Comparative gene expression analysis based on RNA sequencing for the identification of host plant adaptation mechanisms in Chrysomelina leaf beetles

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 Bioinformatikerin Ding Wang

geboren am 5. Sept. 1979

in TianJin (China)

1. **Gutachter:**
2. **Gutachter:**
3. **Gutachter:**

Tag der öffentlichen Disputation:

Contents

1 Introduction

For most non-model insects there is a lack of genomic sequences and prior knowledge of genes, therefore the studies of investigated insects were limited to only a few organisms in the past. In such a case RNA sequencing (RNA-seq), a high-throughput sequencing method enables researchers to also *de novo* assemble transcript libraries, identify genes and use comparative transcriptomics (Conesa *et al.* 2016; Wang *et al.* 2009). The power of this approach is the combination of discovery and quantification of investigated genes without sequenced genomes in a single sequencing assay. However there is no optimal pipeline to analyze the vast amount data. RNA-seq experiment and analysis strategies need to be designed and adjusted in consideration of research goals and the specific organisms in question. In this study, I focus on this aspect to provide a basic understanding of adaptation mechanisms especially for the defensive sequestration and olfactory system between three non-model leaf beetles and their host plants.

The phytophagous leaf beetles have developed a successful defensive system. Their larvae protect themselves against natural enemies by excreting deterrent compounds. These leaf beetles' larvae have adapted to the secondary metabolites of host plants to synthesize the defensive compounds (Agrawal *et al.* 2012; Eben *et al.* 1997; Gross *et al.* 2002; J.M. Pasteels 1983; Oldham *et al.* 1996; Termonia & Pasteels 1999). The toxic deterrents are stored and synthesized in the glandular reservoir that operates as a bioreactor (Blum *et al.* 1978; Meinwald *et al.* 1977; Pasteels *et al.* 1982; Sugawara *et al.* 1979). Before reaching the reservoir, the non-toxic glucosidic precursors from host plants have to pass several membrane barriers inside the body: transfer from gut into hemolymph, afterwards into the glandular system or Malpighian tubules, as a whole forming a transport network (see Figure 1). In this process transmembrane proteins may play an important role. Another topic in this study focuses on the olfactory system in *Chrysomela lapponica*. This system is responsible for the detection of food, mating or determination of oviposition sites. Its adaptation is therefore crucial in a host plant shift from salicin-rich willow to salicin-free birch.

With the aim to understand the adaptation mechanisms, the transcriptome of *Phaedon cochleariae*, *Chrysomela populi* and *Chrysomela lapponica* were investigated by utilization of RNA-seq. The Illumina sequencing platform was used to quantify and compare the expression of transcripts. Transcript libraries were *de novo* assembled with suited methods for count data. To annotate and characterize the investigated transcripts, alignment to public/in-house databases as well as phylogenetic analysis were carried out. In addition, proteomic analysis, real-time quantitative polymerase-chain-reaction (qPCR) and RNAi silencing of investigated transcripts allow verification of comparative and functional analyses in different samples in non-model insect (Bodemann *et al.* 2012; Pabinger *et al.* 2014).

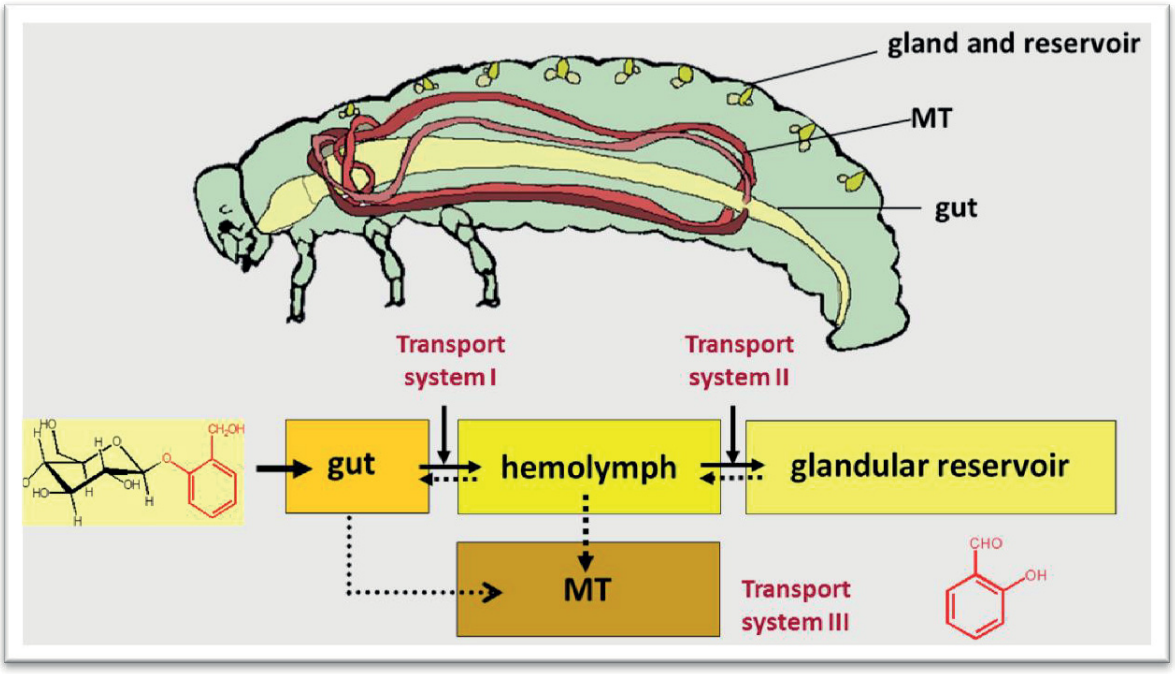

Figure 1. Overview of the transport system I-III in leaf beetle larvae. Plant-derived or *de novo* sequestrated glucosides are transferred from the gut to the glandular reservoir. Malpighian tubules (MT) are required to excrete non-used glucosides.

1.1 Next-generation sequencing (NGS)

In 1953 the double-helix model of DNA was formulated for the first time by Watson and Crick (Watson & Crick 1953). After twenty years of progress of the DNA sequencing technology, chain-termination sequencing (first generation of DNA sequencing) was developed by Sanger

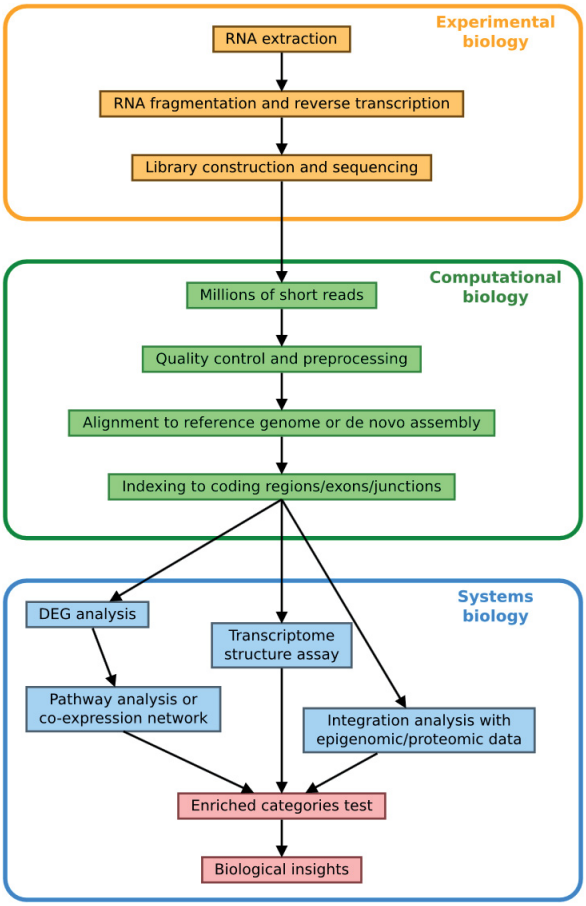

Figure 2. Overview of the typical RNA-seq pipeline. Three main sections are presented: the experimental Biology, the computational Biology and the system Biology. (Han *et al.* 2015)

(Sanger *et al.* 1977). However the time and cost of sequencing still limited research in molecular biology. At the beginning of the 21st century, a massively parallel sequencing technique named next-generation-sequencing (NGS) appeared (Brenner *et al.* 2000). It was cheaper and faster, produced millions of short reads in a single run, and it has revolutionized genomic and genetic research. Using NGS, numerous research results were published, in genome (re)sequencing, transcriptional profiling (RNA-Sequencing), DNA-protein interactions (ChIP-Seq) and epigenome analysis (de Magalhaes *et al.* 2010; Goodwin *et al.* 2016; Morozova & Marra 2008).

1.1.1 RNA sequencing

RNA-seq is an application of NGS that enables rapid and deep research of RNA. The general principle is the sequencing of cDNA, which has been converted by RNA. The experimental protocol of RNA-seq includes three parts (Hou *et al.* 2015): (i) isolation of target RNAs; (ii) conversion of RNAs into cDNAs, binding of adapters to one or both ends of the cDNAs; (iii) amplification using PCR and sequencing of cDNAs. This approach is used to detect alternative splice junctions, mutations, single nucleotide polymorphisms (SNPs), gene fusion as well as post-transcriptional modifications (Conesa *et al.* 2016). Specifically, RNA-seq allows to "*de novo* assemble" the entire transcriptome and to discover genes without any knowledge of the genome. Another powerful function of RNA-seq is the analysis of the global expression level in the case of comparative studies of differential experimental set-ups. RNA-seq has become a standard transcriptome analysis method. In this

thesis, I focus on the RNA-seq of mRNAs. Typically, RNA-Seq libraries are prepared from total RNA using poly(A) enrichment of the mRNA to remove structural RNAs (such as rRNA, tRNA).

1.1.2 RNA-seq data analysis

RNA-seq technology allows the generation of large and complex datasets in a single run and its costs are continuously decreasing. It became one of the most important tools in the life sciences research community. A key challenge is the processing and computational analysis of the huge output data. With the support of bioinformatics, complex computational methods have been developed for the analysis, which involve filtering the RNA-seq reads, reconstructing of transcriptomes, aligning reads with reference sequences and stochastic analyses (Figure 2).

