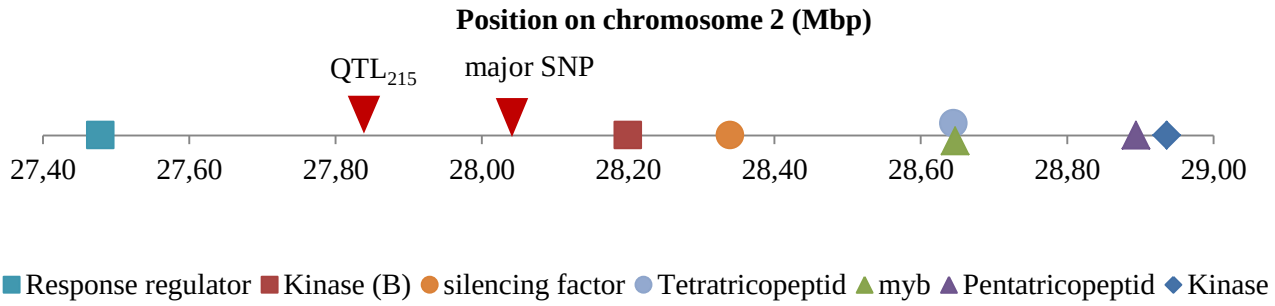

Figure 15: Position of candidate genes on chromosome 2, relative to QTL_{215} marker and the highest-relevance SNP. All candidate genes for QTL_{215} are found in the region between 27,4 and 29 Mbp on chromosome 9. The position of the highest relevance SNP for this QTL is at 28 Mbp.

The following table displays the position and length of the genes, and their suggested functionality from the database, based on similarity to other genes in rice or Arabidopsis. Detailed descriptions of the relatedness to known genes are described in the chapter for each candidate gene.

Name	Position on chromosome 2	Description from database
GRMZM2G035688	27,478,703-27,483,682	RR3 Response regulator
GRMZM2G145051	28,199,812-28,202,949	kinase (named KinaseB here for differentiation)
GRMZM2G340601	28,339,363-28,342,896	silencing factor
LOC100382755 (GRMZM2G702991)	28,644,600-28,645,965	tetratricopeptide
GRMZM5G869984	28,646,632-28,652,080	myb-like with DNA-binding domain
Zm.25643 (GRMZM2G077420)	28,893,901-28,897,200	Pentatricopeptide
LOC100383522 (GRMZM2G032694)	28,936,197-28,940,916	Kinase

Table 5: Candidate genes for QTL_{215}.

All candidate genes were tested for differential regulation after herbivory. A change in expression level is often related to a role in a signal cascade. Expression levels were measured by qRT-PCR, using cDNA as a template, which represent the levels of mRNA in the plant leaves.

4.4.1 Expression of RR3 Response regulator (GRMZM2G035688) after indanone induction, and sequence differences in different maize lines

The RR3 response regulator gene is located in position 27,478,703-27,483,682 bp on chromosome 2 and encoded on the forward strand. It is similar to the RR3 response regulator from *Arabidopsis thaliana* (Urao, et al., 1998). Given this similarity to a known regulatory gene, it was chosen as a candidate gene.

As a first experiment, the expression level of the RR3 Response regulator gene in B73 induced and non-induced plant leaves was measured by qRT-PCR. For a gene that is involved in a signal transduction chain, a change in expression level after activation of the signal chain was often demonstrated (Tsuda, et al., 2014).

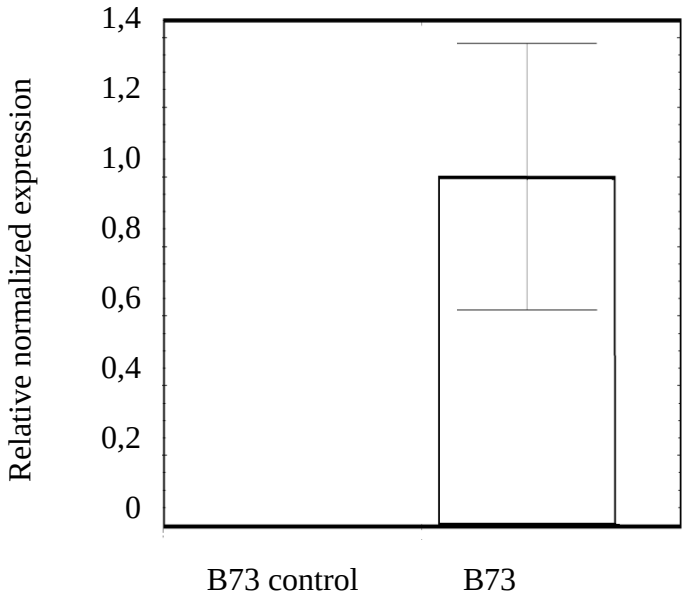

Figure 16: Expression level of the RR3 gene in B73 control and induced plants. The amount of transcript for the RR3 gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

The qRT-PCR is a sensitive method that detects even low concentrations of cDNA for a gene. In indanone-induced leaves of B73, it successfully shows the presence of transcript of the RR3 gene.

In control plants, no transcript of the RR3 gene was detectable. The gene was either not transcribed in non-induced plant leaves, or only at a very low level. PCR experiments using goTaq, or the higher efficient polymerase AdvantageTaq instead, also did not lead to a different result. Agarose gels of the PCR experiments indicated no clear band of the size of the expected gene product.

In genomic DNA of B73, a product was detectable. It was cloned into *E.coli* and sequenced. The resulting sequence matched the database sequence for the RR3 gene, indicating that the primers do bind to the correct target sequence. But when cDNA was used as a template for PCR with the RR3 primers, there was no visible product band in the agarose gel. This means the expression in B73 is too low to result in a visible band in standard PCR and gel electrophoresis.

The sequences of the RR3 gene in B73, Oh7b, CML333 and Tzi8 were compared in HapMap reads. These maize lines exhibit differences in terpene emission, and contribute to the

significance of the QTL. If the presence of a maize line in the QTL calculations changes the significance of the QTL, this means that these maize lines show a link between the investigated phenotype and their genotype in this region. For this work, that means the calculations suggest differences in terpene emission in a context to this QTL. In order to cause a different phenotype, there needs to be a difference in the genetic sequences of the regulatory genes, or in the regulation of these genes by epigenetic variations. These sequence differences can change the activity and specificity of the regulatory factor, thus leading to a different regulation result.

The complete B73 sequence of the RR3 gene was compared to the available set of HapMap reads for this gene in Oh7b, CML333 and Tzi8.
No differences in the RR3 gene were found between the parent lines. The gene sequence, as far as HapMap coverage was available, was completely identical. The HapMap reads did not cover the full sequence. Due to difficulties in amplifying the RR3 gene from cDNA, sequencing experiments could not be performed. The level of transcript was too low to reach a sufficient amount of DNA for sequencing.

4.4.2 Expression of KinaseB (GRMZM2G145051) after indanone induction, and sequence differences in different maize lines

GRMZM2G145051 is a gene located in position 28,199,812-28,202,949 bp on the forward strand of chromosome 2. It is annotated as a kinase. In order to differentiate the several kinase candidate genes at the position of QTL_{215}, this one was labelled KinaseB.

Again, the first experiment was a detection of transcript levels of the candidate gene in induced and control leaves of B73.

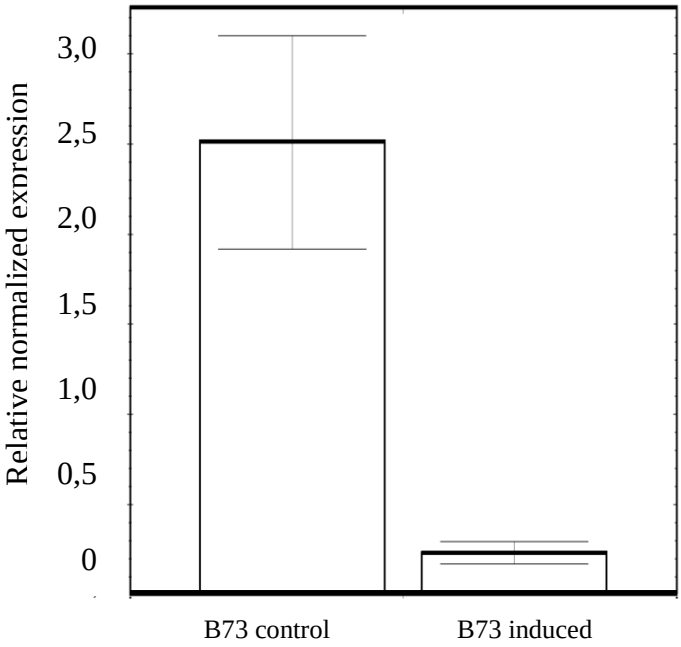

Figure 17: Expression level of the KinaseB gene in B73 control and induced plants. The amount of transcript for the KinaseB gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

Kinase B gene was expressed at a lower level 24 hours after herbivory. The biological variability in non-induced control plants is rather high, while the lowered expression in indanone-induced samples is more constant.

As a next step, HapMap datasets were searched for nucleotide differences between the parent lines. Sequences without significant differences to B73, or missing reads in the non-B73 parent lines are not shown.

Between B73 and Tzi8, no relevant sequence differences were found. Since there might be several regulatory factors for terpene emission within the QTL region, it is possible that some of them are only relevant for one or two lines, and not for other parent lines.

The comparison of HapMap sequences of the gene in the lines B73, CML333 and Oh7b demonstrated the following differences:

```
                                                                1490
B73      TCTTTCAGAT GGATGCCATG ATGTCGATGA GTGCAAGAAG AACAGTCCAT GC-----AGT
CML333   TCCTTCTCAT CATACTGCAG ATGTCGATGN GTGCAAGAAG AACAGTCCAT GCCCTTCAGT
Oh7b     TCCTTCTCAT CATACTGCAG ATGTCGATGA GTGCAAGAAG AACAGTCCAT GCCCTTCAGT
```

```
B73      AGGAGGTGTT TGCCACAACA CGGTC
CML333   AGGAGGTGTA TGCAACACCA CGGTC
Oh7b     AGGAGGTGTA TGCCACAACA CGGT
```

Figure 18: Sequence comparison between B73, CML333 and OH7b for a part of the Kinase B gene. Both CML333 and Oh7b carry a five base pair insert at position 1492, HapMap data contained one read for each maize line in this position.

This result of a five base pair insert in two parent lines that are significant for the QTL suggests that the protein from this candidate gene has an altered activity in both lines. An insert of five base pairs shifts the three-base-pair reading frame of mRNA to protein translation. Such a reading frame shift completely changes the translation of the rest of the sequence beyond this point. It often leads to a complete change in structure, or a stop codon. In both cases, usually the function of the protein is lost.

```
B73      GTGACCCTGA TACCACATTA ATAACAGGTA ACACGCTAGC TGAGATCAGC AGCTGCATGC
Oh7b     TTGACCCTCA TACAACATTA AAAACAGGTA ACACGCTAGC TGAGATCAGC AGCTGCATGC

                                                1671
B73      TACCACTGTT TATTTCTCCC TGTATATATA TATATATATA TATATATATA TATATATATA
Oh7b     TACCACTGTT TATTTCTCCC TGTATATATA TATA------ ---------- ----------

B73      TATATATATA AAAGAAGGAA AGCGCGTGCT CTTATCTGAC GAT
Oh7b     ---------- --AGAAGGAA AGCTCGTGCT CTTATCTGAC GAT
```

Figure 19: Sequence comparison between B73 and OH7b for a part of the Kinase B gene. Oh7b carries a 38 base pair deletion, one read for this sequence part found in HapMap.

A deletion of 38 base pairs in the translated region of a gene creates a big structural difference in the resulting protein. This leads to a frameshift in the 3-base-pair reading frame. Such a large deletion can also delete entire structurally important parts of the resulting protein.

All sequence differences shown here are inside exons of the KinaseB gene. The exons get translated, thus a changed exon will result in a change of the gene product. Shifts of the reading frame will lead to a completely different amino acid sequence, or a stop codon.

4.4.3 Expression of Silencing factor (GRMZM2G340601) after indanone induction, and sequence differences in different maize lines

GRMZM2G340601 is a gene on chromosome 2, forward strand, position 28,339,363 - 28,342,896 bp. It bears similarity to known silencing factors in rice and Arabidopsis (MaizeGDB, 2014).

Transcript levels of the candidate gene in induced and control leaves of B73 were measured using qRT-PCR.

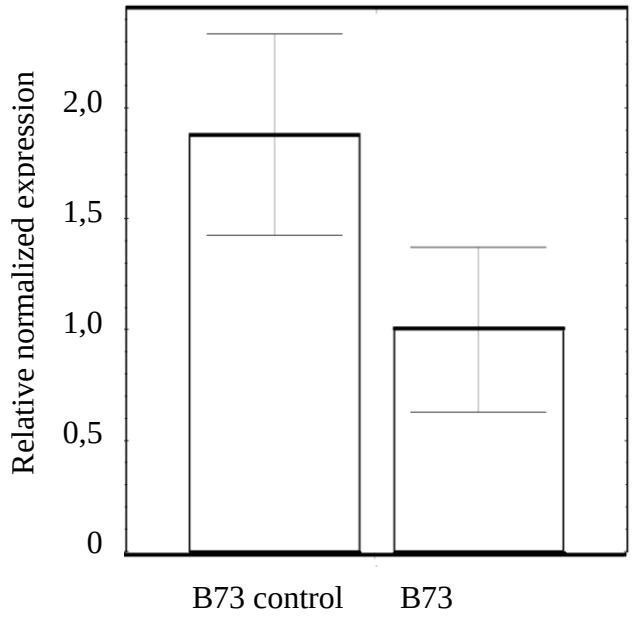

Figure 20: Expression level of the silencing factor gene in B73 control and induced plants. The amount of transcript for the silencing factor gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

The gene for the silencing factor was expressed at a slightly lower level after herbivory when compared to control plants. The biological variability between the replicates - plants of the

same line, grown and induced under the same conditions - in this gene was high, while using the same B73 cDNA as for other qRT-PCRs shown here.

In order to find structural differences that could alter gene function, the gene was then sequenced in Oh7b, CML333 and Tzi8. Oh7b and CML333 did not indicate relevant differences that would alter the function of the protein.
The results for a comparison between B73 and Tzi8 are shown below:

```
        160
B73     GTTGATGAGG TTGGTGTTCC CTTGCCCATG GATTTCGATG ACAGGGAGAT CTTCTTCTCT
Tzi8    GTTGATGAGG TTGGTGTTCC GTTGCCCATG GATTTTGATG ACAGGGAGAT TTTCTTTTCT

        220
B73     ATGGGCCGTC AAGATCTCAA TGTAGAATGC ATTGATGGAT TGGCTGCAGT CCTGAGTGTC
Tzi8    CTGGGTCGTC AAGATCTCAA TGTAGTATGT ATTGATGGAT TGGCTGCAGT CCTGAGTGCC

        280
B73     CAGCACTACG AGAAGTTCAA GGGTGATACC AGATACAACT TGTGGCAACC TTATTTTTGC
Tzi8    CAGCACTACG AGAAGTTCAA GAGTGATACT AGATACAGCT TGTGGCAACC TTATTTTTGC

        340
B73     TGTCGGCAGA TTGATAATGG TGACATCAAA CCTTTTGATG T--------- ----------
Tzi8    TATCGGCAGA TTGATAATGG TGAAGCCAAG CCTTTTGATG TCACACAGCT ACAAGGCTAC

        400
B73     ---------- ---------- ---------- ---------- ---------- ---GCGTTTC
Tzi8    TGGAGCCAGG AAGTACTTAG GACCATGTTC AATGCAACTT CCTGTCTGAA GGTGCGTTTC

        460
B73     AAGGTCCCTA AATCCAGACC CAGTTCAGAT GGAGGGCTGA AGAGGAAGCG TGATGCATTC
Tzi8    AAGGTCCCTA AATCTGGACC CAATTCAGAT GGAGGGCTGA AGAGGAAGCG TGATGCATTC

        520
B73     AACGATGATG CTAATCCACA GAAGTTCATT CCTTCTGGTG CTTCTACTTC TAGGTTTTAT
Tzi8    AACAATGATG CTAATCGACA GAAGTTCATT TGTTCTGGTG CTTTTACTTC TAGGTTTTAT

        580                                                     630
B73     GGTGAGAACA TCCAACAATT GTGTCCTGGT TGCCATGTGG AAGTACTCTC TC
Tzi8    GGTGAGAACA TCGAACACTT GTGTCCTGGT TGCCATGTGG AAGTACTCTC TC
```

Figure 21: Sequence comparison of the silencing factor gene in B73 and Tzi8, one read for Tzi8 in this region found in HapMap.

In Tzi8, there is a large insert of 72 base pairs in a translated region (exon) of the gene. Although this does not alter the reading frame, such a large insert in an exon will likely alter the structure of the encoded protein. Deleting 72 base pairs from the gene amounts to deleting

24 amino acids from the resulting protein – which can lead to a changed fold, as all former interactions with this part of the structure are lost. A different protein fold often results in a loss of binding ability for specific binding partners, which makes the protein unable to function.

4.4.4 Expression of Tetratricopeptide (LOC100382755 (GRMZM2G702991) after indanone induction, and sequence differences in different maize lines

LOC100382755 (GRMZM2G702991) is a gene of the tetratricopeptide repeat family. Its position is on the forward strand of chromosome 2 at 28,644,600-28,645,965 bp. Its function is unknown. Proteins of the tetratricopeptide family can have binding sites for DNA, RNA and proteins. Such a binding mechanism can be part of a regulation process. Tetratricopeptides can be involved in regulation mechanisms on DNA, RNA and protein level. This made them interesting candidate genes for this work. An involvement of tetratricopeptides in the regulation of terpene production has not been proven before.

Transcript levels of the tetratricopeptide gene in induced and control leaves of B73 were quantified using qRT-PCR, as shown in Figure 23. The tetratricopeptide gene is expressed at an increased transcript level after induction.

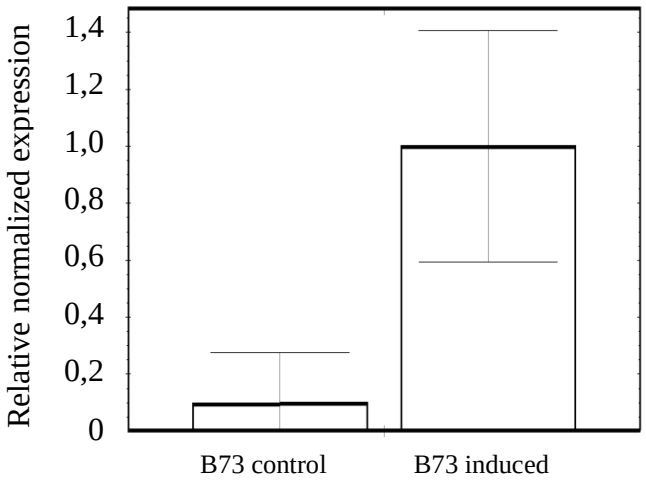

Figure 22: Expression level of the of tetratricopeptide gene in B73 control and induced plants. The amount of transcript for the tetratricopeptide gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates

were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

Next, the sequences of the tetratricopeptide gene in the parent lines relevant for QTL_{215} were compared.
In HapMap sequence reads, the following differences between B73 and Oh7b were found:

```
                      300
B73   TCGGCAAGGG CGAGG---CC GTACTCGCCC CGCTGGCGTC GCCACCCCGC TCCCAGGGAC
Oh7b  TCGGCAGGGG CCAGGGCGCC CTACTCGCCC CGCTGCCGCC TCCACCCCGC TCCCAGGAAC

B73   AGGAAGGGGG TCGGCAGCCT CC
Oh7b  AGGAAGGGGT TCGGCAGCCG CC
```

Figure 23: Oh7b shows a 3 base pair insertion in position 300, confirmed in two reads.

