Genetic diversity of *Lymantria* spp. in relation to their host trees – a cDNA-AFLP approach

Dissertation

zur Erlangung des akademischen Grades

„doctor rerum naturalium“ (Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät

der Friedrich-Schiller- Universität Jena

von Diplom Biologin Kirsten Evertz

geboren am 02.07.1978 in Mönchengladbach

Gutachter:

1. Prof. Dr. David G. Heckel, Max Planck Institut für Chemische Ökologie, Jena

2. Prof. Dr. Annette Reineke, Forschunganstalt Geisenheim, Geisenheim

3. Prof. Dr. Wolfgang W. Weisser, Technische Universität München, Freising-Weihenstephan

Tag der öffentlichen Verteidigung: 29. Oktober 2012

Table of Contents

Summary **1**

Zusammenfassung **3**

Abbreviations **5**

1. General introduction **6**

1.1. Plant-insect interactions 6

1.2. Chemical defence and secondary compounds 6

1.3. Coevolution and genetic changes as a consequence of defence compounds 7

1.4. Investigations into coevolution and genetic changes 8

1.5. Biology and natural history of the study species 8

1.5.1. *Lymantria dispar* 8

1.5.2. *Lymantria monacha* 9

1.5.3. *Tortrix viridana* 9

1.6. Field experiments 10

1.7. Molecular techniques used 10

1.8. Aims and contents of the research study 11

2. Host adaptation of *Lymantria* species: a cDNA-AFLP approach **13**

2.1. Abstract 13

2.2. Introduction 13

2.3. Material and Methods 16

2.4. Results 22

2.5. Discussion 32

3. Differential gene expression in generalist and specialist lepidopteran herbivores in response to feeding on different phenotypes of oak trees **36**

3.1. Abstract 36

3.2. Introduction 37

3.3. Material and Methods 38

3.4. Results 44

3.5. Discussion 55

4. Herbivore (*Lymantria dispar*) induced responses in *Quercus robur* and *Picea abies* on living trees: a natural experiment 58
4.1. Abstract 58
4.2. Introduction 58
4.3. Material and Methods 59
4.4. Results 64
4.5. Discussion 68
5. Identification of SNPs in transcripts of different *Lymantria dispar* populations via 454 pyrosequencing 72
5.1. Abstract 72
5.2. Introduction 72
5.3. Material and Methods 74
5.4. Results 77
5.5. Discussion 86
6. Major Discussion 89
6.1. A synthesis of the research project 89
6.2. *Lymantria* ssp. on different food sources 89
6.3. Two different oak types 90
6.4. Field experiment 92
6.5. SNP-454 project 93
References 95
Acknowledgements 105
Selbständigkeitserklärung 107
Declaration of contributions of the thesis 108
Curriculum Vitae 108

Summary

Plant-insect interactions are a wide and diverse field of study. Abiotic and biotic factors influence the dynamic interaction between plants and their insect herbivores.

The main focus of this thesis is the genetic basis of adaptation of tree-feeding insect larvae in the genus *Lymantria* and *Tortrix* to their host trees. Gene expression at the mRNA level of insect larvae feeding on different tree species was analysed, and similarly gene expression in the host trees undergoing herbivory in the field was measured. Gene diversity at the genomic DNA level was assessed in a population genetic study of one *Lymantria* species. Patterns of adaptation were evaluated in the light of these results.

In the first study, two *Lymantria* species *L. dispar* and *L. monacha* fed on oak, spruce, or artificial diet responded with different patterns of weight gain and of gene expression, as measured by the cDNA-AFLP technique. Sequencing of AFLP fragments revealed genes involved in digestion and detoxification upregulated in growing larvae, and other genes associated with a starvation response in larvae that did not grow.

The second study investigated the response of insects to feeding on two different phenotypes of oak trees. Field observations over a number of years had identified individual "sensitive" trees of *Quercus robur* that were usually heavily defoliated by the specialist *Tortrix viridana* (green leaf oak roller) and other nearby "tolerant" trees that were not. Branches from the two types of trees were offered to the generalist herbivore *L. dispar* and the specialist *T. viridana.* Younger *L. dispar* larvae adapted more successfully to the tolerant oaks than did the older *L. dispar* larvae. No genes with digestive functions were identified in *T. viridana* feeding on tolerant oaks. Only genes with unknown functions were identified in the older larvae of *L. dispar*. The younger *L. dispar* larvae were able to upregulate digestive enzymes in response to whatever defenses were possessed by the tolerant oaks, and this transcriptional response contributed to the overall ability to grow even more on tolerant oaks than on the sensitive oaks, which did not induce such a response.

The third study investigated differential gene expression patterns in intact or detached branches of *Quercus robur* and *Picea abies* exposed to feeding by *L. dispar* under natural conditions in the forest canopy. The cDNA-AFLP technique was used to detect genes differentially expressed in the intact or detached branches. Photosynthesis genes were mainly upregulated and suggest defense response mechanisms switched on after *L. dispar* fed on leaves of intact branches. Only two genes, which had no direct connection to photosynthesis, were identified as upregulated in the detached leaves of oak. Upregulation of an ethylene response factor in the intact treatment suggests operation of a signalling pathway within the intact plant.

The fourth study examined the intraspecific genetic diversity of *L. dispar* populations from different European sites and one Asian site. The 454 pyrosequencing technique in a high

throughput approach was used to produce a large amount of genomic information by sequencing a pool of AFLP fragments. F_{ST} values indicating genetic diversity were calculated from SNP analysis conducted on sequences that were identified from a cDNA library of *L. dispar*. As in a previous study on another widespread lepiodpteran species, the F_{ST} values were very high. Another analysis restricted to genes identified by homology to the domestic silkworm *Bombyx mori* produced somewhat lower F_{ST} values, but not enough data to for pairwise comparison of all populations. The difference may be caused by a high frequency of repetitive DNA in the AFLP fragments due to the particular restriction enzyme used in the AFLP analysis. Candidate genes were identified showing a pattern of high diversity across populations. The comparison between the two methods of analysis indicates that interpretation of high-throughput methods which generate huge amounts of data requires caution.

In general a diverse mixture of techniques and methodologies was used to generate new genetic information on plant-insect interactions. Digestive adaptations in the insects were identified from the herbivory laboratory studies and the host plant responses were identified in the field study. The cDNA-AFLP technique enabled the combination and challenge of field and laboratory experiments to be met and produced a detailed overview in respect of defense strategies on both sides of the interaction.

Zusammenfassung

Pflanzen-Insekten-Interaktionen sind Gegenstand eines breiten und vielfältigen Bereichs. Abiotische und biotische Faktoren beeinflussen die Dynamic der Interaktionen zwischen Pflanzen und Insektenschädlingen.

Der Mittelpunkt dieser Forschungsarbeit liegt in der genetischen Grundlage von Anpassung von Forstschädlingslarven der Art *Lymantria* und *Tortrix* und deren Wirtspflanzen. Die Genexpression in der mRNA von Insektenlarven, die auf verschiedenen Nahrungsquellen gefressen haben, wurde analysiert und ähnliche Genexpressionen in der Wirtpflanze, die von Herbivoren gefressen worden, wurden im Freiland gemessen. Die genetische Vielfalt auf genomischer DNA Ebene wurde im Rahmen einer populationsgenetischen Studie an einer *Lymantria*-Art durchgeführt. Anpassungsmuster wurden anhand dieser Ergebnisse evaluiert.

In der ersten Studie reagierten zwei *Lymantria*-Arten, *L. dispar* und *L. monacha*, die auf Eiche, Fichte oder auf künstlicher Diät fraßen mit verschiedenen Gewichtszunahmen und mit Geneexpressionen, welche mit der cDNA-AFLP-Technik ermittelt wurden. Die Sequenzierung der AFLP Fragmente zeigte in den wachsenden Larven Gene, die involviert in die Verdauungs- und Entgiftungs-Reaktionen der Larven sind. Bei den Larven, die nicht wuchsen, zeigte sich jedoch, dass andere Gene eventuell auf einen Verhungerungseffekt basierend hochreguliert waren.

Die zweite Study untersuchte die Reaktion von Insekten, die an zwei verschieden Phänotypen von Eichen fraßen. Freiland Beobachtungen über mehrere Jahre hatten einzelne „sensitive“ Bäume (*Quercus robur*), welche unter einer schweren Entlaubung des Spezialisten *Tortrix viridana* (Grüner Eichenwickler) litten und andere Bäume, welche fast „tolerant“ demgegenüber waren, identifiziert. Zweige der beiden Baumtypen wurden dem Generalisten *L. dispar* und dem Spezialisten *T. viridana* angeboten. Jüngere *L.dispar* Larven passten sich den toleranten Eichen besser an als ältere Larven. Es wurden keinen Gene mit Verdauungsfunktionen bei *T. viridana* Larven erkannt, die an toleranten Eichen gefressen hatten. Es wurden lediglich Gene mit unbekannter Funktion bei den älteren *L. dispar* Larven gefunden. Die Reaktion der jüngeren *L. dispar* Larven auf die Verteidigungsmechanismen der toleranten Eichen bestand darin, Verdauungsenzyme hochzuregulieren. Diese transkriptionelle Antwort trägt ebenso dazu bei, dass Larven auf den toleranten Eichen besser wachsen als auf den sensitiven Eichen, die diese Induktion nicht herbeigerufen haben.

In der dritten Studie wurden differenzierte Genexpressionsmuster von intakten und abgetrennten *Quercus robur* – sowie *Picea abies* – Zweigen nach Raupenbefall von *L. dispar* unter natürlichen Bedingungen in Baumkronen untersucht. Die cDNA-AFLP-Analysetechnik wurde benutzt, um differentiell exprimierte Gene in intakten und abgetrennten Zweigen zu ermitteln. Nachdem *L. dispar* an Blättern von intakten Zweigen gefressen hatte, wurden in der Pflanze hauptsächlich Photosynthese-Gene hochreguliert, was auf Abwehrreaktions-

Mechanismen schließen lässt. In den Blättern von abgetrennten Zweigen der Eiche wurden nur zwei Gene als hochreguliert identifiziert, diese sind jedoch nicht in die Photosynthese involviert sind. Die Hochregulation des Ethylene Response Factor in den intakten Blättern deutet auf einen Signalmechanismus in der lebenden Pflanze hin.

Das vierte Projekt untersucht die intra-spezifische genetische Diversität von *L. dispar*-Populationen aus verschiedenen europäischen und einem asiatischen Standort. Das 454 Pyrosequencing ermöglichte es durch High-throughput-Verfahren eine große Menge an genomischer Information von gepoolten DNA Fragmenten zu produzieren. F_{ST} –Werte (Anzeiger für genetische Diversität) wurden durch die SNP-Analyse von Sequenzen, die von einer cDNA Bibliothek von *L. dispar* identifiziert worden waren, berechnet. Entsprechend einer vorherigen Studie über andere weitverbreitete Lepidopteren, waren die F_{ST} –Werte sehr hoch. Eine andere Analyse, die nur Gene mit einbezieht, welche eine Homologie zum domestizierten *Bombyx mori* (Seidenspinner) zeigten, produzierte kleinere F_{ST} –Werte. Sie lieferte allerdings nicht genug Daten, um einen paarweisen Vergleich von allen Populationen durchzuführen. Der Unterschied der gewonnen F_{ST} –Werte liegt vielleicht in der sich wiederholenden DNA in den DNA Fragmenten, was durch die benutzten Restriktionsenzyme in der AFLP-Analyse bedingt wurde. Kandidatengene die eine hohe Diversität in den Populationen aufweisen, wurden identifiziert. Der Vergleich der beiden Analysearten verdeutlicht, dass die Interpretation großer Datenmengen mit Hilfe von High-throughput-Methoden mit Vorsicht zu betrachten ist.

Im Allgemeinen wurde eine vielfältige Mischung von Techniken und Methoden verwendet, um neue genetische Informationen über Pflanzen-Insekten-Interaktionen zu generieren. Verdauungsanpassungen in Insekten wurden in Herbivore-Laboruntersuchungen und die der Wirtspflanzen im Freilandversuch identifiziert. Die cDNA-AFLP Technik ermöglichte hier die Kombination von Freiland- und Laborexperimenten, um einen detaillierten Überblick über die Abwehrstrategien beider Seiten zu zeigen.

Abbreviations

AFLP	Amplified fragment length polymorphism
cDNA	complementary Deoxyribonucleic acid
CTAB	Cetyl trimethylammonium bromide
d	*Lymantria dispar* diet
DOR	*Lymantria dispar* oak tolerant
DOS	*Lymantria dispar* oak sensitive
DS	*Lymantria dispar* spruce
EDTA	Ethylenediaminetetraacetic acid
Est	expressed sequence tags
GLVs	green-leaf volatiles
JA	Jasmonic acid
m	*Lymantria monacha* diet
MO	*Lymantria monacha* oak
mRNA	messenger Ribonucleic acid
MS	*Lymantria monacha* spruce
O	Oak
Oc	Oak cut
OS	*Tortrix viridana* sensitive
OT	*Tortrix viridana* tolerant
PCR	polymerase chain reaction
qRT-PCR	quantitative real time polymerase chain reaction
S	Spruce
Sc	Spruce cut
SNP	Single nucleotide polymorphism
VOCs	volatile organic compounds

1. General introduction

1.1. Plant-insect interactions

For more than 60 years, the evolution of plant-insect interactions has been a wide and interesting field of research. Milestones in the development of this field is the publication "Plants and butterflies: a study in co-evolution" (Ehrlich & Raven 1964). In this study herbivore adaptation to plant defences through chemical, mechanical and ecological factors that affect food plant choice are described. They described an evolutionary scenario whereby plants in which mutations arose and chemicals toxic to herbivores were produced left more descendants. Insects which in turn could develop ways to detoxify these chemicals would have an advantage over their competitors, and this reciprocal process of adaptation and counteradaption led to an increase in species diversity of both plants and insects. Investigating the molecular basis of how the herbivore detoxifies such compounds helps in understanding the interaction from both sides.

Plants as sessile organisms have evolved mechanical and chemical defences to protect themselves against herbivores. In response, insects have adapted to these defenses with a wide range of strategies derived from detoxifying or digestive enzymes, feeding morphologies and behaviours. Plants respond either directly by producing toxins, physical barriers or plant defensive traits that prevent herbivory, or indirectly by attracting predators or parasites of the herbivore (Howe & Jander 2010). Another indirect defense strategy is internal plant-plant communication, e.g. leaves which are attacked by herbivores alert other plant parts as well (Yi et al. 2009). To understand these complex interaction processes, it is always helpful to study natural communities, instead of just simple laboratory systems. Laboratory studies are never able to replace natural observations or experiments in the field (Ehrlich & Raven 1964).

1.2. Chemical defence and secondary compounds

Ecology recognises several types of defensive strategies including mechanical defense, indirect defense and chemical defense.

Mechanical defenses (a form of direct defense) are those where, for example, thorns, waxes or hairs on the plant epidermis prevent herbivory to some degree. *Acacia* species have developed thorns and spikes on their twigs and branches to dissuade megaherbivores. At the same time these thorns aid in indirect defense.

Indirect defense belongs to the most fascinating protection mechanism from the plants' point of view. For example, the bull's horn *Acacia cornigera* (Linnaeus: Mimosaceae) of Africa provides shelter in its thorns and food in form from nectar for the ant *Pseudomyrmex ferruginea*. The ants protect the tree against phytophagous organisms. Removal of the ants

results in large increases in herbivory that could over time reduce the fecundity of the *Acacia* plant (Janzen 1966).

This thesis investigates the underlying molecular mechanisms of chemical defense. Chemical defense, like all defense mechanisms in plants, can be induced or constitutive. Induced means actively produced in reaction to damage by herbivores or plant injuries. Induced defense can be production of a phytochemical such as in tobacco, where jasmonic acid (JA) induces the synthesis of the alkaloid nicotine (Baldwin 2001; Kessler & Baldwin 2001). JA has two major roles in the response to wounding. First, it triggers the emission of a variety of volatile compounds, which attracts predators and parasitoids. The plant "informs" the enemies of its herbivores where to find the phytophagous insects. This is an indirect and induced mechanism. Second, JA is a signal for the synthesis of different defensive substances in plants. In herbs and crops it is a signal for the synthesis of a flavonoid (Ishihara et al. 2002; Oikawa et al. 2002), and in Solanaceae and in *Alnus glutinosa*, JA induces the synthesis of proteinase inhibitors (Tscharntke 2001; Tscharntke et al. 2001; Zavala et al. 2008). JA has several more defensive functions in plants and is therefore an evolutionary key factor in plant-insect interactions.

Constitutive defenses are always present and might need a much larger amount of investment by the plant than do induced defences (Howe & Jander 2010). A great variety of defensive compounds, such as tannins, furanocoumarin, cyanogenic glycosides, and glucosinolates which are produced irrespective of whether herbivores are present, are constitutive defenses in plants. Trees such as *Quercus robur* (pedunculate oak) and *Picea abies* (European spruce) use e.g. tannins and terpenoids to protect themselves against forest pest species, e.g. *Lymantria dispar* and *Lymantria monacha*. Tannins are changing through the seasons and are more abundant in older leaves of oak, so early feeding brings advantages for insect development (Van Asch & Visser 2007). Defensive compounds can also be induced by quantity regulation as in cotton, where terpenoid accumulation increases after leaf wounding (Opitz et al. 2008). These plant secondary metabolites are unpalatable to most herbivores. Consumption of the plant by the herbivores is thus inhibited. The toxic substances can be absorbed through the gut membrane and then stored in many sites of the body (Nishida 2002). Secondary compounds are metabolized and eliminated by generalist herbivores such as *L. dispar*, whereas specialist herbivores may sequester the toxic substances (Deml 2004).

1.3. Coevolution and genetic changes as a consequence of defence compounds

Due to chemical compounds in plants a constantly escalating defense arms race between herbivores and plants occurs. As a consequence detoxification and digestion traits in the herbivore and constantly adapting defensive strategies of the host lead to a coevolution of both interaction partners. It has been shown that herbivores mostly adapt to plants rather than plants to insect herbivores (Schoonhoven et al. 2005). Evolutionary and chemical ecological questions have become a combined discipline and the genetic bases of

interactions have become the topic of much recent research. Metabolic processes in insects can be identified by molecular and genetic methods, such as sequence analysis. P450s are involved in biosynthesis by the plants and detoxification by herbivores. The identification of P450s (which oxidize exogenous substances) in *Papilio* is one of these examples. The underlying genetic mechanism of P450s involved in furanocoumarin (secondary plant compound) metabolism helps to explain why some *Papilio* species are specialized on particular species of food plants but others are generalists. The ability to detoxify secondary plant compounds may led to dietary specialization and adaptation of *Papilio* species (Berenbaum et al. 1996).

1.4. Investigations into coevolution and genetic changes

It is predicted that moths feeding on different food sources will exhibit different gene expression and modified protein functions resulting in evolutionary consequences. In this study detoxification and digestion related enzymes are of special interest. Detection of induced defense mechanisms without prior genetic knowledge of the herbivore is analysed with the fingerprinting cDNA-AFLP technique. The technique provides a wide range of genomic information by examining a random selection of expressed fragments. Homologies in comparison with entries in gene databanks can link special traits and organisms, even over different trophic levels (Bailey et al. 2006; Heil 2008; Vidal & Tscharntke 2001). Furthermore candidate genes can be used to construct dominant markers for further studies, particularly in non-model organisms (Meudt & Clarke 2007).

1.5. Biology and natural history of the study species

1.5.1. *Lymantria dispar*

Lymantria dispar (common name: gypsy moth) belongs to the family Lymantriidae. The German common name is “Schwammspinner”, because the eggs are imbedded into a wool-like substance by the laying apparatus of the female, so that the egg masses look like a sponge-German “Schwamm”. The adult is white with a black pattern on the wings. The females measure more than 35mm and the males 27mm in length. Females lay up to 2000 eggs in autumn into tree crevices. The larva develops within the egg and enters directly into diapause in the egg until spring. When the red-blue, hairy larvae hatch in April, the young larvae move about and feed in groups. In later instars they feed solitarily. The larvae move wind-borne from tree to tree or spin their way down with a silkline to get to the next tree. *L. dispar* goes through 5-6 instars before pupation occurs in July. After approximately 2 weeks, the adult emerge and mate (Hannemann 2000).

The origin of *L. dispar* lies in Eurasia, it was introduced from France in 1868-1869 into North America and has spread subsequently over eastern North America and Canada (Keena et al. 2008; Robinet & Liebhold 2009). *L. dispar* is a serious worldwide pest of forest trees and therefore very well studied (Keena et al. 2008). A huge amount of information is available about *L. dispar* particularly on the demographics of populations invading North America. Both *Lymantria* strains spread via anthropogenic transport, airplane and cargo ships overseas to worldwide destinations (Keena et al. 2007; Robinet & Liebhold 2009). The

relatively slow spread in North America appears to be because female European *L. dispar* are not active fliers (Keena 2003; Keena et al. 2008; Reineke & Zebitz 1998). The Asian *L. dispar* can fly much better and is able to invade much more rapidly than the European strain. Females from Russia can fly up to 100 km distances (Keena et al. 2007).

Investigations on genetic variation have been carried out previously using the AFLP technique, a fingerprinting technique for analysing genomic DNA (Reineke et al. 1997, 1999; Reineke & Zebitz 1998). It is postulated that population bottlenecks occurred 10,000 years ago in Eurasia (Reineke et al. 1999). There are a few studies with cDNA-AFLPs (the AFLP technique applied to cDNA made from mRNA) on genetic expression profiles of insects and their food preference (Pathan et al. 2007; Reineke & Löbmann 2005; Reineke et al. 2003). *L. dispar* is a useful model for molecular analysis in many different scientific disciplines because of its coevolution with an increasing range of host plant species (Keena et al. 2008; Keena et al. 2007).

1.5.2. *Lymantria monacha*

Lymantria monacha (common name: nun moth) is a counterpart to *L. dispar*. The 30-50 mm large white moth marked with arching dark lines across the wings is active in July till September. The female lays 250-300 eggs. Just like *L. dispar*, the larva develops and enters diapause directly in the egg till spring. Eggs can survive winter temperatures down to -20 °C. Larvae hatch around April and go through 5 instars. After approximately 2 weeks of pupation in July, the adult emerge and mate (Hannemann 2000).

L. monacha has been mainly studied in respect of host preference and the effect on tree species (Heiermann & Schütz 2008 ; Keena 2003; Withers & Keena 2001). *L. monacha* is also a polyphagous insect responsible for major defoliation in particular in conifer forests. There are large scales monitoring programs to avoid *L. monacha* becoming invasive as *L. dispar* has done in North America. Regular monitoring of the population size of *L. monacha* helps to predict imminent outbreaks (Keena 2003). There is hardly any molecular information available on *L. monacha.*

1.5.3. *Tortrix viridana*

Tortrix viridana (common name: green leaf oakroller) is an oligophagous forest pest insect with a strong preference for *Quercus* spp. (Schroeder & Scholz 2005). Its common name comes from the peculiarity of some larvae rolling the leaves of their hosts. The adult measures between 9-11 mm. The moths appear light green. The up to 50 eggs overwinter. *T. viridana* can have 3-4 generations per year. In outbreak years they can defoliate entire oak stands. In May/June the larvae pupate and 3 weeks later adult moths emerge. Here In this study it is the most specialized herbivore compared to *L. dispar* and *L. monacha*, because it causes damage in pure oak stands over mixed forests (Hannemann 2000). There are only a few studies in which molecular tools such as microsatellites, PCR-RFLPs and AFLP analysis have been used to examine the population structure and the phylogeny of *T. viridana* (Schroeder et al. 2007; Schroeder & Degen 2008, 2008; Schroeder & Scholz 2005). One

observation in North-Rhine-Westphalia made *T. viridana* an ideal comparison to *L. dispar* in studying food preference and their underlying genetic bases. This observation is that in outbreak years of *T. viridana* in this particular forest near Münster, oak trees standing side by side experience different levels of defoliation by *T. viridana*. The differing phenotypes of these trees might reveal differing defense mechanisms against herbivores. The trees were classified into "sensitive" (defoliation was high) and "tolerant" (defoliation was less) to *T. viridana*. The effect of *T. viridana* and *L. dispar* on these possible different "genotypes" is tested by analysing the expression profile of the herbivores via cDNA-AFLPs after feeding on the two types of oaks.

1.6. Field experiments

Most studies about plant-insect interactions are done in the lab or greenhouse under mostly controlled conditions. Through the EU project EVOLTREE, which this study participated in, there was the opportunity to use a canopy crane in one of the project study sites, a forest site in Tampare, Finland. With the 20m high crane, feeding hot spots of the herbivore are reachable. Since much research is done with detached ("cut") material rather than on living plants ("uncut"), statements about expressed anti-herbivory or defense genes have to take this into consideration. Under natural conditions the gene expression profile of plants might look different than under lab situations. The side by side experiments describe here try to answer, whether "cut" and "uncut" oak trees in Finland show differently expressed genes when the polyphagous herbivore, *L. dispar* feeds on them.

1.7. Molecular techniques used

Even though the genomics revolution of the last 10 years has improved understanding the genetic makeup of living organisms, not much is known about non-model species and about their genes relevant in adaptation to their respective host trees or on the extent of nucleotide variation at these loci in different populations. In this study I choose the cDNA-AFLP technique, because working with non-model organism demands a fingerprinting approach, before certain research questions about adaptation and interaction of host partners can be answered. Most research efforts in genomics have been devoted to pathogens that have important economic impacts on crops, whereas non-model insects receive much less attention. Insects such as *L. dispar*, *L. monacha* and *T. viridana* play an important ecological role in terrestrial ecosystems. In Lepidoptera, only functional genomic resources are available and these relate mainly to the domesticated silkworm *Bombyx mori* (Xia et al. 2004). None of the insect model species which are subject to genomic projects is a tree herbivore. In this study the target species are defoliating Lepidoptera, such as *T. viridana, L. dispar* feeding on broadleaves or *L. monacha* on conifers.

Here, the cDNA-AFLP technique is applied mainly in order to identify genes that may be involved in controlling the expression of detoxifying and digestive enzymes in all three species. The use of a rare and a frequent cutting restriction enzyme and the increased

amount of fragments through nucleotide attachments on the corresponding digestive enzymes enlarge the genomic sequence information (Meudt & Clarke 2007).

Furthermore 454 high throughput pyrosequencing technology and conventional Sanger sequencing methods are being used to identify genes that act as candidates for the different insect species. Single nucleotide polymorphisms (SNPs) are being applied to investigate genetic variation in genes relevant to plant-insect interactions between several different populations from Europe and one from China. Whole population dynamics, such as relatedness status within and between populations have been analysed with techniques such as 454 (Bai et al. 2010; Gompert et al. 2010; Santure et al. 2010). In addition, next generation sequencing has recently been used to identify candidate genes in non-model organisms or in marker systems for invasive species (Jun et al. 2011; Pauchet et al. 2009; Pauchet et al. 2010). In combination with these high throughput technologies, AFLPs and SNPs as precise and stable marker systems can help in the understanding of insect genomic phenomena, such as relatedness (Behura 2006; Rothberg & Leamon 2008; Santure et al. 2010). SNPs are good for automatic analysis in high throughput projects, because of a quite simple yes/no type of scoring and thereby reduced error rates (Vignal et al. 2002). Even as codominant markers such as microsatellites have been used for decades, SNPs are more helpful because of possible coding region information and their influence on the phenotype of individuals (Williams et al. 2010).

The power of next generation sequencing and future developments in this quick developing field will enable us to gain a huge amount of genomic information and to sequence the genomes of single individuals and of populations of many species for continuously decreasing costs in increasingly shorter time (Robinson et al. 2011). The alternatives of sequencing projects and microarrays are still financially and technically costly so they are usually applied to model organisms (Meudt & Clarke 2007; Vos et al. 1995).

1.8. Aims and contents of the research study

Biotic interactions between trees and associated species (symbionts, pathogens or herbivores) are a driving factor for evolutionary processes in forest ecosystems. The molecular mechanisms involved in these processes are illuminated by genetic analysis of all system partners. In this thesis, the distinguishing plant-insect interaction will be described prior to an examination of the current understanding of the composition, function and evolutionary development in these relationships.

The thesis is structured into six chapters.
The first chapter documents the general literature associated with the study.

The second chapter aims to understand the genetic diversity of *L. dispar* and *L. monacha* in relation to their host trees. I test the hypothesis that feeding on different food sources has different genetic expression pattern. The hypothesis was tested by offering different food

sources (oaks, spruce and artificial diet) to both species *L. dispar* and *L. monacha*. Afterwards RNA extraction and performing cDNA-AFLP analysis on a gel based and capillary electrophoresis unit was used as a method to look for differentially expressed genes. I addressed the question, which genes are putatively involved in the adaptation of Lepidopteran species to different host plants.

The third chapter is a study designed to examine the hypothesis that *L. dispar* shows different feeding behaviour when feeding on oak trees that are apparently tolerant or sensitive to *T. viridana* observed in a forest near Münster, Germany. The hypothesis is tested by comparing possible different gene expressions in midguts of *L. dispar* feeding on different food sources with the cDNA-AFLP analysis. The aim is to analyse the genetic basis of different defoliation behaviour towards two types of *Quercus robur*. The phenotypes might possibly be different genotypes. The question here is: How does a polyphagous (*Lymantria dispar*) and an oligophagous (*Tortrix viridana*) insect deal with different "types" of trees (tolerant vs. sensitive) and their possible "defence response"? Which genes are putatively involved in adaptation to different host plants in *L. dispar* and *T. viridana*?

The hypothesis that trees react differently when herbivores feed on them under more natural conditions is investigated in the fourth chapter. Feeding on *Quercus robur and Picea abies* samples by *L. dispar* is assayed under more natural conditions (uncut leaves) on the living tree and on detached leaves (cut treatment) side by side in Tampare, Finland. The experiment involves field work using a canopy crane to the region of the crown where herbivory usually occurs. Again the cDNA-AFLP analysis was performed on a capillary electrophoresis unit. The influence on cut and uncut treatment of *Quercus robur and Picea abies* on expression profiles and potential genes in the trees was tested by conducting feeding assays of *L. dispar*. The question here is: Is there a difference in the genes expressed in trees when *L. dispar* feeds on cut and uncut tree samples?

In the fifth chapter candidate genes are screened automatically with next-generation pyrosequencing for SNPs in different worldwide destinations of older samples from a previous study and new samples collected within the network for genetic comparisons. An AFLP-like protocol was used and population pools were individually tagged with a five-base "barcode". The last goal is to look at the genetic variation in genes related to herbivory between different populations of *L. dispar* in European and one in China. With the high throughput application 454, different geographic populations were screened for SNPs. What does the genetic diversity look like in certain candidate genes between different geographic populations of *L. dispar* within Europe and one Asian population?

In the last chapter the thesis concludes with a contextual discussion of the relevance of the project and suggestions for future research

2. Host adaptation of *Lymantria* species: a cDNA-AFLP approach

2.1. Abstract

Biotic interactions between trees and insects are driving factors for evolutionary processes in forest ecosystems. Genetic analysis of both partners in such interactions gives insight into the molecular mechanisms involved. *Lymantria dispar* and *Lymantria monacha* (Lepidoptera: Lymantriidae) are important forest pests in Europe whereas *L. dispar* has a great damaging effect particularly in North America.

Both *Lymantria* species are polyphagous and have to cope with a wide spectrum of different chemical defenses by different host-trees. Adaptation to different hosts could be due to induction of different genes as a response to different hostplant chemistries, or due to different insect genotypes, each of which is well-adapted to a particular host, or both. There is anectodal evidence for genotypes that are specialized for deciduous leaves vs. conifer needles but firm evidence for genetic specialization is lacking. In this study *L. dispar* and *L. monacha* were fed with *Picea abies* (spruce), *Quercus robur* (oak) and artificial diet to see whether different genes are expressed due to different food sources.