1.1.2.1 Preprocessing

In general, parallel sequencing RNA-seq produces large amounts of reads (from millions to billions) (Van Verk *et al.* 2013). The reads are stored in files, typically FASTQ format (Cock *et al.* 2010) that contains the nucleotide sequences and the quality values for each nucleotide. The quality control and the visualization for the raw reads is calculated by using FastQC (Andrews 2016), whose results include sequence quality, GC or N content, sequence length distribution, duplication level.

In order to obtain high quality data, trimming and filtering of raw reads must be done as the first step of the analysis. This involves removing the low quality nucleotides towards the 3' end and the adapters from 5' end of raw reads. Tools like cutadapt (Martin 2016), FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and Trimmomatic (Bolger *et al.* 2014) can be used for trimming.

1.1.2.2 De novo assembly and annotated transcriptome

De novo assembly of transcriptome (using the preprocessed RNA-seq reads) allows analysis in non-model organisms that have neither genome sequences nor transcriptome information (Hornett & Wheat 2012; Marchant *et al.* 2015). In general, RNA-seq assembly merges short reads into longer contiguous sequences, which are named contigs (Ekblom & Wolf 2014). Recently, several tools for *de novo* assembly of transcriptome such as SOAPdenovo-Trans (Xie *et al.* 2014) , Oases (Schulz *et al.* 2012), Trans-ABySS (Grabherr *et al.*

2011) and Trinity (Haas *et al.* 2013) are widely used. At the time of this thesis, Trinity regarding its high performance and high success rate in assembling of complex/high quality transcriptome has been demonstrated to be most effective in several comparative studies (Celaj *et al.* 2014; Chopra *et al.* 2014; Duan *et al.* 2012; Xu *et al.* 2012; Zhao *et al.* 2011). After assembling the RNA-seq reads into contigs, the program builds them into independent Bruijn graphs including splicing isoforms (Grabherr *et al.* 2011). The software TGICL (Pertea *et al.* 2003) for fast clustering and assembly of large EST/mRNA datasets can be chosen to reassemble the output data of Trinity into longer and more complete consensus sequences.

For a biological study, a convincing annotation of transcripts is necessary prior to any analyses. In general, the completely assembled transcript library can be aligned to public databases such as Protein Families (Pfam) (Finn *et al.* 2016) or gene ontology (GO) (Ashburner *et al.* 2000) databases for annotation purpose. It is useful to detect the proteins that share highly conserved domains across species (such as ABC transporters see manuscript 3.1). To verify the identification, all of the determined proteins can be further analyzed by using UniProt (UniProt Consortium 2015), non-redundant RefSeq from NCBI (Pruitt *et al.* 2007).

1.1.2.3 Mapping, counting, normalization and differential expression analysis

Another useful application of RNA-seq is to quantify the abundance of gene expression and to do comparative studies among samples. It has become an important alternative approach to microarrays for the transcript profiling. In comparison with hybridization-based microarrays, RNA-seq offers a number of advantages such as reducing of cross-hybridization artifacts, without needing to determine the sequence a priori, detection of low-abundance transcripts as well as reducing high background noise in the experiment (Kogenaru *et al.* 2012).

Mapping and counting: The abundance level of a transcript signifies the number of reads (fragments) mapped to the transcript. There are numerous programs that support the mapping of RNA-seq reads to reference sequences. For organisms that contain genome and annotation information, package HTSeq (Anders *et al.* 2015), featureCounts (Liao *et al.* 2014) or Tophat (Trapnell *et al.* 2012) are used. For non-model organisms bowtie (Langmead *et al.* 2009)/bowtie2 (Langmead & Salzberg 2012), for example, are widely used tools. The amount of reads aligned to each gene (contig) is counted. Inevitably the raw read counts contain artifacts

and errors. Therefore it is necessary to normalize the raw counts to get convincing gene expression level.

Normalization: There are several popular normalization methods such as RPKM (Mortazavi *et al.* 2008), TPMs, DESeq (Anders & Huber 2010) and TMM (Robinson & Oshlack 2010). RPKM/FPKM (reads/fragments per kilobase per million mapped reads) is used for reducing feature-length and library-size effects within samples. TPMs (transcripts per million) can be used to convert into RPKM/FPKM, but it is more suitable to reduce the errors between samples (Conesa *et al.* 2016). DESeq and TMM (trimmed Mean of M values implemented in package edgeR) (Robinson *et al.* 2010) are the most cited normalization and estimation programs in the study of available expression analysis methods in 2015 by Khang&Lau (Khang & Lau 2015).

Differential expression analysis: By applying statistical methods, differentially expressed genes (DEG) can be identified by comparing the normalized expression level of transcripts among samples. The RNA-seq data follows a positive and skewed discrete distribution, a Poisson distribution or a negative binomial distribution as models are chosen to account for the biological and technical variability in DESeq and edgeR (Rau *et al.* 2015). In this thesis, DESeq and edgeR have been used in manuscript 3.1 and manuscript 3.3, respectively.

In the context of this work, RNA-seq is mainly used to identify/characterize transcriptomes, classify protein families and quantify the expression of genes. This thesis focuses on the computational analysis of transcriptomes from three non-model insect species of leaf beetles.

1.2 Phylogenetic analysis

Phylogenetic trees show likely evolutionary relationships among the molecular sequences. Taking advantage of the cluster of functionally characterized sequences, the researcher is able to estimate the potential function of the sequence which is investigated. Based on the relatively low cost and the easy applicability, it has become popular in the study of non-model organisms.

To generate a phylogenetic tree there are many tools such as BioNJ applying neighbour-joining (NJ) algorithm (Gascuel 1997), PhyML and RAxML estimating maximum likelihood (ML) phylogenies (Guindon *et al.* 2010; Stamatakis 2014), and MrBayes using Bayesian estimation of phylogenies (Ronquist & Huelsenbeck 2003) or the software MEGA, PAUP* and PHYLIP using multiple methods for phylogenetic analysis (Felsenstein 1993; Swofford 2003; Tamura *et al.* 2013). In this thesis RAxML was used to reconstruct a phylogenetic tree with best fit of the evolutional model (see details in the manuscript).

1.3 Investigated species

Currently leaf beetles (family Chrysomelidae) are considered an appropriate model taxon for studying insects' adaptation strategies to host plants and their chemical defenses (Boland 2015). In particular, some species of the leaf beetle subtribe, Chrysomelina, were observed to use sequestered phytochemicals for their own chemical defense. Their larvae have nine pairs of defensive dorsal glands on their back in which the deterrent compounds are synthesized and stored. In the case of predator attack, the larvae release these defensive secretions to the top of the glandular tubercles to overawe the natural enemies (Figure 3). These compounds are *de novo* synthesized by the larvae of ancestral leaf beetles (Oldham *et al.* 1996), or sequestrated from host plants by the more advanced leaf beetles (J.M. Pasteels 1983), or both (Figure 4).

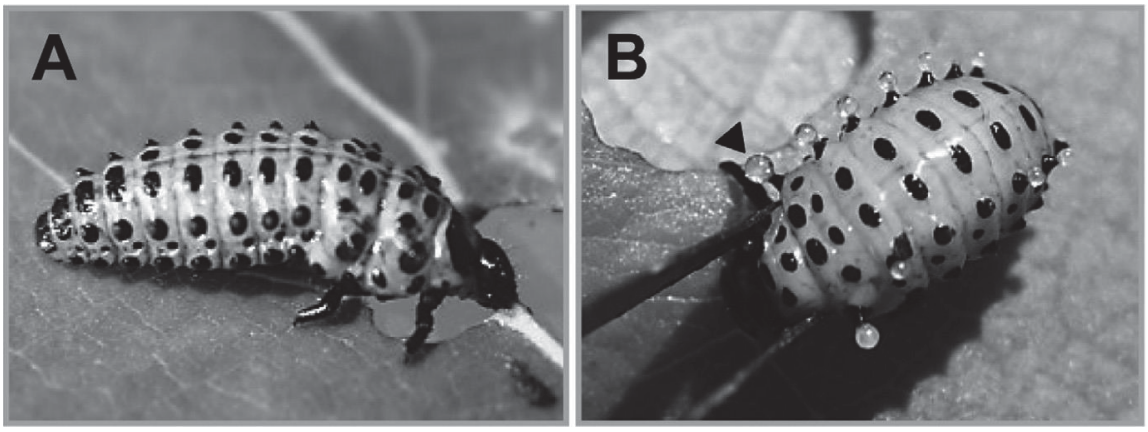

Figure 3. Third instar larva of leaf beetles.
A. Feeding on host plant. **B.** Releasing of defensive secretion after beeing attacked.

In this thesis I focus on three leaf beetles species, *Phaedon cochleariae*, *Chrysomela populi* and *Chrysomela lapponica*. The group of Chrysomelina such as the well-studied *P. cochleariae* is characterized by *de novo* producing iridoids (Veith *et al.* 1994; Veith *et al.* 1997). In the larvae deterrent compounds, iridoid monoterpenes are autogenous synthesized in fat body tissue and the defensive glands independently of the host plant chemistry. The larvae of more advanced species have developed the ability to sequester precursors from their host

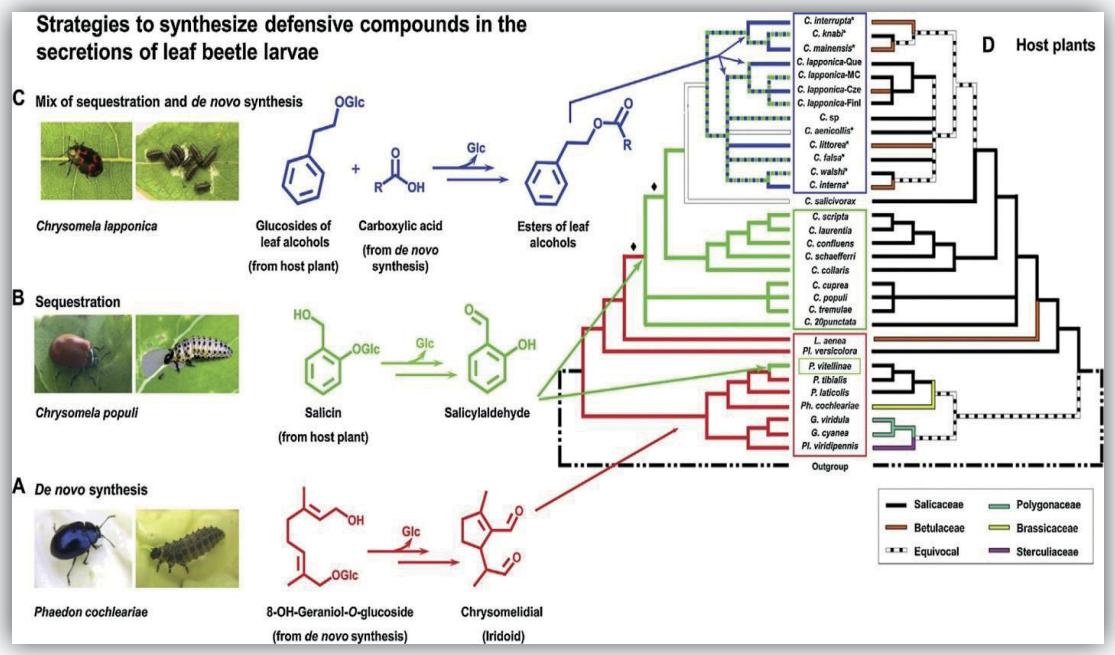

Figure 4. Maximum parsimony reconstruction of the evolution of the taxon *Chrysomelina* considering the synthesis of deterrents in the defensive glands of the larvae (**A**—**C**) and the affiliations of host plants (**D**). Adapted from Termonia et al. (2001).