There is a 3 base pair insert, which does not shift the reading frame for the rest of the sequence. It adds one amino acid to the protein sequence. Adding one amino acid can interrupt a structure like an α-helix or a β-fold in a protein. This part of the sequence is an exon.

```
         898       906                923                                952
B73   AGAATCCAGA TCCAGTACAC CACGCATG-- CATGAACATT CCTAGACTCG ATCGAGGG-C
Oh7b  AGA-TCCAGT T---GTACAC CACACATGTG CATGAACATT CCTAGACTCG ATCGAGGGGC

           960                                990
B73   TCCCCGTCCT TCCTAACGTT GTTGGTTTGT CACT---CTT GCAGCCACGT CCTCGCTGCG
Oh7b  TCTCC----- TCCTAATCGT GTTGGTTTGT CACTCATCTT GCAGCCACGT CCTCGCTGCG

B73   TACGCCAGCT TCTT
Oh7b  TACGCCAGCT TCTT
```

Figure 24: Oh7b carries a 1 base pair insertion in position 898 (confirmed in two reads), three base pair deletion in position 906 (confirmed in two reads), two base pair insertion in position 923 (confirmed in two reads), one base pair insertion in position 952 (confirmed in two reads), five base pair deletion in position 960 (confirmed in four reads), four base pair insertion in position 990 (confirmed in two reads).

Insertions and deletions of base pair numbers not divisible by 3 can shift the reading frame of the rest of the mRNA. This part of the sequence is, according to the database, an intron, so the

differences are not in the final protein sequence. Differences in intron sequences would probably not change the resulting protein, as the introns get cleaved out during mRNA processing. If the mutation in the intron is within a splicing site, it might disturb the mRNA processing. This would lead to an effect on the structure of the resulting protein (Huang, et al., 2015) (Epstein, et al., 1993).

```
                1138                        1160
B73   CGGCCGTCTG A---TCCTT TCCACTCCTG ACAGGGGACA GCGCAGGCGC GCTATAGCAA
Oh7b  CGGCCTTCTG ACGATCCTT TCCACTCCCG ACAG------ -CGCAGGCGC GCTATAGCAA

B73   AATGCAGTAT AAAATGCATG GCGTATCAAC GGTTATAAAG TATGAAATAT AAACTGAATA
Oh7b  AATGCAGTAT AAAATGCATG ACGTTTCAAC GGTTATAAAG TATGAAATAT AAACTGAATA

B73   ATAAGAATAA AAG
Oh7b  ATAAGAATAA AAG
```

Figure 25: Oh7b has a 3 base pair insertion in position 1138 (confirmed in two reads), and a seven base pair deletion at position 1160 (confirmed in three reads).

This part of the sequence is also an intron, which is not translated to protein sequence, but could have an influence on mRNA processing.

Oh7b exhibited a number of insertions and deletions, compared to the parent line B73. However, most of the differences are in introns, which get cleaved out in mRNA processing before translation. They do not influence the gene product, unless they are in part of the introns that have a regulatory role in mRNA splicing.

In CML333 and Tzi8, all differences to B73 were in contradictory HapMap reads - that means a read indicated a difference, and one or more reads said the sequence in this point is the same as in B73. These reads were assumed to be unreliable.

4.4.5 Expression of myb-like protein with DNA-binding domain (GRMZM5G869984) after indanone induction, and sequence differences in different maize lines

The gene GRMZM5G869984 codes for a myb-like protein with a DNA-binding domain. The gene is on position 28,646,632-28,652,080 on the forward strand of chromosome 2. Proteins with a DNA-binding site are often involved in the regulation of gene expression. In this function, the gene can be part of the regulatory cascade for terpene production.

Transcript levels of the candidate gene in induced and control leaves of B73 were measured using qRT-PCR.

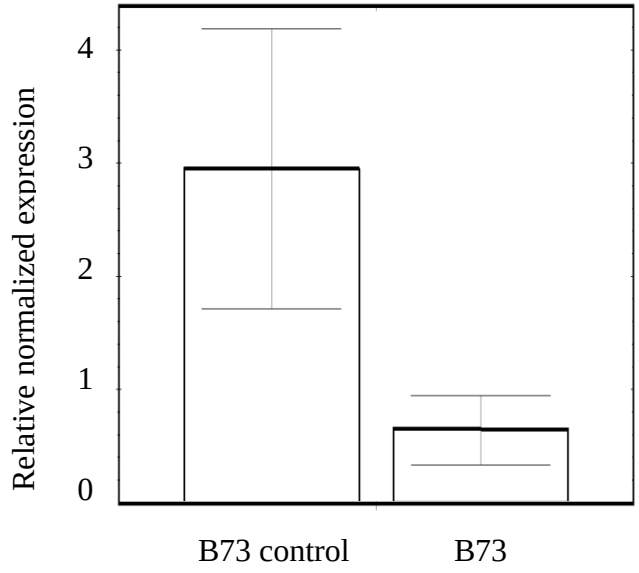

Figure 26: Expression level of the of myb-like gene in B73 control and induced plants. The amount of transcript for the myb-like gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the gene (primers HG5, HG6).

The myb-like candidate gene is expressed at a lower level in induced leaves than in control leaves. The biological variability in the non-induced plant is high than in the induced plants. Variations in non-induced plant samples can occur due to stresses in the metabolism of the plant, such as accidental mechanical damage in plant handling.

The HapMap comparison of sequences of the myb factor resulted in several sequence differences between CML333 and B73.

```
                                                          1303
B73     AGAACCTTCA GGTGGTATTT TCTCTGGACA GTCTCCTGAA GG------CT GTGCAGACCC
CML333  AGAACCTTCA GGTGGTATTT TCTCTGGACT GTCTCCTGAA GGAACCAACT GTGCAGACCC

B73     TGGTGACATT GATTACCGTG TTGA
CML333  TGGTGACATT GATTACCGCG CTGA
```

Figure 27: CML333 shows a six base pair insertion in position 1303 (confirmed in three reads).

This insertion in position 1303 is in an exon of the myb gene, thus leading to a difference in the protein. Adding six base pairs to the gene creates an addition of two amino acids in the protein sequence. Two additional amino acids can interrupt regular secondary structures like α-helices and β-folds. Such changes in protein structure can change the target binding ability of a protein.

```
                                                                3350
B73     TGGTTTAATC TCCCGTTACA TTTTGGTATC TAGTACTTAC AACTTGTCCC TCCTTAAGGC
CML333  TGGTTTAATC TCCCGTTACA TTTTGGTATC TAGTACTTAC AACTTGTCCC T---GAAGGC

B73     TCCATAATAA TAAAACGTG
CML333  TCCAGAATAA TAGAACGTG
```

Figure 28: CML333 carries a three base pair deletion in position 3350 (confirmed in three reads).

The insertion in position 3350 is in an intron of the myb gene, which gets cut out of the mRNA before translation. It does not result in a change in protein sequence and structure. It might lead to a difference in mRNA processing, which would make the gene product inactive.

```
                                                    4733
B73     GATGATGATG CAGTTTCGGA GCCTGAGCCG ATTGGTGGTC GTTATGTACC GGTGCTGGGG
CML333  GATGATGATG CAGTTTCGGA GCCTGAGCCG ATTGGTG--- ---------- ----------

B73     CACGGTGACC GTGAGGACCT GGTGCGTGGT GGTGACCGTG GTGGCCATGA GGAC
CML333  -------ACC GTGAGGACCT GGTGCGAGGT GGTGACCGTG GTGGCCATGA GGAC
```

Figure 29: CML333 with a 30 base pair deletion in position 4733 (confirmed in two reads).

This insertion is in an exon of the myb gene, thus leading to a difference in the protein. 30 base pairs in the gene is equivalent to ten amino acids in the protein. This large deletion leads

to a structural difference in the protein. A loss of ten amino acids can erase binding sites, or make them unavailable for the binding partner of the protein.

All of these inserts and deletions keep the reading frame of three base pairs intact. The translation of downstream sequence parts of the gene is unchanged. So the mutation site is changed locally, but the rest of the protein sequence is unchanged. However, it might fold differently, due to a lack of interaction in the original secondary and tertiary structure.

Oh7b and Tzi8 did not have these insertions and deletions; their sequence was identical to the sequence in B73.

4.4.6 Expression of Pentatricopeptide (Zm.25643 (GRMZM2G077420)) after indanone induction, sequence differences in different maize lines, and the influence of mu transposons in this gene on terpene emission

Zm.25643 (GRMZM2G077420) is a gene of the pentatricopeptide family, on the reverse strand of chromosome 2, position 28,893,901-28,897,200 bp. Like tetratricopeptides, the pentatricopeptides are a family of proteins which can bind DNA, RNA or protein, and are involved in many regulatory processes.

Transcript levels of the pentatricopeptide gene in B73 induced and control leaves were measured using qRT-PCR.

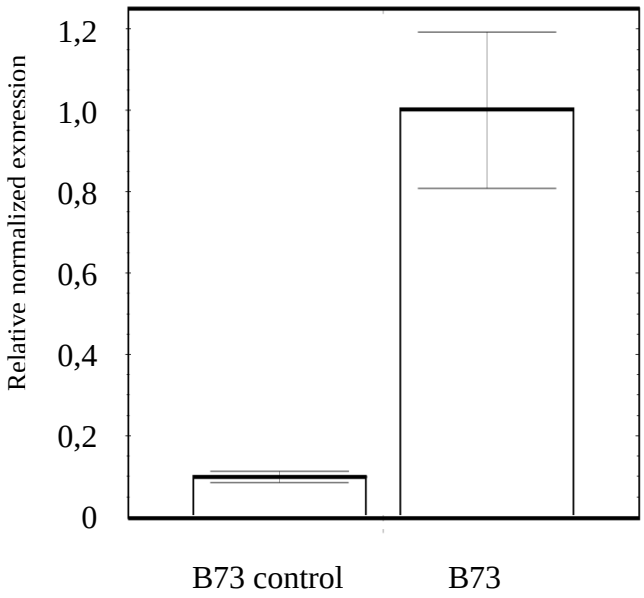

Figure 30: Expression level of the of pentatricopeptide gene in B73 control and induced plants. The amount of transcript for the pentatricopeptide gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

The pentatricopeptide gene is strongly induced in its expression level after indanone induction. The amount of cDNA in the sample, and thus of mRNA in the insect-attacked plant leaf, is about 12 times as high in the induced plant.

The pentatricopeptide gene was sequenced compared in B73, CML333 and Oh7b. No relevant differences were found. In addition to the translated region, an upstream part of 1 kbp was cloned and sequenced to find possible promotor differences. The sequences were found to be identical. A regulatory factor that is identical in all parent lines will have the same functionality in all parent lines, and will not lead to a different phenotype between the maize lines.

UFMU lines for pentatricopeptide

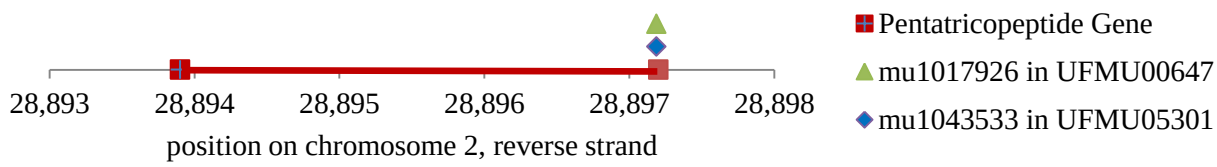

Figure 31: position of Mu transposons in the Tetratricopeptide gene (red)

The pentatricopeptide gene is located on the reverse strand of chromosome 2, position 28,893,901-28,893,200 base pairs. UFMU00647 carries the transposon mu101729 at position 28,897,194-28,897,202 bp right in the beginning of the pentatricopeptide gene first exon. UFMU05301 has the transposon mu1043533 at 28,897,186-28,897,194 bp - so, eight base pairs further downstream in the exon of the pentatricopeptide gene, compared to the other UFMU line. Both transposons are expected to result in a changed gene product, possibly losing its function compared to the wild type without transposon insertions.

For all UFMU lines, the background line is W22. All samples were grown, induced and harvested together, in order to avoid variation in conditions like water availability, temperature variations and age of the plants. Induction and harvesting of W22 and UFMU lines were done following the same protocol as for B73 plants.

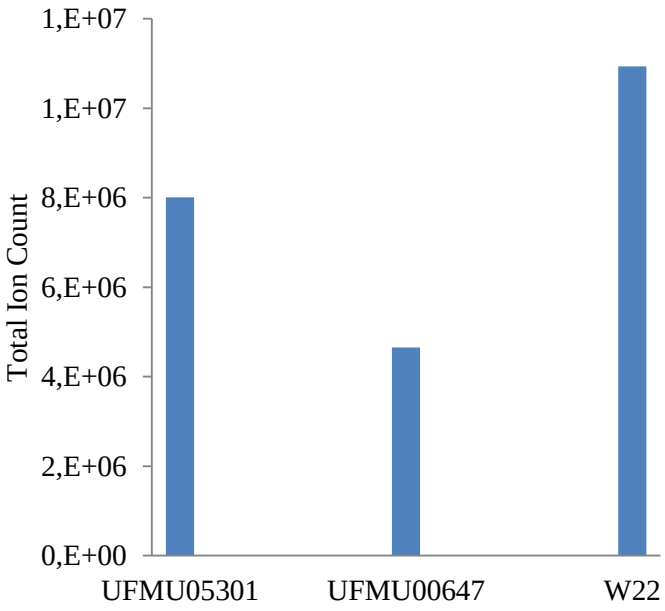

Figure 32: Bergamotene emission of UFMU lines compared to the background line W22.

In Comparison to the background line W22, the emission of bergamotene is slightly lowered in both UFMU lines. The fact that UFMU00647 has the mu transposon closer to the beginning of the gene, leads to a bigger difference to W22 than UFMU05301.

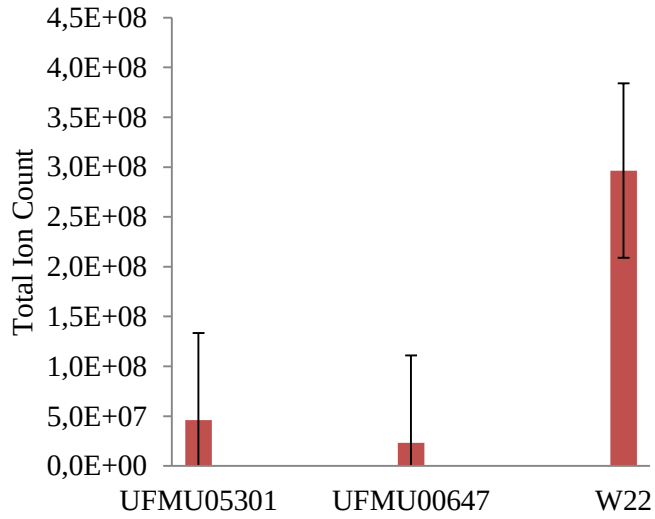

Figure 33: Farnesene emission of UFMU lines compared to the background line W22.

Farnesene is the terpene emitted at the highest amount. Compared to the background line W22, the emission of farnesene is lowered in both UFMU lines, but the biological variation between the UFMU plants within the lines is very high.

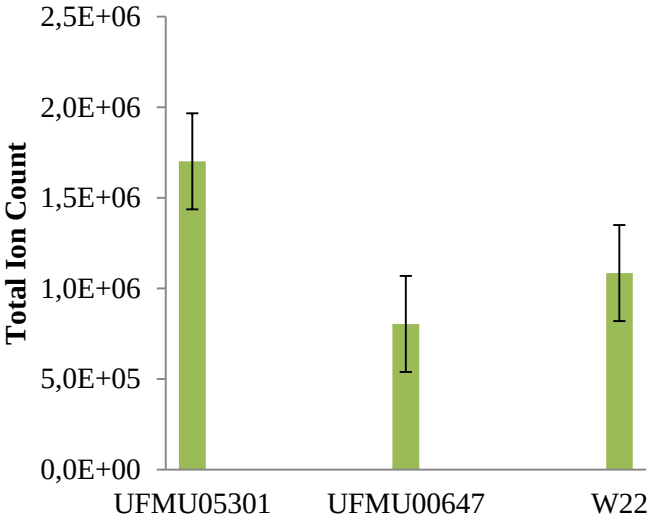

Figure 34: E-β-caryophyllene emission of UFMU lines compared to the background line W22.

Compared to the background line W22, the emission of E-β-caryophyllene slightly lowered in UFMU00647. Again, UFMU00647 emits lower amounts of terpene than UFMU05301, which produces more E-β-caryophyllene than the background line W22.
The total amount of E-β-caryophyllene in all lines is smaller than the amount of bergamotene and farnesene by an order of magnitude.

4.4.7 Expression of LOC100383522 (GRMZM2G032694)) after indanone induction, and sequence differences in different maize lines

LOC100383522 (GRMZM2G032694) is a gene on chromosome 2, position 28,936,197-28,940,916 bp on the reverse strand of chromosome 2.
Transcript levels of the kinase gene in B73 induced and control leaves were quantified using qRT-PCR.

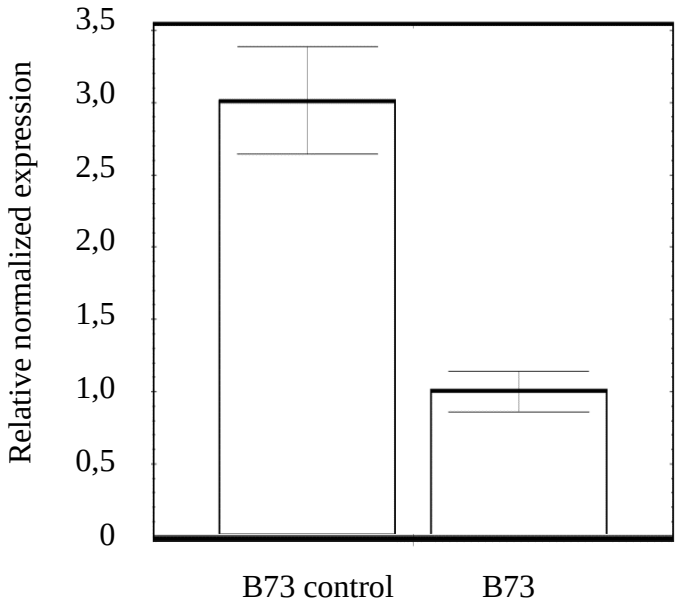

Figure 35: Expression level of the kinase gene in B73 control and induced plants. The amount of transcript for the kinase gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

The expression of the kinase gene decreases significantly within 24 hours after herbivory. The control plant without herbivore-induced reactions has about three times the amount of kinase mRNA of an induced plant.

The kinase gene was sequenced in all three significant lines. Tzi8 had no differences to B73. CML333 exhibited four single-base-pair exchanges, Oh7b shared one of these and was comparable to B73 for the other three. Each of the following sequences was the result for three biological replicates.

```
          580        590
B73       TCACATCACG ACATCCACAG CTTCACTATG AATCAAAACT ATACATGCTT
CML333    TCACATCACG GCATCCACAG CTTCACTATG AATCAAAACT ATACATGCTT
Oh7b      TCACATCACG ACATCCACAG CTTCACTATG AATCAAAACT ATACATGCTT
Tzi8      TCACATCACG ACATCCACAG CTTCACTATG AATCAAAACT ATACATGCTT
```

Figure 36: Comparison between sequence part of the kinase gene in B73, CML333, Oh7b and Tzi8. Single base pair exchange at position 590 in CML333; Oh7b and Tzi8 sequence are identical to B73. This part of the gene is an exon.

```
          1000       1010
B73       TAAATACCCA TCTTGGAGTT GAGCAAAGCA GAAGAGATGA TCTGGAATCT
CML333    TAAATACCCA CCTTGGAGTT GAGCAAAGCA GAAGAGATGA TCTGGAATCT
Oh7b      TAAATACCCA CCTTGGAGTT GAGCAAAGCA GAAGAGATGA TCTGGAATCT
Tzi8      TAAATACCCA TCTTGGAGTT GAGCAAAGCA GAAGAGATGA TCTGGAATCT
```

Figure 37: Comparison between sequence part of the kinase gene in B73, CML333, Oh7b and Tzi8. single base pair exchange at position 1010 in CML333 and Oh7b; Tzi8 sequence is identical to B73. This part of the gene is an exon.

```
                      1202                       1227
B73       CCATCAGAGT TTACTGCATA CTTTCATTAC TGTAGATCAC TACGATTCGG AGGATAAACC
CML333    CCATCAGAGT TTACGGCATA CTTTCATTAC TGTAGATCAT TACGATTCGG AGGATAAACC
Oh7b      CCATCAGAGT TTACTGCATA CTTTCATTAC TGTAGATCAC TACGATTCGG AGGATAAACC
Tzi8      CCATCAGAGT TTACTGCATA CTTTCATTAC TGTAGATCAC TACGATTCGG AGGATAAACC
```

Figure 38: Comparison between sequence part of the kinase gene in B73, CML333, Oh7b and Tzi8. Single base pair exchange at position 1202 and 1227 in CML333; Oh7b and Tzi8 sequences are identical to B73. This part of the gene is an exon.