To identify genes relevant for *Lymantria* species' adaptation to their respective host trees, a differential display approach was employed. RNA was extracted from midguts of larvae that had been fed on different food sources and differences in gene expression patterns were examined using cDNA-AFLPs. Differently expressed fragments were isolated, sequenced and compared with sequences obtained from larval cDNA libraries. Genes involved in controlling the expression of detoxicative and digestive enzymes in both species were identified, and many showed up- or down-regulation on spruce or oak compared with artificial diet. These genes provide some insight into inducible or repressible adaptations to different host plants and will be used as potential candidates to examine populations for genetic variation. The cDNA-AFLP method applied here is particularly useful for application in molecular entomology of forest pests such as *Lymantria* species, since only limited functional genomic resources are available.

2.2. Introduction

Lymantria dispar and *Lymantria monacha* are important forest pest species, consuming a wide spectrum of tree species in their native Europe. *L. dispar* is additionally established in North America since it was deliberately introduced from France in 1868-1869 in a failed attempt to cultivate it for silk production, and has spread in the north of the U.S. and throughout Canada with dramatic success (Keena et al. 2007). Its ability to exploit a wide variety of tree species has contributed to its rapid spread in North America. *L. dispar* can feed on 450 different plant species in the USA alone (Reineke 1998; Reineke & Zebitz 1998). The US Department of Agriculture records economic damage in tens of millions of dollars annually (Reineke 1998; Reineke & Zebitz 1998). The intervention for the damage,

monitoring and the pest control costs around $30 million annually for the United States Department of Agriculture (WSDA 2011).

L. dispar as an oligophagous insect feeds on a wide spectrum of host plants e.g. *Quercus,* Alnus*, Acer, Betula* and *Fagus* (Natural History Museum 2011; Keena 2008). The gypsy moth as it is commonly named, is a generalist forest pest species, although it still prefers, if it has the choice, deciduous trees like *Quercus robur* over conifers (Naidoo & Lechowicz 2001). Adult moths of *L. dispar* fly between June and August. After mating the female lays up to 2000 eggs covered with a wool-like material. The larvae develop within the eggs and then enter diapause over winter until they hatch in spring (April), when food is available. Larvae go through 5-6 instars and feed 6-8 weeks. Pupation occurs in the crevices of the bark or other suitable places. Adults emerge within 3 weeks in July/August. Moths are night active. After the adults have mated and laid their eggs, they die (Hannemann 2000).

L. monacha, commonly called the nun moth, rather prefers conifers, such as *Picea abies* (spruce) (Heiermann & Schütz 2008). The life history of *L. monacha* is very similar to that of *L. dispar*. Eggs of *L. monacha* are laid in masses in July. The larvae also overwinter fully developed in the egg before hatching in the spring. The eggs from *L. monacha* are quite resistant to temperatures down to -20 °C.

Both *Lymantria* spp. are polyphagous insects but still show preferences for certain tree species. Preference studies have shown that *L. dispar* prefers red oak and sweet gum over other deciduous trees in North America (Shields et al. 2003). In experimental choice tests *L. monacha* prefers *Picea* species rather than other conifers such as *Pinus* species, though *Picea abies* is a very suitable host tree (Withers & Keena 2001). *L. dispar* shows a predilection for sunny south-facing leaves rather than for north-facing ones or it chooses full grown leaves over not fully expanded leaves (Schoonhoven et al. 2005). Some studies showed that chestnut hybrids are more preferred than natural seedlings of American chestnut (Rieske et al. 2003).

Foliar chemistry has been found to influence the performance and preference of *L. dispar.* Noticeable are different defoliation stages of trees in outbreak events of the herbivore (Osier et al. 2000). Phytochemicals such as tannins can cause dietary deficiencies in the herbivore by influencing midgut surfaces and reducing nutrient uptake (De Veau & Schultz 1992). Terpenoids, phenolics, cyanogenics and glucosinulates are all secondary plant compounds having a host choice effect on herbivores such as *L. dispar.* In *L. dispar* secondary plant compounds such as alkaloids and tannins were examined in choice tests. Tree species with the highest number of alkaloids were the least acceptable host plant species, whereas tannin containing species were preferred by *L. dispar.* Thus the presence of alkaloids seem to influence the host plant preference in *L. dispar* larvae (Shields et al. 2003). Tannins are thought to have a protective function against insect attack by reducing the nutritive value of leaves in addition to influencing their palatability (Feeny 1970). Controversially, a later

analysis claims no effect between tannins present in the leaves and performance or preference by gypsy moth (Barbosa et al. 1990).

Metabolic mechanisms of insects are slowed down during diapause, so that unfavourable environmental conditions during the winter can be endured (Van Asch & Visser 2007). Emergence from diapause is a crucial step in the life history and proper synchronisation plays an important role in the ability of leaf-feeding insects to exploit their hosts. Foliage age and leaf toughness can have effects on larvae (Hunter & Elkinton 2000). Tannin content increases and nutritional nitrogen declines throughout the year so development of herbivore larvae on mature leaves is slowed down, and emerging too late is disadvantageous. Additionally young larvae cannot penetrate closed tree buds, when they hatch too early in the spring (Feeny 1970). So emerging at the optimal time point is essential for good development, and it is crucial for diapausing insects to synchronize with the host for their development on the plant and their feeding behaviour (Andresen et al. 2001; Gray et al. 2001; Hunter & Elkinton 2000; Regniere & Nealis 2002; Visser et al. 2006). The phenology of the host changes yearly and can be influenced by temperature and photoperiod (Van Asch & Visser 2007). An asynchronisation from the hosts' point of view can be considered as a defense strategy against the herbivore.

In the present study any negative effects of unsynchronized phenology were minimized by offering *L. dispar* young leaves after the oak buds of *Quercus robur* were just opened. *L. monacha* was offered young needles of *Picea abies*. The genetic difference in performance on certain host trees even if both species are polyphagous was investigated, by offering both tree species to both *Lymantria* species.

Previous studies have not examined whether performance on different hosts was related to gene expression. Therefore in this study the question was posed: Which digestion/detoxifying genes are differentially expressed in *L. dispar* and *L. monacha*, when fed on spruce, oak, two different phenotypes of oak and artificial diet? These genes might be involved in adaptation to certain abiotic conditions or to a respective host plant present in the given habitat. We addressed the following questions: 1) Are there differences within a *Lymantria* species in abundance of certain mRNA molecules, when feeding on different plant hosts compared with artificial diet? 2) Are there differences between *L. dispar* and *L. monacha* in their responses to the same host plant? 3) Do the functions of differentially expressed genes show any evidence of adaptation to one host over another?

The cDNA-AFLP technique was chosen to examine differences in gene expression as related to diet, because it does not require previous sequence information and it allows several different treatments to be compared at once. Among the thousands of mRNA types expressed in a given tissue, cDNA-AFLPs display changes in just a small subset of fragments. By varying the combination of primers used, different subsets can be compared. Fragments that are more or less abundant across the different treatments can be cut out of the gel and

sequenced, and these sequences can be compared with public databases to determine their likely function.

The procedure of cDNA-AFLP described as follows: First mRNA is isolated from midgut tissue, then double-stranded cDNA is generated from it. Afterwards digestion with two restriction enzymes is being preformed. The principle is that one restriction enzyme is a rare cutter and one is a frequent cutter to have an increased diversity in fragments of different length. Next a subset of those fragments is amplified by PCR by ligating corresponding adapters followed by a general amplification of fragments with one set of primers. Finally, a particular subset is displayed through amplification with specific primers. The use of different nucleotides at the 3'-hydroxy end of the general primers (Eco+1 and Mse+2) gives the opportunity to test 64 different primer combinations. Separation on a slab gel or capillary apparatus is based on migration of charged particles through a matrix in an electric field. Based on size and charge the PCR products migrate to different distances in the matrix and are visualized by staining producing a differently expressed banding pattern. The different lanes are compared with the naked eye or with the help of scoring programs.

By using cDNA and not genomic DNA, exclusively coding regions will be observed and by sequencing different unknown fragments, genetic information is obtained. Use of the cDNA-AFLP technique has produced achievements in similar projects (Pathan et al. 2007; Reineke & Löbmann 2005). It is a widely established method for gene hunting in species where no prior genetic information is available (Pathan et al. 2007; Vos et al. 1995). Looking at gene expression under certain environmental conditions, such as host tree suitability, makes cDNA-AFLP an appropriate method, since little genetic information about the above mentioned species is available.

2.3. Material and Methods

Insects

Lymantria dispar eggs were initially received from a laboratory rearing kindly provided by Dr. Melody Keena (USDA Forest Service Hamden, CT) and were then continuously reared on gypsy moth diet for the New Jersey Standard Strain. For synchronisation of certain instars insects were kept separately in small round plastic containers with artificial diet under 25°C until the experiment started.

Lymantria monacha was collected in August 2006 in Osie in Poland (figure 1). Females of *L. monacha* moths were collected just from the tree trunk, since they are quite flight-lazy or reluctant to fly (Keena et al. 2008; Reineke et al. 1998). Alternative collection methods were pheromone traps for male moth adults, net trapping and light trapping (figure 2). The light trap was a 400 watt globe hung behind a white linen cloth (1.30 m x 2 m) which attracted the moths in the dark (figure 2). Moths were kept at 20-22°C in kitchen plastic containers (18 x 10 x 14 cm) with a net as lid attached with a rubber band.

Figure 1: Collection site of *L.monacha* in Osie/Poland (2011).

Figure 2: Picture of a light trap set-up in the forest at night. A white linen sheet is connected between two tree trunks and a 400 watt globe is hung behind the linen to attract night active insects. *L. monacha* can be collected directly from the sheet.

Both insect species were reared under controlled conditions in breeding chambers (24°C, day/night cycles of 10 hours day light 14 hours dark). Larvae were kept in plastic containers with artificial diet, till they were separated in small round plastic containers with artificial diet under 25°C until the experiment started (Keena, unpublished diet protocol for *L.dispar* and *L.monacha* (Ingredients: de-ionized water, agar, wheat germ, Wesson salt mix with Fe, sucrose, casein, sorbic acid, methyl parabenzene, ascorbic acid, wheat germ oil, raw linseed oil, choline chloride, cholesterol, ferric citrate, vitamin mixture). Within the feeding experiments artificial diet was chosen as a control treatment for *L. monacha* (treatment denoted as m) and for *L. dispar* (treatment denoted as d), respectively.

Plant material

Oak

In spring 2003 three pedunculate oak (*Quercus robur*) stands in North Rhine-Westphalia, near the city of Muenster (Asbeck, Muenster, Warendorf) (Longitude: 7˚34’, range 8.73’, Latitude: 51˚58', range 6.79’, elevation: 50 - 80 m and total area of about 70 x 20 km) were used as sample collection sites. The terrestrial ecosystem in these stands is a mixed boreal

cultivated forest (used for recreation or as a commercial seed production forest), containing between 60-85% oak trees (*Quercus robur*). In June temperatures were around 24°C at day and 16°C at night.

One observed phenomenon in these stands in Muenster is that trees which stand right next to each other (5-10m apart) may experience totally different defoliation by *Tortrix viridana,* a specialist lepidopteran herbivore and main defoliator of *Q. robur*. The classification T (for tolerant, i.e. not heavily defoliated by *T. viridana*) and S (for sensitive, i.e. heavily defoliated by *T. viridana*) was made by Dr. Hilke Schroeder after several years of monitoring degrees of *T. viridana* defoliation in this forest (Schroeder 2010). In this study the treatments for *L.dispar* larvae feeding on "oak tolerant" are called DOR and on "oak sensitive" DOS, respectively. The same samples which were taken for the oak feeding treatment for *L. dispar* in this study were also taken for a separate research experiment in another context (details in chapter 2).
Collection of oak leaves was conducted with the help of a sling blade, where a bag filled with little metal nuggets connected to a rope was shot into the tree canopy, so that a desired branch could be pulled down. Leaf samples were directly taken from the hot spot of feeding activity of *T. viridana* and were stored in plastic buckets with water for transportation. A total of 6 sensitive trees and 9 tolerant trees were sampled with the sling blade. Feeding assays started soon after leaf sample collections from tree twigs. Other oak samples for *L. monacha* were collected (treatment denoted as MO) in a forest near Jena (Latitude: 50˚55'25''N, Longitude: 11˚37'17''E). Here small branches of *Q. robur* were cut from easy to reach branches and then stored in buckets filled with water, until experiments started. The oak samples in this experiment came from two different locations (Münster, North Rhine Westphalia and Jena, Thuringia), because the experiments concerning DOR and DOS on Münster oak samples were done first. Oak samples of *L. monacha* feeding assays (MO) were done at a later time and sampled in Jena.

Spruce

Spruce twigs were taken from living trees of Norway spruce planted behind the Max Planck institute for Chemical Ecology in Jena, Germany in 2000. The trees came initially from Austrian *Picea abies* clones. Branches were chosen randomly from random trees. Spruce twigs were offered to either *L. dispar* (treatment denoted as DS) or *L. monacha* (MS), respectively.

Feeding experiments

Experiments were conducted in August 2006 at the MPI for Chemical Ecology, Jena at 24 °C under controlled lab conditions. *L. dispar* and *L. monacha* larvae in their 3rd to 5th instar, 48 h post molt, were starved for 24 h before the experiments started. Leaves were chosen randomly from the two tree species and larvae were spread randomly. The Münster oak leaves (DOS, DOR) were offered to one larva of *L. dispar* and the other Jena oak leaves were offered to (MO) one larva of *L. monacha* in petri dishes. The stalk of a branch containing

approximately 3-4 leaves was put through a hole in the lid of a 1.5 ml eppendorf tube filled with water. This oak branch was laid into a standard petri dish (100 x 25 mm) and larvae were put separately into the dish with the lid closed. Spruce twigs were offered to one larva of either species (DS, MS). Insect larvae were put in bag-like nets (10x10cm) securely closed with a rubber band around a branch tip containing 2-3 well expanded young oak leaves. The two control treatments were offered to *L. monacha* and *L. dispar* in the form of 1x1x1 cm sized artificial diet cubes placed in little round plastic containers with a lid with holes in it (d, m). The 24 h feeding assay was conducted with one larva per branchtip (3-4 leaves). Because the starting weight of larvae in this feeding experiment was quite variable between the treatment groups, subgroups were separated according to weight. Larvae with a starting weight of 40 mg were called "DOR S" and "DOS S", respectively. Larvae with a starting weight of 25 mg were called "DOR L" and "DOS L". Due to an accidental mixing of some groups, there was a third group, the DOR/DOS mix group (indicated as DOR mix and DOS mix) consisting of a mixture of 40 mg and 25 mg starting weight larvae. Calculations for the feeding experiments were separated according to starting weight and depicted in graphs. The larvae on DS, MO, MS and m were all 200 mg in average per larvae and were approximately in their 5th instar when the feeding assay started. The number of larvae feeding was for *L. dispar* 60 on DOS mix treatment, 91 on DOR mix, 30 on DS and 11 on d. For *L. monacha* there were 10 on MO, 26 on MS and 8 on m. 50 larvae fed on DOR S, 40 on DOS S, 31 on DOR L and 20 on DOS L, respectively. Larvae were measured before and after feeding with an analytical balance (Denver Instrument SI234, max 230 g).

In all experiments, larvae with obvious signs of feeding (production of feces and leaf damage) were collected, the midguts were dissected from the living insect and then directly shock frozen in liquid nitrogen and stored at -80 °C.

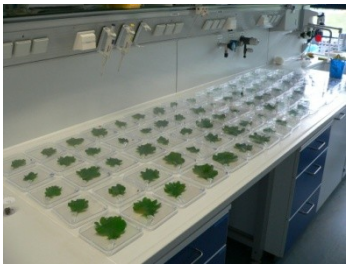
a)  b)  c) 

Figure 3: Experimental set-ups in the lab under controlled conditions. a) For the oak feeding assay one oak leaf was placed in a plastic box sitting in an eppendorf tube filled with water. A single larva was placed for 24h into the box. b) Entire experimental set-up. c) For the spruce feeding assay, larvae were put into net-like bags, which were tied to one little branch. Each branch was placed in a water filled cups. Feeding was conducted for 24h.

RNA extraction and cDNA synthesis

RNA of the midguts from both *L. dispar* and *L. monacha* larvae after feeding on the different treatments was extracted using TRIzol Reagent (Invitrogen, Karlsruhe, Germany) according to the manufacturer's protocol with a slight modification: a Pellet Pestle Cordless Motor (Kontes, VWR labshop, Batavia, il) was used for homogenization of midgut tissues in 2 ml eppendorf tubes. Quality and quantity of RNA was checked via gel electrophoresis and Nanodrop spectrophotometer, respectively. Residual DNA was removed by DNase treatment followed by purification using the RNeasy Minelute Cleanup Kit (Qiagen, Hilden, Germany). RNAs of 1 midgut per larva and treatment were pooled (6 larvae per pool) prior to cDNA analysis. The seven pools of RNAs (1 µg per pool, designated for *L.dispar* as sensitive (DOS), tolerant (DOR), spruce (DS), artificial diet (d), for *L.monacha* as oak (MO), spruce (MS) and artificial diet (m)) were translated into cDNA with the SMART PCR cDNA Synthesis Kit (Clontech). Afterwards cDNA was purified using a PCR Purification Kit (Qiagen). After the accidental mixture of DOR and DOS samples was recognized, four newly synthesized pools of DOR L, DOS L and DOR S and DOS S cDNA were supplemented to the analysis.

cDNA-AFLP analysis for *L. dispar* and *L. monacha*

For cDNA-AFLP reactions, 500 ng purified cDNA was digested with 5 U *Eco*RI and 3 U *Mse*I for 2 h at 37°C followed by enzyme inactivation for 15 min at 65°C. Accordingly, *Eco*RI and *Mse*I double stranded adapters were ligated to the fragments by incubating for 2 h at 16°C. Pre-amplification of cDNA fragments was performed with a 1:10 dilution of the ligation reaction for 20 cycles with 94°C for 30 s, 56°C for 1 min, 72°C for 1 min using primers corresponding to *Mse*I and *Eco*RI adapters (table 1). Resulting PCR products were diluted 50x and were subjected to selective amplification with a combination of different *Mse*I and *Eco*RI primers using a touchdown amplification PCR-profile as described by Vos et al. (1995). *Mse*I and *Eco*RI selective amplification primers each had two extra nucleotides at their 3`-termini. 64 different AFLP-primer combinations were used for selective amplification: 16 *Mse*I-NN primers (NN is either AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG or TT) and 4 *Eco*RI -N primers (N is either A, G, T or C) were used together in different combinations. A subset of primer combinations was used separately for all types of DOR/DOS samples, which were analysed along with samples from *Tortrix viridana* (see Table 2 in Chapter 3). Afterwards, samples were loaded directly on slab gels and visualized by SYBR Safe (DNA gel stain, life technology) staining in UV transilluminator (biostep, wavelength: 312 nm).

For visualization by capillary electrophoresis, it was necessary to attach a fluorescent label to the AFLP products. For this purpose, each selective *Eco*RI primer was modified by placing a M13(-21) tail at the 5'-end of and a fluorescently labelled universal primer M13(-21) was added to the PCR reactions according to the method described by (Schuelke 2000). Different fluorescent dyes (IRD700 or CY5) were used for labelling the universal M13 (-21) primer, enabling the reactions to be loaded on capillary electrophoresis as a multiplex reaction.

Table 1: cDNA-AFLP standard primer sequences

*Eco*RI adapter	forward:	5'-CTCGTAGACTGCGTACC-3'
	reverse:	5'-AATTGGTACGCAGTCTAC-3'
*Mse*I adapter	forward:	5'-GACGATGAGTCCTGAG-3'
	reverse:	5'-TACTCAGGACTCAT-3'
*Eco*RI pre-amplification		5'-GACTGCGTACCAATTC-3'
*Mse*I pre-amplification		5'-GATGAGTCCTGAGTAA-3'

Fragment Analysis and Isolation

The cDNA-AFLP fragments were loaded on two different electrophoresis systems: a submarine gel matrix based electrophoresis system (Elchrom Scientific) and a capillary electrophoresis system (Beckman-Coulter).

Selective amplification PCR reactions containing fragments which were differentially expressed after feeding on different food sources, respectively, were loaded on ready-to-use 9 % PolyNat gels and separated by electrophoresis (Elchrom Scientific). Through SYBR Safe staining differentially expressed bands were visualized and then picked out of the gel with a tool called BandPick (Elchrom Scientific). Desired fragments were directly reamplified with the corresponding selective AFLP primer. PCR products were purified with ExoSAP-IT (Affymetrix) and were directly sequenced.

To verify the results, an alternative cDNA-AFLP visualisation method with capillary electrophoresis (Beckman-Coulter) was carried out for 7 chosen primer combinations. All procedures were repeated for the 4 pools of the additional DOR L/DOS L and DOR S/ DOS L samples described.

Since it was not possible to isolate DNA fragments directly from the capillary electrophoresis unit, after deciding which bands to isolate, an aliquot of the IRD700 or CY5-labelled PCR products was loaded on the slab gel electrophoresis system (Elchrom Scientific) for the isolation of differentially expressed bands as described above.

Sequence Analysis

Sequence analysis was done using the Bioedit sequence alignment editor tool (Ibis Biosciences). Sequences were compared by BLASTN against 23,300 transcripts from *L. dispar* and 21,900 *L. monacha* transcripts from midgut cDNA libraries, which were generated at the MPI for Chemical Ecology (Vogel, unpublished). Sequences were also compared by BLASTN and BLASTX against the non-redundant sequence databases of GenBank at NCBI. To identify

candidates sufficiently a threshold of an *E*-value of at least E-05 was used to filter these BLAST results.

2.4. Results

Performance of *Lymantria* ssp. larvae on different food sources

After establishment of lab populations of *L. dispar* and *L. monacha,* it was possible to do a controlled experiment with larvae feeding on different food sources (figure 4). This was shown by measuring the weight of 3rd /5th instar larvae before and after 24h feeding on different food sources. The following observations have to be interpreted with care, because of uneven starting weights of the larvae examined in this feeding experiment. The different oak treatments by *L. dispar* taken as a general oak food source, larvae gained more weight after feeding on the oak leaves and on diet then on spruce. In contrast to that, *L. monacha* larvae gained more weight on diet and on oak than on spruce. Unexpectedly *L. monacha* gained more weight on oak then on spruce, whereas on spruce the larvae didn't gain weight at all. Furthermore no statistical analysis was done, because mainly the genetic background of feeding preferences were examined and feeding assays were solely done to trigger gene expression levels between the different treatments in the two different species of *Lymantria* for further analysis. Statistical analysis and a detailed comparison was done for DOR L/DOS L and DOR S/ DOS S samples (see chapter 3).

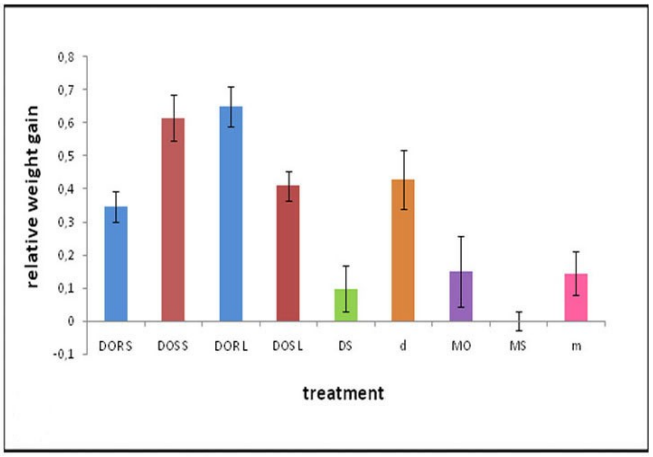


Figure 4: Relative weight gain (mg/mg) of *L. dispar* and *L. monacha* after feeding on different food sources. DOR L/DOS L and DOR S/ DOS L samples are seen as oak food source for *L. dispar.* No comparison between the separate groups was done.

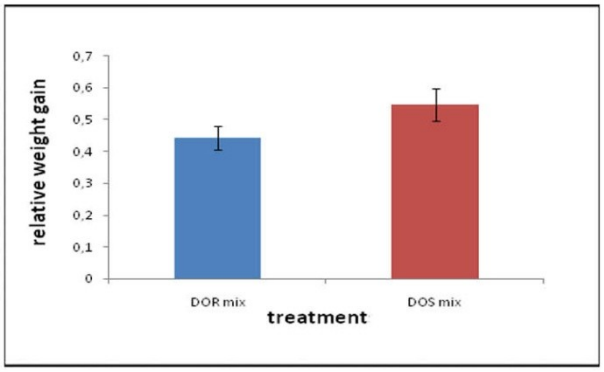


Figure 5: Relative weight gain (mg/mg) of *L. dispar* larvae feeding on the mixed two different oak types (DOR mix/DOS mix). The larvae had a mixture of different starting weights.

Gene expression differences analysed via cDNA-AFLPs

To see whether gene expression profiles showed differences between the treatments after feeding; cDNA-AFLPs were conducted for both different species with a set of selective primer combinations.
On a gel electrophoresis device 44 primer combinations for both *Lymantria* species were tested for the treatments, respectively. A total of 1354 bands for *L. dispar* and 984 fragments for *L. monacha* were detected. On the capillary electrophoresis in 7 primer combinations, 1923 fragments for *L. dispar* and 641 fragments for *L. monacha* were calculated. Fragment differences were found in uniform gel pattern (figure 6) or chromatograms (figure 7) of *L. dispar* and *L. monacha*. Within the primer combinations differences between treatments were clearly detectable, even to the naked eye. To verify results obtained on a gel electrophoresis system, amplifications with a subset of primer combinations was repeated and was loaded on a capillary electrophoresis system. On a capillary electrophoresis device 7 primer combinations for both species were tested, just a subset of the original primer combinations, after it was shown that only these 7 primer combinations contain potentially different expressed genes within the treatments (DOR mix, DOS mix, DS, d, MO, MS, m). For DOR L/DOS L and DOR S/DOS S separately another 37 primer combinations were tested on the capillary electrophoresis system (table 2). Graphical landmarks of each chromatogram were analysed for differences in gene expression patterns between the treatments in a standardized way with the Genome Lab System (Beckman) (table 3). For the gel banding pattern analysis, visualization by eye was done (table 3).

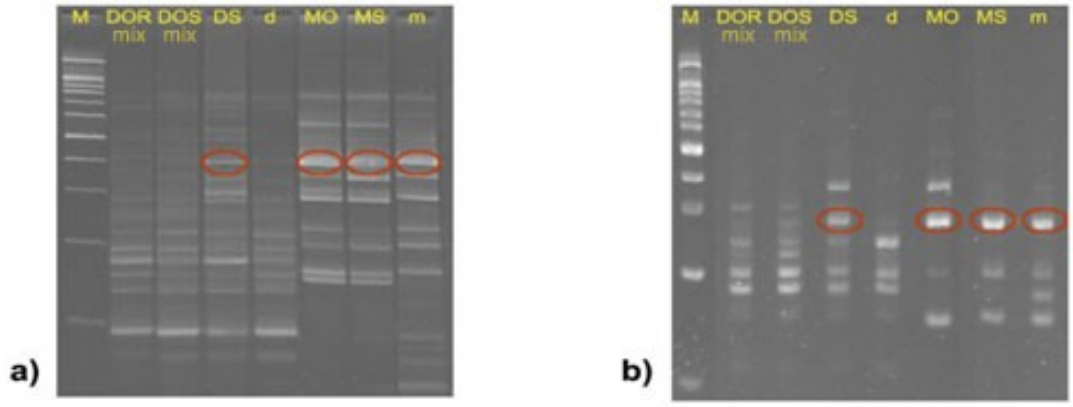


Figure 6: Expression of *L. dispar* and *L. monacha* displayed by cDNA-AFLP. The example shows selective amplification with one primer combination a) (EcoA MseTC) and b) (EcoA MseGA) for the respective treatments.

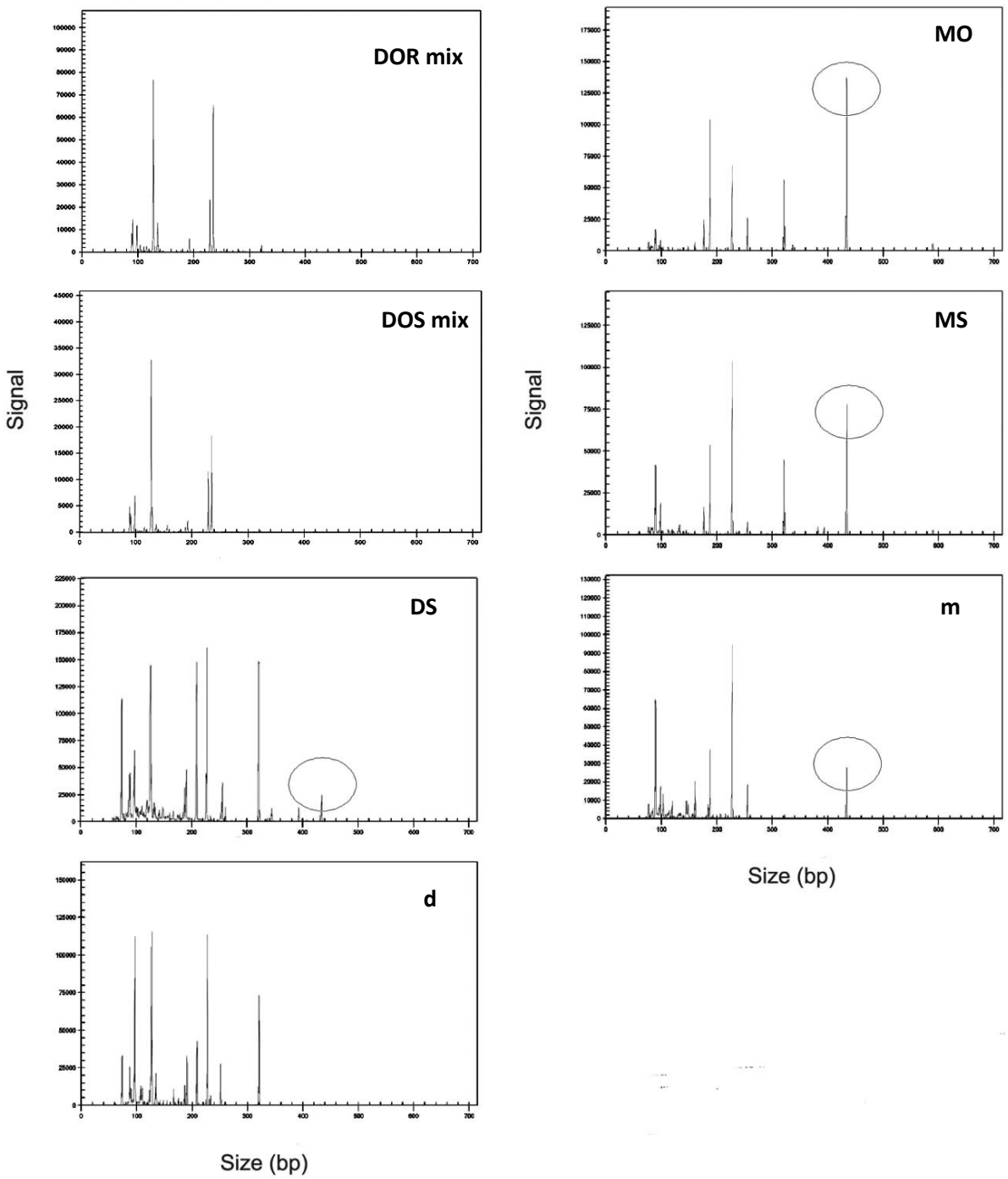


Figure 7: Amplification products visible after capillary electrophoresis obtained with AFLP primer combination Eco-A + MSE-TC in cDNAs from *L. dispar* (left) and *L. monacha* (right) larvae feeding on different food sources. Differentially expressed fragments are indicated by a circle in the respective treatments.

Table 2: Number of bands or fragments in the different treatments and *Lymantria* species are shown. The method and the number of primer combination are indicated.

	species treatment											method	primer combinations
	DOR S	DOS S	DOR L	DOS L	DOR mix	DOS mix	DS	d	MO	MS	m		
fragments	/	/	/	/	317	319	410	308	356	339	289	gel	44
	/	/	/	/	479	479	487	478	214	214	213	capillary	7
	720	721	721	712	/	/	/	/	/	/	/	capillary	37

Table 3: The table shows for which treatments the bands were present and which absent, indicated ("-" for absent and "+" for present). The index number defines a specific banding pattern, identified in the gels or chromatograms. The number and the method of the specific expression profile are given.