plants for defensive purpose. For example *C. populi* consumes phenolic glucosides (e.g., salicin) from poplar and modifies them to volatile salicylaldehyde. To avoid predatory attack and microbial infestation the glucosidically bound precursor of the repellents is transported through the body until it reaches the defensive gland. The strategy of sequestration of plant-derived secondary metabolites makes insects' anti-predator defense more economical. A mixed strategy of *de novo* and sequestration is observed in species *C. lapponica* (Gross *et al.* 2004b).This leaf beetle shifted its host plant from willow to salicin-free birch. The change of the food source provides them with a competition-free ecological niche and allows them to avoid

specialized parasitoids and predators that are attracted by salicylaldehyde (Gross *et al.* 2004b; Stamp 2001; Termonia *et al.* 2001).

Adaptation of host plants has great influence on the synthesis of defensive compounds in larvae. The different precursors of toxic defensive deterrents have to pass through the glandular membrane into the reservoir in which they are catalyzed. To understand this process I focus on the transmembrane proteins that are a key element for the transfer of plant glucosidases. The other focus of this thesis is the adaptation of the olfactory system between *C. lapponica* feeding on willow and feeding on birch. The chemosensory system of adult beetles is responsible for the recognition of food sources and oviposition sites. The adaptation of this system for shifting host plants may play an important role to increase the survivorship particularly for the larvae, which are unable to leave the oviposition plants in their whole life cycle.

1.3.1 ATP-binding-cassette transporters (ABC)

In general, ABC transporters often consist of multiple functional domains, transmembrane domains (TMD) that are anchored to the hydrophobic plasma membrane and nucleotide-binding domains (NBD), on these domains ATP hydrolysis results in import and export of substrates such as amino acids, peptides, lipidmolecules, oligonucleotides, polysaccharides and drugs across membranes in TMD. In insects, ABC transporters are vital for physiological cellular processes (Broehan *et al.* 2013). But they also seem to be frequently implicated in resistance to insecticides (Buss & Callaghan 2008), such as DTT tolerance of the malaria agents transmitting *Anopheles mosquitoes* (Djegbe *et al.* 2014; Fossog Tene *et al.* 2013) or the tolerance against pest control toxins of *Bacillus thuringiensis* which is reported from lepidopterans (Xiao *et al.* 2016). ABC transporters represent one of the largest families of transport proteins. They carry a wide variety of molecules across membranes against a concentration gradient by utilizing ATP (Wilkens 2015). A recent study on insect's defensive research shows that ABC transporters may also play an important role for the efficient sequestration of phytochemicals throughout the larval body (Strauss *et al.* 2013). The larvae of Chrysomelina are able to accumulate defensive compounds up to 500-fold in the reservoir from

hemolymph pool (Feld *et al.* 2001; Kuhn *et al.* 2004). The precursors of defensive compounds have to pass through cell membranes to reach the gland tissue. Manuscript 3.1 describes the transcriptome analysis regarding ABC transporters in the obligate salicin sequestering species.

1.3.2 Glandular beta-glucosidase

Taking advantage of toxic compounds in the defensive glands Chrysomelina larvae deter predators and parasites. The toxic deterrent compounds are stored separately from larval body in the glandular reservoir. Either *de novo* produced or sequestrated non-toxic glucoside precursors are transported to the glandular reservoir of juveniles. In the reservoir toxic deterrent compounds are produced: deterrent iridoids in ancestral species *P. cochlearia*, salicylaldehyde in more advanced *C. populi* (J.M. Pasteels 1983) and esterified alcohols in *C. lapponica* feeding on birch (Hilker & Schulz 1994). To produce the defensive deterrence, the glucosides are activated by hydrolysis of the glucose moiety in gland tissues. This is a typical reaction for proteins of the glycosyl hydrolase family. Beta-glucosidases belong to this family. It was reported that the first activation step in the reservoir was beta-glucosidase catalyzing hydrolysis of glucoses (Laurent *et al.* 2005; Soetens *et al.* 1993). All Chrysomelina larvae possess this same metabolite process in the gland tissue. In manuscript 3.2 I identified the beta-glucosidases to make it possible to analyze their substrate selectivity regarding the adaptation between three leaf beetles and the host plants.

1.3.3 Olfactory chemosensory proteins

The insect chemosensory system is utilized to detect food, mating and oviposition sites. *C. lapponica* is an excellent model for investigating adaptation of chemosensory system after host plant shifts. Shifting host plants from salicin-rich willow to salicin-free birch in *C. lapponica* allows their larvae to escape specialized parasitoids and predators that are attracted by the salicylaldehyde (Gross *et al.* 2004b). The adaptation of the chemosensory system for the novel host plant represents the first barrier to be overcome (del Campo *et al.* 2001). There are two major chemosensory mechanisms, smell and taste. Smell is mainly mediated by the hair-like structures, called sensilla (Hallem *et al.* 2006), found in antennae, mouth parts or legs. The chemical signals from the environment go through the pores of the surface of sensilla into the

lymph until they reach the neuron membrane, where the chemical signals are transduced to electrical signals. The olfactory receptions is mediated by six protein families: The first two families are odorant binding proteins (OBPs) and chemosensory proteins (CSPs) that are small, globular, soluble proteins in sensillar lymph (Pelosi *et al.* 2014). These proteins may help in odor detection, discrimination and coding (Leal 2013). Sensory neuron membrane proteins (SNMPs) are localized in cilia and dendrites of olfactory receptor neuron (ORN) (Nichols & Vogt 2008). The odorant-OBP complexes interact with SNMPs so that they enhance the delivery of odorants to receptors. The remaining three protein families are receptor families: odorant receptors (ORs), gustatory receptors (GRs) and ionotropic receptors (IRs). ORs and GRs are members of G-protein that are anchored in membrane and form heterometric ligand-gated ion channels with their co-receptors (Sato *et al.* 2008; Sato *et al.* 2011; Wicher *et al.* 2008; Zhang *et al.* 2011).

In this thesis I focus on *C.lapponica*, which belongs to the *interrupt*-group. It forms two populations, feeding on willow and birch, respectively (Gross *et al.* 2004a). Manuscript 3.3 described the comparative analysis of expression levels of chemosensory proteins between two populations in olfactory organ (antennae) and non-olfactory organ (legs).

2 Aim of this thesis

Most leaf beetles are adapted to special host plants. They have developed a successful defensive system by using substrates from host plants. Small droplets of repellent compounds are produced in the glandular reservoirs on the back of larvae. For instance, *P. cochlearia* produced the deterrent compounds autogenously (*de novo*). *C. populi* sequestrates defensive precursors from the secondary metabolites of their host plants. A mixed strategy of *de novo* and sequestration was observed in *C. lapponica*.

To produce the deterrent product, the glucosidically bound precursors are transferred from hemolymph into the glandular reservoir. ABC transporters may play an important role to transport the precursors across the membrane against a concentration gradient by utilizing ATP. As soon as the glucoside precursors are secreted into the reservoirs, beta-glucosidase functions in the synthetic pathway of compounds in chrysomelina larvae. In *C. lapponica* shifting of host plants from salicin-rich willow to salicin-free birch benefits the larvae to avoid the natural enemies that are attracted by salicylaldehyde odor. To increase the larval survivor rate, one challenge for the adult beetles is detecting food sources and oviposition sites using smell and taste.

To study sequestration of defensive compounds and the adaptation of chemosensory proteins regarding host plant shift, the first focus of my thesis: *De novo* assembling the transcript libraries of three non-model insects (*P. cochlearia*, *C. populi* and *C. lapponica*) by applying RNA-sequencing technology without genome information. Second focus is (re)identification and characterization of ABC transporters in *C. populi,* beta-glucosidase in three leaf beetles and chemosensory proteins in willow and birch populations of *C. lapponica.* Third focus is tissue expression profiling analysis of the above investigated genes. The comparison analysis among tissues or populations facilitates me to study the high expressed genes that may play an important role in the molecular process. The phylogenetic trees were calculated to characterize the relationships and propose a function for these beetle proteins.

3 Overview of manuscripts

Manuscript 1

"Tissue-Specific Transcript Profiling for ABC Transporters in the Sequestering Larvae of the Phytophagous Leaf Beetle *Chrysomela populi*"

Anja S. Strauss, **Ding Wang**, Magdalena Stock, Rene R. Gretscher, Marco Groth, Wilhelm Boland, Antje Burse

Status: published in *PLoS ONE*, 2014, http://dx.doi.org/10.1371/journal.pone.0098637

Summary:

All 65 putative ABC transporters from *C. populi* were identified from *de novo* assembled transcript library by using RNA-sequencing technology. Aligning with other insects and human phylogeny was used as a diagnostic for functional characterization of the ABC transporters from subfamily A to H. Expression profiling of all putative ABC transporters was studied in four tissues (gut, gland, fat body, malpighian tubules). The result suggested that ABCB, C and G influence the plant metabolite absorption in the gut, ABCC with 14 members is the preferred subfamily responsible for the excretion of these metabolites *via* Malpighian tubules. A combination of tissue-specific transcriptome profiling analysis and phylogenetic analysis showed an ABCC transporter (*Cp*ABC35) in *C. populi* that had extra high expression level in glandular tissues by using R package DESeq. RNAi silencing of this transporter resulted in a lack of deterrent compounds in glandular reservoir. It suggested that *Cp*ABC35 may play a key role in defensive system in *C. populi*.