A single base pair exchange can lead to a different amino acid in the protein sequence stemming from translation of this codon. One changed amino acid can change the protein

structure, if this amino acid was important for interaction in folding the protein. Alternatively, if it was in the active site, the target binding ability can be altered. Due to the degenerated genetic code, it can also be a silent mutation, resulting in the same amino acid as the original gene sequence.

4.5 Identification of Candidate genes for $QTL_{991-996}$

$QTL_{991-996}$ is less precisely defined as QTL_{215} in the NAM calculations. There isn't one sharp peak of iterations for one specific genetic marker. The area demonstrating an influence on terpene regulation spreads over several markers on chromosome 9, position 109.910.040-124.112.712 bp.
For $QTL_{991-996}$, the most significant maize line is Hp301. Without this line in the dataset, the relevance of the QTL is lost. Since there is only one maize line responsible for this QTL, fine mapping could not be done. The fine mapping calculations always require at least three maize lines. Without fine mapping, narrowing down from the large QTL region to a smaller region and specific candidate genes is more difficult.

$QTL_{991-996}$ is involved in the regulation of bergamotene, farnesene, TMTT and nerolidol. The terpenes are products of different enzymes. Bergamotene and farnesene are products of TPS10, TMTT gets produced by a P450 enzyme, and nerolidol is a product of TPS2. In contrast to QTL_{215}, this QTL affects nerolidol as one product of TPS2, but not linalool, the other product of the same terpene synthase. This suggests that there is a regulatory factor for an earlier step of the synthesis of nerolidol, not for the terpene synthase itself. It is also possible that the QTL does have a small influence on the regulation of linalool production, but the calculated value was lower than the threshold that was set for QTL calculations.

The regulatory factor or factory being searched for probably doesn't regulate a single terpene synthase. It is suggested that there are one or more regulatory factors, which influence different steps in terpene synthesis. This means they would act at the level of different enzymes in early reaction steps of the metabolic pathways leading to terpene production, and their regulation. This can include the induced production of an enzyme, or a change in activation of a protein that is present in the cell.

Compared to QTL_{215}, fewer genes with similarities to known regulatory genes in other plants could be identified in the database search.

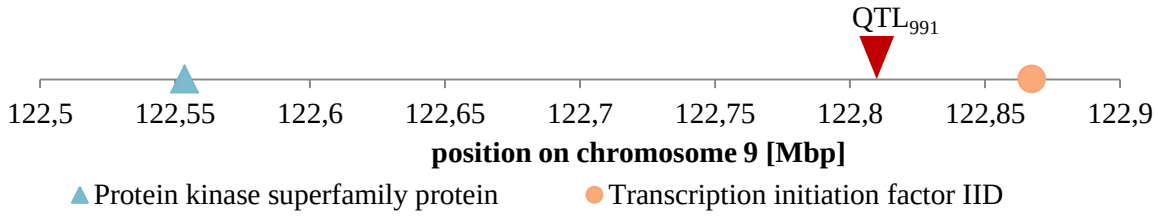

Figure 39: Position of candidate genes on chromosome 9, in relation to the marker for QTL_{991}.

Two candidate genes for this QTL could be identified, their position in relation to the main SNP for QTL_{991} is shown in Figure 40.

Gene name	Position on chromosome 9	Description from database
GRMZM2G044180	122,553,511-122,558,828	Protein kinase superfamily protein
GRMZM2G114312	122,867,393-122,871,672	Transcription initiation factor TFIID subunit 9B

Table 6: Candidate genes for QTL_{215}.

4.5.1 Expression of Kinase (GRMZM2G044180) after indanone induction, sequence differences in different maize lines, and influence of mu transposons in this gene on terpene emission

GRMZM2G044180 is a gene coding for a kinase, on position 122,553,511-122,558,828 bp on the forward strand of chromosome 9.

Transcript levels of the kinase gene in B73 induced and control leaves were measured using qRT-PCR.

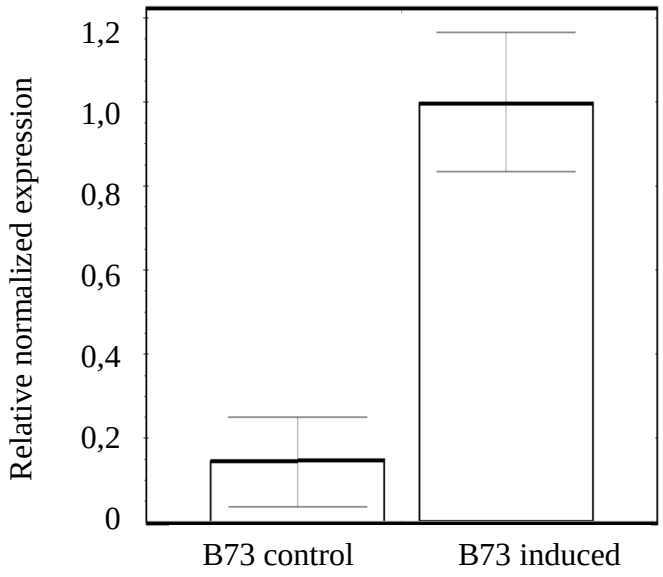

Figure 40: Expression level of the kinase gene in B73 control and induced plants. The amount of transcript for the kinase gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

The expression of the kinase gene was significantly higher after indanone induction. This suggests that the gene is part of the regulatory chain.

Next, the sequence of the candidate gene in B73 and Hp301 was compared. The following graphs show sequences from the gene GRMZM2G044180, in comparison of B73 and Hp301 as found in the HapMap database. Sequence parts not depicted here did not contain differences between the two maize lines, or there were no reads available for this part of the gene in Hp301.

```
B73     ATTGGTTG CATCAAGATA GGGTTATCAC ATCACCGTGC
Hp301   ATTGGTTG CATCAAGATA GGGTTGTCAC GTCACCGTGC

                        1280
B73     CCTACTTTTC AAGGTTCTCT GCTTCGGTCC CTTCCTCTTT CTCATGCA
Hp301   CCTACTTTTC AAGGTTCTT- -CTTTGGTCC CTTCCTCTTT CTCATGCA
```

Figure 41: Deletion of two base pairs at position 1280: confirmed in four reads, this part of the gene is an exon.

Deleting two base pairs in an exon shifts the reading frame for the sequence from this point onward.

```
                          1960
B73     TTCGGGTGTC ACACCATATC TGTGTTCTGC TGGAAGTGTT
Hp301   TTCGGGTGTC ACACCAT--C TGT-TTCTGC TGGAAGTGTT

B73     TTTGGCAGTT CATGGATGCT GGTTAATCAT GACCAAT
Hp301   TTTGGCAGTT CATGGATGCT GGTTAATCAT GACCAAT
```

Figure 42: Deletion of two base pairs at position 1958, deletion of one base pair at position 1964: one read, this part of the gene is an exon.

There are two deletion sites close to each other. In total, the deletion is three amino acids, which means the reading frame for the rest of the sequence remains the same. But the short sequence between the two deletion sites will be translated different from the original sequence. This local difference in the resulting protein can lead to a difference in secondary and tertiary structure of the protein, thus altering its activity.

```
B73                               ACATAT TTAATAAGGG TACAACAGGA
Hp301                             ACATAT TTAATAAGGG TACAACAGGA
             3370
B73     AATCACACAT A--ATTAATC ACTCCTTGGT ATGCTAAATC ATGTCCTAAA
Hp301   AATCACACAT ATAATTAATC ACTCCTTGGT ATGCTAAATC ATGTCCTAAA
```

Figure 43: Insertion of two base pairs at position 3371; confirmed in two reads, this part of the gene is an intron.

This mutation in an untranslated region of the gene will not get translated to a different sequence in the gene product, but may be involved in mRNA processing

```
B73                                                       ATCTTG GAGTTTTCCT
Hp301                                                     ATCTTG GAGTTTTCCT

                                                 3940
B73     AAATAAACTA TTCAATTTTC ATATTTACTC AGTTCAATTG TTGATTTATT GTTTGATATA
Hp301   AAATAAACTA TTCAATTTTC ATATTTACTC AGTTCAGTT- --GATTTATT GTTTGATATA

B73     CGGATATACT TTTTGACAAG ATTCGAATTA TGCGTCGGCT GCGCCATCCA AATATTGTTC
Hp301   CGGATATACT T
```

Figure 44: Deletion of three base pairs at position 3940: one read, this part of the gene is an intron.

This mutation in an untranslated region of the gene might be involved in mRNA splicing. However, in general, introns are not part of the mRNA that gets translated, so their mutations do not get directly translated into a different amino acid sequence.

```
                                                    4120
B73     GGTATAATAG TTACTTTTAC TACACAATCA CTGTTTATAT GGA--ATGTG TGAATTAAAT
Hp301   GGTATAATAG TTACGTTTAC TACACAATCA CTGTTTATAT TGTGGATGTG TGAATTAAAT

              4150
B73     ATCCCACTCT AGACAACACA GAAGGTA
Hp301   ATCCCACTCT AGAC------ GAAGGTA
```

Figure 45: Insertion of two base pairs at position 4124, deletion of six base pairs at position 4155: confirmed in four reads, this part of the gene is an intron.

A mutation in an untranslated region of the gene leaves the sequence of the gene product unchanged. It can be important for mRNA processing, which leads to a different protein product.

```
              4690
B73     CTGCACGAGG T--CGCATGA GTGCAGCCAC GACTAGCCTC ATAACTAAAG TAAAACACAC
Hp301   CTACACGAGG TGTCGCATGA GTGCAGCCAC GACTAGCCTC ATAACTAAAG TAAAACACAC

B73     AAACCACTGT CCACT
Hp301   AAACCACTGT CCACT
```

Figure 46: Insertion of two base pairs at position 4692: confirmed in two reads, this part of the gene is an intron.

The deletion of two base pairs in position 4692 in an untranslated region of the gene will not lead to a difference in the protein sequence, unless it is part of a splicing site.

HapMap indicated several deletions in the kinase gene in Hp301. Only the differences in exons get directly translated into the sequence and structure of the gene product. Differences in introns can make a difference in splicing and regulation processes.

The gene could not be amplified by PCR with Hp301 cDNA or genomic DNA, so the gene sequence in Hp301 might be incomplete or so different that primers are not able to bind. This strongly suggests that the gene product does not have the same activity in Hp301.

UFMU lines for the kinase gene on QTL$_{991\text{-}996}$

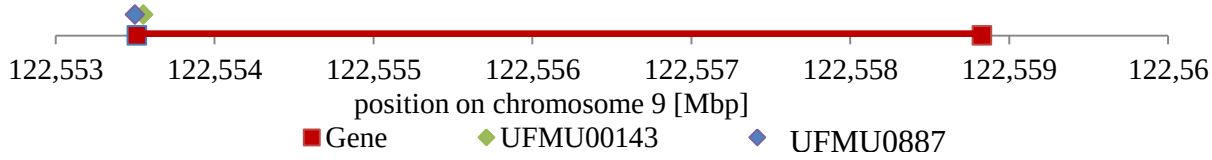

Figure 47: positions of the UFMU transposons in the kinase gene; red: translated region of gene

Both transposons UFMU00143 and UFMU0887 are right in the beginning of the kinase gene first exon. This means they probably interrupt the exon, leading to a change in the resulting protein, and altering its activity. If this deactivation of the kinase gene leads to a difference in terpene emission, this strongly indicates that the kinase is involved in terpene regulation.

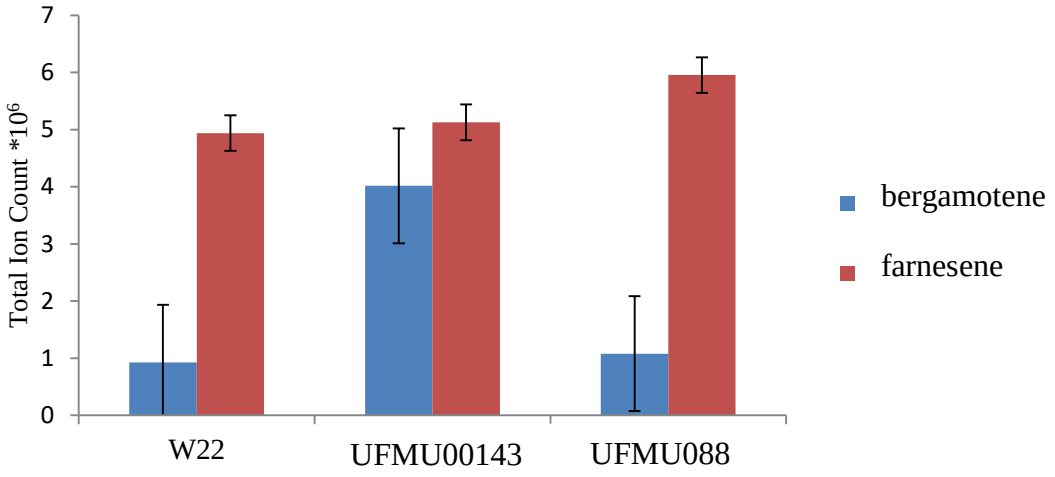

Figure 48: Terpene emission of UFMU lines compared to the background line W22.

The emission of farnesene is similar in all lines. UFMU00143 produces an increased amount of bergamotene, while UFMU08873 emits as much bergamotene as the background line W22.

As a result, the slight difference in the position of the mu transposons within the first exons leads to a different effect on the activity of the kinase.

4.5.2 Expression of Transcription initiation factor TFIID subunit 9B (GRMZM2G114312) after indanone induction, and sequence differences in different maize lines

GRMZM2G114312 is a gene on chromosome 9, forward strand at 122,867,393-122,871,672 bp. It has similarity to a subunit of a transcription initiation factor.
Expression analysis by qRT-PCR was not successful. There was no detectable level of expression of this gene. This means that the expression is very low, or the construction of qRT-PCR primers was not successful. Several sets of primers were tested, none of them showed any positive result in the PCR. This means the gene is probably not expressed in B73. Therefore, sequencing experiments could not be done, and HapMap sequences were used.

```
       640
B73    GGCCGAGAAC CTGGGATTTC GTTACCATTT TTTGTTAATA ATACTAATAA GTGTTTTCCTC
Hp301  GGCCGAGAAC CTGGGATTTC GTTACCATTT TTTGTTAATA ATACTAATAA GTGTTTTCCTC

       700
B73    ACATAGTCAC ATCTCTCTAG TCACGAAGTT CTGAATATTG AAGAAGATGA
Hp301  ACAT------ --CTCTCTAG TCA
```

Figure 49: Deletion of eight base pairs at position 705: confirmed in two reads, this part of the gene is an exon.

A deletion of eight base pairs in the expressed region of the gene causes a reading frame shift in the translation of mRNA to protein. This alters the rest of the protein sequence after this point, probably leading to a loss of function of the protein.

```
       1314
B73    AAAGGCTTTG TTGTTGTTGT TGTTGTCTCTA GTCAAGAAGT AGTATATATT CTATTTAGTT
Hp301  AAAGGCTTTG TTGTTGTTGT ------CTCTA GTCAAGTA-- ---TACTATT CTATGTAGTT

B73    TTTGGGTGAT ATGTATTTTA AATTTATTAA TTAG
Hp301  TTTGGGTGAT ATGTATTTTA AATTTATTAA TTAG
```

Figure 50: Deletion of six base pairs at position 1334, deletion of five base pairs at position 1353: confirmed in three reads, this part of the gene is an intron.

Deletions in untranslated regions of the gene get cleaved out during mRNA processing and do not result in an altered protein. However, they can play regulatory roles in mRNA processing.

The HapMap sequences of the gene GRMZM2G114312 indicate three deletions in the Hp301 sequence. Only the first deletion is in an exon, the other two are in the untranslated region of the gene. The deletion of eight base pairs in the exon will lead to a frameshift, and thus a big change in the translation of the rest of the sequence.

4.6 Investigation of candidate genes not on QTL: ERF transcription factors

4.6.1 Expression of AP2-EREB factor 154 (GRMZM2G026926 (ZEAMMB73_175602)) after indanone induction, and sequence differences in different maize lines

The gene GRMZM2G026926 (ZEAMMB73_175602) codes for a protein referred to as ereb154 - AP2-EREBP-transcription factor 154. Its position is on chromosome 2, from 24,360,914-24,363,125 base pairs on forward strand, which puts it near QTL_{215}.
The following graph represents the position of the gene in comparison to the QTL, its SNP and its candidate genes:

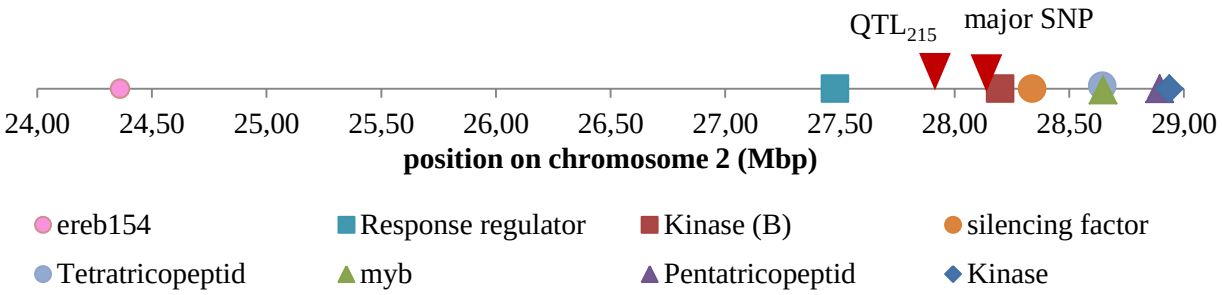

Figure 51: Position of the EREB transcription factor on chromosome 2, compared to QTL_{215} , its candidate genes and markers.

The gene GRMZM2G026926 is not close enough to the QTL marker and SNPs to be a part of the QTL itself. However, it is possible that it has an influence on one of the regulatory factors within the QTL.

As a first step, a qRT-PCR was conducted to test the gene in B73 for induction after indanone treatment of the plants.

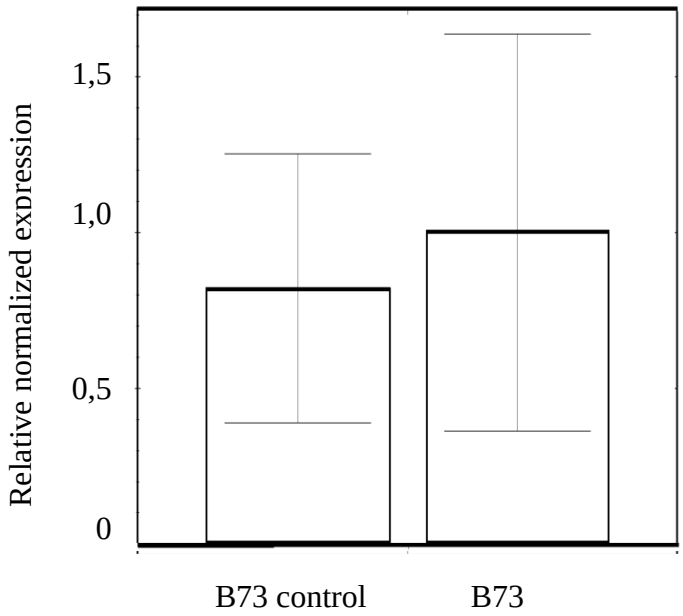

Figure 52: Expression level of the AP2-EREBP-transcription factor 154 gene in B73 control and induced plants. The amount of transcript for the AP2-EREBP-transcription factor 154 gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

For the expression level of the ereb154 transcription factor gene, no change after indanone induction was detected. Transcript levels in control plants and indanone-induced plants are similar. Nevertheless, this does not rule out that the gene plays a role in a regulation cascade.