Index	***L. dispar***				***L. monacha***			**Relevance**	**no. of expression profile , method**
	DOR mix	**DOS mix**	**DS**	**d**	**MO**	**MS**	**m**		
1	-	-	+	-	-	-	-	Ld Spruce	15
2	-	-	+	-	+	+	+	Ld Spruce + Lm all	40 gel , 9 capillary
3	+	-	-	+	-	-	-	Ld diet, Ld DOR	2 gel
4	-	-	-	-	-	+	-	Lm Spruce	5 gel
5	-	-	-	-	+	-	-	Lm Oak	8 gel
6	-	-	-	-	+	+	-	Lm suppressed on diet	10 gel, 1 capillary
7	-	-	-	-	-	-	+	Lm Diet	4 gel
8	-	-	-	-	-	+	+	Lm suppressed on Oak	3 gel
9	-	+	-	-	-	-	-	oak difference	1 gel
10	-	+	+	+	-	-	-	Ld suppressed on DOR	1 gel
11	-	+	+	-	-	-	-	Ld suppressed on diet and DOR	1 gel
	DOR L	**DOS L**	**DOR S**		**DOS S**				
12	++	+	+		+			DOR L more expressed within all treatments	3 capillary
13	+	-	-		-			oak tolerant in Ld L	9 capillary
14	-	-	+		-			oak tolerant in Ld S	2 capillary

Index	*L. dispar*				Relevance	no. of expression profile , method
	DOR L	**DOS L**	**DOR S**	**DOS S**		
15	+	+	++	+	DOR S more expressed within all treatments	1 capillary
16	+	+	+	+	all the same	3 capillary
17	-	+	++	+	DOR S more expressed within 3 treatments	1 capillary
18	+	-	+	+	DOS L absent	2 capillary
19	++	+	+	++	DOR L and DOS S more expressed	1 capillary
20	-	++	++	-	DOS L and DOT S present	3 capillary
21	-	+	-	-	DOS L present	1 capillary
22	-	-	+	+	comparison between L and S	2 capillary
23	++	++	+	+	comparison between L and S, L more expressed	2 capillary

In the table 3 specific banding patterns are listed. The index number indicates the pattern and helps to differentiate between them. Index number 2 appears e.g. 40 times on a gel banding pattern and 9 times on a chromatogram. It is therefore the most frequently identified pattern.

The specific chromatogram patterns were comparable with the pattern of gel electrophoresis analysis of the same primer combinations. In an additional gel electrophoresis analysis of the capillary electrophoresis PCR product (containing dye), different expressed bands of specific length between treatments were spotted for further picking. Within this final gel electrophoresis analysis, interesting bands were picked and reamplified with the associated primer combination before the cleaned up PCR products were sequenced (table 4).

Genes of *L.dispar* and *L.monacha*

In total 20 potential candidate genes were found in all treatments, respectively for both species (table 4). 11 genes were found for capillary electrophoresis, 9 from gel electrophoresis and another 7 genes were identified through the analysis of the DOR L/DOS

L and DOR S/DOS group via capillary electrophoresis. The AFLP reads were matched to the sequence of 23300 transcripts (both 5' and 3') from *L. dispar* and 21900 (both 5' and 3') *L. monacha* midgut cDNA library accordingly.

Sequences of respective AFLPs containing transcripts from the *L. dispar* or the *L. monacha* cDNA libraries were compared to GenBank entries using the BLASTX tool and in those transcripts with significant homologies (E values lower than 1E-04); hits were determined (table 4). The database showed homologies to insects, in particular to lepidopteran model organisms (table 4).

Genes encoding metabolic reactions, such as pyruvate dehydrogenase, were found two times exclusively in different primer combinations in *L. dispar* feeding on sensitive oak (DOS mix), its major host. The index 9 shows the same pattern. Most trypsin-like, putative trypsin precursor and chymotrypsin-like protease proteins were found in *L. dispar* in *L. monacha* in index number 2, the most abundant pattern. It represents feeding on spruce (DS) and feeding on oak (MO), respectively. The juvenile hormone diol kinase was only found in *L. monacha* treatments (MO) and (MS), but still from the same index 2. Enzymes that break down sugar molecules, such as glycoside hydrolase were found in the midgut of *L. dispar* feeding on spruce (DS). Lysosomal acid lipase was also identified in the *L.dispar* on spruce (DS). In (DS), (MO), and (d) treatments, ribosomal proteins were isolated and identified. Carboxypeptidase 3, a digestive enzyme, was discovered in *L.dispar* feeding on oak tolerant (DOR mix). Additionally there was a single domain major allergen protein in (DS), a long-chain-fatty-acid--CoA ligase in (MO) and several proteins with unknown function yet found in (DS), (MO), (m) and (DOR mix).

The comparison between species and treatments has to be treated carefully, because of differences in starting weight of the larvae within the respective species as well as the use of oak samples taken from different locations for a comparison between species. The identified genes can however be regarded as candidate genes in both species, which might have some implications for adaptation to the food source.

Table 4: Candidate genes found for *Lymantria spp*., after isolating differentially expressed fragments in the respective electrophoresis system (either capillary or submarine gel electrophoresis)

Index	**ID**	**Library hit**	**e-value**	**Homology**	**Putative function**	**Electrophoresis system**
2	**1 DS**	no hit	9.00E-12	trypsinogen-like protein 1 [*Manduca sexta*], CAM84320.1	protein breakdown	capillary electrophoresis

Index	ID	Library hit	e-value	Homology	Putative function	Electrophoresis system
2	**2 MO**	*Ld*	5.00E-24	putative trypsin precursor Hz3 [*Helicoverpa zea*], AAF74750.1	protein breakdown	capillary electrophoresis
9	**4 DOS mix**	*Ld*	2.00E-05	pyruvate dehydrogenase [*Tribolium castaneum*], XP_970163.1	oxidative decarboxylation	capillary electrophoresis
2	**5 DS**	*Ld*	0.005	predicted protein [*Nematostella vectensis*], XP_001641516.1	unknown	capillary electrophoresis
2	**7 MO**	no hit	0.001	hypothetical protein LOC100638959 [*Amphimedon queenslandica*], XP_003391567.1	unknown	capillary electrophoresis
2	**8 Mo**	*Lm, Ld*	3.00E-23	juvenile hormone diol kinase [*Manduca sexta*], CAD23378.1	physiology	capillary electrophoresis
7	**9m**	no hit	0.72	hypothetical protein Phum_PHUM095290 [*Pediculus humanus corporis*], XP_002423910.1	unknown	capillary electrophoresis
1	**10 DS**	*Ld*	6.00E-08	lysosomal acid lipase [*Culex quinquefasciatus*], XP_001842037.1	fatty acid metabolism	capillary electrophoresis
2	**11 DS**	*Lm, Ld*	6.00E-26	ribosomal protein S15 isoform A [*Lysiphlebus testaceipes*], AAX62477.1	ribosome	capillary electrophoresis
2	**12 Mo**	*Lm, Ld*	6.00E-24	ribosomal protein S15 [*Manduca sexta*], ACY95351.1	ribosome	capillary electrophoresis
1	**15 DS**	*Lm, Ld*	1.00E-28	chymotrypsin-like protease C1 [*Heliothis virescens*], ABR88231.1	protein breakdown	capillary electrophoresis

Index	ID	Library hit	e-value	Homology	Putative function	Electrophoresis system
3	**21 DOR mix**	*Ld*	3.00E-18	carboxypeptidase 3 [*Mamestra configurata*], ACR16009.1	Protein breakdown	submarine gel electrophoresis
9	**42 DOS mix**	*Lm, Ld*	2.00E-32	pyruvate dehydrogenase [*Tribolium castaneum*], XP_970163.1	oxidative decarboxylation	submarine gel electrophoresis
3	**56 DOR mix**	no hit	no	unknown protein	Unknown	submarine gel electrophoresis
2	**4 DS**	*Lm*	6.00E-17	glycoside hydrolases [*Aedes aegypti*], XP_001659855.1	sugars breakdown	submarine gel electrophoresis
2	**17 MO**	*Lm, Ld*	2.00E-26	putative trypsin precursor Hz2 [*Helicoverpa zea*], AAF74749.1	protein breakdown	submarine gel electrophoresis
2	**18 MS**	*Lm, Ld*	5.00E-30	juvenile hormone diol kinase [*Manduca sexta*], CAD23378.1	physiology	submarine gel electrophoresis
3	**47 d**	*Lm, Ld*	8.00E-39	ribosomal protein S15 [*Heliconius melpomene*], ABS57431.1	ribosome	submarine gel electrophoresis
10	**6 DS**	*Lm, Ld*	4.00E-39	single domain major allergen 2 protein [*Spodoptera frugiperda*], ABX39558.1	detoxification	submarine gel electrophoresis
6	**20 MO**	*Lm, Ld*	6.00E-12	long-chain-fatty-acid--CoA ligase ACSBG2-like [*Bombus terrestris*], XP_003393002.1	fatty acid cycle	submarine gel electrophoresis

Index	ID	Library hit	e-value	Homology	Putative function	Electrophoresis system
12	**2 DORL**	*Ld*	5.00E-17	trypsin Ia precursor (*Sesamia nonagrioides*), AAT95348.1	protein breakdown	capillary electrophoresis
14	**8 DORS**	*Ld*		Lymantria library hit	unknown	capillary electrophoresis
13	**11 DORL**	*Ld*	7.00E-58	beta-glucosidase precursor (*Spodoptera frugiperda*), AAC06038.1	digestion	capillary electrophoresis
18	**17 DORL**	*Ld*	8.00E-06	carboxypeptidase 3 (*Mamestra configurata*), ACR16009.1	protein breakdown	capillary electrophoresis
20	**3 DOSL**	*Ld*	7.00E-20	protein disulfide isomerase (*Bombyx mori*), NP_001037171.1	catalyse protein folding	capillary electrophoresis
22	**9 DOSS**	*Ld*	2.00E-38	projectin (*Pediculus humanus*), ADY75705.1	member of the immunoglobulin superfamily, function unknown	capillary electrophoresis
23	**14 DOSL**	*Ld*	6.00E-27	hypothetical protein EAG_13571 (*Camponotus floridanus*), EFN65448.1	unknown	capillary electrophoresis

2.5. Discussion

In this pilot study the question of which putative candidate genes differ between *L. dispar* and *L. monacha*, when fed on different food sources, might be involved in adaptation to certain host plants has been addressed.

cDNA-AFLP analysis and method

The analysis between the two different electrophoresis systems showed fairly homogeneous expression profiles in both systems. Main fragments on the chromatograms and bands on the gels were comparable and were easily identified. Nevertheless, the different amounts of fragment numbers showed still a huge difference in the quantity of fragments in both systems. The capillary electrophoresis seems to be more sensitive and produces more fragments; even intensity differences could be spotted, whereas in the gel electrophoresis only very noticeable bands are visible. Also the analysis of the gels and the chromatograms are not very standardized. The gel pattern was analysed by eye. The chromatograms were pre-analysed with an automated program, but still needed a visual check by the scientist. All those factors have to been considered when working with different AFLP machinery.

Comparison within one species

It is difficult to compare two closely related species, but still different responses to food types were found even within the species. The feeding experiments were analysed taken the starting weights into account. In respect of the relative weight, *L. dispar* showed an increased weight gain on its preferred host *Q. robur*, as expected. However the growth patterns of *L. dispar* on the different types of oak depended on their starting sizes; the difference between DOR L/DOS L and DOR S/DOS is analysed in detail in the next chapter (chapter 3). *L. dispar* gained more weight on artificial diet than on spruce. Artificial diet was taken as a control in both species. Both species showed a weight gain on this food source, which indicates a nutritional value for the larva (Erlandson et al. 2010; Shen et al. 2006). The adaptation to artificial diet in *L. dispar* might even be a side product of the long lab culture history of these animals. However in *L. monacha* just the nutrient effect can play a role in the weight gain, because larvae were descended from animals collected from the wild and had no opportunity to adapt to artificial diet.

In *L. monacha* the weight gain results were a bit unexpected. Larvae gained more weight on the oak, regarded as a suboptimal host, than on *Picea abies*, regarded as the preferred host. The reason for this could be the reduced fitness condition of the larvae. Many larvae fed poorly or died before the experiment was over. One possible explanation could be the different developmental and physiological conditions of the respective larvae. Both results from the feeding experiments indicate that both *Lymantria* species are polyphagous and reasons for a preference might be factors, such as temperature, overall larvae fitness, chemical composition of the plant material or the different instars of larvae and their different needs in this developmental stage.

Comparison between two species

The banding pattern and the chromatogram profile with equal expression levels in many primer combinations can be compared between the species, under certain limited conditions. If a band of the same size and amplified with the same primers appears in at least one of the *L. dispar* and at least one of the *L. monacha* samples, it probably represents the same (homologous) gene in both species and indicates that the gene is expressed at a high level in both species in at least one treatment. (Sequencing of the same-sized band from both species in a few cases has confirmed this; data not shown.) In this case, variation in the band intensity across treatments can be used to compare expression levels across the two species as they respond to the different treatments. However, if a band appears only in one species, it cannot be concluded that the corresponding gene is absent or not expressed in the other species; for there could be DNA differences leading to failure of the restriction enzymes to cut or the specific primers to amplify the cDNA that is present. In this case, variation in the band intensity can only be used to compare expression levels within the one species. Patterns 4-8 represent *L. monacha*-specific bands and patterns 1, 3, and 9-11 *L. dispar* -specific bands that varied across treatments. The phenomena of different expressed genes after certain treatments have been recorded in other studies (Pathan et al. 2007; Reineke et al. 2003; Reymond et al. 2000). Furthermore particular midgut and metabolism proteins are triggered due to specific diets (Christeller et al. 2010; Lazarevic et al. 2004). As here in this study, midgut enzymes are responsible for direct or indirect function of digestive and detoxifying processes in the moth-host interaction (Peric-Mataruga et al. 2000; Zavala et al. 2008).

Only Pattern 2 indicates bands in both species: *L. monacha* in all 3 treatments (MO, MS, m) and *L.dispar* in one treatment (DS) show expressed genes of the same size within the same primer combination (figures 6a, 6b). No other pattern was observed showing bands in both species. One explanation could be that genes responding to spruce in *L. dispar* are evolutionarily more conserved and hence more likely to be observed at the same size in *L. monacha*, but this does not explain why they are not spruce-specific in *L. monacha*. Another explanation is suggested by the lower weight gains in the feeding experiment, in which both, DS and the more or less entire *L. monacha* group excluding the diet treatment group show very low weight gains. Thus these more highly conserved genes might indicate a starvation stage or stress response of the larva.

Homology of genes in insect species is an often observed phenomenon and also used particularly in studying non-model organisms (Sonoda et al. 2006; Travers et al. 2007). In the present study members of the trypsin family were identified in (DS) and (MO), respectively. The ribosomal protein S15 was also found in both species (DS, MO, d), regardless of treatments. As many basic processes in insects are conserved, homologies for instance from *B. mori* (silkworm) are helpful in studying other Lepidoptera species, where little or no genetic information is available.

Additionally, there was no extra oak treatment prepared when feeding assays began for this project. Unfortunately oak samples came from different locations and trees, so they cannot be taken into consideration in the comparison between *L. dispar* and *L. monacha* feeding on the same food source.

Genes identified

Genes identified through cDNA AFLPs with two separate methods, gel electrophoresis and capillary electrophoresis can however be regarded as candidate genes in both species, which might have some implications for adaptation to food sources. The comparison between species and treatments has to be treated carefully, because of differences in starting weight of the larvae within the respective species as well as the use of oak samples taken from different locations.

The pyruvate dehydrogenase found with both electrophoresis methods, in *L. dispar* feeding on sensitive oak (DOS), belongs to a multi-enzyme-complex, responsible for the metabolism of pyruvate into acetyl coenzyme A (wissenschaft-oneline 2011).

Genes encoding different trypsin or trypsin-like proteins are responsible for the adaptation to host plant proteinase inhibitors and might be controlled through several gene classes (Brioschi et al. 2007). In *Manduca sexta* proteinase inhibitors inhibit midgut digestive proteinase when feeding on *Nicotiana attenuata* (Zavala et al. 2008). Spruce and oak are likely to have different proteinase inhibitors but these are likely absent from the artificial diet; possibly offering an explanation for protease 15 DS showing Pattern 1. The herbivory-tolerant and sensitive oaks may also differ in protease inhibitors, accounting for proteases 2 DOR L and 17 DOR L.

Another interesting candidate gene found is a glycosyl hydrolase which may function in digestion, and is expected to be expressed during feeding on plant material (Allgaier et al. 2010; Henrissat & Davies 1997).

Carboxypeptidase, an enzyme known to play a role in adaptation to several different diets (Erlandson et al. 2010; Wang et al. 2004) was found here when *L. dispar* fed on tolerant oak (DOR). Lysosomal acid lipase (Zschenker et al. 2004) and long-chain-fatty-acid--CoA ligase (Kunieda et al. 2006) are involved in fatty acid metabolism. The single domain major allergen protein (Fischer et al. 2008) is present in the midgut of all insects surveyed. Although its function is unknown, it is the gene from which the nitrile specifier protein (NSP) of *Pieris* butterflies evolved by gene and domain duplication; NSP protects *Pieris* from the toxic effects of glucosinolates. Juvenile Hormone diol kinase functions in the degradation pathway of insect juvenile hormone which is involved in major insect physiological aspects, such as metamorphosis, diapause and reproduction.

However, the majority of the genes found are related to digestion. Particularly the protein breakdown enzymes are an indicator for typically expressed insect midgut genes. Whether

these genes are just generally activated through food intake or triggered by particular food sources is not clear. None of the genes was particularly expressed after special treatment in form from certain food source in this project. Some of the trypsins (1 DS, 2 MO) may have been induced by starvation (appearing in Pattern 2). As mentioned before the reduced weight gain in *L. monacha* and *L. dispar* on spruce could be a consequence of larvae stage and affect the expression of certain fragments and genes.

Outlook

Further research needs to be applied to find out the details about insect host interaction and their genetic bases. Experiments have to be done with more care to really differentiate between food sources preferably just for one species, rather than for two species. The focus has to lie on the exact analysis of differentially expressed genes within one species feeding on different food sources. There through larvae have to be in a similar fitness stage. The cDNA-AFLP technique is a good method to gain information potential candidate genes, when little genetic background is available. It can play a fundamental role in molecular entomology in combination with a cDNA library. However, the genetic underlying mechanism of detoxification and digestion of the two different *Lymantria* species exposed to different food sources remains to be determined.

3. Differential gene expression in generalist and specialist lepidopteran herbivores in response to feeding on different phenotypes of oak trees

3.1. Abstract

Interactions between trees and their pathogens/parasites are a driving factor for developmental and therefore also for evolutionary processes in forest ecosystems.

Genetic analyses of both system partners give insight into such processes. Here, the focus is on interactions between the oak species *Quercus robur* and insect species feeding on it, *Tortrix viridana* (green oak leafroller) and *Lymantria dispar* (gypsy moth). Observations in forests in North Rhine-Westphalia revealed that in years with high *T. viridana* population densities not every oak tree is defoliated to the same extent. In the spring, oaks can be detected growing side by side, one of which is totally defoliated (defined as "sensitive") and the other not (defined as "tolerant"). The interpretation of this observation is that generally oaks are susceptible to herbivory by T. viridana, but individual trees vary in the level of their ability to resist herbivory. Furthermore, since trees exhibit the same susceptibility phenotype over several years, at least some of the phenotype is determined by genetic differences.

Quercus robur – Tortrix viridana is the model system for this study because the green oak leaf roller is an oligophagous moth with a strong dependence on its host and therefore we expect the involvement of coevolutionary processes. For comparison between specialist and generalist herbivores, the generalist *Lymantria dispa*r is used, because it can cope with a wide spectrum of host plants. Two different instars of *L. dispar* are included to examine potential feeding behaviour differences due to developmental factors. In this study the following questions are raised:

How does a generalist (*L. dispar*) and a specialist (*T. viridana*) deal with different phenotypes (genotypes) of oak trees and their possible defence responses? And which genes are putatively involved in adaptation to different host plants in *L. dispar* and *T. viridana*? Furthermore, is there a difference between different instars of *L. dispar* feeding on the two different oak types?

Feeding experiments were conducted on the two different oak phenotypes for both species and two instars (weight groups) of *L. dispar*. Afterwards an expression profiling study (mRNA analyses) via cDNA-AFLP analysis was carried out. Differences in banding patterns revealed by gel electrophoresis or capillary electrophoresis between the treatments and for the different species were documented. Potential candidate genes were isolated and identified. Sequences were compared to cDNA libraries created from *T. viridana* and *L. dispar* larvae. The expression of the identified genes was checked via real-time q-PCR. Evidence for differential expression of some genes in the herbivores was found, which may partly explain

the effects that susceptible vs. tolerant oak trees have on their specialist and generalist herbivores.

3.2. Introduction

Evolution of ecosystems is mediated by trophic interactions such as between trees and their symbionts and/or pathogens/parasites. These dynamic systems can be a driving factor for developmental and therefore also for evolutionary processes in forest ecosystems. Phenology of trees for example governs interactions with entire pollination and herbivore groups. Variation in plant chemistry can also influence herbivory (Osier et al. 2000; Rieske et al. 2003). Well studied examples are phenolic glycosides which are known to have usually negative effects on herbivore development when present in high concentrations (Hemming & Lindroth 2000; Osier & Lindroth 2001).

The plant's response to insect feeding is quite well studied with regard to defence responses of the plant (Arnold & Schultz 2002; Babst et al. 2008). Terpenoids, nicotine and tannins are secondary plant compounds, which make the consumption of plants such as tobacco or oak quite a dangerous adventure to potential herbivores (Agrawal 2000; Baldwin 2001). It has been recognized for a long time that plants have evolved a huge variety of different defence mechanisms to reduce insect attack, which has led to the evolution of certain specialisations in both interacting partners (Fraenkel 1959). As a consequence, insects adapt new strategies to overcome plant defences, which still allow them to feed, grow and reproduce on their host plants. This battle of overcoming defence mechanism and producing new weapons, is an ongoing process in the evolution of plant-insect interactions (Mello & Silva-Filho 2002).

While the plant's response to insect feeding has been studied to a reasonable level in recent years (Allmann & Baldwin 2010; Baldwin 2001; Kessler & Baldwin 2001), reactions of insects, however, feeding on different species, types or clones of host plants are studied to a much lesser extent, in particular on the molecular level (Halitschke et al. 2001). Genes involved in digestion or other biochemical processes in herbivores have been studied so far mainly in a few model organisms, such as *Manduca sexta (Pauchet et al. 2010; Zavala et al. 2008).* It has been shown that *Q. robur* contains inhibitors of trypsin proteases, which might influence the development of herbivores. Interesting is the different adaptation to such inhibitors by specialists or generalists. Investigations have shown that the specialist *T. viridana* produces many trypsin-like proteases and thus is rather maladapted to mature leaves of oak, containing several trypsin –inhibitors. However the generalist *L.dispar* produces mainly non-trypsin proteases, which are inaccessible to the inhibitors of the plant (Ivashov et al. 2001). Another adaption to a possible host defense is the slow starvation rate of gypsy moth larvae, which enables the herbivore to survive even if larvae appear before leaves are available in the spring (Hunter 1993).

The green oak leaf roller *Tortrix viridana* (Lepidoptera: Tortricidae) is an oligophagous herbivorous insect, restricted to members of the genus *Quercus* as host plants (Hunter &

Elkinton 2000; Schroeder & Degen 2008); DU Merle, 1999). Its geographic distribution corresponds to its major host plant *Q. robur* and covers nearly the entire westpalaeartic region. Within the spectrum of other oligophagous herbivores feeding on oak, *T. viridana* can be regarded as one of the most damaging insects, often leading to almost complete defoliation in spring. This, in turn, can lead to oak decline, a disease complex described since the last century (Schroeder & Degen 2008). Interestingly, observations in European forests revealed that in years with high *T. viridana* population densities, not every oak tree is defoliated to the same extent. In spring oaks can be detected growing side by side of which one is totally defoliated (defined as "*T. viridana* sensitive") while the neighbouring oak tree is still well foliated (defined as "*T. viridana* tolerant"). The main cause for this phenomenon is unclear so far, but differences in oak chemical constitution have been suspected as a consequence of different genotypes.

The gypsy moth *Lymantria dispar* (Lepidoptera: Lymantriidae) is a polyphagous insect with a wide range of host plants, among them being oak as a preferred food source (Keena et al. 2008). This species is known as a voracious defoliator of palaearctic and nearctic forests, found in temperate areas from North Africa throughout Eurasia to the Japanese Islands. Since the introduction of moths originating from France into North America in 1869, the gypsy moth has expanded its range as far as Canada, Michigan, and North Carolina and is regarded as one of the major forest pests in the US (Keena et al. 2008).

Both insect species, *T. viridana* as a specialist and *L. dispar* as a generalist, have the capability of causing defoliation in oak forests and tend to outbreaks in large numbers. They also have to cope with plant defence mechanisms, yet so far it is not known whether oaks try to defend themselves in the same manner when attacked by either of the two insect species.

Here, we describe differences in gene expression patterns in larvae of these two insect species after feeding on "*T. viridana* sensitive" or "*T. viridana* tolerant" oak trees. We were particularly interested in asking the following questions: How does a generalist (*L. dispar*) and a specialist (*T. viridana*) deal with different phenotypes of oak trees and their possible defence responses? And which genes are putatively involved in adaptation to different host plants in *L. dispar* and *T. viridana*? As both species are non-model organisms and genomic information is scarce so far, we applied cDNA-AFLPs to identify any genes up-or down-regulated in the two species as a result of feeding on the two different types of oak host plants and tried to verify differences in expression patterns with q-PCR.

3.3. Material and Methods

Insects

Lymantria dispar eggs were initially received from a laboratory rearing colony (Melody Keena, USDA Forest Service Hamden, CT) and were then continuously reared on artificial

gypsy moth diet under controlled conditions in breeding chambers (24°C, day/night cycles of 10 hours day light 14 hours dark). Larvae were kept in plastic containers with artificial diet, then were separated in small round plastic containers with artificial diet under 25°C until the experiment started. Diet ingredients were: de-ionized water, agar, wheat germ, Wesson salt mix with Fe, sucrose, casein, sorbic acid, methyl parabenzene, ascorbic acid, wheat germ oil, raw linseed oil, choline chloride, cholesterol, ferric citrate, vitamin mixture; (Keena, unpublished).

Tortrix viridana larvae were collected in the wild with the help of a sling shot, by which a rope is shot into the tree canopy. Shaking the two ends of the rope caused larvae to fall down onto a plastic sheet where they were collected. Larvae were kept at 20-22°C in white plastic containers (19, 5 x 10, 5 x 5, 5 cm) with a clear lid and were fed daily with fresh oak leaves until pupation. The adult moths were allowed to lay eggs. In the following year the newly hatched first instars were fed with just-opening buds, the later third instars with fresh oak leaves. Feeding lasted for about 10 to 14 days until the feeding experiments started. For the feeding experiments third instars were used.

Plant material

In spring 2006 three pedunculate oak (*Quercus robur*) stands in North Rhine-Westphalia, near the city of Muenster (Longitude: 7˚34', range 8.73', Latitude: 51˚58', range 6.79', elevation: 50 - 80 m and total area of about 70 x 20 km) were used as sample collection sites. The terrestrial ecosystem in these stands is a mixed boreal cultivated forest, containing between 60-85% oak trees (*Quercus robur*). In June temperatures were around 24°C at day and 16°C at night. One observed phenomenon in these stands in Muenster is that trees which stand right next to each other (5-10m apart) experience totally different defoliation by *T. viridana*. The classification T (for tolerant, i.e. not heavily defoliated by *T. viridana*) and S (for sensitive, i.e. heavily defoliated by *T. viridana*) was made by Dr. Hilke Schroeder after several years of monitoring degrees of *T. viridana* defoliation in this forest (Schroeder 2010). In this study the treatments for *L. dispar* larvae feeding on "oak tolerant" are called DOR and on "oak sensitive" DOS, respectively. The same samples are the oak treatment for *L. dispar* in the previous study (see details under chapter 2). For *T. viridana* OS is used for "oak sensitive" and OT for "oak tolerant".
Collection of oak leaves was conducted with the help of a sling blade, where a bag filled with little metal nuggets connected to a rope was shot into the tree canopy, so that a desired branch could be pulled down. Leaf samples were directly taken from the hot spot of feeding activity of *T. viridana* and were stored in plastic buckets with water for transportation. A total of 6 sensitive trees and 8 tolerant trees were sampled with the sling blade. Feeding assays started soon after leaf sample collections from tree twigs.

Feeding experiments

Experiments were conducted in July 2007 at 24 °C in the laboratory under controlled conditions. For the experiment by *L. dispar* a comparison between smaller (3rd instar) and larger (4th instar) larvae was undertaken. Larvae with a starting weight of 40 mg were called "DOR S" and "DOS S", respectively. Larvae with a starting weight of 25 mg were called "DOR L" and "DOS L". Calculations for the feeding experiments were separated by starting weight and depicted in graphs.

Larvae were 48 h post molt and starved for 24 h before the experiments started. Oak leaves were offered in petri dishes to the larvae of *L. dispar.* Branches of oaks from the two different phenotypes (*T. viridana* "sensitive" (OS) and "tolerant" (OT)) having approximately 3-4 expanded leaves were put through a hole in the lid of a 1.5 ml Eppendorf tube filled with water. Single oak branches were laid into a petri dish (100 x 25 mm) and one *L. dispar* larva was put into each dish with the lid closed and was allowed to feed for 24 h on the oak leaves. Before the feeding assay, larval individual weights were recorded. In all experiments, larvae with obvious signs of feeding by production of feces and leaf damage symptoms were collected and their weight was recorded. Statistical analysis on the weight gain of the larvae was conducted using a Tukey HSD Test. In the lighter weight group 20 larvae were fed on (DOS L) leaves and 31 larvae were fed on (DOR L) leaves. In the heavier group 50 individuals were fed on (DOR S) leaves and 40 on (DOS) S leaves, respectively.

The experimental setup was conducted in parallel for *Tortrix viridana* larvae in their 3rd instar, the major feeding stage. In the feeding assay 10 larvae per oak clone were used. In total 4 sensitive and 4 tolerant oak clones were offered to larvae. Altogether 40 larvae each were fed on sensitive and tolerant oak trees for 24h, respectively.

In all experiments, larvae with obvious signs of feeding (production of feces and leaf damage) were collected, the midguts were dissected from the living insect and then directly shock frozen in liquid nitrogen and stored at -80 °C. All larvae were measured before and after feeding with an analytic balance (Denver Instrument SI234, max 230 g).

RNA extraction and cDNA synthesis

RNA of single midguts from both *L. dispar* and *T. viridana* larvae after feeding on the two different oak phenotypes was extracted using TRIzol Reagent (Invitrogen, Karlsruhe, Germany) according to the manufacturer's protocol with a slight modification: a plastic pestle grinder powered by a Pellet Pestle Cordless Motor (Kontes) was used for homogenization of midgut tissues in 2 ml Eppendorf tubes. Quality and quantity of RNA was checked via gel electrophoresis and a Nanodrop spectrophotometer, respectively. Residual DNA was removed by DNase treatment followed by purification using an RNeasy Minelute Cleanup Kit (Qiagen, Hilden, Germany).

In *L. dispar* 6 single RNA preparations of 1 midgut per larvae for each treatment were pooled in equimolar amounts prior to cDNA synthesis. Considering the different weight groups for *L. dispar*, there were 4 pools altogether. In *T. viridana* between 9 midgut RNAs of individual larvae from the OS treatment and 6 from the OT treatment were pooled, respectively. The pools of RNAs (1 µg per pool, designated as OS and OT) for both species separately were translated into cDNA with the SMART PCR cDNA Synthesis Kit (Clontech). Afterwards cDNA was purified using a PCR Purification Kit (Qiagen) and the quantity was checked spectrophotometrically.

cDNA-AFLP analysis

For cDNA-AFLP reactions, 500 ng purified cDNA was digested with 5 U *Eco*RI or *Pst*I (for *L. dispar* and *T. viridana*, respectively) and 3 U *Mse*I for 2 h at 37°C followed by enzyme inactivation for 15 min at 65°C. Accordingly, *Eco*RI or *Pst*I, and *Mse*I double stranded adapters were ligated to the fragments by incubating for 2 h at 16°C. Pre-amplification of cDNA fragments was performed with a 1:10 dilution of the ligation reaction for 20 cycles with 94°C for 30 s, 56°C for 1 min, 72°C for 1 min using primers corresponding to *Eco*RI or *Pst*I and *Mse*I adapters (table 1). Resulting PCR products were diluted 50x and were subjected to selective amplification with a combination of different *Eco*RI or *Pst*I and *Mse*I selective amplification primers using a touchdown amplification PCR-profile as described by Vos et al. (1995), with each primer having one or two extra nucleotides at their 3`-termini.