Author Contributions:

Conceived and designed the experiments: DW (30%) AB MS ASS.

Performed the experiments: RRG MG AB ASS.

Analyzed the data: DW (90%) MS ASS AB.

Contributed reagents/materials/analysis tools: WB.

Wrote the manuscript: DW (20%) ASS MS AB WB.

Manuscript 2

"Glandular beta-glucosidases in juvenile *Chrysomelina* leaf beetles support the evolution of a host-plant-dependent chemical defence"

Peter Rahfeld, Wiebke Haeger, Roy Kirsch, Gerhard Pauls, Tobias Becker, Eva Schulze, Natalie Wielsch, **Ding Wang**, Marco Groth, Wolfgang Brandt, Wilhelm Boland, Antje Burse

Status: published in *Insect Biochemistry and Molecular Biology*, 2015, http://dx.doi.org/10.1016/j.ibmb.2015.01.003

Summary:

The beta-glucosidases from three leaf beetles (*P. cochlearea*, *C. populi* and *C. lapponica* birch-feeder) were identified and characterized by comparing the *de novo* assembled transcript libraries and public databases. A wide substrate spectrum of these catalysts indicated that the specialist of birch host plant did not lead to an "evolutionary dead end". Their phylogenetic tree with other insects supported that the beta-glucosidase from three leaf beetles has a common ancestral.

Author Contributions:

Conceived and designed the experiments: PR RK AB.

Performed the experiments: PR WH RK GP TB WBr AB.

Analyzed the data: DW (10%) PR RK AB.

Contributed reagents/materials/analysis tools: WB.

Wrote the manuscript: DW (10%) PR AB WBo.

Manuscript 3

"Selective adaptation within the chemosensory system of the leaf beetle, *Chrysomela lapponica*, following host plant shift"

Ding Wang, Stefan Pentzold, Maritta Kunert, Marco Groth, Wolfgang Brandt, Jacques M. Pasteels, Wilhelm Boland, Antje Burse

Status: in preparation for *Molecular Ecology*

Summary:

Due to the lack of the genome information, the transcript libraries of *C. lapponica* feeding on willow and birch were *de novo* assembled by using RNA-sequencing technology. In total 113 putative chemosensory genes in six families (OBPs, CSPs, SNMPs, IRs, GRs and ORs) were identified and characterized in both populations applying sequences similarity among published well-studied insects and in-house datasets. The comparison of tissue expression profiling (antennae and legs) analysis between two populations of *C. lapponica* was done by using R package edgeR. Minus-C OBPs and ORs showed up-regulated in different populations. Binding affinities between homology model of minus-C OBPs and the potential chemical cues that are used to discriminate the host plants were calculated. The study showed that minus-C OBPs and ORs may play a role to tolerate the phytochemicals of novel host plants.

Author Contributions:

Conceived and designed the experiments: DW (50%) AB SP.

Performed the experiments: MK AB MG SP.

Analyzed the data: DW (90%) AB SP.

Contributed reagents/materials/analysis tools: WB.

Wrote the manuscript: DW (50%) AB SP MG BW.

4 Manuscripts

4.1 Manuscript 1: "Tissue-Specific Transcript Profiling for ABC Transporters in the Sequestering Larvae of the Phytophagous Leaf Beetle *Chrysomela populi*"

OPEN ACCESS Freely available online

PLOS ONE

Tissue-Specific Transcript Profiling for ABC Transporters in the Sequestering Larvae of the Phytophagous Leaf Beetle *Chrysomela populi*

Anja S. Strauss[1☯], Ding Wang[1☯], Magdalena Stock[1☯], René R. Gretscher[1], Marco Groth[2], Wilhelm Boland[1], Antje Burse[1]*

1 Max Planck Institute for Chemical Ecology, Beutenberg Campus, Hans-Knoell-Str. 8, D-07745 Jena, Thuringia, Germany, **2** Leibniz Institute for Age Research – Fritz Lipmann Institute, Beutenbergstr. 11, D-07745 Jena, Thuringia, Germany

Abstract

Background: Insects evolved ingenious adaptations to use extraordinary food sources. Particularly, the diet of herbivores enriched with noxious plant secondary metabolites requires detoxification mechanisms. Sequestration, which involves the uptake, transfer, and concentration of occasionally modified phytochemicals into specialized tissues or hemolymph, is one of the most successful detoxification strategies found in most insect orders. Due to the ability of ATP-binding cassette (ABC) carriers to transport a wide range of molecules including phytochemicals and xenobiotics, it is highly likely that they play a role in this sequestration process. To shed light on the role of ABC proteins in sequestration, we describe an inventory of putative ABC transporters in various tissues in the sequestering juvenile poplar leaf beetle, *Chrysomela populi*.

Results: In the transcriptome of *C. populi*, we predicted 65 ABC transporters. To link the proteins with a possible function, we performed comparative phylogenetic analyses with ABC transporters of other insects and of humans. While tissue-specific profiling of each ABC transporter subfamily suggests that ABCB, C and G influence the plant metabolite absorption in the gut, ABCC with 14 members is the preferred subfamily responsible for the excretion of these metabolites *via* Malpighian tubules. Moreover, salicin, which is sequestered from poplar plants, is translocated into the defensive glands for further deterrent production. In these glands and among all identified ABC transporters, an exceptionally high transcript level was observed only for *Cpabc35* (*Cpmrp*). RNAi revealed the deficiency of other ABC pumps to compensate the function of *Cp*ABC35, demonstrating its key role during sequestration.

Conclusion: We provide the first comprehensive phylogenetic study of the ABC family in a phytophagous beetle species. RNA-seq data from different larval tissues propose the importance of ABC pumps to achieve a homeostasis of plant-derived compounds and offer a basis for future analyses of their physiological function in sequestration processes.

Citation: Strauss AS, Wang D, Stock M, Gretscher RR, Groth M, et al. (2014) Tissue-Specific Transcript Profiling for ABC Transporters in the Sequestering Larvae of the Phytophagous Leaf Beetle *Chrysomela populi*. PLoS ONE 9(6): e98637. doi:10.1371/journal.pone.0098637

Editor: Gergely Szakacs, Hungarian Academy of Sciences, Hungary

Received January 31, 2014; **Accepted** May 5, 2014; **Published** June 2, 2014

Funding: This work has been supported by the Deutsche Forschungsgemeinschaft (http://www.dfg.de/en/index.jsp; grant number, BU 1862/2-1) and the Max Planck Society (http://www.mpg.de/en). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: aburse@ice.mpg.de

☯ These authors contributed equally to this work.

Introduction

Lipid bilayers form efficient barriers for cellular partitioning. The translocation across these membranous barriers is crucial for many aspects of cell physiology, including the uptake of nutrients, the elimination of waste products, or energy generation and cell signaling. The ATP-binding cassette (ABC) transporters constitute one of the largest families of membrane translocators [1]. The core functional unit of ABC proteins consists of four domains: two cytoplasmic domains containing the highly conserved nucleotide-binding domains (NBDs), which are responsible for the ATP hydrolysis needed to provide energy for the transport cycle, and two transmembrane domains (TMDs), each in most cases composed of six membrane-spanning helices, which impart substrate specificity and translocation [2–4]. The NBDs harbor several conserved sequence motifs from N- to C-terminus. These are the Walker A motif (also called P-loop) which is glycine-rich, a flexible loop with a conserved glutamine residue (Q-loop), the ABC signature (LSGGQ) motif (also called C-loop), the Walker B motif, and a conserved histidine residue (His-switch). The ABC signature motif is diagnostic for this family as it is present only in ABC transporters, while Walker A and B motifs are found in many other ATP-utilizing proteins. The domains are encoded by separate genes, either by genes encoding one NBD and one TMD whose products dimerize to form the functional transporter, or by genes encoding two NBDs and two TMDs on a single polypeptide.

In eukaryotic genomes, ABC genes are widely dispersed and highly conserved between species, indicating that most of these genes have existed since the beginning of eukaryotic evolution [5–8]. ABC transporters can be classified into subfamilies according to sequence homology and domain topology. The existing eukaryotic genes have been grouped into major subfamilies, termed from ABCA to ABCI [1,9]. Both subfamilies H and I are not present in humans. The subfamily ABCH was defined after the analysis of the genome of the fruit fly *Drosophila melanogaster* and was found in other invertebrates and zebrafish to date. The subfamily ABCI is limited to plants [10]. Most ABC proteins transport a wide range of compounds, either within the cell as part of a metabolic process into an intracellular compartment (e.g. endoplasmic reticulum (ER), mitochondria, and peroxisomes) or outside the cell for transport processes to other organs. In humans, the known functions of ABC transporters include cholesterol and lipid transport, multidrug resistance, antigen presentation, mitochondrial iron homeostasis and the ATP-dependent regulation of ion channels [11–14]. Owing to the importance of ABC transporters for cell functions, they are still extensively investigated in many eukaryotes. In insects, one of the best studied ABC proteins is White, which is crucial for pigment transfer in insect eyes [15–20]. As is known for *D. melanogaster*, ABC transporters facilitate translocation of attractants for germ cell migration [21] or participate in the modulation of the molting hormones' (ecdysteroids') signaling in insect tissues [22]. Furthermore, they seem to be frequently implicated in insecticide resistance [23,24], such as in the DTT tolerance of the *Anopheles* mosquitoes which transmit malaria agents [25] or in the tolerance against pest control toxins from *Bacillus thuringiensis* which is reported of lepidopterans [26,27].