Since it is close to QTL_{215}, the gene was sequenced in the lines Tzi8 and CML333, which are relevant for the QTL. In Tzi8, no relevant differences were found. The sequencing results for CML333 and the alignment with the B73 gene are shown below:

```
        1940
B73     AACGGCATCC TCATTAGACG ATTCATGATG TTTTAGATTT TTCCTATATT CTTCAATGTG
CML333  AACGGCATCC TCATTAGACG ATTCATGATG TTTTAGATAT TTTCTATATT CTTCAATGTG

        2000
B73     TTCTAATTTA GCTATGGGTT CCTACTTTTC CTAGTGTCTG GAAAATTATG TTGACCAGTT
CML333  TTCTAGTTTA GCTTTGGGTT CCTACTTT-C CTAGTGTCGG GAAAATTATG TTGACCAGTT

        2060
B73     TATGTACGTT GTACTATATA ATTAGATCTG ATTT-ATTTT CCTGTACTGCT ACCTTCCATA
CML333  TATGTACGTT GTACTATATA ATTACACATG ATTTTATTTT CCTGTACTGTT ACCTTTCATA

        2120
B73     CACAGTTTAG ----GGTTTT CATCCATGCA TTGTTTTGAA ATGTCAGGCA TCAACTTGTG
CML333  CACGGTTTAG CTAGGGTTTT CATCCATGCA TTGTTTTGAA ATGTCAGGCA TCAACTTGTG

        2180
B73     CGTGCATT
CML333  CGTGCATA
```

Figure 53: Sequence of EREB transcription factor gene in B73 and CML333; sequence for B73 taken from [maizegdb], this region is an exon.

Compared to B73, the gene for the Ereb transcription factor in CML333 has some base pair exchanges, and several inserts and deletions in the expressed region of the gene. There is one deletion of one base pair, one insertion of one base pair, and an insertion of four base pairs. The deletion or insertion of base pair numbers not divisible by three results in a shift of the reading frame. This leads to a completely different translation of the downstream sequence or a stop codon. Such a change in gene structure probably alters the resulting protein and its activity.

4.6.2 Expression of Ereb transcription factor (GRMZM2G171179) after indanone induction, and sequence differences in different maize lines

The gene GRMZM2G171179 encodes a transcription factor of the AP2/EREBP transcription factor superfamily. It is positioned on chromosome 9, from 11,535,025 to 11,537,985 base pairs on the reverse strand.

The distance of GRMZM2G171179 to the $QTL_{991-996}$ is too big for it to be associated with the QTL directly.

First, it was tested with qRT-PCR of B73 cDNA for induction after indanone treatment.

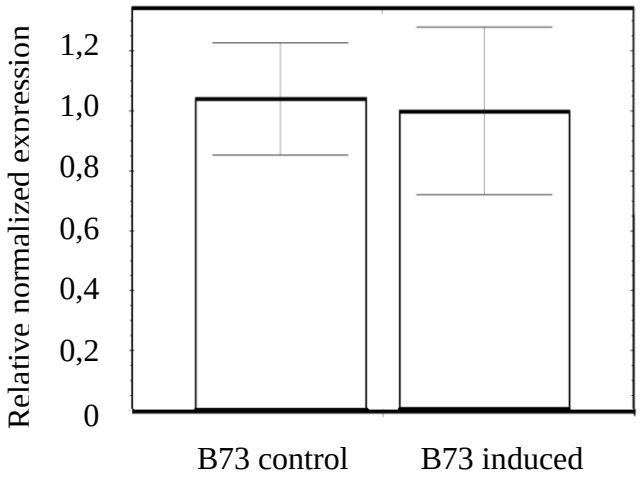

Figure 54: Expression level of the AP2/EREBP transcription factor gene in B73 control and induced plants. The amount of transcript for the AP2/EREBP transcription factor gene was determined by qRT-PCR with B73 cDNA from control plants and indanone-induced plants. Three biological replicates were used, and the transcript levels were normalized in relation to the transcript level of the housekeeping gene (primers HG5, HG6).

Comparing control and indanone-induced plant samples, no induction was proven. Still, a gene that is not induced can be part of a signaling cascade.

As a next step, the gene was sequenced in Hp301 and compared to database-derived sequences in B73. The line Hp301 was chosen because it was the one relevant for $QTL_{991-991}$, and a possible interaction of this transcription factor and the QTL was suggested.

The following sequencing parts were repeated with biological replicates, as well as different primer pairs.

```
            930
B73     GTGAACCTGT CTTCAGACCA AGGCAGCAAC TCGTTCGGTT GCTCGGACTT CAGCCTCGAG
Hp301   GTGAACCTGT CTTCAGACCA AGGCAGCAAC TCGTTCGGTT GCTCGGACTT CAGCCTCGAG

B73     AACGACTCCA GGACCCCTGA CATAACTTCG GTGCCTGCGC CCGTTGCCAC CTTGGCCGCC
Hp301   AACGACTCCA GGACCCCTGA CATAACTTCG GTGCCTGCGC CCGTTGCCAC CTTGGCCGCC

B73     GTTGGCGAGT CTGTGTTCGT CCAGAACACC GCCGGCCATG CTGTGGCGTC TCCTGCGACG
Hp301   GTTGGAGAGT CTGCGTTCGT CCAGAACACG GCCGGCCATG CTGTGGCGCC TCCTGCGACG

B73     GGGAACACTG GTGTTGATCT CGCCGAGTTG GAGCCGTATA TGAATTTCCT GATGGACGGT
Hp301   GGGAACACTG GTGTTGATCT CGCTGAGTTG GAGCCGTATA TGGATTTCCT GATGGACGGT

B73     GGTTCAGACG ACTCGATCAG CACTCTCTTG AGCTGTGATG GATCCCAGGA CGTGGTCAGC
Hp301   GGTTCAGACG ACTCGATCAG CACTCTCTT- AGCTGTGATG GATCCCAGGA CGTGGTCAGC

B73     AACATGGACC TTT-GGAGCT TCGAGGACA- GCCCATGTCT GCTGGTTTCT ACTGAGGCTG
Hp301   AACATGGACC TTTTGGAGCT TCGACGACAT GCCCATGTCT GCTGGTTTCT ACTGAGGCTG


B73     AGGCCCAGCG ACTGGTGCTT --------GT GTACATAGGG GGGGACAAAG GGTAAGAGCC
Hp301   AGGCCCAGCG ACTGGTGCTT GACTGCTTGT GTACATAGGG GGGGACAAAG GGTAAGAGCC

B73     TGCAGTAACA GAGATTGGCT CTTTCTGGTA CTTGCAATTT CTATCCCTTC AACTCTTTCT
Hp301   TGCAGTAACA GAGATTGGCT CTTTCTGGTA CTTGCAATTT CTATCCCTTC AACTCTTTCT

B73     TCCGCCCCCG TGTTTCAGGA ATAATGTTCT GGAGATGAAG AAACGCTTGC GTGGGCGTGC
Hp301   TCCGCCCCCG TGTTTCAGGA ATAATGTTCT GGAGATGAAG AATCGCTTGC GTGGGCGTGC

B73     CTGCAGGCAC GCGTGTAGTA GCTGCGGTAT TAGTATATAT GCTTAGATGT TCAGTCACTT
Hp301   CTGCAGGCAC GCGTGTAGTA GCTGCGGTAT TAGTATATAT GCTTAGATGT TCAGTCACTT


B73     TCTTTAAGTA CAATTTGGCG CTGGACATGT ACCTTATTTT ACTATGTATC CGTGACAACA
Hp301   CCTTTAAGTA CAATTTGGCG CTGGACATGT ACCTTATTTT ACTATGTATC CGTGACAACA


B73     GCTATGTGTC TGCTCCTTTT ATTTTCTTGT CTTTGCTTCA AAAAAATGGC TCTGAACATT
Hp301   GCTATGTGTC TGCTCCTTTT ATTTTCTTGT CTTTGCTTCA AAAAAATGGC TCTGAACATT


B73     GCGAGTTTGT ACTTTGTAGA CAATATATAT ATATATGTGT GTG--TATGT GCTTGGTTTG
Hp301   GCGAGTTTGT ACTTTGTAGA CAATATATAT ATATATATAT GTGTGTATGT GCTTGGTTTG


                                                              1760
B73     CTTCATCCTT CTATGCAATG TGAGTAGTGT TCTGTATGCA GGCTTGCATG TGCTGATATG C
Hp301   CTTCATCCTT CGATGCAATG TGAGTAGTGT TCTGTATGCA GGCTTGCATG TGCTGATATG C
```

Figure 55: Sequence of EREB transcription factor gene in B73 and Hp301; sequence for B73 taken from [maizegdb].

The gene for the Ereb transcription factor GRMZM2G171179, which is located on chromosome 9 but not in the direct vicinity of $QTL_{991\text{-}996}$, also shows nine single base pair exchanges, two single-base-pair insertions, one eight base pair insertion, one two base pair insertion, and a one base pair deletion in the translated region of the gene in the line Hp301 compared to B73. The single base pair exchanges can result in a different codon in translation, which results in a different amino acid in this position of the protein product. The insertions and deletions lead to frame shifts in translation, which means everything downstream from this sequence part gets translated differently, thus altering the structure and activity of the protein.

5. Discussion

5.1 QTL_{215} is involved in the regulation of several terpenes

QTL_{215} is a region on chromosome 2, which is important in the regulation of several terpenes. It plays a role for the regulation of bergamotene, farnesene, TMTT, DMNT, nerolidol and linalool emission. These terpenes are products of different enzymes. Bergamotene and farnesene are products of TPS10, TMTT and DMNT are produced by P450 enzymes, and nerolidol and linalool are products of TPS2. This means the regulation does not apply to just one terpene synthase, but either several enzymes or early steps in the regulation cascade that later on influence several terpene production pathways.

The calculations of the position and significance of this QTL were done with different sets of the NAM parent lines. By adding and removing lines from the dataset, we can demonstrate which lines are the most relevant for this QTL. These are the lines where the regulatory genes in the QTL have a genetic and functional difference, leading to a different phenotype. The calculations link a genetic marker difference to a phenotypic difference in terpene production. This calculation was done for each terpene, as shown in figure 6.

The maize lines that have the biggest influence on the calculation of QTL_{215} are CML333, Tzi8 and Oh7b. If these three maize lines are removed from the calculation, the QTL loses most of its significance. That implies that these lines must have a different regulation in QTL_{215} compared with B73. This different regulation leads to a different amount and proportion of volatiles in the emitted blend.

CML333 displays an overall strongly reduced terpene emission in comparison to B73, especially reduced amounts of linalool, bergamotene, DMNT and E-β-caryophyllene. Compared to B73, the line CML333 presents the biggest difference in products of TPS10 and P450. The difference in E-β-caryophyllene emission is not related to QTL_{215}, as this QTL does not show statistical significance for the production of E-β-caryophyllene.

Tzi8 has a total terpene emission that is slightly lower than B73, but higher than CML333. The strongest reductions are in the amount of nerolidol and TMTT. This means that this maize line shows a big difference to B73 in the amounts of TPS2 and P450 produced.

Oh7b has a total terpene emission amount similar to B73, but the composition is different. There is no linalool in its volatile blend, a much higher amount of and E-β-caryophyllene, and less nerolidol and TMTT than in B73. The products of TPS2, P450 and TPS23 represent a

large difference between Oh7b and B73. Again, the difference in the TPS23 product E-β-caryophyllene is not related to this QTL.

5.2.1 QTL_{215} contains several possible regulatory factors

There are more than one hundred genes within the chromosomal region of the QTL. Candidate genes were chosen by searching for genes that are similar to other regulatory genes in maize, rice or Arabidopsis. Often, structurally similar genes from other plants have similar functions in maize. This can be due to a conserved structure, for example a binding site, resulting in a similar activity. But conserved structures does not necessarily mean the similar genes are part of the same reaction in both plants. For example, a kinase in maize can be similar to a kinase in rice, but its role in a different reaction and signal chain is different in each plant. Throughout evolution, a gene may be modified into a new sequence, which results in a new specificity of the resulting enzyme.

It is expected that most parts of a signaling cascade present a change of their expression level after herbivory (Reymond, et al., 2004). Genes can be upregulated after herbivory, and their increased presence leads to an increase in the activation of the next step in the signal chain. On the other hand, other genes have a repressing function. They need to be downregulated in order to allow for the next step in the cascade to start. The expression level of the candidate genes in this work was quantified by qRT-PCR. As a template, cDNA derived from mRNA was extracted from maize leaves. This template represents the level of transcription in the plant at the time point when the sample was taken. The extracted RNA was then transcribed into cDNA for further analysis. During this process, the amounts of specific sequences in RNA directly result in a proportional quantity of the corresponding cDNA. Gene-specific primers indicate the cDNA level of their target gene by amplification in qRT-PCR. In this kind of PCR, a wavelength-specific dye shows the amount of PCR product in the sample. Absorption is measured after every step, thus documenting the amount of PCR product throughout every step of the PCR. The results are always normalized against a standard housekeeping gene, which is present at constant level in all plants at all times (Tzin, et al., 2015).

The specificity of qRT-PCR primers was checked by cloning and sequencing. It is important to know that they do not bind to a random sequence in the gene, because that would show up as a false higher expression level of the target gene in the qRT-PCR.

In order to have a different regulatory function in maize lines that exhibit a different phenotype, there needs to be a genetic difference in the candidate genes. This genetic difference is what leads to a difference in the function of the gene product, eventually causing a different phenotype. Gene sequences were compared in HapMap and by own sequencing experiments. In both methods, a result was taken as reliable if it could be confirmed by at least two reads. A genetic difference found in a single read could also be caused by an error in PCR, cloning or sequencing.

5.2.2 RR3 Response regulator is induced after herbivory, but shows no sequence differences between maize lines

The gene GRMZM2G035688 was labelled in MaizeGDB as being similar to the RR3 response regulator from *Arabidopsis thaliana* (Urao, et al., 1998)**.** In Arabidopsis, an RR3 response regulator gene is induced by low temperature, dehydration and high salinity. It might play an important role in the cytokinin signal transduction, and regulate gene expression (Osakabe, et al., 2002). Close similarity to a gene that can regulate the expression of other genes, and that is involved in signaling chains in the cell, was the reason to choose this gene as a candidate gene for the herbivory-induced signal chain in maize.

The RR3 response regulator gene in B73 maize plants was induced after herbivory. In control plants, it was not expressed at a detectable level. So it is not transcribed, or only at a very low level, in plants that are not attacked by herbivores. This result was confirmed by an iTaq PCR of B73 cDNA and genomic DNA with primers for a part of the RR3 response regulator gene. The cDNA produced no band, and DNA resulted in a band. Sequencing of said band confirmed that it is the correct fragment. So the primers do bind to the right gene. The expression level in induced B73 plants is high enough to be detected by qRT-PCR, but too low to result in a visible band in iTaq PCR.
The low expression of the gene in control plants, and increased expression in indanone-induced plants suggest a role in herbivore-regulated processes. Nevertheless the expression of

the gene in induced plants was still at a very low level. The low presence of the gene's mRNA even in indanone-induced plants makes its role in the signal chain questionable. Other genes were much more highly induced. The transcriptional regulator is not present in control plants, and present at a very low level in induced plants. This means it could start the transcription of its target as the next step in the chain. The low level of this reaction might point to an earlier step in regulation. Late steps like the expression of terpene synthases show a much stronger induction after herbivory.

The fact that the gene sequence in Oh7b, CML333 and Tzi8 is identical to B73 speaks against a role in the QTL. In order to be relevant for the QTL calculations, there needs to be a difference in the genotypes, as identical sequences between the parent lines would not present in the calculations.

Consequently, due to the low level of expression, and the lack of genetic differences between the maize lines, the RR3 response regulator can be ruled out. On the basis of current evidence, it does not seem to play a role in the insect-induced regulation of terpene synthesis.

5.2.3 KinaseB GRMZM2G145051 is downregulated after herbivory, and has sequence differences in CML333 and Oh7b

Kinases are enzymes that transfer phosphate groups from a high-energy donor molecule, like ATP, to specific substrates. The phosphorylation of a protein can regulate its enzymatic activity, or its ability to bind to other molecules. Thus kinases play a big role in various signaling and regulation processes in the cell.

The maize gene GRMZM2G145051 is similar to the gene "AT1G21230.1 (WAK5) wall associated kinase 5" in *Arabidopsis thaliana* and "LOC_Os04g43730.1 OsWAK51 - OsWAK receptor-like protein kinase" in rice (MaizeGDB, 2014).

In qRT-PCR, the gene demonstrated a strong reduction of the expression level after 24 hours of indanone induction. This suggests a role in downregulation of a process involved in the herbivore-induced signal chain. This indicates that the target of this kinase needs to be less phosphorylated to be activated in its position in the terpene production.

The sequences from HapMap reads of this gene indicated significant differences from B73 in Oh7b and CML333. CML333 has a 5 base pair insertion, and Oh7b has a large deletion of 39 base pairs. The small insertion in CML333 is situated in an intron, thereby suggesting that it is not relevant for the translated gene product. The large 39 base pair deletion in Oh7b is in an exon, which is translated. It keeps the three base pair reading frame, thus not altering the sequence and structure of the resulting protein significantly. 39 base pairs equal a protein sequence of 13 amino acids, and removing these 13 amino acids from a protein can result in a different secondary and tertiary structure of the protein.

It is eye-catching that the concerned region contains a long repetitive TATA sequence in B73. This part of the gene resembles the TATA-box binding domain which is important in gene regulation. Deleting a binding domain in Oh7b would cause a different function or complete loss of function. Losing the activity of a kinase can interrupt a signaling chain, because the target of the kinase no longer gets phosphorylated. This stop in the signaling cascade can alter the phenotype of the plant.

5.2.4 silencing factor (GRMZM2G340601) is downregulated after herbivory, and has a large insert in the Tzi8 sequence

The gene GRMZM2G340601 is closely related to AT1G68580.2 in *Arabidopsis thaliana* and LOC_Os08g33420.1 in rice. Both proteins contain a bromo-adjacent homology domain, which is known to be involved in gene regulation. It is a protein-protein interaction module for gene silencing (Müller, et al., 2010).

Silencing factors are a class of transcription factors that suppress the expression of their target genes. Such a factor could for example play a role in downregulating earlier steps of the insect-induced signaling cascade after the successful terpene production and emission. Consequently, continuing the terpene production longer than necessary would be a waste of resources for the plant.

Another possible function of a silencing factor is the downregulation of a repressing factor. In turn, this then allows for another protein to become active.

The gene GRMZM2G340601 shows a lowered expression after herbivory. A lowered amount of a silencing factor match the idea of upregulation of herbivory-induced processes in the plant. A constant presence of the silencing factor in the non-induced plant blocks the terpene production. After herbivory, the amount of the repressing protein reduces, and so the production of terpenes or terpene synthesis substrates is no longer blocked.

In Tzi8, the sequence of this gene largely differs to the sequence in B73. A 72 base pair insert keeps the reading frame unchanged, but it leads to an addition of 24 base pairs to the protein sequence. Adding 24 base pairs likely alters the secondary and tertiary structure of the resulting protein. Consequently, the whole molecule folds differently. Such a large structural change would change the target binding ability and activity of the protein, especially if it concerns the shape of the substrate-binding site of an enzyme.

The target of this protein is unknown. One approach to find possible targets would be a knock-out mutant of the silencing factor. By comparing gene expression levels in presence and absence of the silencing factor, one could find genes that are regulated by this factor. Complete absence of the factor would mimic a permanent induction of herbivore-regulated processes. In an array of all genes that are expressed differently in the absence of the silencing factor, one needs to identify the gene that is the target of the silencing factor, and those that are later steps in the regulation network.

5.2.5 Tetratricopeptide (GRMZM2G702991) is upregulated after herbivory, and has a sequence difference in Oh7b

GRMZM2G702991 is a gene of the tetratricopeptide family. Tetratricopeptides contain repeat motifs (Tetratricopeptode repeats, TPR) for protein-protein interaction, and participate in cell cycle control, protein transport processes, regulatory phosphate turnover and protein folding (Gregory L. Blatch, 1999). They also take part in specifically binding target RNAs and controlling processing of RNAs (Tillich, et al., 2010). This suggests many possibilities of tetratricopeptides to be involved in the regulation of terpene production. They could work at a transcriptional level, regulating the RNA to protein transcription of target genes. Or they could be involved in phosphorylation and dephosphorylation cycles, thus regulating the

activity of a protein which is present in the cell. It is also possible that their target is present in an inactive form, and the tetratricopeptide protein, that is only produced after herbivory, helps fold its target into an active form. So far, no involvement of TPR proteins in herbivore-induced terpene synthesis has been proven.

The tetratricopeptide gene GRMZM2G702991 is upregulated upon herbivory. This suggests that it does play a role in the signaling process.

The HapMap sequences indicated no relevant differences between B73, Tzi8 and CML333. In order to have a different function or activity, there needs to be a genetic difference. This genetic difference would result in the statistic visibility in the mapping calculations. For the lines Tzi8 and CML333, this gene would not come up as a significant QTL.