A total of 64 *Eco*RI and *Mse*I primer combinations were initially tested for *T. viridana* with a gel electrophoresis system. Subsequently for *T. viridana* 43 *Pst*I and *Mse*I primer combinations and 44 *Eco*RI and *Mse*I primer combinations for *L. dispar* were tested on a capillary electrophoresis apparatus. The *Eco*RI restriction enzyme didn't show sufficient banding patterns on the resulting gels or chromatograms for *T. viridana,* and so switching to the *Pst*I primer was required.

The following 64 different AFLP-primer combinations were used for selective amplification, due to a combination of 16 *Mse*I-NN primers (NN is either AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG or TT) and 4 *Eco*RI/ *Pst*I -N primers (N is either A, G, T or C). For the capillary electrophoresis unit combinations of Eco +1 and Mse +2 or 3 different nucleotides were tested for both species (table 2).

To allow a fluorescent labelling of the generated AFLP products for the capillary electrophoresis, an M13(-21) tail was placed at the 5'-end of each selective *Eco*RI or *Pst*I primer and a fluorescently labelled universal primer M13(-21) was added to the PCR reactions according to the method described by (Schuelke 2000). Different fluorescent dyes (IRD700 or IRD800) were used for labelling the universal M13 (-21) primer, enabling the loading of the reactions as a multiplex reaction.

Table 1: cDNA-AFLP standard primer sequences

*Eco*RI adapter	forward:	5'-CTCGTAGACTGCGTACC-3'
	reverse:	5'-AATTGGTACGCAGTCTAC-3'
*Mse*I adapter	forward:	5'-GACGATGAGTCCTGAG-3'
	reverse:	5'-TACTCAGGACTCAT-3'
*Eco*RI pre-amplification		5'-GACTGCGTACCAATTC-3'
*Mse*I pre-amplification		5'-GATGAGTCCTGAGTAA-3'
*Pst*I adapter	forward:	5'-CTCGTAGACTGCGTACATGCA-3'
	reverse:	5'-CATCTGACGCATGT-3'
*Pst*I pre-amplification		5'-GACTGCGTACATGCAG-3'

Table 2: Subset of primer combinations used for the fragment analysis via capillary electrophoresis for *L. dispar* and *T. viridana*.

L. dispar	***T. viridana***
Eco A Mse TC	Pst G Mse CC
Eco T Mse TC	Pst G Mse CCC
Eco G Mse TC	Pst G Mse CCG
Eco C Mse TC	Pst G Mse CCT
Eco A Mse GT	Pst G Mse CCA
Eco T Mse GT	Pst C Mse AG
Eco G Mse GT	Pst C Mse AGC
Eco C Mse GT	Pst C Mse AGG
Eco A Mse GA	Pst C Mse AGT
Eco T Mse GA	Pst C Mse AGA
Eco G Mse GA	Pst G Mse GC
Eco C Mse GA	Pst G Mse GCC
Eco A Mse CA	Pst G Mse GCG
Eco T Mse CA	Pst G Mse GCT
Eco G Mse CA	Pst G Mse GCA
Eco C Mse CA	Pst A Mse AT
Eco A Mse TG	Pst A Mse ATC
Eco T Mse TG	Pst A Mse ATG
Eco G MseTG	Pst A Mse ATT

L. dispar	*T. viridana*
Eco C Mse TG	Pst A Mse ATA
Eco A Mse AC	Pst C Mse CG
Eco T Mse AC	Pst C Mse CGC
Eco G Mse AC	Pst C Mse CGG
Eco C Mse AC	Pst C Mse CGT
Eco A Mse TA	Pst C Mse CGA
Eco T Mse TA	Pst A Mse AG
Eco G Mse TA	Pst A Mse AGA
Eco C Mse TA	Pst A Mse AGT
Eco A Mse GC	Pst A Mse AGC
Eco T Mse GC	Pst A Mse AGG
Eco G Mse GC	Pst C Mse CC
Eco C Mse GC	Pst C Mse CCA
Eco A Mse AA	Pst C Mse CCT
Eco T Mse AA	Pst C Mse CCC
Eco G Mse AA	Pst C Mse CCG
Eco C Mse AA	Pst A Mse GG
Eco A Mse AT	Pst A Mse GGC
Eco T Mse AT	Pst A Mse GGG
Eco G Mse AT	Pst A Mse GGT
Eco C Mse AT	Pst A Mse GGA
Eco A Mse AG	Pst C Mse AA
Eco T Mse AG	Pst C Mse CT
Eco G Mse AG	Pst C Mse CA
Eco C Mse AG	

Fragment Analysis and Isolation

The cDNA-AFLP fragments were loaded on two different electrophoresis systems: a submarine gel matrix based electrophoresis system (Elchrom Scientific) and a capillary electrophoresis system (Beckman-Coulter). Chromatograms were analysed for differences in gene expression patterns with the GenomeLab software (Beckman). Each PCR reaction was repeated three times to check for reproducibility of gene expression patterns.

Selective amplification PCR reactions containing fragments which were differentially expressed after feeding on “sensitive” or “tolerant” oak phenotypes, respectively, were loaded on ready-to-use 9 % PolyNat gels and separated by electrophoresis (Elchrom Scientific). By SYBR Safe staining differentially expressed bands could be visualized and then picked out of the gel with a tool called BandPick (Elchrom Scientific). Desired fragments were

directly reamplified with the corresponding selective AFLP primer. PCR products were purified with ExoSAP-IT (Affymetrix) and were directly sequenced.

Sequence Analysis

Sequence analysis was done using the Bioedit sequence alignment editor tool (Ibis Biosciences). Sequences were compared by BLASTN against 23,300 transcripts from *L. dispar* and 14,304 (sensitive) and 10,176 (tolerant) transcripts from *T. viridana* from midgut cDNA libraries, which were generated at the MPI for Chemical Ecology (Vogel, unpublished). Sequences were also compared by BLASTN and BLASTX against the non-redundant sequence databases of GenBank at NCBI. To identify candidates sufficiently a threshold of an *E*-value of at least E-05 was used to filter these BLAST results.

q-PCR analysis

In *L. dispar* no q-PCR analysis was conducted. Q-PCR was done with RNA from independent *T. viridana* larvae feeding experiments (OT vs. OS). For efficiency calculation and meltcurve analysis of primers identical quantities of RNA were pooled for each sample and all samples were replicated three times. If quality check of primers and PCR conditions were acceptable, samples were run in 1:100 dilutions in three replicates including housekeeping genes. Single RNA samples were used to determine the threshold cycle (Ct) values. First strand cDNA for pools and single RNAs was synthesised from DNase treated total RNA using BioRad cDNA synthesis kit. Primers were designed after q-PCR guidelines specifically for *T. viridana*. PCR was conducted using BioRad SyberGreen PCR Master Mix in an iQ5- cycler (BioRad). The following thermal cycler profile was used: 95°C for 3 min, 40 cycles with 95°C for 30 s, 60°C for 45 s, 72°C for 30 s ending with one cycle of 95°C for 1 min. Data was analysed with the help of qbase software. Actin and GAPDH were used as reference genes to normalize gene expression levels.

3.4. Results

Performance of *L. dispar* larvae on two different oak phenotypes

In laboratory feeding experiments, heavier *L. dispar* larvae gained significantly (p=0.05) more weight after feeding on “sensitive” leaves than on “tolerant” leaves (Figure 1a). In contrast, lighter *L. dispar* larvae gained more weight on “tolerant” leaves, than after feeding on “sensitive” leaves (Figure 1b).

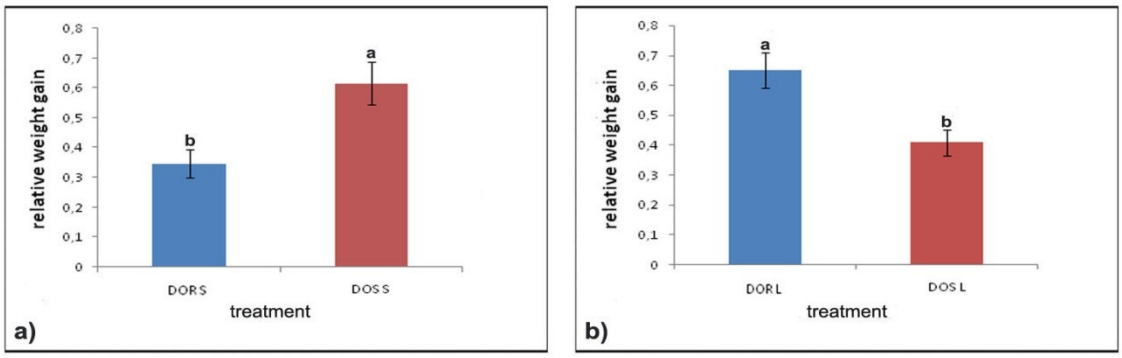


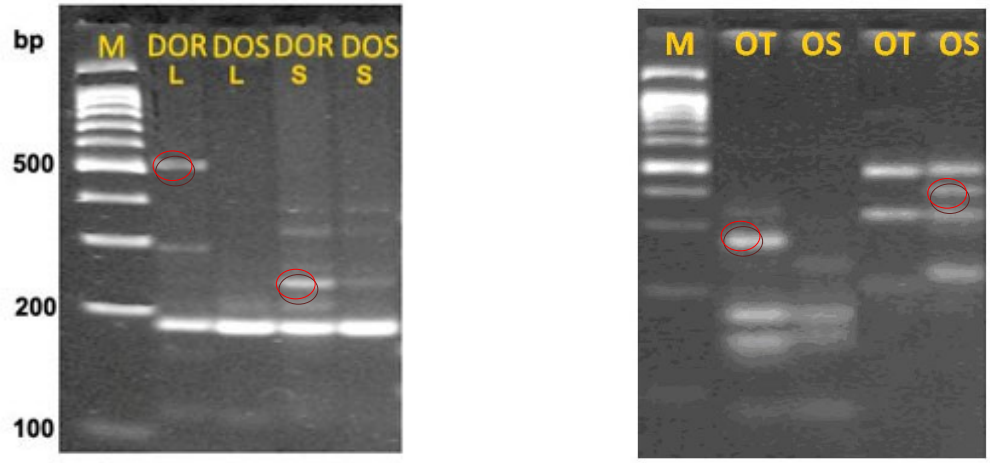
Figure 1 a): Relative weight gain (mg/mg) of heavier *L. dispar* larvae was greater, when feeding on "sensitive" (red) leaves than on "tolerant" (blue) leaves. Starting weight was approximately 42 mg per larva. Significant differences between the variables are indicated by different letters (a/b).

Figure 1 b): Relative weight gain (mg/mg) of lighter *L. dispar* larvae was greater, when feeding on "tolerant" (blue) leaves than on "sensitive" (red) leaves. Starting weight was approximately 25 mg per larva. Significant differences between the variables are indicated by different letters (a/b).

Gene expression differences in *L. dispar* and *T. viridana* larvae after feeding on two oak phenotypes analysed via cDNA-AFLPs

To examine whether gene expression profiles were different in larvae feeding on the two different phenotypes of oak trees, cDNA-AFLPs were conducted for both species with a set of selective primer combinations.

For *T. viridana* 64 primer combinations on gel electrophoresis and 43 primer combinations on capillary electrophoresis were tested for the OS vs. OT treatment, respectively. Fragments were shown by gel analysis, which were specific for each primer combination and each species (figure 2 a,b). Differences between sensitive and tolerant treatment, within one primer combination were clearly detectable, even by the naked eye. These differently expressed gel pattern could be recognized by comparing bands of specific length between treatments. For further sequencing these differences were picked out.



a) b)

Figure 2 a): Expression patterns of *L. dispar* larvae displayed by cDNA-AFLP. The example shows selective amplification with one primer combination (EcoG MseGA) for DOR and DOS (L for light and S for heavy weight group). Differentially expressed fragments are indicated by a circle.

Figure 2 b): Expression patterns of *T. viridana* displayed by cDNA-AFLP. The example shows selective amplification with two different primer combinations (PstA MseAT, PstG MseGC) for OT and OS. Differentially expressed fragments are indicated by a circle.

A subset of 44 primer combinations for both weight groups of *L. dispar* were tested by capillary electrophoresis. Results were repeated up to 3 times and the expression profiles and associated differences were always repeatable. In the capillary profiles of certain genes differences were clearly visible due to the presence or absence of certain peaks in cDNAs from larvae after feeding on “tolerant” and on “sensitive” leaves (Figs. 3 und 4). The specific chromatogram patterns were comparable with the gel electrophoresis analysis of the same primer combinations.

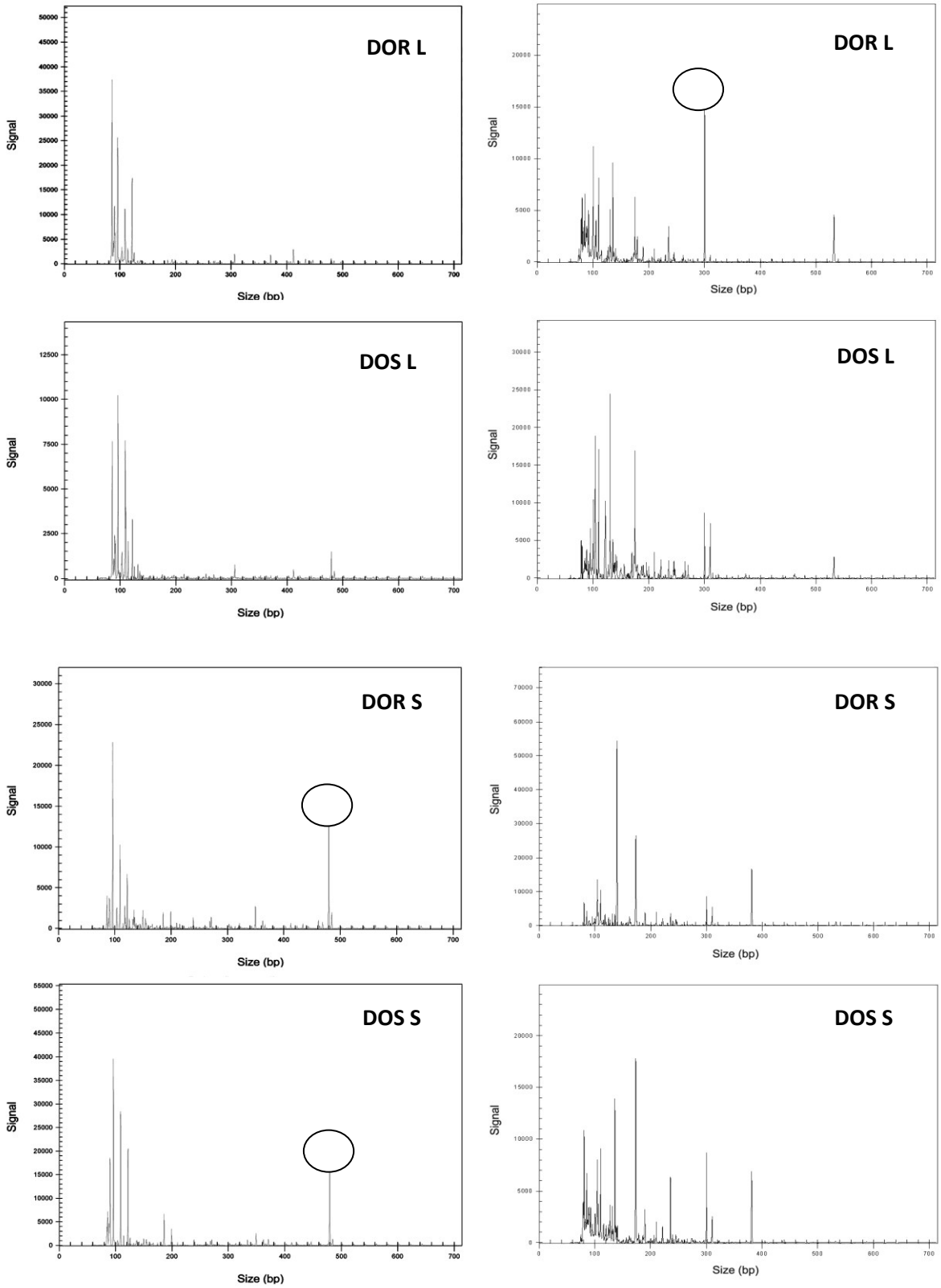


Figure 3: Capillary electrophoresis chromatograms of cDNA-AFLPs from *L. dispar* ((primer combination EcoC MseTA) (left) and EcoA MseGT (right)). Feeding on DOR and DOS oak leaves was performed by larvae of different weight groups (L for light and S for heavy). Differentially expressed fragments are indicated by a circle.

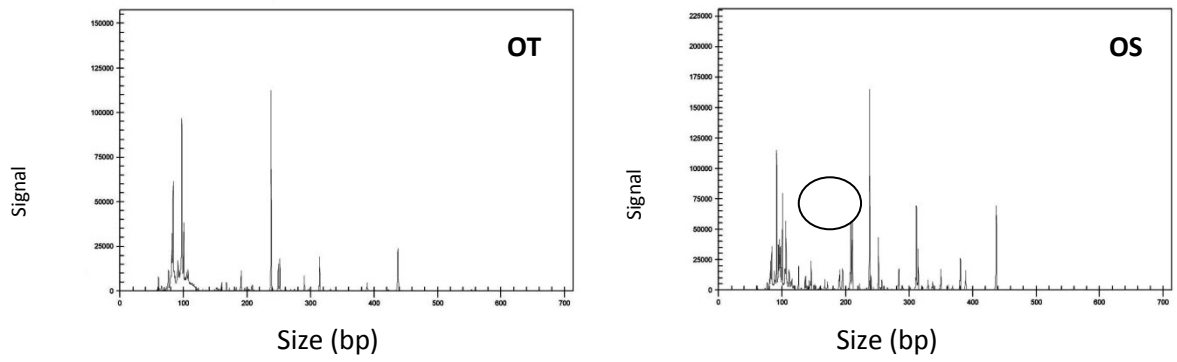


Figure 4: Amplification products visible after capillary electrophoresis obtained with cDNA-AFLPs of *T. viridana* larvae feeding on "tolerant" (left) and on "sensitive" (right) (primer combination: PstG MseGC). Differentially expressed fragments are indicated by a circle in the OS treatment.

In total 1819 fragments were generated in *T. viridana*, 487 in the gel electrophoresis by 64 primer combinations. With 17 primer combinations 1332 fragments were counted in the capillary electrophoresis unit. In *L. dispar* only capillary electrophoresis on 37 primer combinations was applied. Here 1433 fragments in the lighter larvae group and 1441 fragments in the heavier larvae group were produced (table 3).

Table 3: Overview of the number of cDNA-AFLP fragments

	species treatment						**methods**	**primer combinations**
	DOR L	**DOS L**	**DOR S**	**DOS S**	**OT**	**OS**		
fragments	721	712	720	721	/	/	capillary	37
	/	/	/	/	217	270	gel	64
	/	/	/	/	655	677	capillary	17

The expression profiles of both species are demonstrated in table 4 and certain patterns specific for the differences found were sorted in indexes. The expression profile (index 2: present in sensitive, absent in tolerant) appears 49 times compared to the opposite pattern (index 1: present in tolerant absent in sensitive), which is found 26 times in the pattern of *T. viridana* larvae on "tolerant" and on "sensitive" leaves. Also in *L. dispar* the lighter and the heavier groups showed 12 different patterns relevant for comparisons. Among these patterns were differences between the lighter and the heavier group (index: 20, 22, 23) and within the weight groups between tolerant and sensitive treatments (12-15, 17, 18, 21).

Table 4: Table 3: The table shows for which treatments the bands were present and which absent, indicated ("-" for absent and "+" for present). The index number defines a specific banding pattern, identified in the gels or chromatograms. The number and the method of the specific expression profile are given.

Index	***L. dispar***				***T. viridana***		**Relevance**	**no. of expression profile , method**
	DOR L	**DOS L**	**DOR S**	**DOS S**	**OT**	**OS**		
1					+	-	present in tolerant absent in sensitive	26 gel, 14 capillary
2					-	+	present in sensitive, absent in tolerant	49 gel, 36 capillary
3					+	+	present in both treatments	412 gel, 1282 capillary
12	++	+	+	+			DOR L more expressed within all treatments	3 capillary
13	+	-	-	-			oak tolerant in Ld L	9 capillary
14	-	-	+	-			oak tolerant in Ld S	2 capillary
15	+	+	++	+			DOR S more expressed within all treatments	1 capillary
16	+	+	+	+			all the same	3 capillary
17	-	+	++	+			DOR S more expressed within 3 treatments	1 capillary
18	+	-	+	+			DOS L suppressed	2 capillary
19	++	+	+	++			DOR L and DOS S more expressed	1 capillary
20	-	++	++	-			DOS L and DOR S present	3 capillary
21	-	+	-	-			DOS L present	1 capillary
22	-	-	+	+			comparison between L and S	2 capillary
23	++	++	+	+			comparison between L and S, L more expressed	2 capillary

Genes

In *T. viridana*, 16 interesting bands were picked. Bands were successfully reamplified before the cleaned up PCR products were sequenced (table 5). In *L. dispar* 7 interesting differences in both weight groups were picked and successfully reamplified (table 6). Sequence of respective AFLPs containing transcripts from the *L. dispar* or the *T. viridana* cDNA library were compared to GenBank entries using the BLASTX tool and in those transcripts with significant homologies (E values lower than 1E-05), hits were determined. The database showed homologies to insects, in particular to lepidopteran model organism such as *Bombyx mori*.

Potential candidate genes were found in the sensitive and the tolerant treatments, respectively for both species. Upregulated potential genes were more often found in the tolerant treatments 10 out of 16 (index: 1) by *T. viridana* (table 5). In *L. dispar* 4 out of 7 genes (12, 13, 14 and 18) identified were from the tolerant treatment (table 6). Furthermore genes were isolated regarding the comparison between the weight groups (light and heavy). In the lightweight group there were 5 out of 7 genes identified (table 6).

The primary interest of this research was to identify putative candidate genes which differ between *L. dispar* and *T. viridana*, when fed on sensitive vs. tolerant oak leaves and might be involved in adaptation to certain abiotic conditions or to a respective host plant present in the given habitat. The genes can be also engaged in developmental stages of *L. dispar* of the different weight groups.

These genes in *T. viridana* are likely to code for an elongation factor, astacin, flavin-dependent monooxygenase, U2 small nuclear ribonucleoprotein and Arylphorin protein, respectively. Furthermore there are ATP-synthase, a putative lipase, hydroxypyruvate isomerase, U3 small nuclear RNA-associated protein 14-like protein A and fibroin heavy chain representative found in the identified gene pool (table 3).
In *L. dispar* trypsin Ia precursor, beta-glucosidase precursor, carboxypeptidase 3, protein disulfide isomerase, hypothetical protein EAG_13571 and projectin were identified (table 5).

Table 5: Candidate genes found for *T. viridana* after isolating differentially expressed fragments in the respective electrophoresis system (either capillary or submarine gel electrophoresis).

Index	**Sample**	**E-value**	**Homology**	**Putative function**	**Electrophoresis system**
2	**10 OS**	8.00E-13	elongation factor 1-alpha (*Lissorhoptrus sp.*), ACP50167.1	protein synthesis	capillary electrophoresis
1	**11 OT**	3.00E-90	astacin metalloprotease 1 (*Plutella xylostella*), ADX30518.1	proteolysis	capillary electrophoresis

2	**1 OS**	3.00E-05	flavin-dependent monooxygenase FMO1B (*Bombyx mori*), ADH16747.1	oxidation-reduction process	capillary electrophoresis
2	**17 OS**	4.00E-24	U2 small nuclear ribonucleoprotein A' (Bombyx mori), NP_001040446.1	mRNA splicing	capillary electrophoresis
2	**18 OS**	1.00E-33	Arylphorin (Manduca sexta), P14296.1	lipid transport	capillary electrophoresis
1	**b02 OT**	3.00E-123	eukaryotic initiation factor 4A (Plutella xylostella), ABY66383.1	protein biosynthesis	submarine gel electrophoresis
1	**b05 OT**		DUF233 Protein, (Heliothis virescens)	odorant binding protein	submarine gel electrophoresis
2	**b08 OS**	5.00E-138	ATP synthase-like protein (Choristoneura parallela), AAN39695.1	ATP-synthesis	submarine gel electrophoresis
1	**b09 OT**	6.00E-45	putative lipase (Bombyx mori), ADA67928.1	Hydrolase, lipid metabolic process	submarine gel electrophoresis
1	**b10b OT**	6.00E-85	DEAD box polypeptide 5 (Papilio xuthus), BAG30754.1	ATP-binding	submarine gel electrophoresis
2	**b11 OS**	2.00E-135	hydroxypyruvate isomerase (Bombyx mori), ABY57912.1	dicarboxylate metabolism	submarine gel electrophoresis
1	**b13 OT**		Niemann-Pick type C-2h	Sterol binding	submarine gel electrophoresis
1	**b14 OT**	0.004	U3 small nucleolar RNA-associated protein 14-like protein A (Harpegnathos saltator), EFN89503.1	rRNA processing	submarine gel electrophoresis
1	**b17 OT**	8.00E-05	GA26998 (Drosophila pseudoobscura), XP_002134022.1	unknown	submarine gel electrophoresis
1	**b22 OT**	4.1	Conserved oligomeric Golgi complex subunit 2 (*Camponotus floridanus*), EFN61983.1	regulator of the Golgi-glycosylation machinery	submarine gel electrophoresis
1	**b24 OT**	6.00E-06	fibroin heavy chain (*Ephestia kuehniella*), AAP79134.1	core component of silk filament; chemically inert fiber.	submarine gel electrophoresis

Table 6: Candidate genes found for *L. dispar* after isolating differentially expressed fragments in the respective electrophoresis system (either capillary or submarine gel electrophoresis).

Index	Sample	E-value	Homology	Putative function	Electrophoresis system
12	**2 DOR L**	5.00E-17	trypsin Ia precursor (*Sesamia nonagrioides*), AAT95348.1, Lymantria library hit	Proteolysis	capillary electrophoresis
14	**8 DOR S**		Lymantria library hit	unknown	capillary electrophoresis
13	**11 DOR L**	7.00E-58	beta-glucosidase precursor (*Spodoptera frugiperda*), AAC06038.1	digestion	capillary electrophoresis
18	**17 DOR L**	8.00E-06	carboxypeptidase 3 (*Mamestra configurata*), ACR16009.1, Lymantria library hit	Proteolysis	capillary electrophoresis
20	**3 DOS L**	7.00E-20	protein disulfide isomerase (*Bombyx mori*), NP_001037171.1, *L. dispar* library hit	catalyse protein folding	capillary electrophoresis
22	**9 DOS S**	2.00E-38	projectin (*Pediculus humanus*), ADY75705.1, *L. dispar* library hit	member of the immunoglobulin superfamily, function unknown	capillary electrophoresis
23	**14 DOS L**	6.00E-27	hypothetical protein EAG_13571 (*Camponotus floridanus*), EFN65448.1, *L. dispar* library hit	unknown	capillary electrophoresis

q-PCR

The expression level of 4 transcripts of *T. viridana* was analysed via q-PCR to verify the cDNA-AFLP patterns. These genes and the corresponding primers were used, because they showed good melting curves and had an efficiency of 100%. Actin and gapdh were chosen as stable reference genes (figure 5). The tested genes showed a non-validating gene expression in comparison to the cDNA-AFLP methods. None of the tested genes showed a significant difference between the OT and OS treatment (Mann-Whitney-Test, see table 7). For the 4 target transcripts, the distribution of the single RNAs wasn't evenly spread within the treatments, which could have influenced the analysis outcome (figure 6 and 7).

Tabel 7: Statistical results of the Mann-Withney-Test for the comparison between OT and OS treatment in *T. viridana*.

Genes	**Annotation**	**p-Value**
1 OS	astacin metalloprotease 1	n.s.
10 OS	elongation factor 1-alpha	n.s.
11 OT	astacin metalloprotease 1	n.s.
17 OS	U2 small nuclear ribonucleoprotein A'	n.s.

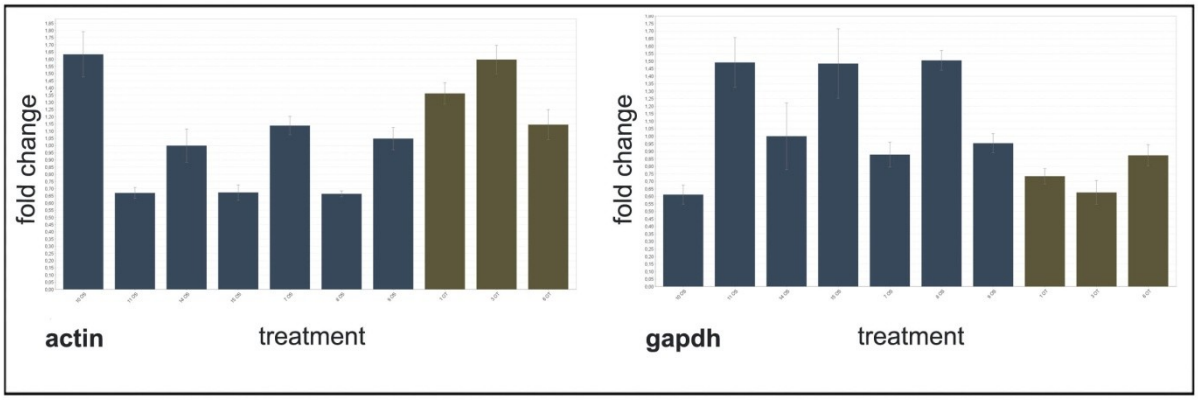


Figure 5: q-PCR analysis of 2 reference transcrips (actin, gapdh) of *T. viridana* on OT (brown) and OS (blue) trees. Data shows fold change of gene expression in „tolerant" and „sensitive" oak trees. Bars represent single RNAs tested.

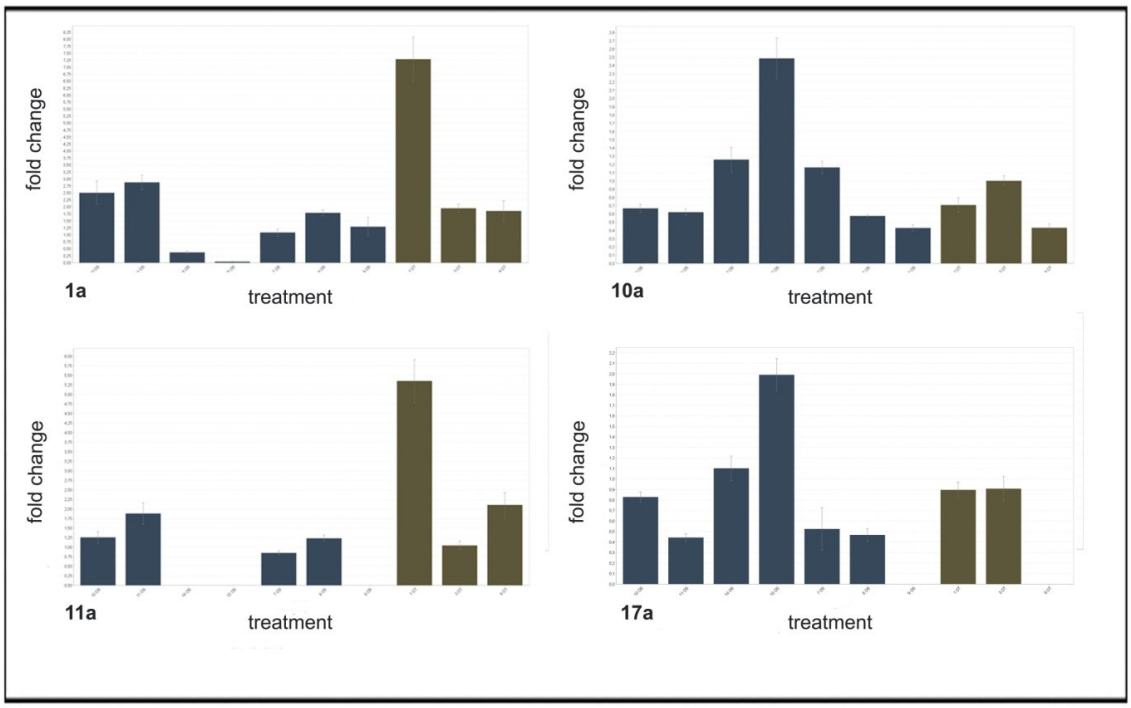


Figure 6: Q-PCR analysis of 4 transcripts (table 7) of *T. viridana* on OT (brown) and OS (blue) trees. Gene annotation correspond to numbers on graphs (table 5). All data were normalized to actin and gapdh expression levels. Data shows fold change of genes expression in „tolerant“ and „sensitive“ oak trees. Bars represent single RNAs tested.