Although ABC transporters were previously analyzed in several insect species at genome-wide level [28,29], profiles of the transcript levels of ABC transporters in non-model insects are not available to date. For this study we analyzed the transcriptomic data with regard to ABC transporters in a phytophagous leaf beetle species. Leaf beetles (Chrysomelidae *sensu lato*; including the seed beetles Bruchidae) constitute together with the Cerambycidae (longhorn beetles) and the Curculionoidea (weevils) the largest beetle radiation. These are known as "Phytophaga" and represent roughly 40% of all the 350,000 described species [30]. Leaf beetles mainly feed on green plant parts. The species of the leaf beetle taxon Chrysomelina, for example, are adapted to use host plants' leaves as a food source during their whole life cycle [31]. Therefore, they have to be protected against both, the noxious effect of plant secondary metabolites and attacks by their enemies. Some species evolved the ability to exploit the phytochemicals for their own chemical defense [32–34]. The larvae of the poplar leaf beetle *Chrysomela populi*, for example, take up the phenolglucoside salicin from salicaceaous food plants. This precursor salicin is transported into nine pairs of exocrine, dorsal glands [35,36], where the compound is converted into salicylaldehyde – a potent, volatile deterrent that repels predators and prevents fatal microbial infections [34,37,38]. This process of sequestration involves a complex influx-efflux transport network which guides plant-derived glucosides through the insect body [39].

Although sequestration is a widespread phenomenon attributed to many insect orders, we recently identified the first example of a transport protein essential for the translocation of phytochemicals in insects [40]. The transporter belongs to the ATP-binding cassette transporter family and functions in the defensive exocrine glands of juvenile poplar leaf beetles. Thus, the comprehensive analysis of putative ABC transporters in the phytophagous *C. populi* larvae provides implications for further studies on the predicted physiological functions of this transporter class in sequestering insects, such as the incorporation and excretion mechanisms of toxic compounds. For this reason, we present a complete inventory of ABC transporters based on available *C. populi* transcriptome sequences. Detailed sequence comparisons of members of each subfamily with those of humans, the red flour beetle *Tribolium castaneum* and other insects reveal their correspondences. We, additionally, studied the expression profiles of ABC encoding transcripts in various tissues by using next-generation sequencing in juvenile *C. populi* and propose a function of ABC pumps in the sequestration process.

Materials and Methods

No specific permissions were required for the locations/activities. *Chrysomela populi* is listed neither as endangered nor as protected species in Germany. Manuela Baerwolff from the Thueringer Landesanstalt für Landwirtschaft, Referat 430 Nachwachsende Rohstoffe, D-07778 Dornburg-Camburg, Apoldaer Str. 4 allowed us to collect the beetles in the area mentioned in the manuscript. GPS coordinates are stated in the method section.

Rearing and Maintaining of *C. populi*

C. populi (L.) was collected near Dornburg, Germany (+51°00′52.00″, +11°38′17.00″) on *Populus maximowiczii* × *Populus nigra*. The beetles were kept in a light/dark cycle of 16 h light and 8 h darkness (LD 16/8) at 18°C±2°C in light and 13°C±2°C in darkness.

RNA Isolation, Library Construction and Sequencing

Tissue samples from five *C. populi* larvae per biological replicate were collected as described by Bodemann *et al.* [41]. Total RNA was extracted from defensive glands, fat body, Malpighian tubules and gut tissue with the RNAqueous Micro Kit (Ambion, Life Technologies, Carlsbad, California, USA) according to the manufacturers' instructions with the exception of 1% (v/v) ExpressArt NucleoGuard (Amplification Technologies, Hamburg, Germany) added to the lysis buffer. The RNA integrity was validated by electrophoresis on RNA 6000 Nano labchips on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, California, USA). RNA concentrations were determined by employing a NanoView (GE-Healthcare, Chalfont St Giles, UK). Up to 5 µg of total RNA was then used for library preparation using TruSeq RNA Sample Prep Kit (Illumina, San Diego, California, USA) according to the manufacturer's description. RNA sequencing (RNA-seq) for three biological samples per prepared tissue was done using next-generation sequencing technique [42] on a HiSeq2000 (Illumina, San Diego, California, USA) in 50-bp single read mode (two or three samples multiplexed in one lane).

Pooled total RNAs from adults (two males, two females), one pupa, and nine first- to third-instar larvae were used for paired-end sequencing. Up to 5 µg of total RNA was then used for library preparation using TruSeq RNA Sample Prep Kit (Illumina Inc., San Diego, USA) according to the manufacturer's description. Afterwards, the fragmentation step during library preparation of these was set to four minutes. This library was sequenced using a GAIIx (Illumina Inc., San Diego, USA) in 150-bp paired-end mode in one sample per lane. All reads were extracted in FastQ format and used for further analysis.

De novo Assembly of *C. populi*'s Transcriptome

To obtain the transcript catalogue of *C. populi*, the paired-end reads were *de novo* assembled by applying the open source tool Trinity v2012–03–17 [43] with the following parameters: minimal

contig length of 300 bp and the paired fragment length limited to 500 bp. In order to reconstruct full-length transcripts, we used the software TGICl (vJan.2009) [44] to reassemble the transcriptome output from Trinity with a minimum overlap length of 100 bp and sequence similarity of 90 percent. A summary of these results is given in Table S1. The raw sequence data are stored in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) with the accession number SRA106166. The corresponding BioProject is PRJNA212154.

Annotation of *De novo* Assembled Transcript Library and Identification of ABC Transporters

We annotated the above mentioned transcript catalogue by translating the cDNAs of the putative transcripts into all six possible open reading frames. This was achieved by applying transeq which is part of the EMBOSS package (v6.3.1). Afterwards, the protein sequences were searched against the Pfam database (update, Jan 2013) with an e-value cut-off of 1e-5 [45,46]. 102 hits were obtained that belong to the protein family "PF00005" (ABC_tran domain). Next, we identified 12 sequences highly similar to obligate intracellular Microsporidia parasites found by BLASTx against the non-redundant protein sequence database (at NCBI). The database search revealed, for example, that the sequence named Msp1 displays 69% identity to the protein of *Nosema ceranae* (XP_002996720.1). Other tested sequences which exhibited similarities between 48% and 93% identity to members of the genus *Nosema* and other Microsporidia were considered as sequences derived from these intracellular parasites. Because the presumed parasite in *C. populi* has not yet been taxonomically classified, we specified these sequences which we have included into our phylogenetic study as Msp1–5 (for Microsporidia). In general, Microsporidia are widespread parasites also reported from Chrysomelidae [47,48], including from *Chrysomela scripta* (a close relative of *C. populi*) whose tissues were infected with *Nosema scripta* [49]. In the future, the probable infection of *C. populi* with Microsporidia and their effects on the host need to be clarified.

For the identification of NBDs in 90 ABC transporters (after the removal of the presumed mircosporidian sequences), firstly, the highly conserved NBDs of the human (ABCA-ABCG, 48 amino acid sequences of NBD) and fruit fly (ABCH, 3 amino acid sequences of NBD) ABC transporters were retrieved from GenBank (NCBI) and chosen as 'homology search targets'. Then, the long coding sequence for each annotated beetle ABC transporter was determined by using getorf of the EMBOSS tools. Afterwards, these longest coding sequences and the chosen 'homology search targets' were aligned by applying the multiple sequence alignment program MAFFT v7.01 [50] (using option E-INS-i). Transcripts containing all five motifs of NBDs with roughly 170 amino acids were kept. Secondly, the remaining ABC transporter transcripts with incomplete motifs were checked again. Their six possible protein sequences were aligned to the chosen 'homology search targets' (with the same parameter E-INS-i, MAFFT). All sequences containing at least four motifs of NBDs and having a sequence length of more than 130 amino acids were selected and added to the other sequences for further studies. These incomplete sequences might be due to the stringent settings while assembling them. For our assembly we have tested different parameter settings including lower stringency. However, we have realized that this let to contiguous sequences consisting of several transcripts that have been assembled together although they do not belong to each other. Therefore, we have decided to choose the stringent parameters which ensure obtaining unique rather than complete sequences. We have aligned our incomplete sequences with the most identical once from *T. castaneaum* and observed that the sequences from *T. castanaeum* encode complete proteins. Based on this, we assumed that our sequences might also encode complete proteins, and that only due to limitations in the *de novo* assembly we did not obtain complete coding regions. Therefore, we did not exclude these sequences from our analyses. After removal of isoforms, the resulting beetle sequences were deposited as Transcriptome Shotgun Assembly project at DDBJ/EMBL/GenBank under the accession GARF00000000. The version described in this paper is the first version, GARF01000000.

Calculation of Phylogenetic Trees

The protein sequences were aligned by the G-INS-i methods from MAFFT with default parameters. To calculate the phylogenetic tree RAxML v7.2.8 [51], a program based on maximum-likelihood inference, was used. In RAxML, the best fit model of protein evolution was RTREVF with gamma distribution for modeling rate heterogeneity. The best fit model was determined by the best likelihood score under GAMMA (perl script ProteinModelSelection.pl, which was downloaded from http://sco.h-its.org/exelixis/hands-On.html). The maximum-likelihood phylogenetic tree was reconstructed with a bootstrap test of 1000 replicates in RAxML.

For phylogenetic analysis of the ABC transporter subfamilies, we used the same methods along with sequences of *T. castaneum*, the most closely related model species to *C. populi*. The ABC transporter protein sequences of *T. castaneum* were retrieved from Broehan *et al.* [52] with the identical designations. Further, we included homologous sequences from human, *Bombyx mori*, *D. melanogaster*, *Apis mellifera*, *Culex quinquefasciatus*, *Dendroctonus ponderosae*, and from a Microsporidia species into our calculations. If not stated in the phylogenetic trees, the accession numbers of these sequences are listed in Table S2.

Expression Profiling of Putative ABC Transporter Transcripts

Each 50-bp single-read dataset of four tissues (gut, defensive glands, fat body, Malpighian tubules) contained three biological replicate samples. The raw sequence data are stored in the SRA of the NCBI with the accession numbers listed in Table S3. The corresponding BioProject is PRJNA212154.

To compare the transcript expression levels of the four tissues, we mapped the RNA-seq reads onto the (*de novo* assembled) transcriptome of *C. populi* with the open source tool Bowtie v0.12.7 [53] using default parameters. Afterwards, the R package DESeq [54,55] (which is part of the Bioconductor package [56]) was used to detect differentially expressed transcripts in the four different tissues.

Based on the Lander/Waterman equation [57], the average coverage per base in each transcript of each biological replicate was separately computed. The mean values of average coverage of each replicate for each tissue, respectively, were compared to show the expression levels of tissues (see Table S4 for normalized data). To compare these results with quantitative real-time PCR measurements, we normalized the output from DESeq to the standards *CpeIF4a* and *CpEF1alpha* (see Table S5 for the accession numbers of normalization genes), which were used in quantitative real-time PCR, as described by Livak and Schmittgen [58].