Between B73 and Oh7b, several differences in the sequence were found. A three base pair insert in the Oh7b sequence does not shift the reading frame, but an extra amino acid in the resulting protein sequence can change the folding of the protein. Inserting an extra amino acid can interrupt structures like a β-fold or an α-helix, or add a new interaction like a hydrogen bond or a charge that leads to a different folding of the surrounding structures. Binding sites for substrates and interaction partners can be changed largely by such a seemingly small sequence difference.

Other deletions and insertions of base pair numbers not divisible by 3 would change the reading frame, which alters the protein structure a lot more. But all of these sequence differences were found in introns, so they are cleaved out before the mRNA is translated. They do not result in changes of the protein, unless they change the splicing of the mRNA.

Therefore, the most key difference in the sequence is a 3 base pair insert in Oh7b, adding one amino acid to the structure. One additional amino acid could interrupt a secondary structure like an α-helix or a β-fold, or the local properties like the charge of a region in the protein or a hydrogen bond. If this happens in a binding site, the target binding ability of the protein can be altered or lost. A loss of target binding, or affinity to a different target, leads to a change in the signal cascade.

5.2.6 myb-like protein with DNA-binding domain (GRMZM5G86998) is downregulated after herbivory, and has sequence differences in CML333

The myb gene family is a family of transcriptional activators, the name being derived from the disease myeloblastosis. Members of this family share an N-terminal DNA-binding domain, a central transactivation domain, and a C-terminal regulation domain (Oh, et al., 1999). Myb proteins were discovered in animals, but plants have their own subfamily of myb factors (Stracke, et al., 2001). They are involved in the regulation of secondary metabolism processes, such as flavonoid production in maize with myb proteins both as activators and repressors (Goettel, et al.), and regulation of anthocyanin production in citrus fruits (Butelli, et al., 2012). An involvement of myb factors in terpene regulation has not been demonstrated before.

The myb-like gene GRMZM5G86998 is downregulated after herbivory. This suggests a role as a suppressor, whose absence allows for the upregulation of terpene production steps. This gene has sequence differences between B73, CML333 and Oh7b. An insertion of six base pairs in an exon in the CML333 sequence keeps the reading frame intact, and adds two amino acids. This can lead to an altered secondary structure, possibly changing the binding site and thus the activity of the protein.

There is a deletion of 30 base pairs in the CML333 sequence of the myb factor gene. A deletion of 30 base pairs results in ten missing amino acids in the CML333 protein. This difference is large enough to alter the 3D structure more significantly than the small insert mentioned before. This can change the shape and interaction sites of an active site to the extent that it cannot bind its target anymore, or binds a different target. A loss of binding activity in a repressing factor would result in a constant increased level of its target. This kind of change in the regulatory chain would be unfavorable to the plant, as a constant upregulation of a metabolic process that isn't constantly needed is a waste of energy and resources. It can be used to identify the target of the myb factor – a panel of the expression level of all genes related to terpene synthesis would show differences between plants with and without the 30 bp insertion in the myb gene. The target of the myb factor, and all downstream factors that are affected by this step, would show an altered expression level.

5.2.7.1 Pentatricopeptide (Zm.25643 (GRMZM2G077420)) is strongly induced after herbivory, but does not have any sequence difference between maize lines

Pentatricopeptides are the biggest class of RNA-binding proteins in plants. They share a motif of 35 amino acids (the pentatricopeptide repeat, PPR) which is repeated up to 30 times. These repeats form a binding platform for targets – the precise binding mechanism and target is as yet unknown. Pentatricopeptides are involved in translation, recruitment of other catalytically active effector proteins, and binding and modifying target RNAs (Beick, 2010). These functions mean the pentatricopeptide could play a role in a signal transduction cascade on the RNA level or protein level. It could directly influence the expression of the next gene in the regulatory cascade by altering the RNA processing or translation. Alternatively, it can recruit another enzyme to start a reaction in the synthesis of terpenes and their precursor molecules.

The expression level of the gene Zm.25643 (GRMZM2G077420) is strongly induced upon herbivory. This hints at a role in the regulatory chain of insect-induced terpene production. Upregulation of a pentatricopeptide can lead to increased translation or local recruitment of other proteins in a signal chain.

The HapMap sequences for this gene were identical between B73, CML333 and Oh7b. Both the translated region and a 1 kbp promotor region were compared. The sequences matching in these lines probably rules out that this gene is relevant for the QTL. In order to come up as a QTL, a genetic difference which is related to a different phenotype would be needed.
This does not mean that the gene is not part of the signal cascade. Rather, the increased expression after herbivory strongly suggests that it does play a role in the terpene production. But the identical sequences mean it is not part of the differential regulation of terpene production between the maize lines.

5.2.7.2 UFMU lines for the pentatricopeptide gene have a different terpene emission than background lines without transposons

There are two UFMU maize lines containing a Mu transposon in the pentatricopeptide gene. These transposons interrupt the genetic sequence, thereby rendering the resulting protein less active or inactive. Both lines carry the transposon at the beginning of the translated region of

the pentatricopeptide gene. A transposon insertion in the beginning of the translated region of a gene is expected to result in a change of functionality of the gene product.

The measurement of volatiles indicates a lowered amount of bergamotene and farnesene in both UFMU lines, compared with the background line without transposon insertion.
E-β-caryophyllene emission is increased in one line, and slightly lowered in the other. For all terpenes, the UFMU lines exhibit a strong biological variability. An UFMU line may contain more than one transposon, so additional transposons having an unforeseen influence on terpene emission cannot be ruled out. Moreover, additional transposons can lead to difference in plant growth and primary metabolism, which can indirectly influence the plants capability to produce secondary metabolites. To exclude this, the whole maize line needs to be sequenced, and the position and function of other transposon inserts needs to be identified.

The fact that inserts in this gene lead to a difference in terpene emission supports the theory that the gene GRMZM2G077420 is part of the signal chain. For bergamotene and farnesene - both products of TPS10 - its function has a positive influence on the level of production. Inactivity of the pentatricopeptide leads to a lowered emission of these two volatiles. This suggests that it plays a role in the regulation of TPS10, or the production of its substrates.
The volatile mix induced after herbivory also contains E-β-caryophyllene. This sesquiterpene is the product of TPS23. The two UFMU lines produce and emit different amounts of E-β-caryophyllene. It is improbable that the few base pair difference in the position of the transposon between the two UFMU lines creates a strong difference in the functionality of the gene product. This probably means that at least one of the UFMU lines contains another transposon influencing the production of E-β-caryophyllene or one of its precursors. This makes a clear statement about the involvement of the pentatricopeptide in E-β-caryophyllene regulation impossible.

5.2.8 Kinase LOC100383522 (GRMZM2G032694) is downregulated after herbivory, and has sequence differences in CML333 and Oh7b

The product of the gene LOC100383522 (GRMZM2G032694) belongs to the same protein family as the previously mentioned GRMZM2G145051. It is similar to Casein kinase 1-like protein 6 CKL6 from *Arabidopsis thaliana* ***(Gramene)***. Kinases are known to be involved in the regulation of enzyme activity, by phosphorylation of specific target proteins. This function suggests a role in a signaling cascade.

The expression level of this gene is lowered after herbivory. This result suggests that it inhibits a factor of the herbivore-induced regulatory cascade, and its absence allows its target to become active. Or it is part of a downregulation of early steps of terpene production, and thus more present 24 hours after induction, to lower the amount of those early steps again and save resources for the plant.

The sequence comparison between B73, CML333, Oh7b and Tzi8 demonstrated that Tzi8 is identical to B73 in this gene. In CML333 and Oh7b, there are base pair exchanges in the exons. One base pair exchange compared with B73 was shared in Oh7b and CML333, and the latter had two more single base pair exchanges that are not present in any of the other lines. The exchange of base pairs in the expressed regions of a gene can be silent due to the degenerated genetic code, or it can lead to a different amino acid in this position. If this amino acid has a different size or different physical properties than the original one, it can change the protein fold in this region, or the shape or affinity of an active site. This leads to a loss of activity, or an altered specificity in binding the substrate. The loss of kinase activity would lead to its target not being phosphorylated anymore, so the target gene does not get activated and the herbivory-induced signal chain is interrupted.

5.3 $QTL_{991\text{-}996}$ is a larger chromosomal region which influences the regulation of several terpenes

$QTL_{991\text{-}996}$ is a region on chromosome 9, that is involved in the regulation of bergamotene, farnesene, TMTT and nerolidol emission. Bergamotene and farnesene are products of TPS10,

while nerolidol is a product of TPS2 and TMTT of a P450 enzyme. So the regulatory factor of this QTL does not directly apply to one terpene synthase.
As opposed to QTL_{215}, this QTL is a less clearly defined region. As indicated by the index 991-996 it spreads over several genetic markers. This can mean that there are several relevant genes, which merge into one QTL due to the marker density and the resolution of statistical calculations for this genetic region.

5.3.1 Candidate genes for $QTL_{991\text{-}996}$

The most significant maize line in the calculation of this QTL is Hp301. Without this line, the QTL loses its statistical significance. Due to only one line being of high significance, fine mapping could not be done for this QTL. This kind of calculation requires at least two maize lines in addition to the common parent B73.
However, it may be noted that candidate genes for this QTL are a Protein Kinase Superfamily protein, and a Transcription Initiation Factor.

5.3.2.1 Kinase GRMZM2G044180 is upregulated after herbivory, and has sequence differences between Hp301 and B73

The gene GRMZM2G044180 codes for a protein of the Protein kinase superfamily. It bears similarity to the genes for AT5G11850.1 Protein kinase superfamily protein from *Arabidopsis thaliana* and LOC_Os03g06410.1 protein kinase domain containing protein from rice (MaizeGDB, 2014). Protein kinases are a family of enzymes that specifically phosphorylate their target proteins. This leads to a shift in activity and binding ability of the target molecule. Kinases are involved in metabolism, signaling, regulation of proteins, transport of molecules in the cell and secretory processes. This suggests many possible levels of involvement in the production and secretion of volatiles in maize.

The level of kinase mRNA in induced plants is about 5-fold increased in comparison to control plants. The strongly elevated expression of this gene after herbivory hints at its role in the signaling cascade. Increased presence of a kinase leads to a much higher level of

phosphorylation of its target protein. Thus, the target protein gets activated and can fulfill the next step in the signal chain.

According to HapMap, there are several genetic differences in the kinase gene between Hp301 and B73. There are two deletions in the translated region. The first one shifts the reading frame. Frame shifts can lead to a completely different amino acid sequence or a stop codon. The second deletion is a 3 base pair deletion, which leads to a difference in one amino acid, but leaves the further translation unchanged. With these two differences, a change in activity or in the binding partner of the protein in Hp301 and B73 is to be expected. A loss of kinase activity implies the target of the kinase does not get phosphorylated. The phosphorylation is a way of regulating the activity of a protein. No phosphorylation by the kinase means no regulation of the target protein activity. The target of this specific kinase is unknown.

Lastly, the other differences found in the HapMap sequence are in introns, which do not lead to a difference in the protein sequence, but can lead to a different regulation in the transcription and translation of the gene.

5.3.2.2 UFMU for kinase GRMZM2G044180 have a different terpene emission than the background line

There are two UFMU lines bearing mu transposons within the kinase gene. Both UFMU00143 and UFMU8873 have a transposon right in the beginning of the gene. The transposon in UFMU00143 is situated a small distance further downstream in the sequence than the transposon in UFMU08873. Both transposons are in a position where they can alter the resulting protein, changing its activity.
Compared with the background line W22, both UFMU lines emit similar amounts of farnesene. In UFMU08873, the amount of bergamotene is also similar to W22. UFMU00143 emits a significantly increased amount of bergamotene compared to W22. This suggests that the kinase gene is involved in the regulation of bergamotene production, and the transposon in UFMU00143 is in a position that is important for its activity or specificity. With that position interrupted, bergamotene regulation is altered. Since bergamotene and farnesene are both products of TPS10, their differential regulation means this regulation step does not happen at

terpene synthase level. Therefore, the role of the kinase is an earlier step in the production of substrates for TPS10.
In further experiments, other proteins involved in terpene production could be analyzed for their phosphorylation level in UFMU00143. With this approach one could find genes which have a different phosphorylation level when the kinase gene is interrupted by a transposon. These genes are potential targets of this kinase, and should be involved in the production of precursors of bergamotene.

5.3.3 Transcription Initiation Factor GRMZM2G114312 has sequence differences between Hp301 and B73

GRMZM2G114312 is a gene on chromosome 9 that codes for subunit 9B of a transcription initiation factor TFIID. These factors play a role in the assembly of the RNA polymerase preinitiation complex (PIC) at the core promoter region of a DNA template, thus starting the synthesis of RNA from the promoter (UniProt, 2015). The target of this factor is not known yet. A factor that starts the transcription of specific target genes can be involved in the terpene regulation by starting the production of other regulatory factors, or enzymes involved in terpene production and emission.

In Hp301, the gene carries three deletions in the sequence compared to the B73 sequence. Two of them are in the untranslated region, so they do not alter the protein. The one that is in an exon can change the reading frame of the translation from mRNA to protein sequence. This difference suggests a loss or an altered specificity of its binding ability, thus changing its function in binding to the promotor of the next step in the signalling cascade. A loss of binding ability means the next step in the signal cascade does not get its transcription upregulation signal.
To find the target, one could create knockout mutants of GRMZM2G114312, and search for genes that are expressed at a lower level than in the wildtype. The inactivity of a transcriptional regulator should result in a lowered transcript level of its target genes, and the factors that follow after it in the signal chain.

5.4 Transcription factors outside of QTL positions: ERF transcription factors

There is a large family of plant-specific transcription factors which are influenced by ethylene. They are called APETALA2-Ethylene responsive factors (AP2/ERF). They are involved in plant development (Kitomi, et al., 2011) (Licausi, et al., 2013), and reaction to abiotic stresses and pathogens (Gutterson, et al., 2004) (Sasaki, et al., 2007) (Xu, et al., 2011) (Mizoi, et al., 2012). In tobacco, they also play a role in herbivore-defense by inducing nicotine synthesis (De Boer, et al., 2011) (Sears, et al., 2014). In Arabidopsis, AP2/ERF were found to be induced upon feeding of *Spodoptera littoralis* on plant leaves (Consales, et al., 2012). Here, they played a role in the insect-saliva induced suppression of the wound response in the plant. The involvement of this class of proteins in insect-induced defense mechanisms in other plants suggests they also have a role in such processes in maize.

In maize, EREB58 is induced upon herbivory, and dependent on jasmonate levels. It was proven to upregulate TPS10, thus increasing TPS10 products E-β-farnesene and E-α-bergamotene (Li, et al., 2015). Other genes of the AP2/ERF family have been found in the maize genome, but their function is not known yet. The high likelihood of these also playing a role in herbivory-induced terpene emission was the reason why two of them were chosen to be a part of this work, despite not being within the QTL regions.

5.4.1 AP2-EREBP-transcription factor 154 near QTL_{215}

GRMZM2G026926 is a gene on chromosome 2, near QTL_{215}. Its product is a protein named ereb154 - AP2-EREBP transcription factor 154. The target of this factor has not been found yet. The distance of the gene to the genetic marker 215, which defines the QTL, suggests that it is not a part of the QTL, but it is close enough to regulate another factor in the QTL.

The factor is not upregulated after herbivory. But this does not mean it cannot be part of the signaling cascade. Not every step needs to be upregulated. It is possible that a protein is present all the time, and gets activated or deactivated in the course of a signalling cascade. For example, phosphorylation of a protein by a kinase can play a role in this case.

In CML333, which is a relevant line for the close-by QTL_{215}, sequencing experiments indicated differences to the B73 sequence. One single-base-pair deletion, one single-base-pair insertion, and a four-base-pair insertion in the Hp301 sequence were found. These differences shift the reading frame in the translation of mRNA to protein, likely to cause an altered activity or target binding ability of the transcription factor. A loss of activity or target binding ability of a transcription factor would lead to a loss of transcription of its target. This means the signal chain would stop here.

In future experiments, knock-out or overexpression lines of the transcription factor can be created to find target genes that have an expression level linked to the presence or absence of the transcription factor. If the factor is part of the regulation of terpene production, its upregulation in an overexpression line should lead to a constant expression of terpenes without further induction by herbivory or indanone.

5.4.2 Ereb transcription factor on chromosome 9

GRMZM2G171179 codes for an AP2/EREBP transcription factor on chromosome 9. At about 11 Mbp on chromosome 9, its position is too far from the QTL_{991} with a position around 122.8 Mbp. It is not part of the QTL, but could be regulating a gene situated in this QTL region.

Moreover, the expression of the gene is not upregulated after herbivory. Again, it could still be part of the signaling cascade despite being expressed at a near constant level.

Since a relation to QTL_{991} was suggested, sequencing experiments of the gene in Hp301 and B73 were done.

Compared with B73, the sequence in Hp301 presented one single-base-pair deletion, two single-base-pair insertions, one insertion of eight base pairs, and another insertion of two base pairs. Since three base pairs code for one amino acid, insertions and deletions of base pair numbers not dividable by 3 lead to a shift in the reading frame of the translation. That means the sequence parts downstream of the insertion or deletion get translated completely different, or there might be a stop codon now. These differences probably cause a different binding specificity or loss of activity. Again, a loss of target binding ability would make the factor nonfunctional, therefore leading to a stop in the signal cascade, since the upregulation of the next step cannot be started.

Knockout or overexpression lines can be used to find target genes whose expression level is influenced by the amount of transcription factor in the cell.

6. Outlook

The herbivore-induced production of terpenes in maize plants is an important protection mechanism for the plants. Different maize lines emit different amounts of several terpenes, which show different efficiency in attracting the natural enemies of the herbivores. Understanding the regulation of terpene production is a big step towards the breeding of maize lines with increased resistance to herbivory, which results in higher yield and less usage of potentially environmentally harmful pesticides.

In this work, a number of factors involved in the regulation of insect-induced terpene production were identified. Due to the complexity of the network, and redundancy of some regulation mechanisms, it proved difficult to unambiguously prove the effect of single proteins in the signal chain. These require further work, for example screening more diverse maize lines for genetic and functional variations in these factors. The analysis of transposon-carrying maize lines has been a useful tool to show which phenotypes result from interrupting a gene sequence, but these results could be consolidated by the creation of knock-out lines for these genes. For such knock-out lines, screening large sets of genes for altered expression levels can help determine the position of that factor in the regulatory network, and its targets and interactions.

The comparison of gene sequences in maize lines with different phenotypes has been used to identify potential candidate genes. The data for the NAM parent lines have some gaps and conflicting entries. Further sequencing of all NAM parent lines, and ruling out contradictory sequence parts within lines, will help make this process more efficient and reliable.

Another tool used in identifying candidate genes was the comparison of homologous genes in *Arabidopsis thaliana* and rice. Keeping up with new publications on those can help find new indications for regulatory factors in maize. Since Arabidopsis is more easily cultivated and genetically modified than maize, it can be used as a model organism to create mutants. These can be helpful in investigating the function of a candidate gene within the regulatory network.

The growing field of so-called interactomics is creating new approaches to research gene-protein or protein-protein interactions. It has the potential to facilitate the investigation of the function of factors in complex interaction networks that are currently still difficult to determine.

Acknowledgements

I want to thank Prof. Jörg Degenhardt and all members of the working group Pharmaceutical Biotechnology for their support throughout this work, and their scientific advice and feedback. Further acknowledgements go to Dr. Justin Lee for his mentoring and helpful input.

Last but not least, I would like to thank my family and friends for their invaluable support.

Literature

Arimura, Gen-Ichiro, Ozawa, Rika und Maffei, Massimo E. 2011. Recent Advances in Plant Early Signaling in Response to Herbivory. *Int J Mol Sci. .* 12, 2011, Bd. (6), 3723–3739.

Aubourg, S, Lecharny, A und Bohlmann, J. 2002. Genomic analysis of the terpenoid synthase (AtTPS) gene family of Arabidopsis thaliana. *Mol Genet Genomics.* Aug, 2002, Bde. 267(6):730-45., Epub 2002 Jun 29.

Becker, EM, et al. 2014. Infection of corn ears by Fusarium spp. induces the emission of volatile sesquiterpenes. *J Agric Food Chem. .* Jun 2014, S. 5226-36.

Beick, Susanne. 2010. *Funktion und Evolution chloroplastidärer PPR-Proteine .* HU Berlin : Dissertation, 2010.

Brandstädter, Jörg, Roßbach, Claus und Theres, Klaus. 1994. The pattern of histone H4 expression in the tomato shoot apex changes during development. *Planta.* Vol. 192, No. 1 (1994), pp. 69-74 , 1994.