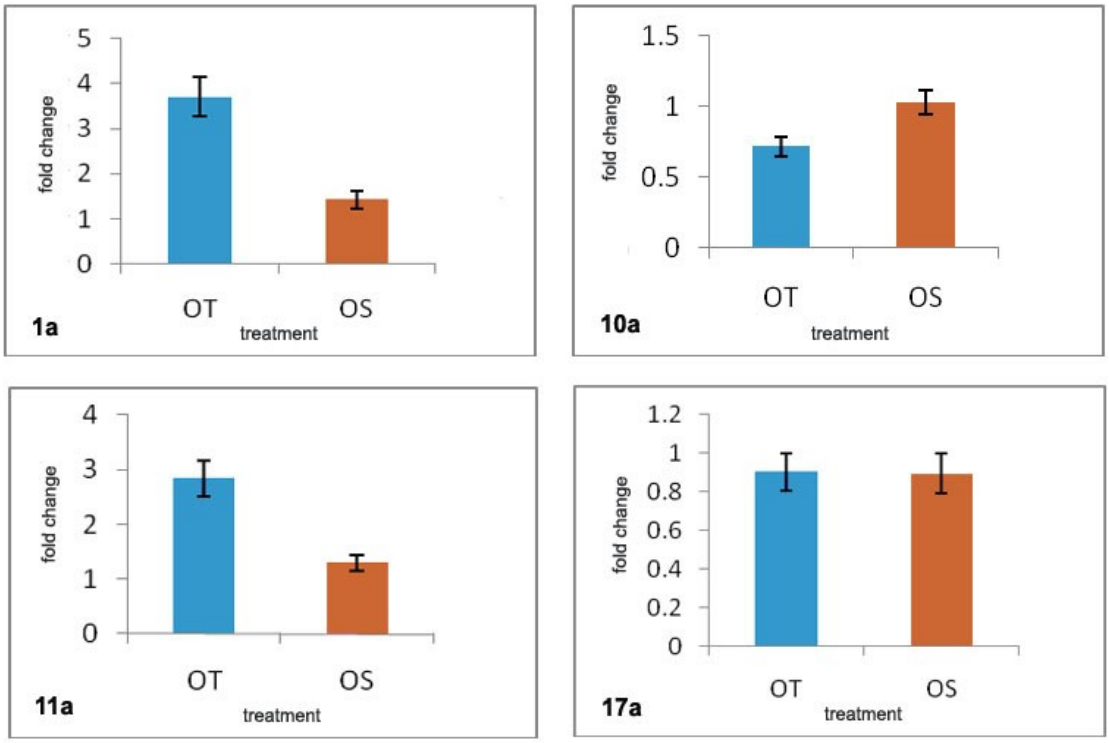


Figure 7: q-PCR analysis of 4 transcripts of *T. viridana* on OT and OS trees. Gene annotation correspond to numbers on graphs (table 7). All data were normalized to actin and gapdh expression levels. Data shows fold change of gene expression in OT and OS trees. Bars represent the global mean of single RNAs tested.

3.5. Discussion

The major question of this approach was why do some trees of the same species show lower levels of defoliation following insect attacks than others located directly next to each other within one forest? A possible explanation is that these trees reveal a different "defence response mechanism" against insect larvae. The question was analysed from the insect point of view.

By experimental circumstances there were two weight groups of *L. dispar* larvae in the comparison of tree phenotypes involved. In this study it was observed that the heavy weight group of the gypsy moth *L. dispar* – a rather generalist species - reacts similarly to the specialist species, the green oak leaf roller *T. viridana*, when feeding on the "sensitive" and "tolerant" oak phenotypes, previously designated. The lighter weight group and therefore also the younger larvae showed rather the opposite effect of feeding behaviour than the older larvae of *L. dispar* or *T. viridana*. They gained more weight on the "tolerant" leaves than on the "sensitive" trees. The reason for this could be the more robust feeding behaviour within younger developmental stages of larvae and the higher resistance to possible defense substances of the "tolerant" trees. It is known from the literature that younger larvae of insects have to cope with a wide spectrum of defence responses, particularly when feeding on fully expanded and mature leaves (Travers et al. 2007). One would suspect that the tree is allocating more defense compounds to buds and young leaves. Why older *L. dispar* larvae show sensitivity towards tolerant trees and gain more weight on sensitive trees cannot be fully explained here. It has been shown that older and younger larvae of *L. dispar* accumulate a different mixture of compounds in the hemolymph, and that the concentration of some of these compounds is affected by the food source (Deml 2004). Polyphagous herbivores have to balance a wide spectrum of plant toxins and nutrition uptake and demonstrate a higher fitness through more moving between hosts than do specialists (Pearse 2011). The fact that the young larvae eclosion of *T. viridana* is adjusted to the bud burst of the oak leaves contributes well to the ability to feed on the young tissue, where for example inhibitors to e.g. trypsin protease are less abundant than in older tissue (Ivashov et al. 1992).

One possible explanation for slower growth of *T. viridana* and the larger *L. dispar* could be the quantity of chemical defense compounds in the different oak types. The presence of symbiotic ectomycorrhizal fungi is known to change plant metabolism which could also influence plant suitability for herbivory (Rieske 2001). Whether the two oak types are differentially colonized by ectomycorrhizal fungi is, however, unknown. Insect performance on plants is greatly influenced by the respective plant's genotype: phytochemical variation among different clones of quaking aspen has been shown to affect performance of gypsy moth larvae (Osier & Lindroth 2001). Certain oak species, which contain higher levels of carbohydrates, tannins and nitrogen and can alter seasonally (Feeny 1970), are better host trees than oaks with lower levels of these compounds (Foss & Rieske 2003). From the plant side, interactions between generalist or specialist lepidopteran larvae with their host plant

have been shown to be influenced by certain larval elicitors (Voelckel & Baldwin 2004). These elicitors in turn may induce the plant to increase its chemical defenses after an initial exposure to herbivory. After previous feeding by gypsy moth larvae, leaves of red oak trees showed increased amounts of phenolics and tannins and greater physical toughness (Schultz & Baldwin 1982). Tannins change seasonally and become more abundant in older leaves in oak, so early feeding brings advantages in insect development (Van Asch & Visser 2007). Moreover the observation of spring feeders, which tend to have more outbreak events, leads to the result, that development and fecundity of the following population have a positive effect when feeding starts in early instars (Hunter 1993). The precise chemical analysis of the two different oak types in this study would help in investigation of the different response of the lighter and heavier generalist larvae and the specialist larvae.

Rather different sets of genes were affected in the two insect species by feeding on the two different phenotypes of oaks. In the specialist *T. viridana* feeding on sensitive oaks, upregulated genes included the U2 small nucleoprotein A' involved in mRNA splicing, and elongation factor 1, involved in protein synthesis. These could promote synthesis of all proteins in the cell in a non-specific manner. Also upregulated were hydroxypyruvate isomerase which participates in intermediary metabolism of dicarboxylates, an ATP synthase involved in energy metabolism, and a flavin-dependent monooxygenase which may be involved in detoxification of xenobiotics. Arylphorin, an abundant hemolymph protein that functions in lipid transport, was also upregulated. These genes would appear to have nothing in common except for general growth-promoting activity.

In *T. viridana* feeding on tolerant oaks, there was somewhat more consistency in the genes upregulated. Eukaryotic initiation factor 4A (eIF4A) is an RNA helicase that mediates binding of mRNA to the ribosome. DEAD box polypeptide 5 is another RNA helicase with unknown function. Most DEAD box proteins function in mRNA or rRNA processing. The U3 small nucleolar RNA associated protein was also upregulated, and U3 functions in ribosomal RNA processing. Thus three of the upregulated genes are involved in RNA interactions. The Niemann-Pick type C2 protein binds and transports cholesterol and other lipids, and the lipase identified is likely involved in intermediary lipid metabolism and not digestion, since homologous ESTs in other lepidoptera come from fat body and not midgut. Thus two of the upregulated genes are involved in lipid metabolism. The oligomeric Golgi complex subunit controls glycosylation of proteins functioning in the Golgi apparatus. Fibroin is a component of silk and its expression in the midgut is mysterious. The astacin is a metalloprotease that functions as a digestive enzyme in crustaceans but not insects, where it may be involved in the control of morphogenesis as in mammals. Overall, it is remarkable that no genes involved in digestion are upregulated by *T. viridana* feeding on tolerant oaks. This suggests that whatever chemical or other types of defenses possessed by the tolerant oaks not only protect them from herbivory in nature, they retard *T. viridana* growth in the laboratory and this specialist is unable to induce responses to overcome them.

In response to feeding on the two types of oaks, *L. dispar* showed different expression patterns as well their different growth responses, depending on their initial size. In the larger larvae, which grew more slowly on the tolerant oaks, changes in only two genes could be documented, neither with a known function. In the smaller larvae, however, three genes with a likely digestive function were upregulated by feeding on tolerant oaks: a trypsin-like protease, a beta-glucosidase, and a carboxypeptidase. This suggests that unlike *T. viridana* and larger *L. dispar* larvae, the smaller *L. dispar* larvae were able to upregulate digestive enzymes in response to whatever defenses possessed by the tolerant oaks, and this transcriptional response contributed to the overall ability to grow even more on tolerant oaks than on the sensitive oaks, which did not induce such a response.

In an additional step during our analysis, we tried to verify gene expression patterns generated through cDNA-AFLPs via q-PCR analysis. Unfortunately genes were either not amplified in q-PCR or expression patterns did not fit to results obtained from cDNA-AFLPs. Accordingly, we were unable to confirm expression patterns of those genes with this second independent technique, as had been accomplished in other studies (Daurelio et al. 2011; Sestili et al. 2011). However, in both species cDNA-AFLP banding patterns showed clearly visible differences between the two samples and patterns were verified through independent repeats. The cDNA-AFLP technique allowed us to find genes involved in the interaction between two different species, a generalist and a specialist and their host tree. This discrepancy observed in the different techniques may be due to the methods employed or to the gene complexity. Reference genes used in both species rather showed a good normalization in our case and didn't indicate a considerable expression themselves for the probes (Polesani et al. 2008). The q-PCR is in general a very sensitive process, where primer design, sample generation and age of the sample (RNA, cDNA) can influence the result immensely. An alternative approach might be the microarray technology to test genes, which are up-or down-regulated, as a supplement to the stable cDNA-AFLP method (Lawrence et al. 2006).

4. Herbivore (*Lymantria dispar*) induced responses in *Quercus robur* and *Picea abies* on living trees: a natural experiment

4.1. Abstract

Lymantria dispar (gypsy moth) (Lepidoptera: Lymantriidae) is an important forest pest species in North America and Europe. *Lymantria* species are polyphagous and have to cope with a wide spectrum of different chemical defenses by different host-trees.

The following study aims to investigate possible defense genes of two species of host tree. Using a canopy crane in a field site in Finland, top feeding areas within the tree were reached to investigate possible defense reactions of *Quercus robur* and *Picea abies*, when *L. dispar* fed on either leaves from living trees or from leaves of cut branches, as used in many lab feeding assays of both tree species. Natural feeding behaviour was closely simulated to test the possible tree defense response to the herbivore. Insect larvae were confined in bags and fed on leaves of "cut" tree branches and "uncut" branches in the living canopy. Leaves which the larvae had fed upon were collected. RNA was extracted from leaves to determine differences in gene expression, using the cDNA-AFLP method. Differently expressed fragments were isolated and sequenced and then compared with sequences in the NCBI database. Genes related to defense activities were identified. The results give an indication of how intact branches may react differently to excised branches in inducing defense genes.

4.2. Introduction

Plant-insect interactions are driving factors in evolutionary biology processes. To understand these processes in more detail, gene expression studies are increasingly employed (Pathan et al. 2007; Reineke & Löbmann 2005; Reymond et al. 2000).

In studying gene expression either in plants or their insect herbivores, experiments are usually done in vitro; either on seedlings (Shields et al. 2003), clones (Schroeder 2010), leaf discs (Shields et al. 2006; Wanner et al. 2000) or eventually in the greenhouse (Travers et al. 2007). These experimental controlled conditions are less intensive compared to field experiments. Field experiments might contain too many variables influencing later statistical analysis (Evertz 2004). At the beginning of plant-insect interaction research, Ehrlich emphasized that experiments in the field can never be replaced entirely by lab studies (Ehrlich & Raven 1964). Studies which are conducted entirely in the field, such as the interaction between North American native tobacco (*Nicotiana attenuata*) and the tobacco hornworm (*Manduca sexta*) are rare and need long-term experience and preparation (Allmann & Baldwin 2010; Baldwin 2001; Kessler & Baldwin 2001; Schultz & Baldwin 1982). Previous studies have proven that combining genomics with field studies can be feasible (Travers et al. 2007). Differences in experimental treatments can be accounted for as shown for the comparison between quaking aspen in the field and in the lab, following herbivory of *L. dispar*. Although nitrogen levels varied between aspen clones in the laboratory and the ones in the field study, the major defensive compound phenolic glycosides were still

comparable (Osier et al. 2000). Most plant-insect interaction studies deal with the genetic bases of direct and indirect defense of plants. Volatiles, terpenoids and tannins are three major plant defenses (Baldwin 2001; Coley et al. 1985; Halitschke et al. 2001; Lawrence et al. 2006; Reymond et al. 2000; Shields et al. 2006). If activation of plant defenses requires the plant to be intact, working with detached plant material could influence the plant defence mechanisms e.g. via jasmonic acid metabolism. Wound induction is the major initiator and injuries not made by the herbivore could mislead the connection (Arnold & Schultz 2002; Lawrence et al. 2006). Subsequently direct or indirect defence mechanisms might not function in the same manner as they do in intact plant material, possibly because primary metabolism would be affected when a plant is detached from its roots.

There is currently no genetic information about the effect of "cut" (detached samples in form from leaves) and "uncut" (attached samples on living trees) samples used for experiments to date. Since the focus of this PhD thesis lies in digestive physiology of lepidopteran herbivores e.g. the gypsy moth (*Lymantria dispar*) (Lepidoptera: Lymantriidae) feeding on different food sources with the aim to find genes which are putatively involved in adaptation to different hostplants, there was the need to test for the "sample effect" that might affect all experiments. The use of detached leaves for feeding in all the previous studies does not allow a test of whether different genes are expressed in "cut" material or in "uncut" material when *L. dispar* larvae feed on both types of leaves. The genetic response of both the plant and the larvae could be different. Hence the objective of the study reported here was to evaluate whether there are different genes expressed in forest trees *Quercus robur* and *Picea abies*, when the polyphagous herbivore *L. dispar* feeds on either leaves from living trees or from leaves of cut leaves, as commonly done in many lab feeding assays. Given that *L. dispar* is a forest pest species with a broad host range in deciduous and conifer trees it is appropriate as the study object in this project.

Through the EU project EVOLTREE, there was the opportunity to use one specific field site in Finland, where feeding experiments could be conducted under natural conditions. The Finnish field site provides a canopy crane to facilitate study of larvae feeding in the canopy of *Q. robur* and *P. abies* trees in a mixed boreal forest.

4.3. Material and Methods

Insects

Lymantria dispar larvae were received from a laboratory rearing (Melody Keena, USDA Forest Service Hamden, CT) as synchronized 2nd instar larvae. Larvae were kept in plastic container with artificial diet, separated in small round plastic containers with artificial diet under controlled conditions in breeding chambers (24°C, day/night cycles of 10 hours day light 14 hours dark) until the experiment started (Keena, unpublished diet protocol for *L. dispar* (Ingredients: de-ionized water, agar, wheat germ, Wesson salt mix with Fe, sucrose, casein, sorbic acid, methyl parabenzene, ascorbic acid, wheat germ oil, raw linseed oil, choline chloride, cholesterol, ferric citrate, vitamin mixture)).

Feeding experiments

Experiments were conducted in July 2008 at the METLA research station (Longitude: 29˚19', range 12' - 56', Latitude: 61˚48', range 61˚34' - 62˚05', Elevation: 80 - 160 m) in Punkaharju, Finland. The terrestrial ecosystem there consists of a mixed boreal, untouched forest with a total area of around 1500 ha. Temperatures were around 24°C at day and 16°C at night. *L. dispar* larvae in their 3rd instar, 48 h post molt, were starved for 24 h before the experiment started. Two different experimental set-ups were conducted in parallel on two different tree species, i.e. oak (*Quercus* robur) and spruce (*Picea abies*), respectively:
i) in the first experimental set-up (hereafter called "uncut") larvae were allowed to feed for 24 h on the top (ca. 30-60 m in height) of spruce and oak tree canopies, respectively. Using a canopy crane of the research facility, top feeding areas within the trees could be reached, which allowed creating a simulated natural feeding environment. Single larvae were bagged in netlike linen material squares (10x10cm) tied up around a branch tip containing 2-3 well expanded either oak leaves or spruce twigs with a cord (figure 1). Experiments were conducted with a total of 20 larvae per tree species (oak or spruce), with 4 larvae per tree and 5 trees as replicates.
ii) In a second experimental set-up (hereafter called "cut") branches (ca. x 20cm in length, containing about 2-3 leaves) were cut from the same trees used for the "uncut" experiment. Plant material was taken from a canopy area close to the one used for the first experimental set-up. Cut branches were put in 300 ml Erlenmeyer flasks with water at the base of the same trees where larvae were feeding on in the top in order to have created almost same experimental conditions (figure 2). Therefore again 4 larvae per tree were taken and 5 trees as replicates, in total 20 larvae per tree species, similar as the experiment in the canopy. This side by side experiment helped to compare molecular responses of uncut or cut trees to herbivore attack.

Figure 1: Experimental set up in the canopy of oak stands in Finland, approximately 30 m above ground. The bagged-in larvae were positioned through the help of a canopy crane. Red arrows show bags with one larva each hanging in the trees, evenly distributed. This is the "uncut"-treatment.

Figure 2: Experimental set-up on the ground next to the trees in the oak stands in Finland. Detached oak branches were put into 300 ml Erlenmeyer flasks with water. Bags with one larva each are fixed on the tree branches. This is the “cut”-treatment.

In both experiments, leaves with obvious signs of larval feeding and feces detection were collected, shock frozen in liquid nitrogen directly on the spot and stored for transportation in dry ice pellets in styrofoam boxes. Long-term storage of leaves was conducted at -80 °C back in the local laboratory.

RNA extraction and cDNA synthesis

RNA of oak and spruce leaves was extracted using TRIzol Reagent (Invitrogen, Karlsruhe, Germany) according to the manufacturers’ protocol including an initial grinding of the leaves with mortar and pestle in liquid nitrogen. Quality and quantity of RNA was checked via gel electrophoresis and Nanodrop spectrophotometer, respectively. Residual DNA was removed by DNase treatment followed by purification using RNeasy Minelute Cleanup Kit (Qiagen, Hilden, Germany). RNAs of two leaves per plant species and treatment were pooled prior to cDNA analysis. These pools of RNAs (1 µg per pool, hereafter called: uncut spruce (S), cut spruce (SC), uncut oak (O) and cut oak (OC) were then translated into cDNA with the SMART PCR cDNA Synthesis Kit (Clontech).

cDNA-AFLP analysis

CDNA-AFLPs were conducted using capillary electrophoresis (Beckmann). The *Eco*RI restriction enzyme didn’t show sufficient banding patterns on the resulting chromatograms for *Q. robur*, so switching to *Pst*I primer was required. For *P. abies Eco*RI enzymes and primers were suitable. For cDNA-AFLP reactions, 500 ng purified cDNA was digested with 5 U *Eco*RI or 5 U *Pst*I and 3 U *Mse*I for 2 h at 37°C followed by enzyme inactivation for 15 min at 65°C. Accordingly, *Eco*RI or *Pst*I and *Mse*I double stranded adapters were ligated to the fragments by incubating for 2 h at 16°C. Pre-amplification of cDNA fragments was performed with a 1:10 dilution of the ligation reaction for 20 cycles with 94°C for 30 s, 56°C for 1 min, 72°C for 1 min, using primers corresponding to *Mse*I and *Eco*RI or *Pst*I adapters, respectively (table 1). Resulting PCR products were diluted 50x and were subjected to selective amplification with a combination of different *Mse*I and *Eco*RI or *Pst*I primers (table 1 and 2) using a touchdown amplification PCR-profile as described by Vos (1995) (Vos et al. 1995).

A total of 64 different AFLP-primer combinations were tested for selective amplification in *P.abies* and *Q.robur*: 16 *Mse*I-NN primers (NN is either AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG or TT) and 4 *Eco*RI or *Pst*I -N primers (N is either A, G, T or C).

To allow a fluorescent labelling of the generated cDNA-AFLP products, a M13(-21) tail was placed at the 5'-end of each Eco or Pst primer and a fluorescently labelled CY5 universal primer M13(-21) was added to the PCR reactions according to the method described by Schuelke (2000).

Table 1: cDNA AFLP standard primer sequences

*Eco*RI adapter	forward:	5'-CTCGTAGACTGCGTACC-3'
	reverse:	5'-AATTGGTACGCAGTCTAC-3'
Pst adapter	forward:	5'-CTCGTAGACTGCGTACATGCA-3'
	reverse:	5'-CATCTGACGCATGT-3'
*Mse*I adapter	forward:	5'-GACGATGAGTCCTGAG-3'
	reverse:	5'-TACTCAGGACTCAT-3'
*Eco*RI pre-amplification		5'-GACTGCGTACCAATTC-3'
Pst pre-amplification		5'-GACTGCGTACATGCAG-3'
*Mse*I pre-amplification		5'-GATGAGTCCTGAGTAA-3'

Fragment analysis and isolation

Amplified cDNA-AFLP fragments were separated by capillary electrophoresis on a Beckman-Coulter GenomeLab GeXP Genetic Analysis System and chromatograms were analysed for differences in gene expression patterns with the GenomeLab software (Beckman). A selection of randomly chosen PCR reactions was repeated up to three times to check for reproducibility of gene expression patterns.

After several repeats and quality checks, a subset of 34 primer combinations was chosen to include into analysis with software (Beckman) for resulting sequencing of potential genes for both species *Q. robur* and *P. abies* (table 2). In *Q. robur* selective amplification results with Pst+Mse variation and for *P.abies* Eco+Mse variation were chosen to analyse (table 2).

Table 2: Subset of primer combinations used for fragment analysis and sequencing potential genes for both species *Q.robur* and *P.abies*.

P. abies	***Q. robur***
Eco A Mse TC	Pst A Mse TC
Eco A Mse GA	Pst A Mse GA
Eco A Mse GC	Pst A Mse GC
Eco C Mse GC	Pst C Mse GC
Eco G Mse GC	Pst G Mse GC

P. abies	*Q. robur*
Eco T Mse GC	Pst T Mse GC
Eco A Mse TA	Pst A Mse TA
Eco C Mse TA	Pst C Mse TA
Eco G Mse TA	Pst G Mse TA
Eco T Mse TA	Pst T Mse TA
Eco A Mse TG	Pst A Mse TG
Eco C Mse TG	Pst C Mse TG
Eco G Mse TG	Pst G Mse TG
Eco T Mse TG	Pst T Mse TG
Eco A Mse TT	Pst A Mse TT
Eco G Mse GG	Pst G Mse GG
Eco C Mse GT	Pst C Mse GT
Eco T Mse GT	Pst T Mse GT
Eco A Mse CA	Pst A Mse CA
Eco G Mse CA	Pst G Mse CA
Eco T Mse CA	Pst T Mse CA
Eco C Mse CC	Pst C Mse CC
Eco A Mse CG	Pst A Mse CG
Eco C Mse CG	Pst C Mse CG
Eco G Mse CG	Pst G Mse CG
Eco T Mse CG	Pst T Mse CG
Eco C Mse CT	Pst C Mse CT
Eco A Mse AA	Pst A Mse AA
Eco A Mse AC	Pst A Mse AC
Eco C Mse AC	Pst C Mse AC
Eco G Mse AC	Pst G Mse AC
Eco T Mse AC	Pst T Mse AC
Eco T Mse AG	Pst T Mse AG
Eco A Mse AT	Pst A Mse AT

Selective amplification PCR reactions containing fragments which were differentially expressed after feeding on cut and uncut treatments, respectively, were loaded onto a 9 % PolyNat Gel (Elchrom Scientific) and were separated on an Origins gel electrophoresis system (Elchrom Scientific). Differentially expressed bands were picked using the BandPick tool (Elchrom Scientific) and desired fragments were directly reamplified from the gel plug with the corresponding selective AFLP primers. Success of amplifications was checked via agarose electrophoresis. Obtained PCR products were purified with ExoSAP-IT (Affymetrix) and were then sequenced with the corresponding primer.

Sequence analysis

Obtained sequences were trimmed and annotated. Then sequences were BLASTed and compared to GenBank entries using the blastx function. To identify candidates sufficiently a threshold of an *E*-value of at least E-05 was used to filter BLAST results.

4.4. Results

From the subset of 34 primer combination chosen for both species, only 30 in spruce and 29 in oak could actually be used for analysis, due to technical losses. Within these primer combinations a total number of 934 fragments for oak and 2665 fragments for spruce were generated (table 3). These fragments were displayed in chromatograms, which were specific for each primer combination and each species (figure 3, figure 4). Differences between cut and uncut treatments, within one primer combination were clearly detectable, even by the naked eye. These graphical landmarks of each chromatogram were analysed for differences in gene expression patterns between the treatments. Within the oak treatment there were 490 fragments for uncut and 444 fragments for cut and within the spruce treatment were 1339 fragments for uncut and 1326 fragments for cut detected (table 3).

In the end there were 33 potential differently expressed genes picked, of which 24 were reamplified respectively with the matching selective primer combination. From the total number of 33 differences within both tree species and treatments, 5 appeared in the OC, 20 in O, 6 in the SC and 2 in the S treatment, respectively. In total 24 sequences were reamplified successfully and were used for sequencing procedure. Finally only 12 sequences from oak could be identified in the NCBI genebank (table 5).

Table 3: Number of fragments in the different tree species and treatments are shown. The method and the number of primer combination are indicated.

tree species/ treatment	***Q. robur* uncut**	***Q. robur* cut**	***P. abies* uncut**	***P. abies* cut**
method	capillary electrophoresis			
primer combinations analysed	29		30	
fragments	490	444	1339	1326
total no. of fragments	934		2665	

Table 4: The table shows which treatments the bands were present and which absent, indicated ("-" for absent and "+" for present). The index number defines specific banding patterns identified in the chromatograms. The number and the type of the certain expression profile are given.

Index	**species**	**cut**	**uncut**	**Relevance**	**no. of expression profile**
1	*Quercus robur*	+	-	oak cut present, oak uncut absent	5
2		-	+	oak uncut present, oak cut absent	51
3		+	+	present in both treatments	878
4	*Picea abies*	+	-	spruce cut present, spruce uncut absent	24
5		-	+	spruce uncut present, spruce cut absent	11
6		+	+	present in both treatments	2630

In Table 4 specific banding patterns are listed. The index number indicates the pattern and helps to differentiate between them. Index number 2 *(*oak uncut present, oak cut absent) appears e.g. 51 times on a chromatogram. It is therefore the most frequently identified pattern between uncut and cut treatment in oak, because oak cut is only present 5 times when uncut is absent in the treatments. Compared to that, the ratio between cut and uncut in spruce is the opposite of oak. Here fragments are present 24 times in the cut spruce and only 11 times in the uncut spruce treatment.

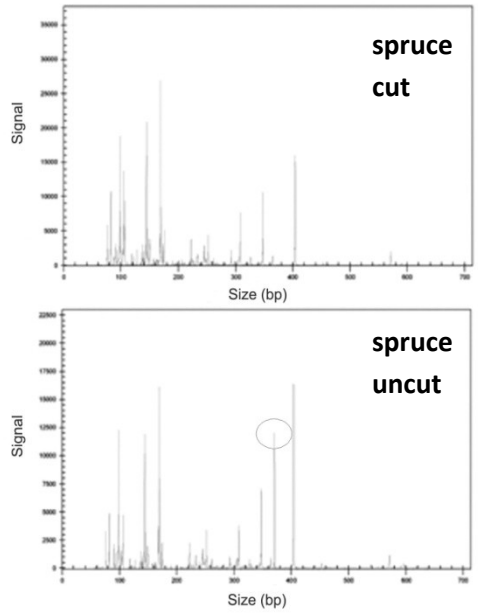


Figure 3: Amplification products visible after capillary electrophoresis obtained with cDNA-AFLPs (Eco-A + MSE-TA) from cut and uncut *P. abies* samples after *L. dispar* herbivory. Differential expressed fragments are indicated by a circle.

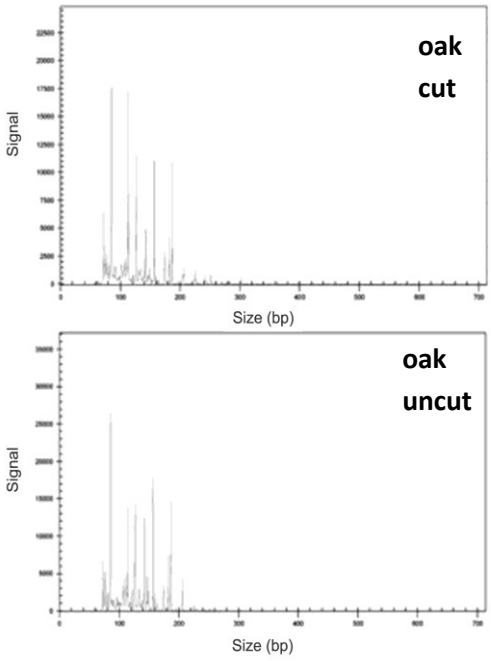


Figure 4: Amplification products visible after capillary electrophoresis obtained with cDNA-AFLP (Eco-A+ MSE-TG) in cut and uncut *Q. robur* samples after *L. dispar* herbivory.

Table 5: Potential candidate genes of *Q. robur* after feeding of *L. dispar* on OC (cut) and O (uncut) leave samples generated by cDNA-AFLP capillary electrophoresis.