Quantitative Real-time PCR (qPCR)

Total RNA was extracted from larval tissue using an RNeasy MINI kit (Qiagen, Hilden, Germany). Complementary DNA was

synthesized from DNA-digested RNA using SuperScript III reverse transcriptase (Invitrogen, Life Technologies, Carlsbad, California, USA). Real-time PCR was performed using Brilliant II SYBR Green qPCR Master Mix (Agilent Technologies, Santa Clara, California, USA) according to the manufacturer's instructions and the Mx3000P Real-Time PCR system (Agilent Technologies, Santa Clara, California, USA). *CpeIF4a* and *CpEF1alpha* expression were used to normalize transcript quantities (see Table S5 for primer sequences). Running conditions: 3′ 94°C, 40 cycles (30″ 94°C; 30″ 60°C), melting curve with 1°C increase 60–95°C. Analyses were performed according to the MIQE-guidelines [59].

RNA Interference of *Cpabc35* (*CpMRP*) in *C. populi* Larvae

The most abundant ABC transporter derived from the glandular tissue (*Cpabc35* (*Cpmrp*)) [40] was analyzed *via* RNAi experiments. The sequence-verified plasmid pIB-*Cp*MRP was used to amplify a 730-bp fragment of *Cpabc35* dsRNA. As control, a *gfp* sequence was amplified from pcDNA3.1/CT-GFP-TOPO (Invitrogen, Life Technologies, Carlsbad, California, USA). The amplicons were subjected to *in vitro* transcription assays according to the instructions of the MEGAscript RNAi kit (Ambion, Life Technologies, Carlsbad, California, USA; see Table S5 for primer sequences). The resulting dsRNA was eluted three times with 50 µl of injection buffer (3.5 mM Tris-HCl, 1 mM NaCl, 50 nM Na_2HPO_4, 20 nM KH_2PO_4, 3 mM KCl, 0.3 mM EDTA, pH 7.0) after nuclease digestion. The quality of dsRNA was checked by TBE-agarose-electrophoresis.

First-instars of *C. populi* (3–4 days after hatching) with 3–5 mm body length (chilled on ice) were injected with 0.25 µg of dsRNA by using a nanoliter microinjection system (Nanoliter 2000 Injector, World Precision Instruments, Sarasota, Florida, USA). Injections were made into the hemolymph next to the ventral side between the pro- and mesothorax. Differential expression in the glandular tissue was analyzed 10 days after RNAi treatment by using RNA-seq. Two biological replicates (pool of glandular tissue of 3 larvae, each) compared to two biological replicates of *gfp*-control samples [42] were sequenced on a HiSeq2500 (Illumina, San Diego, California USA) in 50-bp single read mode (two or three samples multiplexed in one lane). The raw sequence data are stored in the SRA of the NCBI with the accession numbers listed in Table S3. The corresponding BioProject is PRJNA212154. All short reads again were extracted in FastQ format for further analysis.

Analysis of Differentially Expressed Genes in the Glandular Tissue of Rnai-silenced *C. populi* Larvae

The short reads (sequenced in 50 bp single-mode) from the glandular tissue of the RNAi-silenced (2 samples) as well as ds*gfp*-injected (2 samples) *C. populi* larvae were mapped onto *C. populi*'s transcriptome using Bowtie [53]. The mapping results for the ABC transporter transcripts were subjected to DESeq statistical analysis [54,55] by reading them into R statistics software. Transcript counts were normalized to the effective library size. Afterwards, the negative binomial testing was carried out to identify differentially expressed genes (DEGs). All those genes were determined as differentially expressed when having an adjusted p-value less than 0.1. From all DEGs, the annotated ABC transporters were selected and checked for co-regulation.

Results and Discussion

Identification of Putative ABC Transporters Encoded in the Transcript Catalogue of *C. populi*

In our study, we focused on the distribution of ABC transporters in the different tissues of juvenile *C. populi* to assign a function to each transcript related to a certain tissue. For this purpose, we first identified potential ABC transporters in the *de novo* assembled transcript catalogue of the poplar leaf beetle. The transcriptome sequences were translated into all possible amino acid sequences and further processed as described in the method section. As a result, we predicted 65 ABC transporters for *C. populi*. This corresponds with previous studies on insects reporting, for example, 73 ABC transporter genes in the genome of *T. castaneum* [52], 44 in *Anopheles gambiae* [29], 56 in *D. melanogaster*, 43 in *A. mellifera*, or 51 in *B. mori* [28]. The *C. populi* sequences were given temporary designations as numbered series in the form of *Cp*ABCxx (Table S4).

Phylogenetic Analysis of the Putative ABC Transporters

Based on structural and functional similarity, ABC transporters in general can be grouped into subfamilies. In order to predict the subfamilies for the 65 identified ABC proteins in *C. populi*, we used their extracted NBDs for the multiple sequence alignments and then calculated the phylogenetic tree. Similarly to other insects and eukaryotes, we were able to show a division of the predicted transporters into 8 subfamilies (A–H) (Figure 1; Table 1). Members of ABCA, ABCE/F, ABCG and ABCH form distinct branches (bootstrap value ≥75 percent). ABCH forms a sister group of ABCA. The ABCC family segregates into two groups: ABCC1 contains NBDs1 and shows a similarity to the ABCD subfamily; ABCC2 contains NBDs2 and shows a similarity to the ABCB subfamily. Among the 65 putative ABC transporters from *C. populi* we identified full, half and incomplete transporters. The distribution of domains in the sequences is shown in detail for each subfamily in Table 2 and for each sequence in Table S6.

Next, we integrated human and other insect sequences into our phylogenetic trees. This allowed us to group the putative *C. populi* ABC transporters with functionally characterized proteins and, thus, to propose a substrate for the beetles' proteins.

In the case of **subfamily A**, its members in humans are full transporters and implicated in the transport processes of phospholipids, sterols, sphingolipids, bile salts, retinal derivatives (restricted to ABCA4) and other lipid conjugates indispensable for many biological processes [11,60–63]. In insects, both full and half transporters were identified whose physiological function, however, is not yet understood [28]. In *C. populi* we predicted five transporters. According to our phylogenetic analysis, ABCA proteins segregate into one branch containing NBD1 and one branch with NBD2 (Figure 2). Human ABCAs form three groups (I, ABCA1-4, 7; II, ABCA5, 6, 8-10; III, ABCA12 and 13) which are particularly distinguishable in the NBD2 branch with bootstrap values ≥76 percent. Considering the beetles' homologs, the tree shows that the majority of *C. populi* and *T. castaneum* sequences seem to cluster to human ABCA3 which results in an expansion of group I.

The **ABCB** subfamily contains ABCB1 (MDR1/P-glycoprotein) which is the first characterized human ABC transporter to confer multidrug resistance (MDR) in cancer cells [64–66] and which has been intensively studied ever since the discovery of cross-resistances after selection with chemotherapeutics [67–71]. Later studies revealed additional ABCB transporters as MDR proteins. Besides xenobiotic extrusion (ABCB1, 5, 8) [72–74], ABCB members are also known in human biology for the

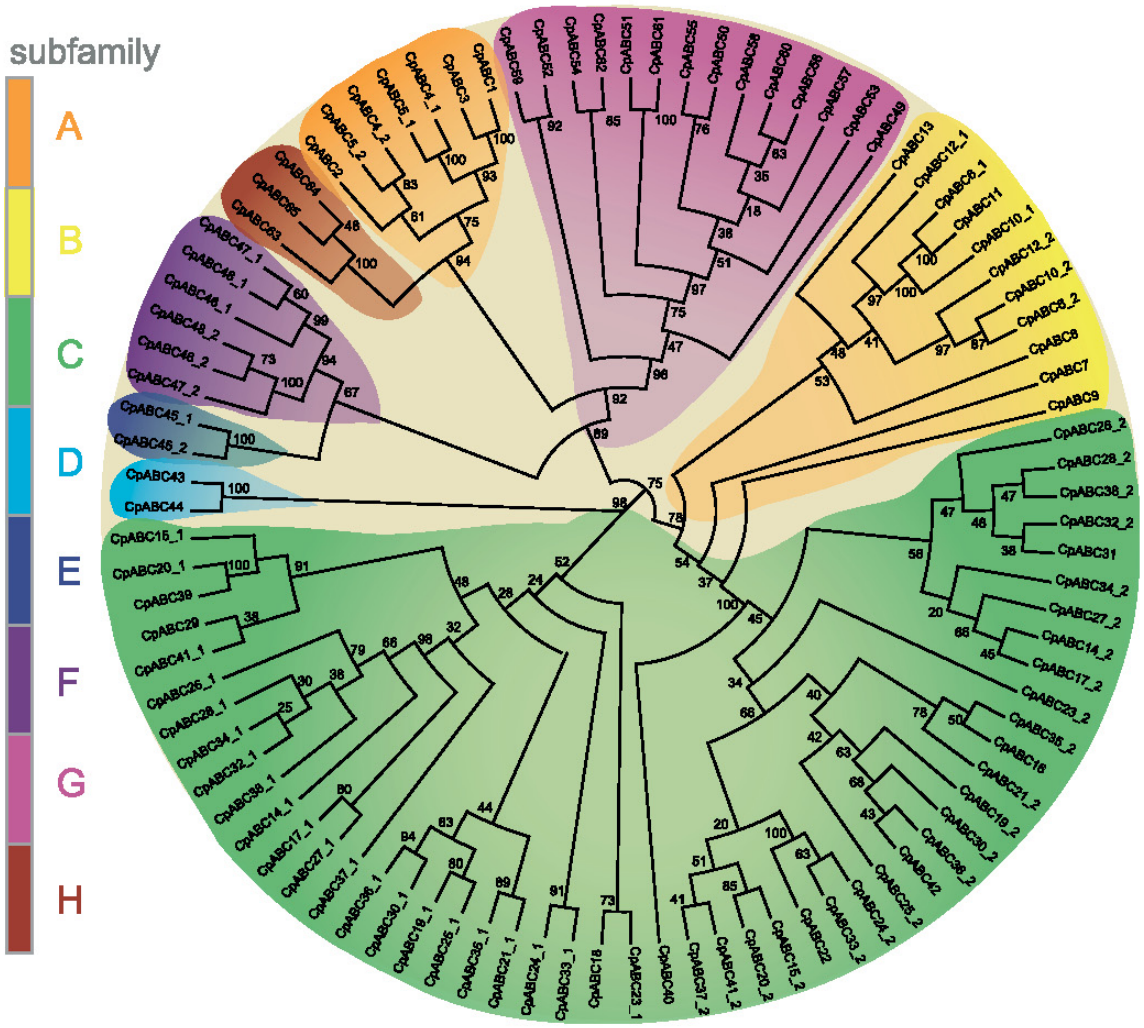

Figure 1. Eight subfamilies of 65 putative ABC transporters of *C. populi*. Some transporters contain two NBDs (NBD1 as *Cp*ABCX_1 and NBD2 as *Cp*ABCX_2), others contain only one NBD. Numbers at nodes represent bootstrap values.
doi:10.1371/journal.pone.0098637.g001