Butelli, E, et al. 2012. "Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges". *Plant Cell.* 5 2012, S. 1242–55.

Capra, E, et al. 2015. Protein profiling and tps23 induction in different maize lines in response to methyl jasmonate treatment and Diabrotica virgifera infestation. *J Plant Physiol.* 1. Mar 2015, S. 68-77.

Chemistry and Technology, Corn. 2016. worldofcorn. [Online] 2016. [Zitat vom: 07. feb 2017.] http://www.worldofcorn.com/#components-of-yellow-dent-corn.

Chezem, William R. und Clay, Nicole K. 2016 . Regulation of Plant Secondary Metabolism and Associated Specialized Cell Development by MYBs and bHLHs. *Phytochemistry. .* Nov; , 2016 , Bd. 131: , 26–43. .

Consales, F, et al. 2012. Insect oral secretions suppress wound-induced responses in Arabidopsis. *J Exp Bot.* Jan 2012, S. 727-37.

De Boer, K, et al. 2011. APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. *Plant J.* Jun 2011, S. 1053-65.

De Vos, M, et al. 2005. Signal signature and transcriptome changes of Arabidopsis during pathogen and insect attack. *Mol Plant Microbe Interact. .* Sep, 2005, Bde. 8(9):923-37.

Degen, T, et al. 2012. Differential performance and parasitism of caterpillars on maize inbred lines with distinctly different herbivore-induced volatile emissions. *PLoS One ePub.* 7 2012.

Degenhardt, J, et al. 2003 . Attracting friends to feast on foes: engineering terpene emission to make crop plants more attractive to herbivore enemies. *Curr Opin Biotechnol. .* Apr;14, 2003 , Bde. (2):169-76.

Degenhardt, Jörg. 2009. Indirect Defense Responses to Herbivory in Grasses. *plantphysiol.* Jan, 2009.

Dicke, M und van Loon, JJ, Soler, R. 2009. Chemical complexity of volatiles from plants induced by multiple attack. *Nat Chem Biol. .* May, 2009, Bde. ;5(5):317-24. , doi: 10.1038/nchembio.169.

Dicke, M, et al. 1990. Plant strategies of manipulating predatorprey interactions through allelochemicals: Prospects for application in pest control. *J Chem Ecol.* Nov 1990, S. 3091-118.

Dicke, Marcel und Sabelis, Maurice W. 1987. How Plants Obtain Predatory Mites as Bodyguards. *Netherlands Journal of Zoology.* 1987, S. 38, vol.2.

Dudareva, Natalia, et al. 2006. Plant Volatiles: Recent Advances and Future Perspectives. . *Critical Reviews in Plant Sciences.* 25: 417–440, , 2006.

Epstein, Douglas, Vogan, Kyle und Trasler, Daphne. 1993. A mutation within intron 3 of pax-3 gene produces aberrantly spliced mRNA transcripts in the Splotch mouse mutant. *Proc Natl Acad Sci.* 90, 1993, S. 532-536.

Eurofins. Eurofins. [Online] https://ecom.mwgdna.com.

Eyles, A, et al. 2003. Host responses to natural infection by Cytonaema sp. in the aerial bark of Eucalyptus globulus.. *Forest Pathology.* 33:317-331, 2003.

Farmer, E. E. und Ryan, C. A. 1992. Octadecanoid Precursors of Jasmonic Acid Activate the Synthesis of Wound-Inducible Proteinase Inhibitors. *Plant Cell.* 1992.

Fatouros, Nina E., et al. 2016. Prospects of herbivore egg-killing plant defenses for sustainable crop protection. *Ecology and Evolution.* 2016, Bd. 19.

Fontana, A, et al. 2011. Attractiveness of constitutive and herbivore-induced sesquiterpene blends of maize to the parasitic wasp Cotesia marginiventris (Cresson). *J Chem Ecol. .* Jun 2011, S. 582-91.

Fontana, Anna. 2011. *The Roles of Vegetative Volatiles in Plant Defense and .* FSU Jena : Dissertation, 2011.

Fuchs, Stephan. PrimerFox. [Online] Ernst-Moritz-Arndt-Universität Greifswald. http://www.primerfox.com/.

Fürstenberg-Hägg, Joel, Zagrobelny, Mika und Bak, Søren. 2013 . Plant Defense against Insect Herbivores. *Int J Mol Sci.* 14, 2013 , Bd. (5), May, S. 10242–10297.

Gandolfi, M, Mattiacci, L und Dorn, S. 2003. Preimaginal learning determines adult response to chemical stimuli in a parasitic wasp. *Proc Biol Sci. .* 22. Dec 2003, S. 2623-9.

Gholami, Azra. 2012. *Identification of Potential Regulators of Jasmonate-Modulated Secondary Metabolism in Medicago trunculata.* Ghent : Department of Plant Systems Biology, 2012.

Giamakis, A, et al. 2001. Eucalyptus camaldulensis: volatiles from immature flowers and high production of 1,8-cineole and beta-pinene by in vitro cultures. *Phytochemistry. .* Sep;, 2001, Bde. 58(2):351-5.

Gillett-Kaufman, Jennifer L. 2017. Entomology and Nematology. [Online] University of Florida, 2017. [Zitat vom: 03. 04 2017.] http://entnemdept.ufl.edu/creatures/misc/wasps/cotesia_marginiventris.htm.

Goettel, Wolfgang und Messing, Joachim. Structural And Transcriptional Analysis Of The Complex P1-wr Cluster In Maize. *Plant & Animal Genomes XVI Conference.*

Göhre, V und Robatzek, S. 2008. Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu Rev Phytopathol. .* 46:189-215. , 2008, Bd. doi: 10.1146/annurev.phyto.46.120407.110050.

Gramene. Gramene. [Online] [Zitat vom: 11. 10 2015.] gramene.org.

Gregory L. Blatch, Michael Lässle. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioAssays.* Nov 1999, S. 932–939.

Gutterson, N und Reuber, TL. 2004 . Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr Opin Plant Biol. .* Aug 2004 , S. 465-71.

HapMap-Project. Hapmap. [Online] http://cbsuss05.tc.cornell.edu/hdf5/select.asp.

Hernández, Ángel und Falcó, José Vicente. 2014. Quantitative Parameters and Ecological Implications of a Specialized Tritrophic Interaction Involving a Seed-Feeding Tortricid, Pseudargyrotoza conwagana, a Braconid Parasitoid, Bracon otiosus, and the Wild Privet, Ligustrum vulgare. *J Insect Sci.; 128.* 14, 2014, Bd. 128.

Hernandez-Cumplido, J, Glauser, G und Benrey, B. 2016. Cascading effects of early-season herbivory on late-season herbivores and their parasitoids. *Ecology.* May 2016, S. 1283-97.

Huang, Jun, et al. 2015. Comparative transcriptomics uncovers alternative splicing changes and signatures of selection from maize improvement. *BMC Genomics.* May, 2015, Bd. 8.

Huffaker, A, et al. 2013. Plant elicitor peptides are conserved signals regulating direct and indirect antiherbivore defense. *Proc Natl Acad Sci U S A.* Apr 2013, S. 5707-12.

Kappers, IF, et al. 2005 . Genetic engineering of terpenoid metabolism attracts bodyguards to Arabidopsis. *Science.* Sep 23, 2005 , Bd. 309(5743), 2070-2.

Kitomi, Y, et al. 2011. The auxin responsive AP2/ERF transcription factor CROWN ROOTLESS5 is involved in crown root initiation in rice through the induction of OsRR1, a type-A response regulator of cytokinin signaling. *Plant J.* Aug 2011, S. 472-84.

Köllner, TG, et al. 2008. A maize (E)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *Plant Cell.* 20. Feb 2008, S. 482-94.

Körber-Grohne, Udelgard. 1995. *Nutzpflanzen in Deutschland – Von der Vorgeschichte bis heute,.* Hamburg : Nikol, 1995. ISBN 3-933203-40-6, S. 88–93..

Külheim, Carsten, et al. 2015. The Eucalyptus terpene synthase gene family. *BMC Genomics. .* 16(1): 450., 2015.

Lauchli, Ryan und Boland, Wilhelm. 2003. Indanoyl amino acid conjugates: Tunable elicitors of plant secondary metabolism. *The Chemical Record.* 24 January , 2003.

Li, S, et al. 2015. The maize transcription factor EREB58 mediates the jasmonate-induced production of sesquiterpene volatiles. *Plant J. .* Oct;84(2):296-308. doi: 10.1111/tpj.12994., 2015.

Licausi, F, Ohme-Takagi, M und Perata, P. 2013. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytol.* Aug 2013, S. 639-49.

Lin, F, et al. 2014. Genome-wide identification of housekeeping genes in maize. *Plant Mol Biol.* Nov 2014, S. 543-54.

Macherey-Nagel. 2014. Nucleospin gel manual. [Online] 7 2014. http://www.mn-net.com/Portals/8/attachments/Redakteure_Bio/Protocols/DNA%20clean-up/UM_PCRcleanup_Gelex_NSGelPCR.pdf.

—. 2015. Nucleospin Manual. [Online] 12 2015. http://www.mn-net.com/Portals/8/attachments/Redakteure_Bio/Protocols/Plasmid%20DNA%20Purification/UM_pDNA_NS.pdf.

MaizeGDB. 2014. [Online] 2014. [Zitat vom: 10. 12 2016.] http://www.maizegdb.org.

Martin, DM, et al. 2010. Functional annotation, genome organization and phylogeny of the grapevine (Vitis vinifera) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biol. .* Oct 21, 2010, Bd. 10:226., doi: 10.1186/1471-2229-10-226.

Mattern, Roland. 2017. Wikipedia. [Online] 2017. https://de.wikipedia.org/wiki/Festphasenmikroextraktion.

McGarvey, D. J. und Croteau, R. 1995. Terpenoid metabolism. *Plant Cell.* 1995, Bd. 7, 1015-1026.

McMullen, Michael D., et al. 2009. Genetic Properties of the Maize Nested Association Mapping Population. *Science.* Vol 325. Issue 5941 2009, S. 737-740.

Mizoi, J, Shinozaki, K und Yamaguchi-Shinozaki, K. 2012. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim Biophys Acta.* Feb 2012, S. 86-96.

Müller, P, et al. 2010. The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. *Genes Dev.* 1. Jul 2010, S. 1418-33.

Oh, IH und Reddy, EP. 1999. The myb gene family in cell growth, differentiation and apoptosis. *Oncogene.* 13. May 1999, S. 3017-33.

Osakabe, Y., et al. 2002. Overexpression of Arabidopsis response regulators, ARR4/ATRR1/IBC7 and ARR8/ATRR3, alters cytokinin responses differentially in the shoot and in callus formation. *Biochem. Biophys. Res. Commun.* 2002.

Ozawa, R, et al. 2012. Temperature-dependent, behavioural, and transcriptional variability of a tritrophic interaction consisting of bean, herbivorous mite, and predator. *Mol Ecol. .* Nov;21(22):5624-35., 2012, Bd. doi: 10.1111/mec.12052. , Epub 2012 Oct 8.

Piesik, D, et al. 2011. Fusarium infection in maize: volatile induction of infected and neighboring uninfected plants has the potential to attract a pest cereal leaf beetle, Oulema melanopus. *J Plant Physiol.* Sep 2011, S. 1534-42.

Pinto-Zevallos, DM, Pareja, M und Ambrogi, BG. 2016. Current knowledge and future research perspectives on cassava (Manihot esculenta Crantz) chemical defenses: An agroecological view. *Phytochemistry.* Oct 2016, S. 10-21.

Piperno, Dolores R., et al. 2009. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences of the United States of America.* 2009.

Poulter, C. D. und Rilling, H. C. 1981. Prenyl transferases and isomerase. *Biosynthesis of Isoprenoid Compounds.* New York : John Wiley & Sons, 1981.

Qiagen. 2017. pDrive vector map. [Online] 2017. https://www.qiagen.com/es/resources/download.aspx?id=4d546584-2550-4806-a663-995fef1431d9&lang=en.

Rao, A, et al. 2010. Mechanism of antifungal activity of terpenoid phenols resembles calcium stress and inhibition of the TOR pathway. *Antimicrob Agents Chemother.* Dec 54(12):5062-9. doi: 10.1128/AAC.01050-10., 2010, Bd. Epub 2010 Oct 4.

Reymond, P, et al. 2004. A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell.* Nov;16(11):3132-47., 2004, Bd. Epub 2004 Oct 19.

Richter, A, et al. 2016. Characterization of Biosynthetic Pathways for the Production of the Volatile Homoterpenes DMNT and TMTT in Zea mays. *Plant Cell.* 28. october 2016, S. 2651-2665.

Richter, Annett. 2014. *Identifizierung von QTLs der Terpenbiosynthese mittels „Nested Association Mapping“ (NAM) & „Genome Wide Association Study“ (GWAS).* MLU Halle-Wittenberg : Dissertation, 2014.

Robert S. Peabody Museum, of Archaeology. maize breeding. s.l., Andover, Massachusetts : Phillips Academy.

Robert, CA, et al. 2013. Genetically engineered maize plants reveal distinct costs and benefits of constitutive volatile emissions in the field. *Plant Biotechnol J.* Jun 2013, S. 628-39.

Romeo, John T. 2009. New SPME Guidelines. *J Chem Ecol .* 35:1383, 2009.

Rosenthal, G.A. und Berenbaum, M.R. 1992. *Herbivores: Their Interactions with Secondary Plant Metabolites. Ecological and Evolutionary Processes. : . ;* San Diego, CA, USA : Academic Press, 1992.

Sasaki, K, et al. 2007. Two novel AP2/ERF domain proteins interact with cis-element VWRE for wound-induced expression of the Tobacco tpoxN1 gene. *Plant J.* Jun 2007, S. 1079-92.

Schifrin, Alexander, et al. 2016. A single terpene synthase is responsible for a wide variety of sesquiterpenes in Sorangium cellulosum SOCE65. *Org. Biomol. Chem.* 2016, S. 3385.

Schmelz, EA, et al. 2003. Quantitative relationships between induced jasmonic acid levels and volatile emission in Zea mays during Spodoptera exigua herbivory. *Planta.* Feb 2003, S. 665-73.

Schnable, Patrick S., Ware, Doreen und Robert S. Fulton, Joshua C. Stein, Fusheng Wei, Shiran Pasternak, Chengzhi Liang, Jianwei Zhang, Lucinda Fulton, Tina A. Graves, Patrick Minx, Amy Denise Reily, Laura Courtney, Scott S. 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science.* Vol 326, Issue 5956, 2009, 20 November.

Schnee, C, et al. 2006. The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. *Proc Natl Acad Sci.* 24. januar 2006, S. 1129-34.

Schüler, Göde, Görls, Helmar und Boland, Wilhelm. 2001. 6-Substituted Indanoyl Isoleucine Conjugates Mimic the Biological Activity of Coronatine. *European Journal of Organic Chemistry.* 9 April, 2001.

—. 2001. 6-Substituted indanoyl isoleucine conjugates mimic the biological activity of coronatine. . *European Journal of Organic Chemistry .* Nr. 9 2001, S. 1663–1668.

Sears, MT, et al. 2014. NtERF32: a non-NIC2 locus AP2/ERF transcription factor required in jasmonate-inducible nicotine biosynthesis in tobacco. *Plant Mol Biol.* Jan 2014, S. 49-66.

Seidl-Adams, I, et al. 2015. Emission of herbivore elicitor-induced sesquiterpenes is regulated by stomatal aperture in maize (Zea mays) seedlings. *Plant Cell Environ.* Jan 2015, S. 23-34.

Sigma-Aldrich. 2017. Biofiles. [Online] 2017. [Zitat vom: 03. 01 2017.] http://www.sigmaaldrich.com/technical-documents/articles/biofiles/dietary-terpenes.html.

Sobhy, IS, Erb, M und Turlings, TC. 2015. Plant strengtheners enhance parasitoid attraction to herbivore-damaged cotton via qualitative and quantitative changes in induced volatiles. *Pest Manag Sci.* May 2015, S. 686-93.

Srividya, N, Lange, I und Lange, BM. 2016. Generation and Functional Evaluation of Designer Monoterpene Synthases. *Methods Enzymol.* 2016, S. 147-65.

Strable, Josh und Scanlon, Michael J. 2010. Maize (Zea mays): A Model Organism for Basic and Applied Research in Plant Biology. *Cold Spring Harbour Laboratory Press.* 2010.

Stracke, R, Werber, M und Weisshaar, B. 2001. "The R2R3-MYB gene family in Arabidopsis thaliana". *Curr. Opin. Plant Biol. .* October 2001, S. 447–56.

Tamiru, Amanuel, et al. 2017. A maize landrace that emits defense volatiles in response to herbivore eggs possesses a strongly inducible terpene synthase gene. *Ecol Evol.* Apr;, 2017, Bd. 7(8), 2835–2845.

ThermoFisher. Thermo Fisher. [Online] [Zitat vom: 31. 01 2017.] https://www.thermofisher.com/order/catalog/product/450641.

—. Thermo Fisher. [Online] [Zitat vom: 31. 01 2017.] https://www.thermofisher.com/order/catalog/product/K1231.

ThermoScientific. 2008. Technical Bulletin. [Online] 2008. http://www.nanodrop.com/Library/T009-NanoDrop%201000-&-NanoDrop%208000-Nucleic-Acid-Purity-Ratios.pdf.

Tillich, Michael, Beick, Susanne und Schmitz-Linneweber, Christian. 2010. Chloroplast RNA-binding proteins: Repair and regulation of chloroplast transcripts. *RNA Biology.* 2010, S. 172-178.

Tsuda, Kenichi und Somssich, Imre E. 2014. Transcriptional networks in plant immunity. *New Phytologist.* 9 December , 2014.

Tumlinson, James H., Lewis, W. Joe und Vet, Louise E. M. 1993. How Parasitic Wasps Find Their Hosts. *SCIENTIFIC AMERICAN.* 1993.

Turlings, T. C. J., Tumlinson, J. H. und Lewis, W. J. 1990. Exploitation of Herbivore-Induced Plant Odors by Host-Seeking Parasitic Wasps. *Science .* 30 Nov , 1990.

Turlings, TC, et al. 1995. How caterpillar-damaged plants protect themselves by attracting parasitic wasps. *Proc Natl Acad Sci U S A.* 9. May 1995, S. 4169-74.

Turlings, Ted C. J., et al. 1998. Timing of induced volatile emissions in maize seedlings. *Planta.* 1998, S. 146.

Tzin, Vered, et al. 2015. Dynamic Maize Responses to Aphid Feeding Are Revealed by a Time Series of Transcriptomic and Metabolomic Assays. *Plant Physiol.* Nov; 169(3): 1727–1743., 2015.

UniProt. 2015. UniProt. [Online] 2015. http://www.uniprot.org/uniprot/B6TMH7.

Urao, T, et al. 1998. Stress-responsive expression of genes for two-component response regulator-like proteins in Arabidopsis thaliana. *FEBS Lett.* 1998.

Wang, Q, et al. 2016. Identification of a Dolabellane Type Diterpene Synthase and other Root-Expressed Diterpene Synthases in Arabidopsis. *Front Plant Sci.* Nov 2016, S. 1761.

Wang, Y, et al. 2016. A comprehensive meta-analysis of plant morphology, yield, stay-green, and virus disease resistance QTL in maize (Zea mays L.). *Planta.* Feb 2016, S. 459-71.

Xiong, L und Zhu, JK. 2001. Abiotic stress signal transduction in plants: Molecular and genetic perspectives. *Physiol Plant. .* Jun, 2001, Bde. 112(2):152-166.

Xiong, Liming, Schumaker, Karen S. und Zhu, Jian-Kang. 2002. Cell Signaling during Cold, Drought, and Salt Stress. *The Plant Cell.* May , 2002, DOI: https://doi.org/10.1105/tpc.000596 .

Xu, ZS, et al. 2011. Functions and application of the AP2/ERF transcription factor family in crop improvement. *J Integr Plant Biol.* Jul 2011, S. 570-85.

Yamauchi, Atsushi, et al. 2015 . Cry-wolf signals emerging from coevolutionary feedbacks in a tritrophic system. *Proc Biol Sci. .* Nov 7, 2015 , Bd. 282(1818), 20152169.

Yoshikuni, Y, et al. 2006. Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chem Biol.* Jan 2006, S. 91-8.

Yoshinaga, N, et al. 2010. Fatty acid-amino acid conjugates diversification in lepidopteran caterpillars. *J Chem Ecol.* Mar 2010, S. 319-25.