Index	**Sequence no.**	**E-value**	**homology**	**Function**
1	1 OC	8 E-26	oxygen-evolving enhancer protein 2 [*Arabidopsis lyrata subsp. lyrata*], XP_002889614.1	chloroplastprotein, Photosystem ll
1	2 OC	2 E-02	ribosome-associated protein p40-like [*Solanum tuberosum*], ABA81872.1	cell functions
2	3 O	6 E-20	subunit of photosystem I [*Cucumis sativus*], BAA11677.1	photosynthesis, photosystem l
2	7 O	1 E-22	Calcium-binding allergen Ole e, putative [*Ricinus communis*], XP_002534401.1	reacts/stimulates antibodies, immune response, pollen
2	8 O	4 E-11	chloroplast chlorophyll a/b-binding protein [*Cucumis sativus*], ABK55666.1	photosynthesis
2	10 O	1E-57	light harvesting chlorophyll a /b binding protein [*Hedera helix*], CAA48410.1	photosynthesis
2	11 O	1 E-37	chlorophyll a/b-binding protein [*Fagus crenata*], BAA24493.1	photosynthesis
2	12 O	3 E-24	chlorophyll a/b-binding protein-C [*Nicotiana plumbaginifolia*], P12469.1	photosynthesis

Index	Sequence no.	E-value	homology	Function
2	13 O	4 E-24	similar to putative ethylene response factor 4 [*Vitis vinifera*], XP_002282253.1	regulator of defense response genes, ethylene signalling pathways
2	14 O	5 E-33	chlorophyll a/b-binding protein [*Fagus crenata*], BAA24493.1	photosynthesis
2	15 O	0.009	ATP synthase epsilon chain, mitochondrial [*Arabidopsis lyrata subsp. lyrata*] , XP_002891649.1	energy deliverer for respiratory chain
2	23 O	8 E-29	carbonic anhydrase [*Populus tremula x Populus tremuloides*], AAC49785.1	respiratory regulator

The majority of genes identified in the remaining OC and O treatments are photosynthesis genes, mainly chloroplast chlorophyll a/b-binding proteins. In total there are 7 out of 12 identified genes directly related to photosynthesis. Most of the genes come from expression profile pattern index 2; just one comes from index 1. This reflects the fact that most genes were sequenced from the index 2 pattern (table 5).

Two genes are known as photosystem proteins, carrying out photosynthesis, one belongs to the photosystem ll in the chloroplast, the other one fits into the photosystem l.

There are two genes identified having respiratory chain functions, an ATP synthase as energy deliverer within the mitochondrial membrane and carbonic anhydrase precursor, which is a zinc-containing enzyme that is involved in catalyzation of carbon dioxide. Both enzymes are also involved in processes like photosynthesis, respiration, pH homeostasis and ion transport (NCBI group 2012). Furthermore, there is ribosome-associated protein p40-like gene with general cell functions and there are two genes known for defensive functions, a immune stimulating putative Calcium-binding allergen Ole e and a defense response gene regulator: similar to putative ethylene response factor 4.

4.5. Discussion

This project aimed to determine whether there are expression differences between samples from cut and from uncut trees, when feeding experiments on *Q. robur* a preferred host and *P. abies* a suboptimal host tree of *L. dispar* are conducted. For this, side by side experiments

in the tree canopy and on the ground of a Finnish forest were carried out. Through gene expression analysis via cDNA-AFLP, differences in both tree species and treatments were recognized. Subsequently, reamplified PCR products were sequenced and identified through homologies of already known NCBI databank entries.

Chromatograms showed differences between cut and uncut leaf samples in both tree species after feeding of *L. dispar*. Within differently expressed fragments in spruce there were more patterns of index 4 (present in cut, absent in uncut) than index 5 (present in cut, absent in uncut). In oak surprisingly the pattern 2 (present in uncut, absent in cut) appeared more often than the pattern 1. The reversed ratio between both species for cut and uncut treatment was not expected.

It was observed that within uncut oak, more genes show increased expression after *L. dispar* feeding than in the cut treatment. This finding supports the hypothesis that intact plant material, here the living tree, reacts more strongly to herbivory than the cut leaves. For the leaves on the intact tree a defense induction might be more accomplished, than for the already cut leaves. Plant-plant interactions or internal plant signals to undamaged tree parts seem to play an important role in defense strategies (Yi et al. 2009). One type of warning signal is the emission of volatile organic compounds (VOCs), an indirect defense mechanism (Ballhorn et al. 2008). VOCs induced by herbivory have been found to attract parasites and predators of the herbivore, thus the benefits to the plant are indirect, as opposed to any direct negative effect of VOCs on herbivores. Due to jasmonic acid (JA) release, the plant responds with VOCs. In one study the total volatile emission increased after artificial JA application, and different allocations of defense strategies to primary or secondary leaves of lima beans were detected. Primary leaves obtained mainly indirect defense allocation in the form of VOCs and secondary leaves responded rather by direct defense through cyanogenesis, toxic to herbivores. These differences in primary and secondary leaves were interpreted as being due to adaptations to changing enemy pressures. Furthermore in these experiments, a lower resistance level was found in damaged than in undamaged plants (Ballhorn et al. 2008). This implies that defensive traits are not as strongly expressed in previously damaged samples. The number of genes with increased expression directly due to herbivory was not measured in this experiment. A comparison to control leaves not exposed to larval feeding in either the cut and uncut treatments would have provided additional information on genes induced by actual insect feeding damage.

Most of the genes discovered in oak play a role in the photosynthetic process of plants (table 5). There are differences in genes found between cut and uncut treatments. The genes in the uncut treatment are general photosynthesis genes, belonging to life-sustaining functions of the plant. Despite the expectation that more primary metabolism genes are switched on in cut samples, because of the additional stress the tree suffers from, these genes were more frequent in the uncut treatment. Interestingly the gene ribosome-associated protein p40-like, which might be involved in defence mechanism against insect herbivores (Giri et al. 2006; Zavala et al. 2008) was found in the cut treatment, whereas photosynthetic genes

were expressed in the uncut treatment. This finding supports the possibility of increased photosynthetic rates due to a compensation process of the plant. It has been shown that many small injuries damage the plant in the same manner than a big defoliation event would do; photosynthesis increases and reallocation of nutrients occurs (Schoonhoven et al. 2005).

Within the uncut samples of oak, two respiratory system genes were found. They might be involved in an initiating protection processes using additional energy for potential defense mechanism of the tree. As reported in other research, genes related to jasmonic acid, a key molecule initiating plant defense responses and proteinase inhibitors, are activated by insect herbivores feeding on plants (Howe & Jander 2010; Ivashov et al. 2001; Kellogg et al. 2005; Lawrence et al. 2006; Reymond et al. 2000; Tooker & De Moraes 2005). In this study, an ethylene response factor was identified from expression differences in the uncut oak treatment. Ethylene is a plant hormone, responsible for a wide variety of plant functions sometimes regulating defense response genes, belonging to the ethylene signalling pathway. This pathway is largely separate from the jasmonic acid pathway but possibilities for cross-talk between the two exist (Lorenzo et al. 2003). An increase in ethylene signalling pathway operations could thus be a consequence of herbivore attack (Schoonhoven et al. 2005). Upregulation of the ethylene response factor in the uncut treatment would also be consistent with the greater importance of the volatile ethylene for signalling within the intact plant.

Although more cDNA-AFLP fragments were surveyed for spruce, fewer spruce genes were found to be differentially expressed in the treatments than for oak, and more were found in the cut treatment. Since no sequence information was obtained for the spruce fragments, the two treatments cannot be compared on a functional basis. Perhaps the variation between the species might be a consequence of *L. dispar*'s preferred food source. It could be that oak trees have already adapted to one of its major pest species *L. dispar* rather than spruce trees. Even if *L. dispar* is polyphagous, spruce is rather a suboptimal host to *L. dispar*, where oak is mostly preferred (Foss & Rieske 2003). Spruce samples reacted with gene expression differences slightly more strongly when branches were detached from the tree and not when feeding on the living tree was conducted. Spruce as a conifer is an evergreen tree and foliage is growing back all year long. A loss of foliage might not as alarming as cutting a whole branch. The effect of the single larvae on the needle tip wasn't as drastic as on oak.

Outlook/Trend

Overall, whether using cut or uncut plant material matters for analysing plant responsive expression profiles of defensive or detoxifying genes in response to insects feeding on trees, cannot be conclusively answered by this pilot study. However, an actual direct defensive strategy has not been indicated by any the genes found to differ between these two treatments. More warning and preparing processes showed a transcriptional response when feeding was conducted on the uncut and cut leaves of the oak species. This indicates that

any genes in this species with a more obvious connection to herbivore defense, that are discovered in experiments using cut plant material with or without herbivory, are probably not induced as an artefact of cutting. Investigations into long term experiments and analysis of the response of the herbivores themselves on cut and uncut trees will bring more insights into the insect-plant interaction.

5. Identification of SNPs in transcripts of different *Lymantria dispar* populations via 454 pyrosequencing

5.1. Abstract

The subject of this study is the genetic diversity at the nucleotide level in the gypsy moth *Lymantria dispar* through a high-throughput sequencing approach.

By using a high throughput pyrosequencing technology, sequences were generated to compare levels of diversity between *L. dispar* populations, which are geographically distant. In *L. dispar* a set of loci was obtained by sequencing a subset of the genome contained in an AFLP fragment, in order to identify SNPs for characterising levels of between population polymorphisms, as well as the effects of migration, genetic drift, and selection. 10 *L. dispar* populations, collected at sites across Europe and Asia were included in this approach, each population sample containing at least DNAs from 10 individuals.

Bioinformatic analyses of sequences selected on the basis of representation in a previously constructed cDNA library produced estimates of population differentiation, Fst values, which were very high. Single nucleotide polymorphisms (SNPs) were common among these sequences and the function of many of these genes could be identified by similarity to genetic databases. However, most SNPs occurred not in the coding regions but in introns or noncoding regions of these genes. Another analysis approach used single-copy conserved coding sequences identified by comparison to the genome sequence of the domesticated silkmoth, and produced less data for comparison but lower Fst values where comparisons were possible. The difference between the two approaches may be due to occurrence of the restriction site used to generate the AFLP fragment within a highly repeated DNA sequence such as a transposable element, leading to Fst values that were not based on single-copy genes. For further research, particularly with next-generation sequencing technologies, different restriction enzymes should be tested to increase the data necessary for high-precision estimation of population genetic parameters.

5.2. Introduction

Within the past five years, novel technological approaches for sequencing DNA have become commercially available and widely used. The two leading technologies are Illumina sequencing and 454 pyrosequencing. 454 pyrosequencing has been used in many different fields, such as human genome projects, such as Alzheimer, HIV cure research and for food protection. Nonhuman applications are mainly studied model organism in the plant and animal kingdoms (Behura 2006; Patin et al. 2009; Rounsley et al. 2009; Toth et al. 2007). This novel method, which is suitable to study evolutionary ecology, has advantages and disadvantages compared to the Sanger sequencing system (Bekel et al. 2009). The quality of Sanger sequencing is higher, but pyrosequencing produces much higher coverage that can compensate for its higher error rate (Gompert et al. 2010). The cost per basepair is far less, due to the high sample throughput in the patented picotiter plates from Roche, producing

millions of independent sequence reads. The main difficulty is the bioinformatic processing of such a large amount of data.

The pattern of distribution of genetic variation across populations can give information about how their evolutionary history has been affected by their geographical relationships. The genetic diversity in populations is influenced by selection, mutation, immigration rates and population sizes (Patterson et al. 2006). Up till now most studies using next generation sequencing technologies have put their emphasis on transcriptome characterization, distinguishing species, strain identification and sample characterization (Rothberg & Leamon 2008; Vera et al. 2008). Not much has been done using these new techniques in analysing whole population relationships, such as relatedness status within and between populations (Gompert et al. 2010; Santure et al. 2010). Additionally next generation sequencing has been useful in identifying candidate genes in non-model organisms with little or no prior sequence information (Pauchet et al. 2009; Pauchet et al. 2010).

As the interest in molecular information for individuals and populations increased, the use of molecular markers has also changed over time. Firstly fingerprinting methods, such as RFLPs, AFLPs and RAPDs gave an overall insight into the genome structure of organisms, detecting rare polymorphisms and analysing paternity (Reineke et al. 1999; Vos et al. 1995). Comparing the degree of relationship between two individuals and between populations was mainly studied by using the microsatellites, a codominant marker system. However, microsatellites have disadvantages; they are more difficult to develop initially, more complicated to analyse, and furthermore they can be unreliable when the occur within repetitive DNA (Smouse 2010). In comparison, single nucleotide polymorphisms (SNPs) are simpler to automatically analyse in high throughput projects, because of a quite simple yes/no scoring, leading to reduced error rates (Vignal et al. 2002). SNPs can also provide more information because many occur within coding regions and may thus influence the phenotypic appearance and adaptation of individuals to their environment (Williams et al. 2010). In combination with the high throughput sequencing technologies, SNPs can reveal patterns of genetic variation associated with natural phenomena in the insect world (e.g. extinction of species and pest outbreaks) (Rothberg & Leamon 2008; Santure et al. 2010).

The central aim of this study is to characterize patterns of genetic variation in candidate genes that may affect environmental adaptation, in widespread populations of the non-model insect *Lymantria dispar*. This species is a polyphagous forest pest which has been studied in regards to plant-animal interactions, especially in respect to host plant choice and detoxification mechanisms (Keena 2003; Lazarevic et al. 2003; Lazarevic et al. 2002; Lazarevic et al. 2004; Lobinger & Skatulla 2001; Withers & Keena 2001). SNPs are an established tool to study populations, which are geographically separated from each other (Schroeder & Scholz 2005). SNPs can vary between insect taxa (Kruse & Sperling 2001). Through high-throughput sequencing of genomic DNA from several populations, we

investigated SNP variation in specific candidate genes, which were represented in a previously-constructed *L. dispar* cDNA library and identified by comparison to the NCBI databank. An alternative set of more evolutionarily conserved genes was identified based on comparisons to the genome sequence of another lepidopteran, the domesticated silkmoth. We examined the following questions: How much genetic variation is there between populations of *L. dispar* in different geographic regions in Europe and Asia? And in which candidate genes for local population adaptation does this variation occur?

5.3. Material and Methods

Insects

Male *Lymantria dispar* adults were collected via pheromone traps in different European countries (table 1). One set of samples was collected during 1994 and 1996 and has been used for population genetic analysis before (Reineke et al. 1999; Reineke & Zebitz 1999). The other set of samples was collected within the framework of the EU Network of Excellence EVOLTREE in the year 2008. Dead adults were removed carefully from the sticky inserts of the traps and stored frozen at -20 °C until further processing.

Table 1: *L. dispar* populations, their geographic locations and multiplex identifier (MID) used for 454 pyrosequencing

MID	Country	Latitude	Longitude	Year collected	Sample size
6	Shenyang, China	41°50 N	123°26 E	1994	13
9	Kaplná, Slovakia	48°17 N	17°27 E	1996	11
7	Florence, Italy	43°47 N	11°15 E	1996	13
5	Puszcza Swietokrzyska, Poland	51°02 N	20°42 E	2008	12
1	Mt. Ventoux, France	44°10 N	5°17 E	2008	10
4	Valais, Switzerland	46°29 N	7°63 E	2008	13
8	Leipzig, Germany	51°20 N	12°22 E	1996	13

MID	Country	Latitude	Longitude	Year collected	Sample size
10	Jena, Germany	50°55 N	11°35 E	2008	17
12	Neuenstadt, Germany	49°15 N	9°20 E	1994	12
11	Grossbottwar, Germany	49°00 N	9°19 E	1994	14
3	Grossbottwar, Germany	48°47 N	9°12 E	2008	8

DNA extraction and sample preparation

DNA was extracted from approximately 10 mg of male adult body tissue according to a CTAB protocol (Reineke 1998). Quality and quantity of DNA was checked via gel electrophoresis and in a NanoDrop spectrophotometer, respectively. DNA samples prepared between 1994 and 1996 were stored as DNA pellets in 70 % ethanol and were precipitated with isopropanol prior to use in SNP analysis. Residual RNA was removed from all DNA samples by treatment with RNase A followed by a further purification step using the RNeasy Minelute Cleanup Kit (Qiagen, Hilden, Germany). Individual DNA samples were diluted to a concentration of 10 ng/µl and samples of 8 to 17 individuals per population were pooled in equal amounts (table 1), resulting in 11 DNA pools each representing one *L. dispar* population which were used in subsequent reactions.

Preparation of DNA samples for pyrosequencing

Templates for 454 pyrosequencing were prepared on the basis of an AFLP-like protocol (Gompert et al. 2010). Initially, 400 ng DNA was digested with 5 U *Eco*RI and 5 U *Mse*I for 18 h at 37°C. Accordingly, *Eco*RI and *Mse*I double strand adapters were ligated to the fragments by incubating for 2 h at 16°C, followed by enzyme inactivation for 10 min at 65°C. Pre-amplification of DNA fragments was performed with a 1:10 dilution of the ligation reaction in 0.1 x TE buffer for 30 s at 98°C, followed by 30 cycles of 30 s at 98°C, 30 s at 56°C and 120 s at 72°C, and a final extension for 10 min at 72°C using primers corresponding to *Mse*I and *Eco*RI adapters (table 2). For all amplifications a high-fidelity polymerase (Phusion DNA polymerase, Biozyme) was used to minimize PCR error rates. Preamplification PCR products were separated on a 2% agarose gel, and a range of fragments between 400 and 550 bp was cut out of the gel and purified using GeneClean Purification Kit (MP Biomedicals, LLC). In addition, a second preamplification PCR under identical conditions was conducted and products were again loaded on a gel. This time, fragments between 200 and 550 bp were purified from the gel to increase the number of fragments suitable for 454 pyrosequencing.

Both purified PCR products were combined and these amplicons were used as templates in a third final PCR under the conditions described above. In this PCR, amplicons from each pool (population) were individually labelled with a distinct 10 bp barcode-sequence added at the 5'-end of each primer (multiplex identifier, MID), allowing a later identification of sequences amplified in individual populations. Sequences of MIDs were according to (Gompert et al. 2010). The final PCR products were purified with GeneClean Purification Kit (MP Biomedicals, LLC) and concentrations were measured in a NanoDrop spectrophotometer. Each sample was diluted to a concentration of 100 ng/µl and amplicons from all 11 populations were subsequently pooled in equal amounts to obtain a final pool with a concentration of 100 ng/µl, which was used as a template for pyrosequencing. Next generation 454 pyrosequencing was performed commercially (LGC Genomics GmbH, Berlin, Germany) on a Roche GS-FLX DNA Sequencer in ¼ PicoTiterPlate as described (Margulies et al. 2005).

Table 2: Sequences of primers and adapters used for preparation of templates for pyrosequencing

Adapter/Primer	**fwd/rev**	**Sequence**
*Eco*RI adapter	fwd	5'-CTCGTAGACTGCGTACC-3'
	rev	5'-AATTGGTACGCAGTCTAC-3'
*Mse*I adapter	fwd	5'-GACGATGAGTCCTGAG-3'
	rev	5'-TACTCAGGACTCAT-3'
*Eco*RI pre-amplification		5'-GACTGCGTACCAATTC-3'
*Mse*I pre-amplification		5'-GATGAGTCCTGAGTAA-3'

Assembly of sequences, SNP detection and population genetic analysis

Data processing and initial bioinformatic analysis was conducted at LGC Genomics GmbH. Initial quality filtering of *L. dispar* genomic 454 sequences was performed using Roche proprietary analysis software. Sequences were sorted according to their MIDs, followed by trimming sequences of their barcodes and primer sequences. In addition, sequences containing no primer sequence or only one primer at the 3'-end were removed, as these sequences are known to contain high error rates. Accordingly, sequences of individual MIDs (populations) were BLASTed against sequences from 6236 transcripts from a *L. dispar* midgut cDNA library. To identify transcripts sufficiently covered by the pyrosequencing reads a threshold of an *E*-value of at least E-05 was used to filter BLAST results. Average coverage of the cDNA library sequence by the reads was also determined. Transcript sequences from the *L. dispar* cDNA library and reads obtained through 454 pyrosequencing were aligned using BLAT. Only sequences of 90% and greater similarity and of length of 40 bases or more

were included in further analysis. Alignments were parsed for SNPs using BioPerl. A locus was identified as containing a putative SNP if it had a frequency of at least 33.33% and a coverage of at least 30x. If a given SNP occurred in 33.33 to 66.66% of the sequence reads obtained from one population, the respective population was classified as being heterozygous for this SNP.

Sequences of transcripts containing SNPs were compared with those available in GenBank database using the BLASTX tool, with the significance cut off for amino acid homology set at an *E*-value of at least E-03. Functional categories for genes of interest were further identified via searches on the UniProt database.

Basic population genetic parameters were calculated using the programs SNPator (Morcillo-Suarez et al. 2008) and Arlequin ver.3.11 (Excoffier et al. 2005). In particular, pairwise F_{ST} coefficients for all pairs of the 11 *L. dispar* populations were calculated to measure levels of genetic differentiation between the different populations and SNP allele frequencies within each population were estimated. Genetic distance between pairs of populations was estimated by by F_{ST} /(1 - F_{ST}). Isolation by distance was calculated using the Isolation By Distance Web Service, v.3.16 (Jensen et al. 2005) to compare genetic distances with geographic distances.

Because F_{ST} values were very high, a second analysis approach was applied. The second analysis approach was done partly manually, and selected only conserved coding regions for analysis by aligning all sequences assorted for certain quality criteria to the *Bombyx mori* predicted protein sequences from the full genome sequencing project (Xia et al. 2004). We focused on SNP-containing sequences for analysis. Then pairwise F_{ST} values were calculated separately for each gene in Dnasp5 and then averaged across genes.

5.4. Results

In total, 304,000 sequence reads were obtained in one run of 454 pyrosequencing, of which 9842 reads were discarded as they contained either no primer or only the reverse primer. The resulting 294,482 sequences had an average length of 157 base pairs. The number of reads obtained per population varied to quite an extent, with a maximum of 68,340 reads obtained for MID06 (Shenyang, China) and a minimum of 4622 reads for MID09 (Kaplná, Slovakia).

These reads were matched to the sequence of 6236 transcripts from a *L. dispar* midgut cDNA library. Around 2.4% of the cDNA library was covered by genomic pyrosequencing reads with good hits (BLASTN *E*-value <1E-05). Coverage was lowest for MID11 (1.9%) and highest for MID4 (3.3%). As SNPs were analysed only for those transcripts, which were matched by high quality reads obtained from all 11 populations, in summary, 97 *L. dispar* transcripts were included in subsequent SNP analysis. Under these criteria, we identified 791 loci matching to these transcripts containing a SNP of which on average 761 loci per population were usable for further analysis (table 3). In the majority of transcripts, on average 3-5 SNPs per transcript were detected, however, in 7 transcripts more than 20 SNP loci were present. In

these SNPs, the proportions of transition substitutions were A⁄G, 29.6%, and C⁄T, 34.2%, compared to transversions A⁄C, 7.9%; G⁄T, 6.9%; A⁄T, 16.1% and C⁄G, 5.4%. This corresponds to a transition:transversion ratio of 1.76:1. There was a substantial variation in the number of polymorphic SNPs detected among populations, with 163 polymorphic loci identified in MID4, while only 3 polymorphic SNPs were present in MID9 (table 3). Number of polymorphic loci was highly correlated with the number of assembled reads for each population (r=0.8507, P=0.000352).

Table 3: Sample data obtained for 454 pyrosequencing in 11 *L. dispar* populations

MID	Population	Reads used	Average sequence length after clipping	No. of usuable loci	No. of heterozygous loci
1	Mt. Ventoux, F	13302	156 bp	762	12
3	Grossbottwar08, D	33354	143 bp	766	64
4	Valais, CH	55662	151 bp	757	163
5	Puszcza Swietokrzyska, PL	48683	168 bp	760	54
6	Shenyang, CN	68340	165 bp	755	95
7	Florence, I	17098	160 bp	763	28
8	Leipzig, D	22462	134 bp	767	38
9	Kaplná, SK	4622	152 bp	766	3
10	Jena, D	9496	144 bp	762	10
11	Grossbottwar94, D	14029	184 bp	757	20
12	Neuenstadt, D	7434	173 bp	763	11
	Mean		157 bp	761	45

Pairwise F_{ST} values for all populations are given in table 4. In general, F_{ST} values suggest a very high genetic differentiation between the analysed *L. dispar* populations. In this data set, the lowest, yet still very pronounced level of genetic differentiation was evident between two populations collected both in 2008 (Switzerland, Valais and Germany, Grossbottwar) as well as two populations from Germany, one collected in 1994 (Neuenstadt) and the other in 2008 (Jena). A very pronounced level of genetic differentiation was also obvious between two populations, collected at the same geographic location but with a period of 14 years in between collection dates (Grossbottwar, Germany, MID03 and MID11, F_{ST} value of 0.80). There was no correlation between genetic distance and geographic distance detected in the data set (r= -0.012, P=0.406).

Table 4: Pairwise F_{ST} values (below the diagonal) between 11 *L. dispar* populations from a worldwide collection based on the analysis of 791 SNPs identified in cDNA transcripts and their corresponding distances (km) between populations (above the diagonal). Distance between the Chinese and European populations was estimated.

MID	1	3	4	5	6	7	8	9	10	11	12
1		610	291	1614	7500	477	937	982	883	610	614
3	0.7654		338	1060	7500	590	360	600	270	5	14
4	0.6245	0.3848		929	7500	455	540	506	630	338	290
5	0.6765	0.7457	0.6291		7500	1395	780	688	819	1060	1070
6	0.7762	0.6858	0.6245	0.7529		7500	7500	7500	7500	7500	7500
7	0.7826	0.6667	0.5901	0.7530	0.7273		830	714	800	590	587
8	0.7664	0.4769	0.5390	0.7416	0.7194	0.5571		480	70	360	348
9	0.8000	0.7481	0.6438	0.7533	0.7650	0.7232	0.7389		515	600	612
10	0.7885	0.7208	0.6279	0.7500	0.7482	0.6042	0.6619	0.5667		270	285
11	0.8656	0.8033	0.7077	0.8078	0.7081	0.8255	0.8204	0.8671	0.8477		14
12	0.7810	0.7148	0.6266	0.7461	0.7500	0.6214	0.6573	0.5484	0.3438	0.8434	

In the second analysis with Dnasp5, only a few MIDs had more than 2 genes for a reliable pairwise comparison (e.g. MID5 and MID1) (table 5). These pairwise comparisons show smaller F_{ST} values than the first analysis. Instead a F_{ST} value of 0.6765 between MID5 and MID1 there was a F_{ST} value of 0.2103, which indicates already a major difference in both calculations approaches. The overall average pairwise F_{ST} is 0.4441. The overall average pairwise an F_{ST} based on the LGC calculations is 0.7002. Nevertheless, F_{ST} s of most of the calculations showed similar high F_{ST} values, since not enough genes were available for a sufficient analysis. For example on the pairwise comparison between MID7 and MID5 were quite similar; the F_{ST} was 0.7530 in the first and 0.6667 in the second approach (table 4, 5).

Table 5: Pairwise F_{ST} values for the populations indexed by MID number. Below the diagonal are the average pairwise F_{ST} values for populations with 2 or more genes for comparison. Above the diagonal is the number of genes that could be compared for each pair of populations.

Mid	1	3	4	5	6	7	8	9	10	11	12
1		0	2	10	5	1	2	2	0	1	0
3	--		0	0	0	0	0	0	0	0	0
4	0.9167	--		11	1	3	6	6	1	2	1
5	0.2103	--	0.1854		15	6	8	9	2	4	0
6	0.1900	--	--	0.0653		1	2	4	0	2	0
7	--	--	0.2192	0.5475	--		2	3	0	1	2
8	0.5538	--	0.2266	0.1694	0.8230	0.3495		6	1	3	1
9	0.9167	--	0.0000	0.3645	0.9151	0.6667	0.2094		1	4	0
10	--	--	--	0.3750	--	--	--	--		0	0
11	--	--	0.4706	0.3612	0.8750	--	0.4189	0.5000	--		0
12	--	--	--	--	--	0.5746	--	--	--	--	

The primary interest of this research was to identify putative candidate genes which differ between worldwide *L. dispar* populations and might be involved in adaptation to certain abiotic conditions or to respective host plants present in the given habitat. Thus, from the 791 SNPs identified in the first study approach, a sub-set of 28 transcripts was chosen where the respective SNP occurred in 2 to 5 of the 11 *L. dispar* populations. Sequence of respective SNP containing transcripts from the *L. dispar* cDNA library was compared to GenBank entries using the BLASTX tool and in those transcripts with significant homologies (E values lower than 1E-04) locations of SNPs were determined. For the majority of these genes, SNPs were not located in the putative open reading frame, but in noncoding regions (exon) of the transcript (table 6). However, for 6 of the 28 transcripts a couple of SNPs was located in the coding region, causing in most of the cases a different codon and thus a change of the amino acid sequence (nonsynonymous mutation). These genes are likely to code for a transposon, different members of the reverse transcriptase family, and a coiled-coil domain containing protein, respectively.

Table 6: Sequence homologies and GenBank accession numbers of selected *L. dispar* transcripts from a midgut cDNA library containing one or more SNPs in coding (c) or non coding (nc) regions of the transcript with an indication of their putative function

SNP	**Transcript**	**SNP in x pops**	**region**	**Sequence homology**	***E*-value**	**Putative Function**
0044	Ld0745	5	nc	Heat shock protein 25.4 (*Bombyx mori*, ACA25336.1)	2E-07	Stress response
0048 - 0055	Ld0971	2	nc	Ethanolaminephosphotransferase 1 (*Harpegnathos saltator*, EFN89734.1)	2E-89	Phospho-transferse activity
0056	Ld0979	2	nc	Multiple inositol polyphosphate phosphatase (*Tribolium castaneum*, XP_972932.1)	2E-33	Acid phosphatase activity
0520	Ld3638	3	nc	Antennal esterase CXE11 (*Spodoptera littoralis*, ACV60238.1)	4E-45	Hydrolase
0589	Ld4084	5	nc	35kDa protease (*Bombyx mori*, AAK06410.1)	3E-04	Hydrolase
0743 - 0751	Ld6039	2	nc	N-acetylneuraminate lyase (*Harpegnathos saltator*, EFN85455.1)	1E-39	Lyase
0373 - 0374	Ld2949	3	nc	Glutathione S-transferase sigma 2 (*Bombyx mori*, NP_001036994.1)	1E-71	Transferase activity
0383 - 0387	Ld3003	2	nc	3-dehydroecdysone 3alpha-reductase (*Spodoptera littoralis*, AAF70499.1)	4E-54	Oxidation-reduction process
0657 - 0658	Ld4835	3	nc	NADH-ubiquinone reductase (*Bombyx mori*, ABF51426.1)	9E-23	Oxidation-reduction process

SNP	Transcript	SNP in x pops	region	Sequence homology	*E*-value	Putative Function
0678 - 0679	Ld5093	5	nc	Peroxinectin (*Ixodes scapularis*, EEC14558.1)	9E-23	Oxidation-reduction process
0733 - 0742	LD6034	3	nc	hypothetical protein TcasGA2_TC014056 (*Tribolium castaneum*, XP_974328.1)	1E-07	Oxidation-reduction process
0089 - 0096	Ld1315	2	nc	Calmodulin (*Tribolium castaneum*, XP_969708.1)	7E-66	Calcium ion binding
0105 - 0107	Ld1348	5	nc	Chitin deacetylase 5b (*Helicoverpa armigera*, ADB43612.1)	4E-08	Carbohydrate metabolic process
0065 - 0068	Ld1081	3	nc	Lebocin-4 precursor (*Bombyx mori*, NP_001119731.1)	2E-11	Immune response
0249 - 0252	Ld2408	4	nc	Megourin-2 (*Megoura viciae*, P83418.2)	6E-06	Immune response
0630	Ld4594	2	nc	Proline-rich protein (*Galleria mellonella*, ACQ99193.1)	1E-24	Antimicrobial activity
0431 - 0435	Ld3283	4	c	Planarian mariner-9 gene (*Girardia tigrina*, CAA56856.1)	6E-47	Transposon
0205 - 0207	Ld2118	5	c	Reverse transcriptase (*Ostrinia nubilalis*, ABO45237.1)	3E-09	RNA-dependent DNA replication
0224	Ld2346	2	nc	Reverse transcriptase (*Ostrinia nubilalis*, ABO45231.1)	1E-07	RNA-dependent DNA replication
0787 - 0791	Ld6186	2	c	Reverse transcriptase (*Ostrinia nubilalis*, ABO45239.1)	6E-38	RNA-dependent DNA replication
0420 - 0423	Ld3207	5	nc	Endonuclease-reverse transcriptase (*Bombyx mori*, ADI61811.1)	2E-30	RNA-dependent DNA replication

SNP	Transcript	SNP in x pops	region	Sequence homology	*E*-value	Putative Function
0510	LD3600	3	c	Endonuclease-reverse transcriptase (*Bombyx mori*, ADI61810)	6E-04	RNA-dependent DNA replication
0561 - 0566	Ld3870	2	c	Reverse transcriptase/integrase (*Lymantria dispar*, AAO72078.1)	0.08E-03	Nucleotidyltransferase, RNA-directed DNA polymerase
0168 - 0170	Ld1790	2	nc	Hypothetical protein AaeL_AAEL011551 (*Aedes aegypti*, XP_001655423.1)	3E-15	unknown
0315 , 0320	Ld2660	2	c	Coiled-coil domain-containing protein 130 homolog (*Drosophila melanogaster*, NP_611383.1)	9E-40	unknown
0404 - 0406	Ld3119	5	nc	Putative uncharacterized protein (*Tribolium castaneum*, EFA11633.1)	1E-08	unknown
0427 - 0428	Ld3215	2	nc	Similar to CG12918-PA (*Apis mellifera*, XP_394366.1)	4E-29	unknown
0793 - 0794	Ld6214	2	nc	Hypothetical protein (*Nasonia vitripennis*, XP_001603290.1)	1E-42	unknown

Table 7: Distribution of SNPs identified via 454 pyrosequencing in 28 transcripts in 11 *L. dispar* populations. Presence of polymorphisms compared to a reference sequence from a cDNA library is indicated. For locations and details on MIDs see table 1.