translocation, for example, of phosphatidylcholine (ABCB4) [75], bile acids (ABCB11) [76], peptides (TAP1:TAP2 (antigen processing in the adaptive immune system), TAPL, mitochondrial ABCB10) [77], metabolites of the heme synthetic pathway (ABCB6) [78], or iron (mitochondrial ABCB7 and 8) [79–82]. In insects, several examples suggest the involvement of P-glycoproteins in the resistance to insecticides used for crop protection [23,24,83–91]. However, only few P-glycoprotein-like genes have been linked to a xenobiotic substrate such as *Mdr49* and *Mdr65* of *D. melanogaster* with tolerance against colchicine and α-amanitin, respectively [92,93]. Mdr49 can act also as transporter for a germ cell attractant in fruit flies [21]. Recent studies on lepidopteran species revealed that a P-glycoprotein-like transporter mediates the efflux of cardenolides in the nerve cord and thereby prevents the interactions of these toxins with the susceptible target site of Na+/K+-ATPases [94].

Similar to other insects, the eight sequences from *C. populi* encode full and half transporters. Bootstrapping of the ABCB phylogenetic tree in Figure 3 and Figure S1 (together with ABCC) produced nodes weakly supporting segregation of this subfamily containing human and insect ABCB sequences. Based on our phylogenetic analysis, we found no homologs to TAP sequences (bootstrap value of 100 percent) in the insects, but insect homologs to the other human peptide transporters were identified. In accordance with the literature, we can also speculate that TAPL is the ancestor of the TAP family [95].

Full transporters of striking functional diversity are found in the **ABCC** subfamily. In humans thirteen ABCC members were identified, nine of which are referred to as multidrug resistance proteins (MRP) ('short' MRPs (ABCC4, 5, 11 and 12); 'long' MRPs (ABCC1, 2, 3, 6 and 10) [13,96,97]. Some ABCC members not considered as MRPs have unique functions. The

Table 1. Subfamilies of ABC transporters in eight eukaryotic species (Numbers were derived from [115]).

Species	ABCA	ABCB	ABCC	ABCD	ABCE	ABCF	ABCG	ABCH	total
S. cerevisiae	0	4	6	2	2	6	10	0	30
C. elegans	7	24	9	5	1	3	9	0	58
D. pulex	4	7	7	3	1	4	24	15	65
T. urticae	10	4	39	2	1	3	23	22	103
D. melanogaster	10	8	14	2	1	3	15	3	56
T. castaneum	10	6	35	2	1	3	13	3	73
C. populi	5	8	29	2	1	3	14	3	65
H. sapiens	12	11	12	4	1	3	5	0	48

doi:10.1371/journal.pone.0098637.t001

Table 2. Distribution of domains in the eight ABC transporter subfamilies of *C. populi*.

	Full-trans	Half-trans	2*NBD	2*NBD+1*TMD	1*NBD	1*NBD+2*TMD	total
ABCA	2	1	0	0	2	0	**5**
ABCB	3	5	0	0	0	0	**8**
ABCC	18	4	0	3	3	1	**29**
ABCD	0	2	0	0	0	0	**2**
ABCE	0	0	1	0	0	0	**1**
ABCF	0	0	3	0	0	0	**3**
ABCG	0	12	0	0	2	0	**14**
ABCH	0	3	0	0	0	0	**3**
total	23	27	4	3	7	1	**65**

Full-trans, full transporters; Half-trans, half transporters; NBD, nucleotide-binding domain; TMD, transmembrane domain; 2*NBD+1*TMD, two NBDs and one TMD (example).
doi:10.1371/journal.pone.0098637.t002

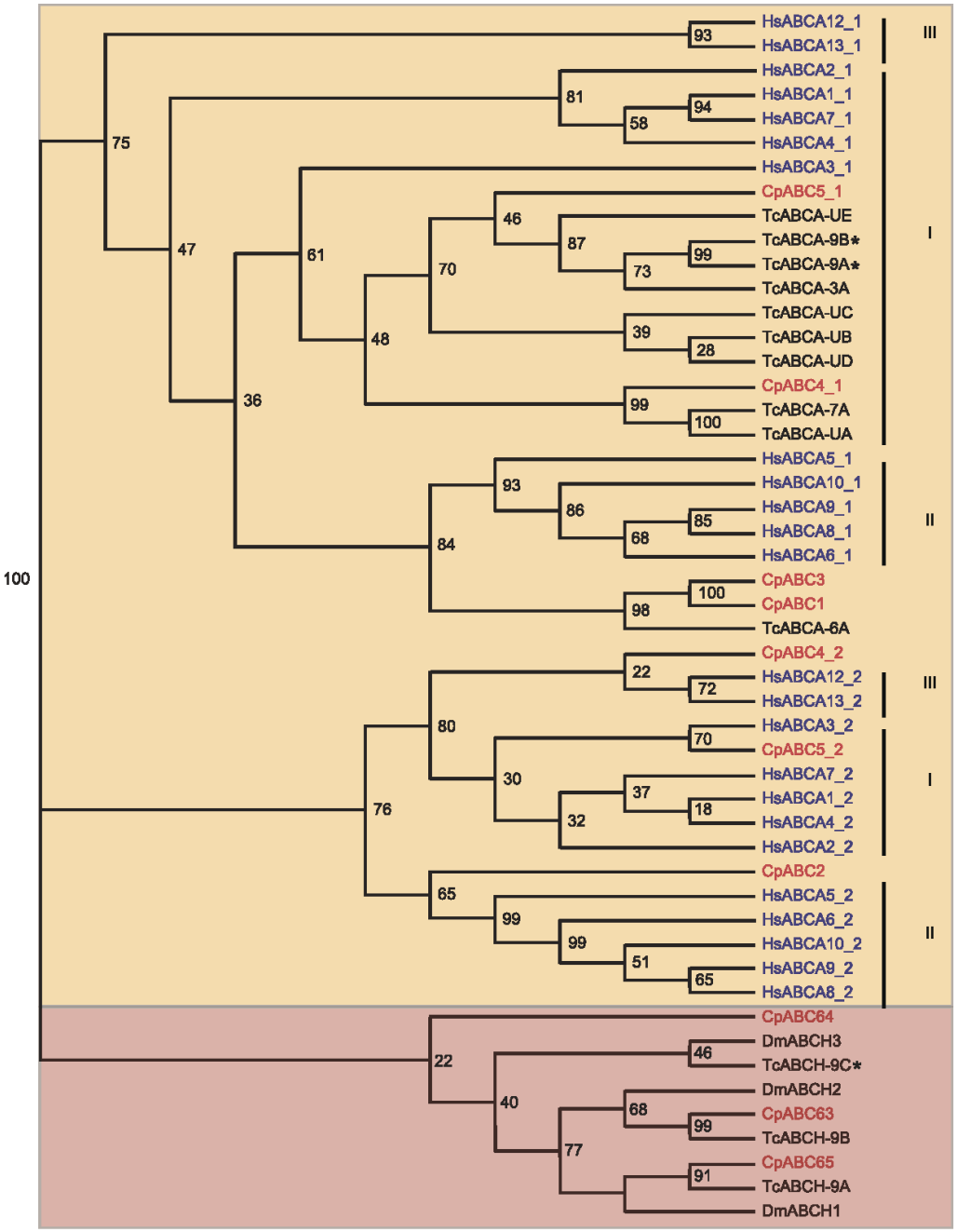

Figure 2. Phylogenetic tree of subfamilies ABCA (orange) and ABCH (pink). Some transporters contain two NBDs (NBD1 as *Cp*ABCX_1 and NBD2 as *Cp*ABCX_2), others contain only one NBD. Red, *C. populi* (Cp); blue, *H. sapiens* (Hs); black, *T. castaneum* (Tc), *D. melanogaster* (Dm). *, *T. castanaeum* with phenotype after RNAi. Numbers at nodes represent bootstrap values.
doi:10.1371/journal.pone.0098637.g002

cystic fibrosis transmembrane conductance regulator (CFTR/ABCC7), for example, functions as an epithelial ATP-gated chloride channel [98,99]. ABCC8 and ABCC9 are assembled as sulfonylurea receptors (SUR) into ATP-sensitive K^+ channels and are coupled to the gating mechanism of the ion-conducting pore [100]. In insects, ABCC members are thought to be involved in