Yu, Jianming, et al. 2008. Genetic Design and Statistical Power of Nested Association Mapping in Maize. *Genetics.* Jan 2008, S. 539-551.

Zangerl, Arthur R. und Rutledge, Claire E. 1996. The Probability of Attack and Patterns of Constitutive and Induced Defense: A Test of Optimal Defense Theory. *The American Naturalist.* 147, 1996, Bd. 4, 599-608.

Appendix

Primers

QTL215		
RR3 (GRMZM2G035688)	101_RR3_F	CACCACAGAGATCCATTTCG
	102_RR3_R	CTTTGGTGATGGAGAAGTCG
	105_RR3_F	CTGAAGCTAGATAGCTAGCG
	106_RR3_R	ATGATCATGCTCACGTTGGG
	26_RR3 Fw	CGTGAGCATGATCATTACCG
	27_RR3 Rev	CTTCCTCCAGGCACCTGTT
KinaseB GRMZM2G145051	9_Kin F	TCGAAGTTCCTGTCCGCTAT
	10_Kin R	CAGCGAAAGCTTCCTCAATC
	70_kinB Start	ATGATCCTCCAGAAGAGA
	71_kinB Ende	CTACCTCGCATAGTACGA
	64_KinB F	TCTTGACGGCAACTACATGG
	65_KinB R	CTTCATGGATGGTCTGTTGC
silencing factor GRMZM2G340601	11_silencing F	GGATTTCGATGACAGGGAGA
	12_silencing R	TTGAAACGCACATCAAAAGG
	74_silencerStart	ACAAGCAAGCAGCATTCTGC
	75_silencerEnde	TTATGGGTGTGCCTGTGCAG
	66_silencer F	GTTGATGAGGTTGGTGTTCC
	67_silencer R	GAGAGAGTACTTCCACATGG
Tetratricopeptide GRMZM2G702991	13_tetratrico F	CTGTTCCTGCGCAACTACG
	14_tetratrico R	GCACCGACTGCTGGAAGTA
	68_tetratrico F	CTCCCCTTCCTCTCCTCCT
	69_tetratrico R	CCTTGCCGAAGGTGTTGTAG
	72_tetratrico Start	GAGCTCGTCCACGCCGTTCCC
	73_tetratrico Ende	GACGCCAGCTCCATGGCTTGCC
myb factor GRMZM5G869984	15_myb F	GTCATGGCAACATGATGAGC
	16_myb R	TGGAACTTCAGAAGGGTGCT
	76_mybF	TACCATGAACAACGCAGAGG
	77_mybR	TTACACGATGCCTTGGATCG
	78_mybF	TGATGATGCAGTTTCGGAGC
	79_mybR	TTTCCCAATCCATGGTAGCC
	80_myb F	TTGGTGGTCGTTATGTACCG
	81_myb R	CCTCTAGCTACTCTCTATGG

	82_myb F	CATCTATGGCACCATCTTCC
	83_myb R	TGCCGAAATTCATGTGCTGC
Pentatricopeptide GRMZM2G077420	19_pentatrico F	TGATGGTGTGATCTGGGCTA
	20_pentatrico R	CCTTTTTCTCGCATTGCTTC
	30_pentatrico F	GACGATGCAAGGAGGATGTT
	31_pentatrico R	TGCTTGCAAGAGTGGAAATG
	32_pentatrico Start	CCTAAGCCGTGGCCCGAGCAG
	33_pentatrico Ende	AAACAGGGTCATTGTCCC
	34_pentatrico F	CATGTAGACGCAAGCAGCAT
	35_pentatrico R	CCCCGAAAGAATGCTTGTAA
	38_pentatrico R	TCAGCTCTGAATGGTAACTGGTCA
	39_pentatrico R	TGTGATTCTTGCATCTTCTGG
	40_pentatrico R	CAAAACCAGATTTTACAATGAGGC
	41_pentatrico R	TTGTGATCTGGCACCTCCTC
	42_pentatrico R	CCAACAGCATGGTTTGAGCAG
	43_pentatrico R	GAGCATTCCAGGATACGCT
	44_pentatrico R	ACTCACAAATGTAAAATCATCAGCCT
	45_pentatrico R	TGAGTGAACTGCCCACA
	46_pentatrico R	GCAATCATGCTTGTCCAGCA
	47_pentatrico R	GACAGAACCACCGCAATGCC
	48_pentatrico R	TCGCCGAGGCGACCCCGG
	49_pentatrico R	TTCGAAGTTGCCGGAAAGA
	50_pentatrico F	CTTCT CTTCCAGAAG ATGCAA
	51_pentatrico F	CCATTGTTACACACTGAAATCTGC
	52_pentatrico F	CTCATTCCAGTTAAGGATAGCGT
	53_pentatrico F	A ACTAGGTCCA CTTTTGC
	54_pentatrico F	TTGGTAGCACGCCGGACCAGTT
	60_Penta5UTR	ACCGAACGGACGGCACATG
	61_Penta3UTR	TCAATCTAGGCCCTTGATTA
Kinase GRMZM2G032694	21_Kin F	GAAAGCTGGCACAAAGAAGC
	22_Kin R	ACCCCTCGCGGATAAATAAG
	36_kinase start	ACAGAATGGAGCACGTAGTCGG
	37_kinase ende	TGTCAGCACCAAGTGAGAGCAGCTC
	55_kin F	ACTGCAGCAGAAAGTTCTCC
	56_kin F	CTGGAATCAGTCACATCACG
	57_kin R	TTGTGACCAGAACCAGATCC
	58_Kin 5UTR	ACACACCTCTTTAACTCGCG
	59_Kin 3UTR	CAGTGATCTACTAGGACAGC
	62_CML333Kinf	CGGGATACTTTCCTTACTGG
	63_CML333Kinr	CGCGGATAAATAAGTCACGG

QTL991		
Protein Kinase Domain GRMZM2G044180	107_PKD_F	TTGGTTGATCTGATGGCTGC
	108_PKD_R	TGTGGTCTTCCAGCATTAGG
	109_PKD_F	TGGAATGGAACCGAAGTAGC
	110_PKD_R	CCTGATGCATATTCTGGTCG
	111_PKD_F:	GATTCGAATTATGCGTCGGC
	112_PDK_R:	GGTCGAGTGTGATTGAATCG
	113_PKD_F:	TTATGGGTGCTGTTACACGC
	114_PDK_R:	TAGCAAAGAACCTGTGGTCG
Transcription Initiation Factor GRMZM2G114312	127_TIF_St	ACCATCTCTTTGTCCAACGC
	128_TIF_End	AGAACATGTAGAAGACCCGG
	129_TIF_F	TGGTTTGTGGCTTCTGTTGG
	130_TIF_R	TTCAGAATGCAACCAGGTGG
	95_TIF_F	AAATCAATCGCTCCTCCTGG
	96_TIF_R	AGAACATGTAGAAGACCCGG

EREB transcription factors		
GRMZM2G026926 ereb154	115_F	TCATTCGCCTTCCTTCATGG
	116_R	GACGGATGGTTCTTGAATGC
	117_F	AACGGCATCCTCATTAGACG
	118_R	ATGCACGCACAAGTTGATGC
	131_F	GTTCGTTCGTTCGTTAGTCC
	132_R	ATGCACGCACAAGTTGATGC
	133_F	CCACATCACGCGCACACGCT
	134_R	AAGAATTCTAAAGTAATC
	135_f	ACATCGTCGCAGCTAGAAGG
	136_r	TCTGAAACCTGCGTGCATCT
	137_f	AAGGTGTGACGCCCATACAG
	138_r	CCTTCTAGCTGCGACGATGT
	139_f	TTCGTTCGTTAGTCCGTCCG
	140_r	CTGTATGGGCGTCACACCTT
	141_f	AGATGCACGCAGGTTTCAGA
	142_r	GGTTGCCCTAGTAGCTGACC
	143_f	GCTACATCGTCGCAGCTAGA
	144_r	GGCTCTGGCGAGTGAATCAT
	145_f	ATGATTCACTCGCCAGAGCC
	146_r	GATGCCGTTGCAGATCATGG
	147_f	GCGCATTCAAGAACCATCCG

	148_r	CATGCATGGTGTCCTGGTGA
	149_f	CCTTGGCAGATTCGGGTCAA
	150_r	TGCATGGACGGATGGTTCTT
	151_f	GGTCAGCTACTAGGGCAACC
	152_r	ACGCACAAGTTGATGCCTGA
	153_f	AGCTCCTTCTGGGACTACGA
	154_r	GTGTGATGCCGCAAGTTGTC
GRMZM2G17117910. 11.2015 Putative AP2/EREBP transcription factor superfamily protein	119_F	AGTGAACCTGTCTTCAGACC
	120_R	AGATCAACACCAGTGTTCCC
	121_F	AATTTCCTGATGGACGGTGG
	122_F	TGTTACTGCAGGCTCTTACC
	123_F	AGAGGTGGACTAAACACTGG
	124_R	GCATATCAGCACATGCAAGC
	125_Start	CTCTCCCACGCACTCCGCAA
	126_Ende	TTATGAAGAAAATTTGATTG

other genes		
TPS8	tps8fwd	ATGGTACGTCTCAGCGCATGGCGCCG AAGACTGTGTGG
	tps8rev	ATGGTACGTCTCATATCAGCAGAGGG GAACATGGTTGACG
Housekeeping gene: putative APT1A	HG5	AGGCGTTCCGTGACACCATC
	HG6	CTGGCAACTTCTTCGGCTTCC
TPS2 for Oh7b	97_TPS2_F	AGATGATTGAGGGCATGTGG
	98_TPS2_R	GTTGTTCGACGGTTTCATGG
TPS2-specific, by Annett Richter	S33	AACCTCGTGCAGGTCTGCTT
	S35	GAAGGAGCATGGATCTAACCATG

Hapmap data

The Hapmap V2 at http://cbsuss05.tc.cornell.edu/hdf5/select.asp was used. To conduct a search, select the dataset HapMap V2, and "Query Genotyping Data". Select a chromosome and position, and submit the search. The results will be displayed in HapMap format. There are complete gene sequences for B73, and short reads for other NAM parent lines.

The following examples are results for the Kinase B gene in QTL_{215}:

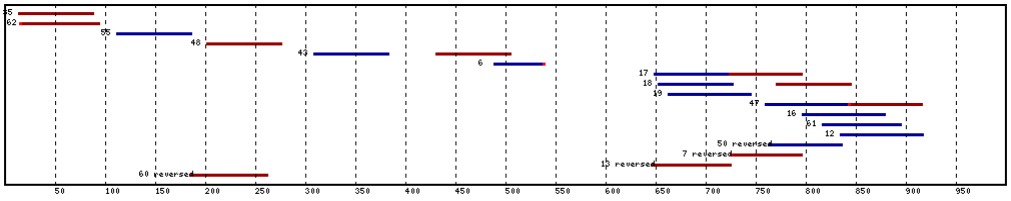

Figure 56: coverage of reads in CML333 for the selected gene position of Kinase B

Reads for B73 for Kinase B:

CAAGTATGTATATGTGTTTCTGCGCAGCACGTCTGCAACATATATAATACGGTTTTTTT

>2_28199812_28200812
GAGGGAACAGAAACCAGATGAACAACTCCCACAGACCGTGTGGTTGACAAATTAAAACATGAAGCAATCAGA
CAAATACAGGCATGCAACATCTCAAAAAAGAAAAGAAAAAAGATGCAGGCATGCAGCATGTTGGAGCAGAGC
TGACCAGGCTAGCACTAAATCAAGTGAACAAGCCACTGATAGCATGCACATATAATTAAGTAGTCATGCCAATT
GATTGAATGAGTTCAATGCCACATTCAGGGAACCAATTAGAATGCTATGGGTTATGTATTTTGTGGATCTGGAC
TTGTGTTTTTGTCTTCAGAATTCGGTGTGAAAGGTAGATATAGGAGCTGTGCTCTCCTTGTCAACAAAGGGGAA
AAACTAAAAGAATACGACCACTAATGGTTTCTATTTTCTGGTTATTATATCCAGCAAAGCAAAAAGGACAGAAT
GGAAGAAGTGCTCTCTGTACACCTTGCTGTTGCTGTTGCTCTTCTACTTCTGTTGGCGGCAAAGCGCACCCCCG
CTGTTGCGGTTCCTAGCCCTCAGTGCCAACGGCAATGCGGCGGCGTTGACATTGTGTTTCCGTTCGGCATCGGT
GACAACTGCTCGCTATCAGCAGGATTTAACGTCAGCTGCCAGGAGGTCCAAGTCCAAGGTGGCGTCGTCTACA
GGCCATTCCTCGGCAACGTTGAGCTGCTCAACATCTCCTTGATACATGGCACGATCCGGGAGCTGAACCACATC
TCGACATACTGCTACGACTCCTCCTCGAGCTCCATGGAGCTCAGTACTTGGTGTTTTGACGCGAGCGAAACCCC
GTTCCGGTTCTCGGACGTCCAGAACAAGTTCACCGCCATAGGGTGCCAGACCCTCGCCTACATCATGGACAAC
ACCGACAAGAGCTACCAGAGTGGGTGCGTCTCGACGTGCCAGAGTCTGTCAAACCTAGCGGACGGATCCTGC
TCCGGCATAGGCTGCTGCCAGACAGACATACCCAAGGGGATGGGCTT
>2_28200813_28201813
CTACAACGTAAGCTTCGACACAGGCCAAATTTCGCCGTCGGGCCTCGGCCGCTGCAGCTACGCCGTGCTGATG
GAGGCGGCGGCGTTCAGCTTCAGGACCACGTACATCGACACGACAGATTTCAACGACACGAGCGAGTACTGG
GCAGGTACCCGTGGTGATGGACTGGGCGATAAGAGATGGCCTGCCGTCGTGTGAGCTCGCCACAAGGAACGA
AACGGGCACTTACGCGTGCCGCAGCGGCAACAGCAAGTGTGTGGATTCCCCCAATGGGCCAGGGTATCGGTG
CAACTGCTCCGATGGGTACGAAGGCAACCCATATCTTTCAGATGGATGCCATGGTGAGTGACAAACTATCAAC

TGTCCCACAAAATATACAGTAATATATTCCCACAAAATATACAGTAATATATTCAATCTATAGACTTTGAAATCT
ATCCTTCTCATCATACTGCAGATGTCGATGAGTGCAAGAAGAACAGTCCATGCAGTAGGAGGTGTTTGCCACA
ACACGGTCGGAGCGCACAAGTGTTCTTGTCGAGCAGGAAGAAGGCTTAACAAGCAAAACAATACGTGTGACC
CTGATACCACATTAATAACAGGTAACACGCTAGCTGAGATCAGCAGCTGCATGCTACCACTGTTTATTTCTCCCT
GTATATATATATATATATATATATATATATATATATATATATATATATATATAAAAGAAGGAAAGCGCGTGCTCTTATC
TGACGATGCTGCTACGCATGGGTTACAATCGGCTTCCTTGTTCTCGTAATCTTCTCCTCCTTCGGATACATGATC
CTCCAGAAGAGAAAACTGAACCAGGTTAAGCAAGATCATTTTCGGCAGCACGGAGGCATGATTTTGTTCGAGA
GGATGAGATCAGAAAATGGCCTTGCTTTCACGGTGTTCAGTGAAGCTGAGCTTGTAAAAGCTACAGACAGCTA
CGACAAGAGCAGAATAATTGGGAAGGGAGGCCACGGGACAGTCTACAA
>2_28201814_28202814
AGGGATAGTCAAGGGCAACGTGCCGATCGCGATTAAGCGATGTGCGCTTATTGACGAGAGGCAGAAGAAAG
AGTTTGGTCAGGAGATGCTGATACTCTCCCAGATCAACCACAAGAACATCGTCAAGCTCGAGGGCTGTTGCCT
CGAGGTGGAAGTTCCAATGCTGGTCTACGAGTTCGTTCCAAATGGCACCCTGTACGAACTCATCCATGGCAAG
AACCAAGCGCTACAGATCCCCTTCAGCACCCTACTGAGGATCGCCCATGAAGCAGCAGAGGGCCTCAGTTTTC
TGCACTCGTACGCGTCTCCTCCGATCATCCATGGCGACGTGAAGAGCGCCAACATACTTCTTGACGGCAACTAC
ATGGCCAAAGTGTCAGATTTTGGGGCCTCCATACTAGCGCCGTCCGACAAAGAGCAGTATGTCACGATGGTTC
AAGGTACCTGTGGATACCTCGACCCTGAATACATGCAGACATGCCAACTGACTGAGAAGAGTGACGTCTACAG
CTTCGGCGTCATCCTTCTTGAGGTACTCACCGGCCAAGAGCCTCTCAAGTTGGATGGGCCTGAGACGCAGAGA
AGCTTGTCATCGAAGTTCCTGTCCGCTATGAAGGAGAACAATCTTGATGTGATATTGCCGAGCCACGTGAATG
GTGGACAAGGGAGCAATGAACTGATCAGAGGGCTCGCAGAGCTAGCCAAGCAGTGCCTGGACATGTGTGGC
TGCAACAGACCATCCATGAAGGAGGTTGCCGATGAGCTTGGTAGATTGAGGAAGCTTTCGCTGCATCCTTGGG
TACAGATCGATGCAGAGATGATAGAGTCCCAAAGCCTTCTTAGTGGGACAACGACTGCTAGCTTTGAAATAGA
AGTTGGTACAACTGGGTATCCTACACAGGAAGCTGAGAACCTGCCCATGAACCCAAGAAGTTCGTACTATGCG
AGGTAGGTGATTAGCGTATGTTGTTGTGTACTCCGTTGAAATTTCTGCTGCTAGG

Reads for CML333 for Kinase B:

>35_1
TTGCATGCCTGTATTTGTCTGATTGCTTCATGTTTTAATTTGTCAACCACACAGTCTGTGGGTGTTGTTCATCTGG
>35_2
GTACGTACTCTATTTCCAAACAAATGTGCTTTGTTTTCAAATGAACTAGTGCTTTGTTTTCAAAGGCAAACAAGG
C

>62_1
GAGATGTTGCATGCCTGTATTTGTATGATTGCTTCATGTTTTAATTTGTCAACCACACAGTCTGTGGGTGTTGTT
CATTTGGCT
>62_2
AATGTGCTTTGTTTTCAAATAAACTAGTGCTTTGTTTTCAAAGGCAAACAAGGCATCACACAGAGGGAACAGAA
ACCNGATGAA

>55_1
AATCAATTGGCATGACTACTTAACTATATATGCATGCTCTGGTTGTGAACTCATTCAATGAATAACTAAGGTTTT
A
>55_2

AAGATGCAGGCATGCAGCATGTTGGAGCAGAGCTGACCAGGCTAACACTAAATCAAGTGAACAAGCCACTGA
TAGC

>48_1
ATACATAGCCCATGGCATTCTAATTGGTTCCCTGAATGTGGCATGGAACTCATTCAATCAATTGGCATGACTAC
TT
>48_2
CTGACCAGGCTAACACTAAATCAAGTGAACAAGCCACTGATAGCATGCACATAAAAAATAAAACCTTAGTTATT
CA

>43_1
AGAATTCAGTGTGAAAGGTAGATATAGGAGCTGTGCTCTCCTTGTCAACAAANGGGAAAAACTAAAAGAATA
CGAC
>43_2
CGCTTTGCCGCCAACAGAAGTAGAAGAGCAACAGCAACAGCAAGGTGTGCAGAGAGCACTTCTTCCATTCTGT
CCT

>6_1
TCTGTTGGCGGCAACGCGTACCCGCGCTGTTTCGGGTCCTAGNCCTCAGGTTCAAACGCAAAGTGTCGGGGCG
CTGNGTGTGTT
>6_2
CAGCTCAACGTTGCCGATGAATGGCNTGNAGACGACGCCACCACATTGGACTTGGACCTCCTGGCGNGTGCC
GTTAAGACCTTC

>17_1
GGCGTCGTCTACAGGCCATTCCTCGGCAACGTTGAGCTGCTCAACATCTCCTTGATACATGGCACGATCCGGGA
GC
>17_2
GCGTCAAAACACCAAGTACTGAGCTCCATGGGGCTCGAGGAGGGGTCGGAGCGGTAGGTGGGGATGTGGGT
CAGAT