SNP	Homology	MID 01	MID 03	MID 04	MID 05	MID 06	MID 07	MID 08	MID 09	MID 10	MID 11	MID 12
0044	Heat shock protein 25.4	x	x	X	x			x				

SNP	Homology	MID 01	MID 03	MID 04	MID 05	MID 06	MID 07	MID 08	MID 09	MID 10	MID 11	MID 12
0048 - 0055	Phospho-transferase 1					x					X	
0056	Inositol phosphatase					x					x	
0520	Antennal esterase		x	x		x						
0589	35kDa protease	x		x	x	x		x				
0743 - 0751	Lyase					x					x	
0373 - 0374	Glutathione S-transferase		x	x		x						
0383 - 0387	Dehydro-ecdysone reductase		x	x								
0657 - 0658	NADH-ubiquinone reductase		x	x				x				
0678 - 0679	Peroxinectin		x	x		x	x	x		x		
0733 - 0736	hypothetical protein					x					x	
0740 - 0742	hypothetical protein	x		x	x							
0089 - 0096	Calmodulin					x		x				
0105 - 0107	Chitin deacetylase		x	x	x	x		x				
0065 - 0068	Lebocin-4 precursor		x	x				x				

SNP	Homology	MID 01	MID 03	MID 04	MID 05	MID 06	MID 07	MID 08	MID 09	MID 10	MID 11	MID 12
0249 - 0252	Megourin-2		x	x		x		x				
0630	Proline-rich protein		x					x				
0431 - 0435	Planarian mariner-gene		x	x			x	x				
0205 - 0207	Reverse transcriptase		x	x			x	x				x
0224	Reverse transcriptase					x					x	
0787 - 0791	Reverse transcriptase		x					x				
0420 - 0423	Reverse transcriptase		x	x		x		x				
0510	Reverse transcriptase		x	x			x					
0561 - 0566	Reverse transcriptase					x					x	
0168 - 0170	Hypothetical protein			x			x					
0315 , 0320	Coiled-coil cont. protein				x	x						
0404 - 0406	Uncharacter-ized protein		x	x		x	x	x				
0427 - 0428	Similar to CG12918-PA			x		x						
0793 - 0794	Hypothetical protein					x					x	

5.5. Discussion

Here, we report on the utility of using 454 pyrosequencing to identify SNPs in transcripts of a world-wide collection of 11 *L. dispar* populations. Using this system, we obtained 294,482 sequence reads, with up to 68,000 reads for a single *L. dispar* population. Around 2.4% of a *L. dispar* midgut cDNA library was covered with reasonable significance by these reads, resulting in 97 *L. dispar* transcripts which were included in the present SNP analysis. In these transcripts, 791 SNP loci were identified in coding and non-coding regions. As discussed in other publications, minimum of 500 SNPs should be sufficient for good statistical statement (Santure et al. 2010).

The number of SNPs detected varied between populations and was clearly positively correlated with the number of sequence reads obtained for a given populations. This polymorphism has been already shown in previous studies (Williams et al. 2010). Thus, results obtained in particular for MID09 (Kaplná, Slovakia) have to be taken with care, as a particular low number of reads was obtained for this population and none of the SNPs identified in the 28 transcripts was detected in this population as well. Currently, we cannot explain this uneven distribution of sequence reads between the 11 analysed *L. dispar* populations.

Genetic distances

The 454 SNP data set generated in the present study showed evidence of strong population genetic structure in the world-wide sample of *L. dispar* populations, as it has been described in previous studies on this forest insect. *L. dispar's* quite large number of DNA polymorphisms has been already established with fingerprinting methods such as AFLPs (Reineke & Karlovsky 2000; Robinet & Liebhold 2009). However, the F_{ST} values found in the present study are much greater than observed in the previous studies. A recent publication which pioneered the use of pyrosequencing of AFLP fragments from pooled samples (Gompert et al. 2010) also found very high F_{ST} values in populations of *Lycaeides* butterflies. That study also found a similar potential source of bias, namely a positive correlation between population polymorphism and the number of sequence reads obtained from that population. Both studies subsampled the genome by sequencing a set of AFLP fragments between 200 and 550 basepairs; and this subsample may have contained a higher proportion of repetitive elements than the genome as a whole.

All populations were quite diverse, so it is difficult to interpret the finding that genetic distances of nearby populations (such as Grossbottwar and Neuenstadt in Germany) can be greater than more distant populations (such as Valais in Switzerland and Stuttgart in Germany) . One assumption could be the low level of flight ability or entirely flightless females in most European *L. dispar* populations as an explanation for low admixture in the populations close to each other (Keena et al. 2008; Keena et al. 2007; Reineke & Zebitz 1998). One study, where a retrotransposon *Lydia* was identified showed that the Asian

populations which are strong fliers and the European flightless populations have been isolated for many years and therefore are genetically differentiated (Pfeifer et al. 2000).

Because we pooled our individuals according to population prior to pyrosequencing, population genetic analysis of parameters within populations was not possible (Behura 2006; Robinet & Liebhold 2009). This would have required the use of additional barcodes to also identify individuals within pools. In future, studying the genetic diversity of *L. dispar* individuals it might be useful to investigate whether diversity levels vary between individuals (Gompert et al. 2010). Studies have shown that inter-specific variability between populations from different geographic regions comes from phenotypic variation among individuals within populations (Haig et al. 1994), which we cannot say about the *L. dispar* individuals we used.

Conclusion

In the only previous paper which used the pyrosequencing technique on AFLP fragments for population studies, Gompert et al. (2010) also found unusually high F_{ST} values. To test whether this could be due to a sampling bias leading to a higher occurrence of multi-copy transposable elements, we chose for analysis only those sequences with high similarity to protein-coding genes from the only moth species with a fully sequenced genome, *Bombyx mori*. However, this may have led to a bias in the opposite direction, because these highly conserved genes would tend to have less polymorphism. It might be that the subset of genes compared to *Bombyx mori* is experiencing stronger purifying selection. On the other hand, a nonrandom distribution of restriction enzyme sites could lead to a bias in certain selection of transcripts. It has been shown already in the cDNA-AFLP studies that not all restriction enzymes cut randomly. In the cDNAs, sometimes fragments generated with *Eco*RI did not show enough variability in the expression pattern and had to be replaced by *Pst*I. Perhaps we see here a similar phenomenon, in that the sites for restriction enzyme used (EcoRI) are not randomly distributed and are located in a repeated non coding region or in the coding region of a transposable element. The higher copy number of these elements would predominate among the 454 high throughput sequences obtained, and a correspondingly smaller fraction of the fragments would be derived from single copy genes for more accurate estimation of genetic distances.

Before population genetic conclusions can be made, coding regions need to be filtered from the existing cDNA libraries to make a better matrix for comparisons and alignments. Furthermore in these high throughput approaches, different restriction enzymes need to be tested for randomness.

Function of transcripts

This study profits from a high coverage of potential candidate genes useful for further research questions in many other fields. In our case, we focused mainly on transcripts containing SNPs. The resulting transcripts cover in particular respiratory genes, immune response genes and protein coding genes. A reverse transcriptase, a peptidase, a coiled-coil-

domain-containing protein, a proline-rich protein, an RNA-directed DNA polymerase and an acid phosphatase activity are among the interesting candidate genes with essential functions for *L. dispar*.

The reverse transcriptase is encoded by the pol gene of retroviruses and by certain retrovirus-like elements (Garner & Slavicek 1999; Pfeifer et al. 2000). It functions in a DNA replication process that uses RNA as a template for DNA synthesis (e.g. reverse transcriptase) (Robertson & Lampe 1995). Reverse transcriptases are used in the replication cycle of retrotransposons and retroviruses (Behura 2006; Robertson & Lampe 1995). Thus the reverse transcriptase gene might originate from a virus infecting some of the samples, or from a retrotransposon occurring in some of the populations. In the former case,it would indicate more about the population structure of the virus than of its host.

We also found genes likely to code for a transposon of the Mariner family, so-called jumping genes, which can cause mutations by inserting into host genes and which have been suggested to mediate horizontal gene transfer across species. In general transposons are indicated as "junk DNA", so it might be the case for these *L. dispar* populations. Mobile elements are novel marker systems which can be useful to track and follow up different strains and subpopulations of pest species, such as *L. dispar* (Robertson & Lampe 1995; Terzian et al. 2000). However, since they exist in multiple copies in each genome, differences between copies such as SNP variation may inflate the overall estimate of genetic variability and produce higher F_{ST} values, relative to single-copy genes.

Outlook for 454 pyrosequencing in non model organisms

454 pyrosequencing projects are a fast and efficient way of defining transcriptomes of specific insect tissues particularly of non model organism, as shown already in *Manduca sexta,* and *Chrysomela tremulae* (Pauchet et al. 2009; Pauchet et al. 2010). It's a fairly new promising method to inexpensively obtain a large number of reads and their thorough genetic information about many organisms. A previous paper illustrated the utility of pyrosequencing of AFLP fragments for population studies (Gompert et al. 2010; Pauchet et al. 2009) and this approach is likely to be more widely used in the future.

Scoring genetic variation in candidate genes on a large scale is costly with traditional sequencing methods. SNPs provide polymorphism information for defined loci, in contrast to anonymous methods such as AFLPs and RAPD marker systems (Vignal et al. 2002). In comparison to other DNA markers, SNPs are of low cost and simple to handle in the lab (Vignal et al. 2002). Using the 454 sequencing method, four times more data can be generated then with conventional sequencing systems (Rothberg & Leamon 2008). So analysis of SNPs obtained by pyrosequencing is a useful method to obtain genome information about non-model organism in a cost efficient way (Behura 2006). In our study, genomic information of the non-model organism *L. dispar* emerged, where little resources were previously available.

6. Major Discussion

6.1. A synthesis of the research project

It has been shown that when *L. dispar*, *L. monacha* and *T. viridana* feed on certain food sources, different genes are induced, either directly or indirectly by compounds in the host plant. Additionally, genes were expressed in cut (detached from the tree) and uncut (attached on the living tree) treatments of *Quercus robur*, when *L. dispar* feeding was conducted. The project with the 454 technique was an unexpected technical challenge. Within the project, different angles could be illuminated concerning insect-plant interaction processes. The cDNA-AFLP analysis is a useful technique, working on non-model organisms, where little genetic information is previously known.

6.2. *Lymantria* ssp. on different food sources

First of all a general overview could be accomplished, by exploring *L. dispar* and *L. monacha* on different food sources. Because of certain preferences of both polyphagous moth species for either decicious vs. conifer trees, feeding assays with natural food and a control assay with artificial diet were conducted.

Genes were identified via cDNA-AFLP analysis. With gel and capillary electrophoresis different expression profiles were visualized between treatments and species.
The major findings of this experiment are that *L. dispar* larvae gained weight on the tolerant oak, sensitive oak, and diet treatments, respectively. Even if *L. monacha* gained overall less weight on all treatments than *L. dispar*, the larvae still gained more weight on oak and diet than on spruce. The clear trend of *L. dispar* putting on less weight on spruce and the general low weight gain in the total *L. monacha* group is shown in the later expressed genes of both species. The groups which were losing weight often showed the same banding or fragment pattern in the expression profiles. These common patterns likely due to stress were picked out, reamplified and sequenced more often in comparison to the less common patterns. Why larvae across species in these different treatments always showed this obvious pattern, can be explained by the decreased weight gain shown in all *L. monacha* and in *L. dispar* on spruce as a consequence of the larval condition which might affect the expression of certain genes (Christeller et al. 2010). The larvae of both species losing weight had indeed the same starting weight, but gained little to hardly any weight in the respective treatments. Within *L. monacha* larvae still gained little weight on diet and oak. Even if feeding deterrents, such as terpenoids in *Picea abies* could play a major role in this phenomenon, there is no proof of a food repellent or sole larva condition effect. Most terpenes are volatile compounds, which can repel herbivores even before feeding starts. Genes found in this particular Pattern 2 could be responsible for digestive processes, but also for starvation processes due to intake of indigestible substances. Larvae could actually have not fed (*L. monacha* on spruce) or fed very little (*L. dispar* on spruce) on the needles, and switched on enzymatic, sugar and protein breakdown genes, as identified by the expression analysis. Conifers contain diterpenoids e.g.

resin acid. Certain species can have complex mixtures of terpenoids, as in *Picea abies*, which has 23 different terpenoid compounds found in the resin extract (Martin et al. 2002). Developing trees such as the ones used in this experiment produce more tepenoids than older ones ((Schoonhoven et al. 2005)and references therein). *L. monacha* larvae were collected in a mature old forest in Poland. For *L. dispar* spruce is rather an unsuitable host. Under these circumstances, *L. monacha* and *L. dispar* larvae perhaps were not used to feeding on these particular plants with their concentrations of compounds. Insect performance on plants is greatly influenced by the plant's genotype and foliar chemistry (Osier & Lindroth 2001). Within one species, *L. dispar* gains more weight feeding on oak than on the other treatments. This preference can be explained by the reduced tannin content in leaves which have just opened from their buds. Tannin accumulates through the year, when leaves are older. The increase in proteins and water could explain the increased variety of oak spring-herbivores (Feeny 1970). Other defense compounds in *Q. robur* are trypsin-like protease inhibitors, which can cause developmental disorders (Ivashov et al. 1992), inhibiting gut proteinases and affecting larval performance in a very effective way. However, herbivore larvae can minimize the effect by feeding on leaves with lower concentrations (Zavala et al. 2008).

The feeding experiment conducted in this study showed a reasonable weight gain on the artificial diet treatment, which has been reported in the literature several times before (Erlandson et al. 2010; Shen et al. 2006). Genes found in *L. dispar* and *L. monacha* are of similar functions. No clear conclusion about food sources and types of genes identified through cDNA-AFLP analysis can be drawn. Genes affected protein breakdown, enzymatic functions, sugar breakdown, ribosomal functions, the fatty acid cycle and respiratory pathways. Even if data has to be taken with caution because of deficiencies in the experimental set up (insects and tree samples taken for the feeding assays), all genes identified were participants in detoxification or digestion. Conclusively, here in this experiment these genes can serve as candidates for further research, in this respect.

6.3. Two different oak types

The next experiment can additionally be regarded as a pilot experiment to test the suitability of the cDNA-AFLP method for identifying genes putatively involved in the interactions between oliphagous *T. viridana* or highly polyphagous *L. dispar* and two different types of *Quercus robur* . Observations in a European forest revealed that in years with high *T. viridana* population densities, not every oak tree is defoliated in the same extent. In May oaks can be detected growing side by side of which one is totally defoliated (defined as "sensitive") and the other is less defoliated (defined as "tolerant"). Our interpretation of this observation was that generally oaks are sensitive to insect larvae, but single trees show a more or less extended sensitivity or tolerance to herbivory (Schroeder 2010).

In the experiment two different weight groups of *L. dispar* larvae were tested on "tolerant" and "sensitive" oak samples. The result of this study shows that lighter *L. dispar* (L) gained

more weight on the tolerant than on the sensitive trees. Compared to this, heavier *L. dispar* (S) showed the opposite response. Larvae gained more weight on sensitive than on tolerant trees. Under the assumption that gaining more weight is equated with feeding more on the respective tree, the feeding performance is comparable with the feeding behaviour of *T. viridana*. The difference between sensitive and tolerant oaks could be due to various reasons, such as the chemical composition. The chemical composition may vary through abiotic and biotic factors, such as wind, sunlight or herbivore pressure. We assume that the tolerant trees have a less favourable chemical constitution than the sensitive ones. That is why *T. viridana* might feed more on the sensitive trees. It has been recorded that for example reduced foliage quality of undamaged neighbours could be a reason for the unpalatability (Tscharntke et al. 2001), which could apply to the tolerant trees

T. viridana and *L. dispar* are all members of *Q. robur's* natural herbivore community, but show partly different reactions to the tolerant and sensitive oak types. This phenotypic plasticity may be caused by the herbivore itself. It could be that in the past the tolerant trees suffered more during outbreak events or experienced other fitness losses. Trees adjusted to these situations by defending themselves more strongly and becoming a "tolerant" tree. This could even be derived by mutations and genetically fixed for the moment, until new environmental changes cause alterations (Schoonhoven et al. 2005). Phenotypic plasticity can also derive from plant-signalling of neighbouring plants which are attacked by herbivores. As the oaks are standing side by side, one of the trees could suffer due to the herbivore attack and send out long-distance signals, such as green-leaf volatiles (GLVs), which prime or even induce defenses in neighbouring plants (Heil & Ton 2008). The sensitive trees might be the emitters and the tolerant receiver trees the ones increasing their defense readiness. The airborne signals are expressed through ethylene and jasmonate pathways (Heil & Ton 2008). Variation of plant genes in response to variation to environmental, physiological and genetic bases such as phenology of trees, changing foliar chemistry and secondary metabolic compound production might be responsible for the herbivore's choice (Schultz & Baldwin 1982).

The difference within *L. dispar* between the younger and the older larvae perhaps needs to be seen as a separate phenomenon. It is known that the synchronized phenology between interaction partners favours young larvae to feed on the more edible leaves of oaks (Hunter & Elkinton 2000; Van Asch & Visser 2007). It is known that tannin accumulation in oak increases throughout the year, when leaves get older. Young leaves consist of less tannin, more protein and water, which are preferred by young larvae (Feeny 1970). Nevertheless, younger larvae particularly of a polyphagous herbivore have to cope with a huge amount of chemicals, so they might rather feed on the tolerant than on the sensitive leaves, because they are used to moving between hosts. The heavier/older larvae of *L. dispar* feeding on the sensitive trees seem to behave more like the specialist *T. viridana*, which moves less from host to host (Pearse 2011). Nevertheless, the expression patterns and genes differing between the younger and the older larvae demand further investigations. It was recently

observed that in feeding experiments of *Helicoverpa armigera* on protease inhibitor diets, larvae feeding 12 hours (younger) showed different expressed genes than larvae feeding 72 hours (older) even in the control treatments (Suyog Kuwar, unpublished data). This is consistent with our results in respect of the age of the *Lymantria* larvae.

The quantitative expression profile of the tested genes via q-PCR did not match the cDNA-AFLP results. Parameters for q-PCR were not tested sufficiently before candidate genes could be verified. More primers designed from the analysis need to be tested. Also RNA samples did not meet all the guidelines for q-PCR experiments (Nolan et al. 2006; Vandesompele et al. 2002).

Through this cDNA-AFLP approach, potential candidate genes of *L. dispar* and *T. viridana* could be identified. More genes in *T. viridana* were present in the sensitive treatment and absent in the tolerant one. This fits with the feeding behaviour of *T. viridana*. In *L. dispar*, only two genes differed between the light and heavy group. All the other genes came from different fragment intensities in the tolerant and sensitive treatments, not their status as absent vs. present. This result could also fit into the polyphagous feeding behaviour of *L. dispar*. Whether younger larvae show a different expression pattern than older larvae hasn't been examined yet.

Even though there is a need for a chemically and genetically analysis of the different types of oaks, this subset of genome information of herbivores feeding on certain food sources enriches our knowledge about the genetic bases of herbivore-host interactions, particularly in non-model organisms (Reineke et al. 1997, 1999; Schroeder & Degen 2008). The information can be used in further research projects, particular in microarray-technique based approaches (Lawrence et al. 2006; Travers et al. 2007).

6.4. Field experiment

When the polyphagous herbivore *L. dispar* feeds on trees, the trees respond with certain defenses. Plants react with different expression profiles when oligophagous or polyphagous lepidopteran larvae feed on them, influenced by certain larval elicitors (Voelckel & Baldwin 2004). In general, plant defences are either direct or indirect. Caused by the herbivore pressure on the plant, different defenses are for different herbivores or plant parts. Indirectly plants can emit VOCs as infochemicals towards neighbouring plants or to other undamaged parts of the same plant (Kost & Heil 2008). They can also provide shelter and food to insect enemies (Kost & Heil 2005; Radhika et al. 2008). Direct defenses include thorns, waxes or toxins (Ballhorn et al. 2009).

With this background information, we wanted to test whether leaves of trees show different defense responses to feeding when detached from the tree (cut), or when feeding is conducted on the living tree (uncut). It is known from the literature that plants react differently to mechanical wounding than to herbivory (Lawrence et al. 2006; Reymond et al. 2000). When tree samples are taken for feeding experiments, they are normally cut into

little branches. The question here was do *Quercus robur* and *Picea abies* show a different pattern of gene expression in cut and uncut samples after herbivory by *L. dispar*. Only oak sequences were left to analyse. The study showed that the major gene activity was found in the uncut treatment of oak. Two genes were identified in the cut treatment, which may be a sample too small to be representative of the cut treatment. In general many photosynthetic related genes were expressed, which indicates an increased energy production, perhaps to compensate for herbivory (Schoonhoven et al. 2005). The ethylene response factor indicates that indirect defense pathways, such as communication signals are switched on (Heil & Ton 2008; Lorenzo et al. 2003). Plant signals are long-distance signals, which may not only reach predators and parasites of the herbivores as an indirect defense response, but also neighbouring plants (Ballhorn et al. 2011). Compared to that there are also short-distance signals, which only reach internal undamaged plant parts (Heil & Adame-Alvarez 2010). Interesting in this respect is the systemic transport of either airborne signals, or vascular information molecules such as proteinase inhibitors and toxic secondary compounds (Heil & Ton 2008). Because vascular information transport cannot reach all plant parts and takes more time to deliver the information to upper leaves, a combination of both provides a good defence reaction after herbivory. A low cost option could be priming through an airborne signal, which follows a defensive reaction after confirmed vascular signals (Heil & Ton 2008).

6.5. SNP-454 project

The last part of this study examines single nucleotide polymorphisms (SNP) in *L. dispar* populations throughout Europe and in Asia. Populations were collected at different times, one set was from 1994-1996 (Reineke 1998) and was used in previous population genetic analysis (Reineke et al. 1999; Reineke & Zebitz 1999). The other set of samples was collected within the framework of the EU Network of Excellence EVOLTREE in the year 2008.

Two analysis approaches were applied resulting in different F_{ST} values (index for genetic distance of two populations). The first analysis was adopted from a study using the 454 sequencing SNP analysis to investigate the historical phylogeography of North American butterfly *Lycaeides* (Gompert et al. 2010). SNPs were detected in aligned transcripts and used for F_{ST} analysis. In our determination of F_{ST} values, sequences were aligned to homologous genes in the *L. dispar* cDNA library by LGC GmbH. Then detection of SNPs was done using software of the company and further estimation of F_{ST} was done with Arlequin software (Excoffier et al. 2005). This approach did not account for coding vs. non-coding regions and transposable elements of the genome and this may have resulted in inflated F_{ST} values between the populations.

The second analysis approach was done partly manually, by ensuring coding regions for analysis by aligning all sequences assorted for certain quality criteria to the *Bomby mori* predicted proteins of the full genome sequencing project. In *Bomby mori* most transposable elements are identified and were sorted out. Then F_{ST} values were calculated separately for all populations in using the Dnasp5 software and then manually combined. Unfortunately

only a few populations could be compared for each gene and F_{ST} values were much smaller than the F_{ST} values of the prior analysis approach.

The distribution of restriction enzyme sites in the genome of *L. dispar* could introduce a bias, because they might not occur randomly but instead preferentially within transposable elements. In the cDNA-AFLP studies it was observed that not all restriction enzymes cut randomly. In the cDNAs, sometimes *Eco*RI was not showing enough variability in the expression pattern and had to be replaced by *Pst*I. Perhaps we see here the same phenomenon, in that the restriction sites (EcoRI) are not randomly distributed in *L. dispar* and are preferentially located in non coding regions or in transposable elements. In 454 high throughput analyses, these genes could be over-represented and that is the reason for the much lower number of single copy genes available to estimate genetic distances. Especially the huge amount of data in the new techniques entails risks of multiplying false data (Gompert et al. 2010).

Before population genetic statements can be made, coding regions need to be filtered from the existing cDNA libraries to make a better matrix for comparisons and alignments. Furthermore in these high throughput approaches, different restriction enzymes need to be tested for randomness.

The discovered genes cannot be related directly to the geographic population differences, because they have to be treated with care. Still, this study enables us to profit from a great coverage of potential candidate genes useful for further research question in many other fields. The power of next generation sequencing and other subsequent improvements of this rapidly developing field will enable us to gain a huge amount of genomic information and sequenced genomes in single individuals and populations of many species for continuously decreasing costs in increasingly shorter times (Robinson et al. 2011).

References

NCBI Group, 2012. Conserved domains [WWW document]. URL http://www.ncbi.nlm.nih.gov/Structure. [13.1.2012, 2.30 Uhr]

Washington State Department of Agriculture, 2011. Gypsy moth facts [WWW document]. URL http://agr.wa.gov/PlantsInsects/InsectPests/GypsyMoth/ControlEfforts/docs/GM2011FactSheet.pdf. [5.1.2012, 12.48 Uhr]

The Natural History Museum, 2011. Host-Database of the world's Lepidopteran [WWW document]. URL http://www.nhm.ac.uk/research-curation/research/projects/hostplants/search/index.dsml. [5.1.2012, 15.30 Uhr]

Wissenschaft-online.de, 2011. Pyruvatdehydrogenase-Komplex [WWW document]. URL http://www.wissenschaft-online.de/abo/lexikon/biok/9585. [9.8.2011, 9.00Uhr]

Polen, 2011. Wikipedia Die freie Enzyklopädie: Polen [WWW document]. URL http://de.wikipedia.org/wiki/Polen. [9. 11.2011, 20:06 Uhr]

Agrawal AA 2000. Benefits and costs of induced plant defense for *Lepidium virginicum* (Brassicaceae). Ecology **81** (7), 1804-1813.

Allgaier M, Reddy A, Park JI, Ivanova N, D'Haeseleer P, Lowry S, Sapra R, Hazen TC, Simmons BA, VanderGheynst JS, Hugenholtz P 2010. Targeted Discovery of Glycoside Hydrolases from a Switchgrass-Adapted Compost Community. PLoS ONE **5** (1), e8812.

Allmann S, Baldwin IT 2010. Insects Betray Themselves in Nature to Predators by Rapid Isomerization of Green Leaf Volatiles. Science **329**, 1075-1077.

Andresen JA, McCullough DG, Potter BE, Koller CN, Bauer LS, Lusch DP, Ramm CW 2001. Effects of winter temperatures on gypsy moth egg masses in the Great Lakes region of the United States. Agricultural and Forest Meteorology **110** (2), 85-100.

Arnold TM, Schultz JC 2002. Induced sink strength as a prerequisite for induced tannin biosynthesis in developing leaves of Populus. Oecologia **130** (4), 585-593.

Babst BA, Ferrieri RA, Thorpe MR, Orians CM 2008. *Lymantria dispar* herbivory induces rapid changes in carbon transport and partitioning in *Populus nigra*. Entomologia Experimentalis et Applicata **128**, 117-125.

Bai X, Zhang W, Orantes L, Jun T-H, Mittapalli O, Mian MAR, Michel AP 2010. Combining Next-Generation Sequencing Strategies for Rapid Molecular Resource Development from an Invasive Aphid Species, *Aphis glycines*. Plos One **5** (6).

Bailey JK, Wooley SC, Lindroth RL, Whitham TG 2006. Importance of species interactions to community heritability: a genetic basis to trophic-level interactions. Ecology Letters **9**, 78-85.

Baldwin IT 2001. An Ecologically Motivated Analysis of Plant-Herbivore Interactions in Native Tobacco. Plant Physiology **127**, 1149-1458.

Ballhorn DJ, Kautz S, Heil M, Hegeman AD 2009. Cyanogenesis of Wild Lima Bean (*Phaseolus lunatus* L.) Is an Efficient Direct Defence in Nature. PLoS ONE **4** (5).

Ballhorn DJ, Kautz S, Jensen M, Schmitt I, Heil M, Hegeman AD 2011. Genetic and environmental interactions determine plant defences against herbivores. Journal of Ecology **99** (1), 313-326.

Ballhorn DJ, Kautz S, Lion U, Heil M 2008. Trade-offs between direct and indirect defences of lima bean (*Phaseolus lunatus*). Journal of Ecology **96** (5), 971-980.

Ballhorn DJ, Schiwy S, Jensen M, Heil M 2008. Quantitative Variability of Direct Chemical Defense in Primary and Secondary Leaves of Lima Bean (*Phaseolus lunatus*) and

Consequences for a Natural Herbivore. Journal Of Chemical Ecology **34** (10), 1298-1301.

Barbosa P, Gross P, Provan GJ, Pacheco DY, Stermitz FR 1990. Alleochemicals in foliage of unfavored tree hosts of the gypsy moth, *Lymntria dispar* L.1. Alkaloids and other components of *Liriodendron-Tulipifera* L (Mangoliaceae), *Acer-Rubrum* L (Aceraceaa), and *Cornus-Florida* L (Cornacea). . Journal Of Chemical Ecology **16** (5), 1719-1730.

Behura SK 2006. Molecular marker systems in insects: current trends and future avenues. Molecular Ecology **15**, 3087-3113.

Bekel T, Henckel K, Kuester H, Meyer F, Runte VM, Neuweger H, Paarmann D, Rupp O, Zakrzewski M, Puehler A, Stoye J, Goesmann A 2009. The Sequence Analysis and Management System - SAMS-2.0: Data management and sequence analysis adapted to changing requirements from traditional sanger sequencing to ultrafast sequencing technologies. Journal of Biotechnology **140** (1-2), 3-12.

Berenbaum MR, Favret C, Schuler MA 1996. On Defining "Key Innovations" in an Adaptive Radiation: Cytochrome P4505 and Papilionidae. The American Naturalist **148**, 139-155.

Brioschi D, Nadalini LD, Bengtson MH, Sogayar MC, Moura DS, Silva-Filho MC 2007. General up regulation of *Spodoptera frugiperda* trypsins and chymotrypsins allows its adaptation to soybean proteinase inhibitor. Insect Biochem. Mol. Biol. **37** (12), 1283-1290.

Christeller JT, Poulton J, Markwick NM, Simpson RM 2010. The effect of diet on the expression of lipase genes in the midgut of the lightbrown apple moth (*Epiphyas postvittana* Walker; Tortricidae). Insect Molecular Biology **19** (1), 9-25.

Coley PD, Bryant JP, Chapin FS 1985. Resource Availability and Plant Antiherbivore Defense. Science **230** (4728), 895-899.

Daurelio LD, Petrocelli S, Blanco F, Holuigue JO, Orellano EG 2011. Transcriptome analysis reveals novel genes involved in nonhost response to bacterial infection in tabacco. Journal of Plant Physiology **168**, 382-391.

De Veau IEJ, Schultz JC 1992. Reassessment of interaction between gut detergents and tannins in Lepidoptera and significance for gypsy moth larvae Journal of chemical Ecology **18** (8), 1437-1453.

Deml R 2004. Effects of age and food source on secondary chemistry of larvae of *Lymantria* species (Lepidoptera : Lymantriidae). Bulletin of Entomological Research **94** (2), 137-143.

Ehrlich PR, Raven PH 1964. Butterflies and Plants: A Study in Coevolution. Evolution **18** (4), 586-608.