>18_1
TCGTCTACAGGCCATTCCTCGGCAACGTTGAGCTGCTCAACATCTCCTTGATACATGGCACGATCCGGGAGCTG
AA
>18_2
GGCGGTGAACTTGTTCTGGACGTCCGAGAACCGCAACGGGGTTTCGCTCGCGTCAAAACACCAAGCACTGAG
CTCC

>19_1
GCCATTCCTCGGCAACGTTGAGCTGCTCAACATCTCCTTGATACATGGCACGATCCGGGGGATGAACCACATCT
CGACATACTG
>19_2
TGGCGGACAAACTCTTCCTGGCTGCCCACAAACGGAACAAGGGGGGTCCCGCCGCCACAACCACAACAATCA
GCGCCCTCGCCG

>47_1

TCGAGCTCCATGGAGCTCAGTACTTGGTGTTTTGACGCGAGCGAAACCCCGTTCCGGTTCTCGTACGTCCAGAA
CAAGATCACC
>47_2
GTCGAGACGCACCCACTCTGGTAGCTCTTGTCGGTGTTGTCCCTGATGTAGGCGAGGGGCTGGCACCCTATGT
GGGCTGGGCAG

>16_1
CGAGCGAAACCCCGTTCCGGTTCTCGGACGTCCAGAACAAGTTCACCGCCATAGGGGGCCAGACCCTCGCCTA
CATCATGGACA
>16_2
CCGGAGCAGGATCCGTCCGCTAGGTCTGNCAAACTCTGGCACGTCGAGACGCACCCACTCTGGTAGCTCTTGT
CGGTGGTGGTC

>61_1
TTCTCGGACGTCCAGAACAAGTTCACCGCCATAGGGTGCCAGACCCTCGCCTACATCATGGACAACAACGACA
AGAGCTAACAG
>61_2
GAAGCCCATCCCCTTGGGTATGTCTGTCTGGCAGCAGCCTATTCCGGAGCAGGATCCATCCACGAGGCCTGAC
AAACCACTGAA

>12_1
AAGTTCACCGCCATAGGGTGCCAGACCCTCGCCTACATCATGGACAACACCGACAAGAGCTACCAGAGTGCGT
GCGTCTCGACG
>12_2
CTGTCTGGCAGCAGCCTATGCCGGAGCAGGATCCGTACGCTAGGTCTGACAAACTCTGGCACGTCGAGACGCA
CCCACTCTGGT

>50_1
GCAGCCTATGCCGGAGCAGGATCCGTCCGCTAGGTCTGACAAACTCTGGCACGTCGAGACGCACCCACTCTGG
TAG
>50_2
AGCTCCATGGAGCTCAGTACTTGGTGTTTTGACGCGAGCGAAACCCCGGTCCGGGTCTCGGACGTCCAGAACA
AGG

>7_1
CGTCAGCTGCCAGGAGGTCCAAGTCCAAGGTGGTGGCGTTGTCCACAGGCCACTACTCCGCAACGTTGCGCG
GGTGACAATCTC
>7_2
GCGTCAAAACACCAAGTACTGAGCTCCATGGAGCTCGAGGAGGAGTCGTAGCAGTATGTCGAGATGTGGTCA
AGCCCCCGGGCG

>13_1
GCCAACGGCAATGTGGCGGCGTTGACATTTTGTTTTCGTTTCGCGTCGGTGGTGACTGCTCCCCATCACCAGGG
ATTTACGCAC
>13_2

CAGCTCCCGGATCGTGCCATGTATCAAGNAGATGTTGAGCAGCTCAACGTTGCCGAGGAATGGCCTGTAGAC
GACGCCACCACC

>60_1
CCAGGCTAACACTAAATCAAGTGAACAAGCCACTGATAGCATCCACATAAAAAATAAAACCTTACTTATTCATT
GAATGAGTTC
>60_2
GGCATTCTAATTGGTTCCCTGAATGTGGCATGGAANTCATTCAATCAATTGGCATGACTACTTAACTATATATG
CATGCTCTGT

>37_1
TGAATATATTACTGTATATTTTGTGGGACAGTTGATAGTTTGTCACTCACCATTGCATCCATCTGCAAGATATGG
GTTGCCTTT
>37_2
GCGTGCCGCAGCGGCAACAGCAAGTGTGTGGAATCCCCCAATGGGCCAGGGTATCGGTGCAACTGCTCCGAT
GGGTACGAAGGC

>1508_1
TCATCGACATCTGCAGTATGATGAGAAGGATAGATTTCAAAGTCTATAGATTGAATATATTACTGTATATTTTGT
G
>1508_2
GGCCAGGGTATCGGTGCAACTGCTCCGATGGGTACGAAGGCAACCCATATCTTTCAGATGGATGCCATGGTGA
GTG

>481_1
GAAGGCAACCCATATCTTTCAGATGGATGCCATGGTGAGTGACAAACTATCAACTGTCCCACAAAATATACAGT
AA
>481_2
CTGTTCTTCTTGCACTCATCGACATCTGCAGTATGATGAGAAGGATAGATTTCAAAGTCTATAGATTGAATATAT
T

>1606_1
GGTATCAGGGTCACACGTATTGTTTTGCTTGTTAAGGTTTCTTCCTGCTCGACAAGAACACCTGTACGCTCCGAC
C
>1606_2
CAAAATATACAGTAATATATTCAATCTATAGACTTTGAAATCTATCCTTCTCATCATACTGCAGATGTCGATGAG
T

>1662_1
TTCTTCCTGCTCGACAAGAACACCTGTACGCTCCGACCGTGTTGTGGCATACACCTCCTACTGAAGGGCATGGA
CT
>1662_2
ATATACAGTAATATATTCAATCTATAGACTTTGAAATCTATCCTTCTCATCATACTGCAGATGTCGATGAGTGCA
A

>1155_1

AAACCTTAACAAGCAAAACAATACGTGTGACCCTGATACCACATTAATAACAGGTAACACGCTAGCTGAGAAT
GAG
>1155_2
ACAAGGAAGCCGATTGTAACCCCTGCGTAGCAGCATCGTCAGATAAGAGCACGCGCTTTTCTTCTTCTCTATAT
AC

>2054_1
GTAGCTGTCTGTAGCTTTTACAAGCTCAGCTTCACTGAACACCGTGAAAGCAAGGCCATTTTCTGATCTCATCCT
C
>2054_2
CCTTCGGATACATGATCCTCCAGAAGAGAAAACTGAACCAGGTTAAGCAAGATCATTTTCGGCAGCACGGAGG
CAT

>600_1
GATCATTTTCGGCAGCACGGAAGCATGATTTTGTTCGAGAGGATGAGATCAGAAAATGGCCTTGCTTTCACGG
TGT
>600_2
CCCATGGCCTCCCTTCCCAATTATTCTGCTCTTGTCGTAGCTGTCTGTAGCTTTTACAAGCTCAGCTTCACTGAAC

>236_1
AGATCAGAAAATGGCCTTGCTTTCACGGTGTTCAGTGAAGCTGAGCTTGTAAAAGCTACAGACAGCTACGACA
AGAGCAGAATA
>236_2
TATCAGCATCTCCTGACCAAACTCTTTCTTCTGCCTCTCGTCAATAAGCGCACATCGCTTAATCGCGATCGGCAC
GTTGCCCTT

>53_1
GGTGGTGATGGACTGGGCGATAAGACATGATGGCGCGCCGTCGTGTGAGCTCGCCACAAGGAACGAGAGGG
GCACTTACGCGCG
>53_2
TCGTACCCATCGGAGCAGTTGCACCGATACCCTGGCCCATTGGGGGATTCCACACACTTGCTGTTGCCGCTGCG
GCAACGCGTA

>437_1
AGCTGATCTCATTCTCAGCTAGCGTGTTACCTGTTATTAATGTGGTATCAGGGTCACACGTATTGTTTTGCTTGT
TACGGTTTT
>437_2
CCTTCTCATCATACTGCAGATGTCGATGNGTGCAAGAAGAACAGTCCATGCCCTTCAGTAGGAGGTGTATGCA
ACACCACGGTC

>2207_1
NAGGCCATTTTCTGATCTCATCCTCTCGAACAAAATCATGCCTCCGTGCTGCCGAAAATGATCTTGCTTAACCTG
G
>2207_2
TGTATATATATAAGAAGGAAAGCGCGTGCTCTTATCTGACGATGCTGCTACGCAGCGGTTACAATCGGCTTCCT
TG

>176_1
CAAACTCTTTCTTCTGCCTCTCGTCAATAAGCGCACATCGCTTAATCGCGATCGGCACGTTGCCCTTGAATATAC
CTTTGTAGA
>176_2
GAAGCTGAGCTTGTAAAAGCTACAGACAGCTACGACAAGAGCAGAATAATTGGGAAGGGAGGCCATGGGAC
AGTCTACAAAGGG

>1711_1
AGTGGTAGCATGCAGCTGACCTCATTCTCAGCTAGCGGGTTACCTGTTATTAATGTGGTATCAGGGTCACACGT
AT
>1711_2
CCATGCCCTTCAGTAGGAGGTGTATGCCACAACACGGTCGGAGCGTACAGGTGTTCTTGTCGAGCAGGAAGA
AACC

>159_1
GATACTCTCCCAGATCAACCACAAGAACATCGTCAAGCTCGAGGGCTGTTGCCTCGAGGTGGAAGTTCCAATG
CTGGTCTACGG
>159_2
TACGAGTGCAGAAAACTGAGGCCCTCTGCTGCCTCATGGGCGATCCTCAGTAGGGTGCTGAAGGGGATCTGT
AGCGCTTGGTTC

>35_1
GCCCACGAAGCAGCAGAGGGCCTCAGTTTTCTGCACTCGCACGCGTCTCCTCCGATCATCCATGGCGACGCGA
AGAGCGCCAAC
>35_2
GGGTCGAGGTATCCACAGGTACCTTGAACCATTGTGACATACTGCTCTTTGTCGGATCGCGCTAGTATGCACGC
CCCAAAATCT

>65_1
AAGCAGCAGAGGGCCTCAGTTTTCTGCACTCGTACGCGCCTCCTCCGATTATCCATGGCGACGTGAAGAGCGC
CAA
>65_2
CAAGAAGGATGTTGGCGCTCTTCACGTAGCCATGGATGAACGGTGGAGACGCGTAGGAGCGCAGAAGAATG
TGGCC

>74_1
CGGTGAGGACCTCAAGAAGGATGACGCCGAAGCTGTAGACGTCACTCTTCTCAGTCAGTTGGCATGTCTGTAT
GTA
>74_2
TACATGGCCAAAGTGTCAGATTTTGGGGCCTCCATACTAGCGCCGTCCGACAAAGAGCAGTATGTCACGATGG
TTC

>138_1
TTCGATGACAAGCTTCTCTGCGTCTCAGGCCCATCCAACTTGAGAGGCTCTTGGCCGGTGAGGAGCTCAAGAA
GGATGACGCCG

>138_2
CTCCATACTAGCGCCGTCCGACAAAGAGCAGTATGTCACGATGGTTCAAGGTACCTGGGGATACCTCGACCCT
GAATAGATGCA

>59_1
AGAGCAGTATGTCACGATGGTTCAAGGTACCTGGGGATACCTCGACCCTGAATACATGCAGACATGCCAACTG
ACT
>59_2
TTCTCTGCGTCTCAGGCCCATCCAACTTGAGAGGCTCTTGGCCGGTGAGGACCTCAAGAAGGATGACGCCGAA
GCT

>91_1
AGGAACTTCGATGACAAGCTTCTCTGCGTCTCAGGCCCATCCAACTTGAGAGGCTCTTGGCCGGTGGGGACCT
CAA
>91_2
AANGTACCTGGGGATACCTCGACCCTGAATACATGCAGACATGCCAACTGACTGAGAAGAGTGACGTGTACA
GCTT

>22_1
ACGGATGCAGCGAAAGCTTCCTCAATCTACCAAGCTCATCGGCAACCTCCTTCATCGATGGTCTCTTGCAGCCA
CACATGTGCT
>22_2
CCGAGCCACGTGAATGGTGGACAAGGGAGCAATGAACTGATAAGAGGGCTCGCAGAGCTAGCCAAGAAGTG
CCTGGAGATGTGG

>39_1
GCAGCGAAAGCTTCCTCAATCTACCAAGCTCATCGGCAACCTCCTTCATGGATGGTCTGTTGCAGCCACACATG
TCCAGGCACC
>39_2
GAGCCACGTGAATGGTGGACAAGGGAGCAATGAACTGATCAGAGGGCTCGCAGAGCTAGCCAAGCAGTGNC
TGGACATGTGTGG

>107_1
AGAGCTAGCCAAGCAGTGCCTGGACATGTGTGGCTGCAACAGACCATCCATGAAGGAGGTTGCCGATGAGCT
TGGG
>107_2
GCTAGCAGTCGTTGTCCCACTAAGAAGGCTTTGGGACTCTATCATCTCTGCATCGATCTGTACCCAAGGATGCA
GC

>87_1
ACCAACTTCTATTTCAAAGCTAGCAGTCGTTGTCCCACTAAGAAGGCTTTGGGACTCTATCATCTCTGCATCGAT
CTGTACCCA
>87_2
GCCAAGCAGTGCCTGGACATGTGTGGCTGCAACAGACCATCCATGAAGGAGGTTGCCGATGAGCTTGGGAGA
TTGAGGAAGCTT

>114_1
ANGTGTGGCTGCAACAGACCATCCATGAAGGAGGTTGCCGATGAGCTTGGTAGAGGGAGGAAGCTTTNNNN
NNNNN
>114_2
CAGTTGTACCAACTTCTATTTCAAAGCTAGCAGTCGTTGTCCCACTAAGAAGGCTTTGGGACTCTATCATCTCTG
C

>126_1
CATGAAGGAGGTTGCCGATGAGCTTGGTAGATTGAGGAAGCTTTCGCTGCATCCTTGGGTACAGATCGATGCA
GAG
>126_2
GTTCATGGGCAGGTTCTCAGCTTCCTGTGTAGGATACCCAGTTGTACCAACTTCTATTTCAAAGCTAGCAGTCG
TT

>143_1
CGCATAGTACGAACTTCTTGGGTTCATGGGCAGGTTCTCAGCTTCCTGTGTAGGATACCCAGTTGTACCAACTT
CTATTTCACA
>143_2
TGGTAGATTGAGGAAGCTTTCGCTGCATCCTTGGGTACAGATCGATGCAGAGATGATAGAGTCCCAAAGCCTT
CTTGATGGGAC

>67_1
CTTGGGTACAGATCGATGCAGAGATGATAGAGTCCCAAAGCCTTCTTAGTGGGACAACGACTGCTAGCTTTGA
AAT
>67_2
AGTTATCCTAGGAGTAGAAATTTCAACTGAGTACACAACAACATACGCTAATCACCTACCTCGCATAGTACGAA
CT

>64_1
AGATCGATGCAGAGATGATAGAGTCCCAAAGCCTTCTTAGTGGGACAACGACTGCTAGCTTTGAAATAGAAGT
TGG
>64_2
AGTACACAACAACATACGCTAATCACCTATCTCGCATAGTACGAACTTCTTGGGTTCATGGGCAGGTTCTCAGC
TT

>129_1
NTAGGAGTAGAAATTTCAACTGAGTACACAACAACATACGCTAATCACCTACCTCGCATAGTACGAACTTCTTG
GG
>129_2
CCAAAGCCTTCTTAGTGGGACAACGACTGCTAGCTTTGAAATAGAAGTTGGTACAACTGGGTATCCTACACAG
GAA

>149_1
GGGACAACGACTGCTAGCTTTGAAATAGAAGTTGGTACAACTGGGTATCCTACACAGGAAGCTGAGAACCTG
ACCATGAACCCA
>149_2

TAACAGAACTGCTAGCTTGCAGTTATCCTAGGAGTAGAAATTTCAACTGAGTACACAACAACATACGCTAATCA
CCTACCTCGC

>81_1
GATTTGTGTCAATATGCAGTATACCTTTGATAACAGAACTGCTAGCTTGCAGTTATCCTCGCACTAGAAATTTCA
CCTGAGTAC
>81_2
GAGAACCTGCCCATGAACCCAAGAAGTTCGTACTATGCGAGGTAGGTGATTAGCGTATGTTGTTGTGCACTCA
GTTGAACTTTC

>17_1
CAAACTCTTTCTTCTGCCTCTCGTCAATAAGCGCACATCGCTTAATCGCGATCGGCACGTTGCCCTTGAATATAC
CTTTGTAGA
>17_2
GAAGCTGAGCTTGTAAAAGCTACAGACAGCTACGACAAGAGCAGAATAATTGGGAAGGGAGGCCATGGGAC
AGTCTACAAAGGG

>95_1
AAGGATGACGCCGAAGCTGTAGACGTCACTCTTCTCAGTCAGTTGGCATGTCTGCATGTATTCAGGGTCGAGG
TATCCCCAGGA
>95_2
TGAAGAGCGCCAACATCCTTCTTGATGGCAACTACATGGCCAAAGTGTCAGATTTTGGGGCCTCCATACTAGC
GCCGTCCGACG

>117_1
AGAACAATCTTGATGCGATCTTGCCGAGCCACGTGAATGGTGGACAAGGGAGCAATGAACTGATCAGAGTGC
TCGC
>117_2
AGCTTCCTCAATCTACCAAGCTCATCGGCAACCTCCTTCATGGATGGTCTGTTGCAGCCACACATGTCCAGGCA
CT

>49_1
ATCATCTCTGCATCGATCTGTACCCAAGGATGCAGCGAAAGCTTCCTCAATCTACCAAGCTCATCGGCAACCTC
CT
>49_2
TCTTGCCGAGCCACGTGAATGGTGGACAAGGGAGCAATGAACTGATCAGAGGGCTCGCAGAGCTAGCCAAGC
AGTG

>115_1
CTGTACCCAAGGATGCAGCGAAAGCTTCCTCAATCTACCAAGCTCATCGGCAACCTCCTTCATGGATGGTCTGT
TG
>115_2
CTTGCCGAGCCACGTGAATGGTGGACAAGGGAGCAATGAACTGATCAGAGGGCTCGCAGAGCTAGCCAAGCA
GTGC

>56_1

GTAGACCAGCATTGGAACTTCCACCTCGAGGCAACAGCCCTCGAGCTTGACGATGTTCTTGTGGTTGATCTGG
GAGAGAATCAG
>56_2
CATGGGACAGTCTACAAAGGGATAGTCAAGGGCAACGTGCCGATCGCGATTAAGCGATGTGCGCTTATTGAC
GAGAGGCAGAAG

>32_1
GATGGTCTGTTGCAGCCACACATGTCCAGGCACTGCTTGGCTAGCTCTGCGAGCCCTCTGATCAGTTCATTGCT
CCCTTGTCCC
>32_2
ACGCAGAGAAGCTTGTCATCGAAGTTCCTGTCCGCTATGAAGGAGAACAATCTTGATGCGATCTTGCCGAGCC
ACGTGAATGGG

>119_1
AGCTTGTCATCGAAGTTCCTGTCCGCTATGAAGGAGAACAATCTTGATGCGATCTTGCCGAGCCACGTGACTG
GCG
>119_2
TTGCAGCCACACATGTCCAGGCACTGCTTGGCTAGCTCTGCGAGCCCTCTGATCAGTTCAATGCTCCCTTGTCCA
C

Declaration of Authorship

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.

I declare that this thesis has not been submitted for a higher degree to any other University or Institution.

Curriculum Vitae

Name	Franziska Irmer
Adresse	Alte Wiesenstr 12 07749 Jena Franziska.irmer@web.de
Date and Place of Birth	24.02.1988, Jena
Citizenship	German

Schools	1993-1997 Grundschule Heinrich Heine, Jena 1997-2006 Gymnasium Am Anger, Jena Abitur: final grade 1,5
University	October 2006 – September 2009 Biotechnology FH Jena: Bachelor of Engineering , final grade 2.5 Bachelor thesis: *„Untersuchungen zum Mechanismus der Regeneration des aktiven Zentrums von B-Typ-Methioninsulfoxidreduktasen durch Thioredoxine“* October 2009 – January 2012 „Pharmaceutical Biotechnology“ MLU Halle-Wittenberg Master of Science, final grade 2.2 Master Thesis: *“Investigations on the supposed autoproteolysis of Microbial Transglutaminase”*
Work experience	July 2012 – December 2016 PhD student at the University of Halle-Wittenberg Field of dissertation: Biotechnology Since March 2017 Lab Technician at the Max Planck Institute for the Science of Human History, Jena

I hereby certify that I have no criminal record or ongoing investigations.