Erlandson MA, Hegedus DD, Baldwin D, Noakes A, Toprak U 2010. Characterization of the *Mamestra configurata* (Lepidoptera: Noctuidae) larval midgut protease complement and adaptation to feeding on artificial diet, *Brassica* species, and protease inhibitor. Archives of Insect Biochemistry and Physiology **75** (2), 70-91.

Evertz K, 2004. Effects of increased atmospheric CO2 concentration on the composition and quantity of extra floral nectar by *Acacia holosericea*. In: Department of Botany Rheinische Friedrich Wilhelms University, Bonn.

Excoffier L, Laval G, Schneider S 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol. Bioinform. **1**, 47-50.

Feeny P 1970. Seasonal Changes in Oak Leaf Tannins and Nutrients as a cause of spring feeding by winter moth Caterpillars. Ecology **51** (4), 565-581.

Fischer HM, Wheat CW, Heckel DG, Vogel H 2008. Evolutionary Origins of a Novel Host Plant Detoxification Gene in Butterflies. Molecular Biology and Evolution **25** (5), 809-820.

Foss LK, Rieske LK 2003. Species-specific differences in oak foliage affect preference and performance of gypsy moth caterpillars. Entomologia Experimentalis Et Applicata **108**, 87-93.

Fraenkel GS 1959. The Raison d'Être of Secondary Plant Substances. Science **129** (3361), 1466-1470.

Garner KJ, Slavicek JM 1999. Identification of a non-LTR retrotransposon from the gypsy moth. Insect Molecular Biology **8** (2), 231-242.

Giri AP, Wünsche H, Mitra S, Zavala JA, Muck A, Svatos A, Baldwin IT 2006. Molecular Interactions between the Specialist Herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and Its Natural Host *Nicotiana attenuata*. VII. Changes in the Plant's Proteome. Plant Physiology **142** (4), 1621-1641.

Gompert Z, Forister ML, Fordyce JA, Nice CC, Williamson RJ, Buerkle CA 2010. Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of *Lycaeides* butterflies. Molecular Ecology **19**, 2455-2473.

Gray DR, Ravlin FW, Braine JA 2001. Diapause in the gypsy moth: a model of inhibition and development. Journal of Insect Physiology **47** (2), 173-184.

Haig SM, Rhymer JM, Heckel DG 1994. Population differentiation in randomly amplified polymorphic DNA of red-cockaded woodpeckers *Picoides borealis*. Molecular Ecology **3**, 581-595.

Halitschke R, Schittko U, Pohnert G, Boland W, Baldwin IT 2001. Molecular Interactions between the Specialist Herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and Its Natural Host *Nicotiana attenuata*. III. Fatty Acid-Amino Acid Conjugates in Herbivore Oral Secretions Are Necessary and Sufficient for Herbivore-Specific Plant Responses. Plant Physiology **125**, 711-717.

Hannemann HJ, 2000. Ordnung Lepidoptera-Schuppenflügler oder Schmetterlinge. In: Die grosse farbige Enzyklopädie Urania Tierreich Insekten, Urania Veralg, Berlin, 712-715.

Heiermann J, Schütz S 2008 The effect of the tree species ratio of European beech (*Fagus sylvatica* L.) and Norway spruce (*Picea abies* (L.) Karst.) on polyphagous and monophagous pest species - *Lymantria monacha* L. and *Calliteara pudibunda* L. (Lepidoptera: Lymantriidae) as an example. Forest Ecology and Management **255**, 1161-1166.

Heil M 2008. Indirect defence via tritrophic interactions. New Phytologist **178** (1), 41-61.

Heil M, Adame-Alvarez RM 2010. Short signalling distances make plant communication a soliloquy. Biology Letters **6** (6), 843-845.

Heil M, Ton J 2008. Long-distance signalling in plant defence. Trends in Plant Science **13** (6), 264-272.

Hemming JDC, Lindroth RL 2000. Effects of phenolic glycosides and protein on gypsy moth (Lepidoptera : Lymantriidae) and forest tent caterpillar (Lepidoptera : Lasiocampidae) performance and detoxication activities. Environmental Entomology **29**, 1108-1115.

Henrissat B, Davies G 1997. Structural and sequence-based classification of glycoside hydrolases. Current Opinion in Structural Biology **7** (5), 637-644.

Howe GA, Jander G 2010. Plant immunity to to insect herbivores. Annual Review of Plant Biology **59**, 41-66.

Hunter AF 1993. Gypsy moth population sizes and the window of opportunity in spring. Oikos **68**, 531-538.

Hunter AF, Elkinton JS 2000. Effects of synchrony with host plant on populations of a spring-feeding Lepidopteran. Ecology **81** (5), 1248-1261.
Ishihara A, Ogura Y, Tebayashi S, Iwamura H 2002. Jasmonate-induced changes in flavonoid metabolism in barley (*Hordeum vulgare*) leaves. Bioscience Biotechnology and Biochemistry **66** (10), 2176-2182.
Ivashov AV, Boyko GE, Simchuk AP 1992. Modification and ultilization of the phenolic-compounds in the pubescent oak leaves by caterpillars of the oak leaf-roller moth and the gypsy moth Zhurnal Obshchei Biologii **53** (3), 384-393.
Ivashov AV, Simchuk AP, Medvedkov DA 2001. Possible role of inhibitors of trypsin-like proteases in the resistance of oaks to damage by oak leafroller *Tortrix viridana* and gypsy moth *Lymantria dispar* Ecological Entomology **26** (6), 664-668.
Janzen DH 1966. Coevolution of Mutualism Between Ants and Acacias in Central America. Evolution **20** (3), 249-275.
Jensen JL, Bohonak AJ, Kelley ST 2005. Isolation by distance, web service, v.3.16. BMC Genetics **6**, 13.
Jun T-H, Michel AP, Mian MAR 2011. Development of soybean aphid genomic SSR markers using next generation sequencing. Genome **54** (5), 360-367.
Keena MA 2003. Survival and development of *Lymantria monacha* (Lepidoptera : Lymantriidae) on North American and introduced Eurasian tree species. Journal of Economic Entomology **96** (1), 43-52.
Keena MA, Côté MJ, Grinberg PS, Wallner WE 2008. World Distribution of Female Flight and Genetic Variation in *Lymantria dispar* (Lepidoptera: Lymantriidae). Environmental Entomology **37** (3), 636-649.
Keena MA, Grinberg PS, Wallner WE 2007. Inheritance of Female Flight in *Lymantria dispar* (Lepidoptera: Lymantriidae). Environmental Entomology **36** (2), 484-494.
Kellogg SK, Hebard FV, Rieske LK 2005. Interactions between a fungal pathogen, foliar properties, and generalist herbivores. Entomologia Experimentalis Et Applicata **117** (3), 209-219.
Kessler A, Baldwin IT 2001. Defensive Function of Herbivore-Induced Plant Volatile Emissions in Nature. Science **291**, 2141-2144.
Kost C, Heil M 2008. The defensive role of volatile emission and extrafloral nectar secretion for lima bean in nature. Journal Of Chemical Ecology **34** (1), 2-13.
Kost C, Heil M 2005. Increased availability of extrafloral nectar reduces herbivory in Lima bean plants (*Phaseolus lunatus*., Fabaceae). Basic and Applied Ecology **6** (3), 237-248.
Kruse JJ, Sperling FAH 2001. Molecular Phylogeny Within and Between Species of the Archips argyrospila Complex (Lepidoptera: Tortricidae). Annals of the Entomological Society of America **94** (2), 166-173.
Kunieda T, Fujiyuki T, Kucharski R, Foret S, Ament SA, Toth AL, Ohashi K, Takeuchi H, Kamikouchi A, Kage E, Morioka M, Beye M, Kubo T, Robinson GE, Maleszka R 2006. Carbohydrate metabolism genes and pathways in insects: insights from the honey bee genome. Insect Molecular Biology **15** (5), 563–576.
Lawrence SD, Dervinis C, Novak N, Davis JM 2006. Wound and insect herbivory responsive genes in poplar. Biotechnology Letters **28** (18), 1493-1501.
Lazarevic J, Peric-Mataruga V, Prolic Z, Tucic N 2003. Behavioural response to an unsuitable host plant in the gypsy moth (*Lymantria dispar* L.). Folia Biologica-Krakow **51** (1-2), 129-131.
Lazarevic J, Peric-Mataruga V, Stojkovic B, Tucic N 2002. Adaptation of the gypsy moth to an unsuitable host plant. Entomologia Experimentalis Et Applicata **102** (1), 75-86.

Lazarevic J, Peric-Mataruga V, Vlahovic M, Mrdakovic M, Cvetanovic D 2004. Effects of rearing density on larval growth and activity of digestive enzymes in *Lymantria dispar* L. (Lepidoptera : Lymantriidae). Folia Biologica-Krakow **52** (1-2), 105-112.

Lobinger G, Skatulla U 2001. Studies on the survivability and development of gypsy moth larvae of two different origins, *Lymantria dispar* L. (Lep.: Lymantriidae) as a function of the host plant. Anzeiger Fur Schadlingskunde-Journal of Pest Science **74** (4), 89-93.

Lorenzo O, Piqueras R, Sanchez-Serrano JJ, Solano R 2003. ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. Plant Cell **15** (1), 165-178.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen ZT, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu PG, Begley RF, Rothberg JM 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437** (7057), 376-380.

Martin D, Tholl D, Gershenzon J, Bohlmann J 2002. Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway spruce stems. Plant Physiology **129**, 1003-1018.

Mello MO, Silva-Filho MC 2002. Plant-insect interactions: An evolutionary arms race between two distinct defense mechanisms. Brazilian Journal of Plant Physiology **14** (2), 71-81.

Meudt HM, Clarke AC 2007. Almost forgotten or latest pratice? AFLP applications, analyses and advances. Trends in Plant Science **12** (3), 106-117.

Morcillo-Suarez C, Alegre J, Sangros R, Gazave E, de Cid R, Milne R, Amigo J, Ferrer-Admetlla A, Moreno-Estrada A, Gardner M, Casals F, Perez-Lezaun A, Comas D, Bosch E, Calafell F, Bertranpetit J, Navarro A 2008. SNP analysis to results (SNPator): a web-based environment oriented to statistical genomics analyses upon SNP data. Bioinformatics **24** (14), 1643-1644.

Naidoo R, Lechowicz MJ 2001. Effects of gypsy moth on radial growth of deciduous trees. Forest Science **47** (3), 338-348.

Nishida R 2002. Sequestration of defensive substances from plants by Lepidoptera Annual Review of Entomology **47**, 57-92.

Nolan T, Hands RE, Bustin SA 2006. Quantification of mRNA using real-time RT-PCR. Nature protocols **1** (3), 1559-1582.

Oikawa A, Ishihara A, Iwamura H 2002. Induction of HDMBOA-Glc accumulation and DIMBOA-Glc 4-O-methyltransferase by jasmonic acid in poaceous plants. Phytochemistry **61** (3), 331-337.

Opitz S, Kunert G, Gershenzon J 2008. Increased terpenoid accumulation in cotton (Gossypium hirsutum) foliage is a general wound response. Journal Of Chemical Ecology **34** (4), 508-522.

Osier TL, Hwang SY, Lindroth RL 2000. Effects of phytochemical variation in quaking aspen *Populus tremuloides* clones on gypsy moth *Lymantria dispar* performance in the field and laboratory Ecological Entomology **25**, 197-207.

Osier TL, Lindroth RL 2001. Effects of genotype, nutrient availability, and defoliation on aspen phytochemistry and insect performance. Journal Of Chemical Ecology **27**, 1289-1313.

Pathan AAK, Devi KU, Vogel H, Reineke A 2007. Analysis of differential gene expression in the generalist entomopathogenetic fungus *Beauveria bassiana* (Bals.)Vuillemin grown on different insect cuticular extracts and synthetic medium through cDNA-AFLPs. Fungal Genetics and Biology **44**, 1231-1241.

Patin E, Laval G, Barreiro LB, Salas A, Semino O, Santachiara-Benerecetti S, Kidd KK, Kidd JR, Van der Veen L, Hombert J-M, Gessain A, Froment A, Bahuchet S, Heyer E, Quintana-Murci L 2009. Inferring the Demographic History of African Farmers and Pygmy Hunter–Gatherers Using a Multilocus Resequencing Data Set. PLoS Genet **5** (4), e1000448.

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature **441**, 1103-1108.

Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, Ffrench-Constant RH 2009. Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. Insect Biochem. Mol. Biol. **39** (5-6), 403-413.

Pauchet Y, Wilkinson P, Vogel H, Nelson DR, Reynolds SE, Heckel DG, Ffrench-Constant RH 2010. Pyrosequencing the *Manduca sexta* larval midgut transcriptome: messages for digestion, detoxification and defence. Insect Molecular Biology **19** (1), 61-75.

Pearse IS 2011. The role of leaf defensive traits in oaks on the preference and preformance of polyphagous herbivore, *Orgyia vetusta*. Ecological Entomology **36**, 635-642.

Peric-Mataruga V, Blagojevic D, Lazarevic J, Spasic M 2000. Host plant effects on the presence of superoxide dismutase isoforms in gypsy moth midgut. Biologia **55** (5), 525-531.

Pfeifer TA, Ring M, Grigliatti TA 2000. Identification and analysis of Lydia, a LTR retrotransposon from *Lymantria dispar*. Insect Molecular Biology **9** (4), 349-356.

Polesani M, Desario F, Ferrarini A, Zamboni A, Pezzotti M, Kortekamp A, Polverari A 2008. cDNA-AFLP analysis of plant and pathogen genes expressed in grapevine infected with *Plasmopara viticola*. BMC Genomics **9** (142), 1-14.

Radhika V, Kost C, Bartram S, Heil M, Boland W 2008. Testing the optimal defence hypothesis for two indirect defences: extrafloral nectar and volatile organic compounds. Planta **228** (3), 449-457.

Regniere J, Nealis V 2002. Modelling seasonality of gypsy moth, *Lymantria dispar* (Lepidoptera : Lymantriidae), to evaluate probability of its persistence in novel environments. Canadian Entomologist **134** (6), 805-824.

Reineke A 1998. Differenzierung von verschiedenen Populationen des Schwammspinners, *Lymantria dispar* L. (Lepidoptera, Lymantriidae), mit Hilfe molekularer Marker., 115.

Reineke A, 1998. Differenzierung von verschiedenen Populationen des Schwammspinners, *Lymantria dispar* L.(Lepidoptera, Lymantriidae), mit Hilfe molekularer Marker. In: Institut für Phytomedizin, Universität Hohenheim, Stuttgart.

Reineke A, Karlovsky P 2000. Simplified AFLP protocol: Replacement of primer labeling by the incorporation of alpha-labeled nucleotides during PCR. Biotechniques **28** (4), 622-623.

Reineke A, Karlovsky P, Zebitz CPW, 1997. AFLP: A novel DNA fingerprinting technique for detecting genetic variation in insects. In: Mitteilungen Der Deutschen Gesellschaft Fur Allgemeine Und Agewandte Entomologie, Band 11, Heft 1-6, Dezember 1997 -

Entomologists Conference. Ed. by K H Hoffmann, W Volkl, Deutschen Gesell Allgemeine & Angewandte Entomologie, Bremen 33, 509-512.
Reineke A, Karlovsky P, Zebitz CPW 1999. Amplified fragment length polymorphism analysis of different geographic populations of the gypsy moth, *Lymantria dispar* (Lepidoptera : Lymantriidae). Bulletin of Entomological Research **89** (1), 79-88.
Reineke A, Karlovsky P, Zebitz CPW 1998. Preparation and purification of DNA from insects for AFLP analysis. Insect Molecular Biology **7** (1), 95-99.
Reineke A, Löbmann S 2005. Gene expression changes in *Ephestia kuehniella* caterpillars after parasitization by the endoparasitic wasp *Venturia canescens* analyzed through cDNA-AFLPs. Journal of Insect Physiology **51**, 923-932.
Reineke A, Schmidt O, Zebitz CPW 2003. Differential gene expression in two strains of the endoparasitic wasp *Venturia canescens* identified by cDNA-amplified fragment length polymorphism analysis. Molecular Ecology **12**, 3485-3492.
Reineke A, Zebitz CPW 1998. Flight ability of gypsy moth females (*Lymantria dispar* L.) (Lep., Lymantriidae): A behavioural feature characterizing moths from Asia? Journal of Applied Entomology-Zeitschrift Fur Angewandte Entomologie **122** (6), 307-310.
Reineke A, Zebitz CPW 1999. Suitability of polymerase chain reaction-based approaches for identification of different gypsy moth (Lepidoptera: Lymantriidae) genotypes in Central Europe. Annals of the Entomological Society of America **92**, 737-741.
Reymond P, Weber H, Damond M, Farmer EE 2000. Differential Gene Expression in Response to Mechanical Wounding and Insect Feeding in Arabidopsis. The Plant Cell **12**, 707-719.
Rieske LK 2001. Influence of symbiotic fungal colonization on oak seedling growth and suitability for insect herbivory. Environmental Entomology **30**, 849-854.
Rieske LK, Rhoades CC, Miller SP 2003. Foliar chemistry and gypsy moth, *Lymantria dispar* (L.), herbivory on pure American chestnut, *Castanea dentata* (Fam : Fagaceae), and a disease-resistant hybrid. Environmental Entomology.
Robertson HM, Lampe DJ 1995. Distribution of transposable elements in arthropods. Annual Review of Entomology **40**, 333-357.
Robinet C, Liebhold AM 2009. Dispersal polymorphism in an invasive forest pest affects its ability to establish. Ecological Applications **19** (7), 1935-1943.
Robinson GE, Hackett KJ, Purcell-Miramottes M, Brown SJ, Evans JD, Goldsmith MR, Lawson D, Okamuro J, Robertson HM, Schneider DJ 2011. Creating a Buzz About Insect Genomes. Science **331**, 1386.
Rothberg JM, Leamon JH 2008. The development and impact of 454 sequencing. Nature Biotechnology **26** (10), 1117-1124.
Rounsley S, Marri PR, Yu Y, He R, Sisneros N, Goicoechea JL, Lee SJ, Angelova A, Kudrna D, Luo M, Affourtit J, Desany B, Knight J, Niazi F, Egholm M, Wing RA 2009. De Novo Next Generation Sequencing of Plant Genomes. Rice **2** (1), 35-43.
Santure AW, Stapley J, Ball AD, Birkhead TR, Burke T, Slate J 2010. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. Molecular Ecology **19**, 1439-1451.
Schoonhoven LM, Van Loon JJA, Dicke M 2005. Isect-Plant-Biology. Oxford University Press, United States, New York.
Schroeder H 2010. Eine Erfolgsgeschichte: Sommerveredelung bei Eichen. AFZ-DerWald **5**, 16-17.

Schroeder H, Arens P, Smulders MJM 2007. Autosomal and sex-linked microsatellite loci in the green oak leaf roller *Tortrix viridana* L. (Lepidoptera, Tortricidae). Permanent genetic resources note, 809-811.

Schroeder H, Degen B 2008. Genetic structure of the green oak leaf roller (*Tortrix viridana* L.) and one of its hosts, Quercus robur L. Forest Ecology and Management **256**, 1270-1279.

Schroeder H, Degen B 2008. Spatial genetic structure in populations of the green oak leaf roller, *Tortrix viridana* L. (Lepidoptera, Tortricidae). European Journal of Forest Research **127**, 447-453.

Schroeder H, Scholz F 2005. Ansteigende Eichenwickler-Kalamität in Nordrhein-Westfalen. AFZ-DerWald **4**, 172-173.

Schroeder H, Scholz F 2005. Identification of PCR-RFLP Haplotypes For Assessing Genetic Variation in the Green Oak Leaf Roller *Tortrix viridana* L. (Lepidoptera, Tortricidae). Silvae Genetica **54** (1), 17-24.

Schuelke M 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology **18** (2), 233-234.

Schultz JC, Baldwin IT 1982. Oak leaf quality declines in response to defoliation by gypsy moth larvae. Science **217** (4555), 149-151.

Sestili S, Polverari A, Luongo L, Ferrarini A, Scotton M, Hussain J, Delledonne M, Ficcadenti N, Belisario A 2011. Distinct colonization patterns and cDNA-AFLP transcriptome profiles in compatible and incompatible interactions between melon and different races of *Fusarium oxysporum* f. sp. *melonis*. BMC Genomics **12** (122), 1-21.

Shen TC, Tseng CM, Guan LC, Hwang SY 2006. Performance of *Lymantria xylina* (Lepidoptera : Lymantriidae) on artificial and host plant diets. Journal of Economic Entomology **99** (3), 714-721.

Shields VDC, Broomell BP, Salako JOB 2003. Host selection and acceptability of selected tree species by gypsy moth larvae, *Lymantria dispar* (L.). Annals of the Entomological Society of America **96** (6), 920-926.

Shields VDC, Rodgers EJ, Arnold NS, Williams D 2006. Feeding responses to selected alkaloids by gypsy moth larvae, *Lymantria dispar* (L.). Naturwissenschaften **93** (3), 127-130.

Smouse PE 2010. How many SNPs are enough? Molecular Ecology **19**, 1265-1266.

Sonoda S, Fukumoto K, Izumi Y, Yoshida H, Tsumuki H 2006. Cloning of heat shock protein genes (hsp90 and hsc70) and their expression during larval diapause and cold tolerance acquisition in the rice stem borer, *Chilo suppressalis* Walker. Archives of Insect Biochemistry and Physiology **63** (1), 36-47.

Terzian C, Pelisson A, Bucheton A, 2000. Retrotransposons and retroviruses in insect genomes. In: Insect Transgenesis: Methods and Applications. Ed. by A M Handler, J A.A., CRC Press, 191-201.

Tooker JF, De Moraes CM 2005. Jasmonate in lepidopteran eggs and neonates. Journal of Chemical Ecology **31** (11), 2753-2759.

Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, Willoughby DA, Simons JF, Egholm M, Hunt JH, Hudson ME, Robinson GE 2007. Wasp Gene Expression Supports an Evolutionary Link Between Maternal Behavior and Eusociality. Science **318** (5849), 441-444.

Travers SE, Smith MD, Bai J, Hulbert SH, Leach JE, Schnable PS, Knapp AK, Milliken GA, Fay PA, Saleh A, Garrett KA 2007. Ecological genomics: making the leap from model systems in the lab to native populations in the field. Frontiers in Ecology and the Environment **5** (1), 19-24.

Tscharntke T 2001. New evidence for talking trees? Herbivory and induced resistance in *Alnus glutinosa*. Zeitschrift fuer Pflanzenkrankheiten und Pflanzenschutz **108** (6), 635-636.

Tscharntke T, Thiessen S, Dolch R, Boland W 2001. Herbivory, induced resistance, and interplant signal transfer in *Alnus glutinosa*. Biochemical Systematics and Ecology **29**, 1025-1047.

Tscharntke T, Thiessen S, Dolch R, Boland W 2001. Herbivory, induced resistance, and interplant signal transfer in *Alnus glutinosa*. Biochemical Systematics and Ecology **29** (10), 1025-1047.

van Asch M, Visser ME 2007. Phenology of Forest Caterpillars and Their Host Trees: The Importance of Synchrony. Annual Review of Entomology **52**, 37-55.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology **3** (7), 1-12.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology **17** (7), 1636-1647.

Vidal S, Tscharntke T 2001. Special feature: Multitrophic plant-insect interactions. Basic and Applied Ecology **2** (1), 1-2.

Vignal A, Milan D, SanCristobal M, Eggen A 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution **34**, 275-305.

Visser ME, Holleman LJM, Gienapp P 2006. Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. Oecologia **147** (1), 164-172.

Voelckel C, Baldwin IT 2004. Generalist and specialist lepidopteran larvae elicit different transcriptional responses in *Nicotiana attenuata*, which correlate with larval FAC profiles. Ecology Letters **7** (9), 770-775.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Homes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research **23** (21), 4407-4414.

Wang P, Li K, Kain W 2004. Characterization and cDNA cloning of midgut carboxypeptidases from *Trichoplusia ni*. Insect Biochem. Mol. Biol. **34**, 831–843.

Wanner KW, Helson BV, Harris BJ 2000. Laboratory and field evaluation of spinosad against the gypsy moth, *Lymantria dispar*. Pest Management Science **56** (10), 855-860.

Williams LM, Ma X, Boyko AR, Bustamante CD, Oleksiak MF 2010. SNP identification, verification, and utility for population genetics in a non-model genus. BMC Genetics **11** (32), 1-14.

Withers TM, Keena MA 2001. *Lymantria monacha* (nun moth) and *L.dispar* (gypsy moth) survival and development on improved *Pinus radiata*. New Zealand Journal of Forestry Science **31** (1), 66-77.

Xia QY, Zhou ZY, Lu C, Cheng DJ, Dai FY, Li B, Zhao P, Zha XF, Cheng TC, Chai CL, Pan GQ, Xu JS, Liu C, Lin Y, Qian JF, Hou Y, Wu ZL, Li GR, Pan MH, Li CF, Shen YH, Lan XQ, Yuan LW, Li T, Xu HF, Yang GW, Wan YJ, Zhu Y, Yu MD, Shen WD, Wu DY, Xiang ZH, Yu J, Wang J, Li RQ, Shi JP, Li H, Li GY, Su JN, Wang XL, Li GQ, Zhang ZJ, Wu QF, Li J, Zhang QP, Wei N, Xu JZ, Sun HB, Dong L, Liu DY, Zhao SL, Zhao XL, Meng QS, Lan FD, Huang XG, Li YZ, Fang L, Li DW, Sun YQ, Zhang ZP, Yang Z, Huang YQ, Xi Y, Qi QH, He DD, Huang HY, Zhang XW, Wang ZQ, Li WJ, Cao YZ, Yu YP, Yu H, Li JH, Ye JH, Chen H, Zhou Y, Liu B, Ye

J, Ji H, Li ST, Ni PX, Zhang JG, Zhang Y, Zheng HK, Mao BY, Wang W, Ye C, Li SG, Wong GKS, Yang HM 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science **306** (5703), 1937-1940.

Yi H-S, Heil M, Adame-Alvarez RM, Ballhorn DJ, Ryu C-M 2009. Airborne Induction and Priming of Plant Defenses against a Bacterial Pathogen. Plant Physiology **151** (4), 2152-2161.

Zavala J, Giri A, Jongsma MA, Baldwin IT 2008. Digestive duet: midgut digestive proteinases of *Manduca sexta* ingesting *Nicotiana attenuata* with manipulated trypsin proteinase inhibitor expression. PLoS ONE **3** (4), 1-10.

Zschenker O, Oezden D, Ameis D 2004. Lysosomal Acid Lipase as a Preproprotein. Journal of Biochemistry **136** (1), 65-72.

Acknowledgements

First of all I would like to thank my supervisors Prof. David G. Heckel and Prof. Annette Reineke for giving me this very diverse topic and the opportunity to learn many new techniques, EU project experiences and patience.

The thesis work was supported by the European Union within the 6th framework programme in the network of excellence EVOLTREE. I would like to thank the entire EVOLTREE community, particularly direct collaboration partners from the insect group. Dr Hilke Schröder, I would like to thank for many scientific conversations, the possibility to do this excellent fieldtrip to Münster and the opportunity to make new laboratory-experiences in the Grosshansdorf research centre. Thank you for the warm welcome by the members of the Grosshansdorf institute. Another EVOLTREE-collaboration was with Dr Egbert Beuker from the Finnish Forest Research Institute Metla. He enabled me to fulfill my big dream of conducting experiments within a tree canopy with the help of a canopy crane. Thank you for great hospitality in the Finland station. Thank you very much to Dr Andrzej Oleksa and his effort for giving me the opportunity to collect *L. monacha* in Osie, Poland. I have gained so many experiences in this fieldtrip.
I am also very grateful to Dr Melody Keena from USDA for her useful rearing information and for sending eggs to me to start my own lab culture. Unfortunately we never met in person.

I would like to thank the Max Planck Institute for Chemical Ecology and especially the department for Entomology for financial support. I would like to acknowledge Dr Heiko Vogel for generating the cDNA libraries. Thank you to Domenica Schnabelrauch for a always friendly, quick and helpful sequencing service. Katrin Salzmann-Böhmer was always very helpful with all kind of questions. Thank you to Bianca Wurlitzer for teaching me the great dissection method of larvae midguts. Thank you to Regina Seibt with whom it was easygoing to work within the rearing lab. I really enjoyed the many conversations about travelling and foreign countries. Thanks also to Henriette Ringys-Beckstein for help with molecular methods and instruments. Thank you very much to my Hiwis and friends, Anja and Alex. Lisa Knolhoff, who I met much too late, thank you for being a friend and help with reviewing chapters of my thesis. Thank you to English experience service, which reviewed also parts of my thesis.
Many people in the administration and service groups were in a friendly way very supportive throughout my PhD. Martin Nibergall, I would like to thank for always solving my computer problems and for nice chats. Thanks to Linda Maack for answering many Endnote questions even on the phone. Frau Boland, I would like to thank for the smooth processing of the travel billings, which were not always easy with all my business travel. Listening to our region dialect made me feel home. I would like to thank particularly Christel Hugk for making ordering in external labs easy. Additional I would like to thank Eleonore Lobenstein and Silvia Lindemann for much support and always a nice word, when I was in Geisenheim.

I would like to acknowledge the nice crowd around the International Max Planck Research School in Jena, especially Karin Groten, who was always helpful and supportive!

My colleagues at the department of Phytomedicine of the Research centre Geisenheim I would like to thank for welcoming me so friendly and supporting me. I always felt home in Geisenheim throughout my stay. Especially I would like to thank Mrs Dolezal for overall help and Mirijam Hauck for introducing me to the laboratory and statistics. Elsmarie Mello, I would like to thank for always having an open ear for me and my sorrows, you were like a mother to me, especially when I broke my sternum.
Of course, I also thank all the PhD students of Geisenheim for company and many happy moments.
I would like to thank my nice neighbour Mareille Baier for her versatile support.
The next Thank you is devoted to the "Men" who were always on my side. Oliver Martinez and Stefan Klärner (designer of this book cover) gave me support, companionship and so many nice moments. Special thanks to Seung-Joon Ahn a good friend for life, who always believed in me. Through his calming, friendly and kind character he always settled conflicts. I owe special thanks to Dustin Kulanek for his endless optimism, patience, support in hard times and for giving me a not always work related life. He also gave me lots of good suggestions and helped me improve the figures in this thesis.

Above all I thank my parents, who are always there for me and without them I would have never come so far in my career. You gave me the power and love to always stand up again.

Selbständigkeitserklärung

Entsprechend der geltenden, mir bekannten Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena erkläre ich, dass ich die vorliegende Dissertation eigenständig angefertigt und alle von mir benutzten Hilfsmittel und Quellen angegeben habe.

Personen, die mich bei der Auswahl und Auswertung des Materials sowie bei der Fertigstellung der Dissertation unterstützt haben, sind im Text oder in der Danksagung genannt. Es wurde weder die Hilfe eines Promotionsberaters in Anspruch genommen, noch haben Dritte für Arbeiten, welche im Zusammenhang mit dem Inhalt der vorliegenden Dissertation stehen, geldwerte Leistungen erhalten. Die vorgelegte Dissertation wurde außerdem weder als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung noch als Dissertation an einer anderen Hochschule eingereicht.

Jena, den 9.2.2012

Kirsten Evertz

Declaration of contributions of the thesis

Prof. David G. Heckel supervised the thesis and was involved with the planning of the experiments and edited the thesis chapters. He did the second analysis of the chapter 6.
Prof. Annette Reineke supervised the thesis was involved with the planning of the experiments and edited the thesis chapters. She did the first analysis and wrote respective parts in materials/ method and result section of the chapter 6.
Dr. Heiko Vogel constructed the cDNA libraries for *Lymantria dispar*, *Lymantria monacha* and *Tortrix viridana*.
Dr. Hilke Schröder was involved in the planning and collection of the plant material for chapter 4. She edited the chapter 4.
Domenica Schnabelrauch preformed all sequencing of the cDNA-AFLP studies.
Dustin Kulanek preformed all extractions for chapter 5 and carried out all Beckman capillary electrophoresis.
LGC GmbH, Berlin preformed the 454 pyrosequencing and bioinformatic preparations for the SNP analysis.