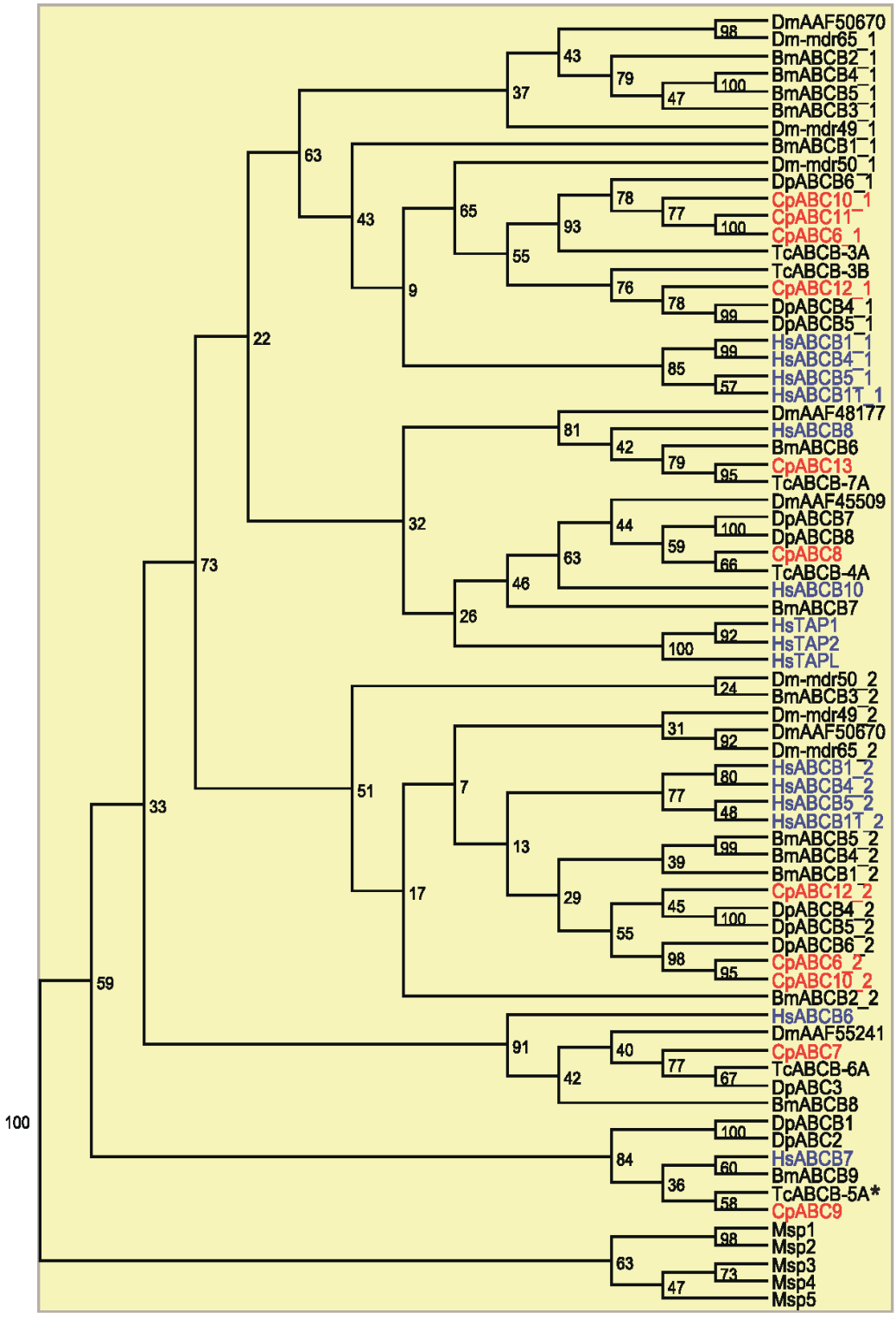

Figure 3. Phylogenetic tree of subfamily ABCB. Some proteins contain two NBDs (NBD1 as *Cp*ABCX_1 and NBD2 as *Cp*ABCX_2), others contain only one NBD. Red, *C. populi* (Cp); blue, *H. sapiens* (Hs); black, *T. castaneum* (Tc), *D. ponderosae* (Dp), *D. melanogaster* (Dm), *B. mori* (Bm); Microsporida (Msp). *, *T. castaneum* with phenotype after RNAi. Numbers at nodes represent bootstrap values.
doi:10.1371/journal.pone.0098637.g003

the translocation of xenobiotics and phytochemicals [23–26,40,91,101]. As observed in the red flour beetle *T. castaneum* and the spider mite *Tetranychus urticae*, the ABCC subfamily in *C. populi* with 29 putative members has undergone an expansion (Table 1). In our phylogenetic tree, the NBDs1 and NBDs2 form distinct branches (bootstrap value of 100 percent; Figure S1). The human 'short' MRPs ABCC5, 11, and 12 are clearly separated from all other tested sequences (Figure 4). The vast majority of insect sequences cluster together with human CFTR, SURs and multidrug resistant proteins, such as 'long' MRPs and ABCC4 implying a broad substrate spectrum of these proteins (Figure 4 and S1). Into this group falls also *Cp*ABC35 (*Cp*MRP) which is known to translocate phytochemicals including salicin. A substrate for any other insect homolog in this group has not been determined to date.

Members of the **ABCD** subfamily are involved in the translocation of fatty acids into peroxisomes [102]. The ABC transporters are half-size and assemble mostly as a homodimer after posttranslational transport to peroxisomal membranes. ABCD4 is not a peroxisomal membrane protein but an ER-resident protein that mediates translocation of lipid molecules essential for lipid metabolism in the ER [103]. Unlike humans but like all other sequenced insects *C. populi* contains two ABCD half transporters. Because they are homologous to the human peroxisomal and *T. castaneum* transporters, a similar function can be inferred in poplar leaf beetles. No insect sequence could be grouped to ABCD4 (Figure S2).

The **ABCE** and **ABCF** proteins comprise a pair of linked NBDs but lack TMDs. Therefore, they are not involved in molecule transport, but they are active in a wide range of other functions pivotal for cell viability. For example, the human ABCE1 not only acts as a ribonuclease L inhibitor, it also regulates RNA stability, viral infection, tumor cell proliferation, anti-apoptosis, translation initiation, elongation, termination, and ribosome recycling [104]. In *D. melanogaster*, the ABCE homolog Pixie plays a catalytic role in the assembly of protein complexes required for translation initiation [105]. All genomes of multicellular eukaryotes analyzed to date possess one ABCE gene [106]. In the transcript catalogue of *C. populi*, one complete ABCE protein has been predicted. The NBDs of *Cp*ABC45 are highly conserved with the respective NBDs of the human ABCE1 and *T. castaneum Tc*ABCE-3A (Figure S2). Among the subfamily ABCF involved in translation initiation and elongation in humans [106], we found in *C. populi* three putative members each with two NBDs that are highly similar to the transporters of human and *T. castaneum* suggesting functional proteins used in similar physiological processes in the cell.

The **ABCG** subfamily in humans is comprised of five half transporters. While the homodimer ABCG2 is a multidrug transporter with a wide substrate specificity [72], the homodimers ABCG1 and ABCG4 and the heterodimer ABCG5:ABCG8 translocate cholesterol and other sterole derivatives [107–110]. In insects, ABCG transporters are essential for the translocation of ommochromes for the pigmentation of eyes and body coloration. In *D. melanogaster*, for example, the half transporter White forms heterodimers with Scarlet or Brown, each of which is responsible for the transport of another type of ommochrome precursor to pigment granules [15–17]. In silkworms, White-orthologs (Bm-ok) are responsible for the translocation of uric acid for accumulation in urate granules in epidermal cells, resulting in opaque white coloration of the larval skin [20,111]. In *D. melanogaster*, E23 encodes a transporter capable of modulating the ecdysone response with consequences for the circadian transcription of clock genes [22,112].

The phylogenetic analysis revealed that the majority of the chosen insect sequences, including predicted ABCG proteins from *C. populi*, cluster together with the human ABCG1 and ABCG4 (bootstrap value of 98 percent) (Figure 5). Several insect ABCG candidates form a branch with the human ABCG5:ABCG8. Also E23 from *D. melanogaster* clusters in this branch together with *Tc*ABCG-8A. Silencing of *Tc*ABCG-8A resulted in molting defects, premature compound eye development, aberrant wing development and lethality, suggesting a function in the regulation of ecdysteroid-mediated effects [52]. Because *Cp*ABC49 is homologous to *Tc*ABCG-8A and *Dm*-E23, it allows the expectation of a similar function for this protein in *C. populi*. In addition, the insect ABCG proteins (White, Brown, and Scarlet) involved in the transfer of ommochrome precursors form a separate branch (bootstrap value of 93 percent). In accordance with the observation of *T. castaneum* [52], in *C. populi* a Brown ortholog is also missing. Interestingly, not a single analyzed insect sequence clusters with the human multidrug efflux transporter ABCG2.

The transporters of the **ABCH** subfamily are missing in humans [6,8]. To date, the only vertebrate member has been found in zebrafish [113,114]. All other ABCH proteins have been identified from invertebrate species [29,52,115–117]. The ABCH subfamily of *C. populi* includes three putative ABC transporters that are highly similar to those of *T. castaneum* (Figure 2).

RNA-seq Analyses Reveal Tissue-specific Expression of ABC Transporters in Juvenile *C. populi*

To link the above suggested functions for the *C. populi* ABC proteins to those which are differentially expressed in the larval tissues of *C. populi*, we carried out a comprehensive transcriptome sequencing of different tissues dissected from the poplar leaf beetle. All raw sequence data (in the following called reads) are listed in Table S3 and S4. The resulting expression patterns of all identified ABC transporters in intestinal tissue, Malpighian tubules, fat body and defensive glands is depicted in Figure 6. It shows that among the 65 predicted ABC transporters, 43 are expressed at least in one of the tested tissues with more than 25 normalized read counts per base (25-fold sequence coverage). As previously demonstrated in the literature [118,119], evaluation of the RNA-seq data with quantitative real-time PCR data shows also in our experiments the comparability of the two methods (Figure S3).

Five transcripts were found to be abundant in all tested tissues which suggest their essential role in cellular processes. Among them is, for example, *Cp*ABC4 which was classified as member of the ABCA subfamily. According to our phylogenetic analysis, the closest human homologs, which are involved in lipid translocation, are clustered into group I of the NBD1 branch (Figure 2). Although the NBD2 of *Cp*ABC4 clusters to ABCA12 and 13, the sequence comparison (using BLAST) of the complete sequence supported the homology of *Cp*ABC4 to human ABCA members of group I. Additionally, *Cp*ABC44 an ABCD candidate was highly expressed in all larval tissues, as well. It is homologous to the human ABCD1 and 2 and, therefore, presumably linked to the transport of very long chains of fatty acids in peroxisomes (Figure S2) [102]. Furthermore, we detected in all larval tissues abundantly expressed transcripts encoding soluble ABC proteins: *Cp*ABC45 as a member of the ABCE and *Cp*ABC47 and *Cp*ABC48 members of the ABCF subfamily. Also, in the red flour beetle, the *Tcabce-3a* and *Tcabcf-2a* transcripts were abundant throughout all life stages and highly abundant in the adult intestinal/excretory tissues and carcass [52]. Furthermore, the silencing of *Tcabce-3a* as well as *Tcabcf-2a* resulted in growth arrest and mortality of the beetles. Thus, ABCE and ABCF proteins are

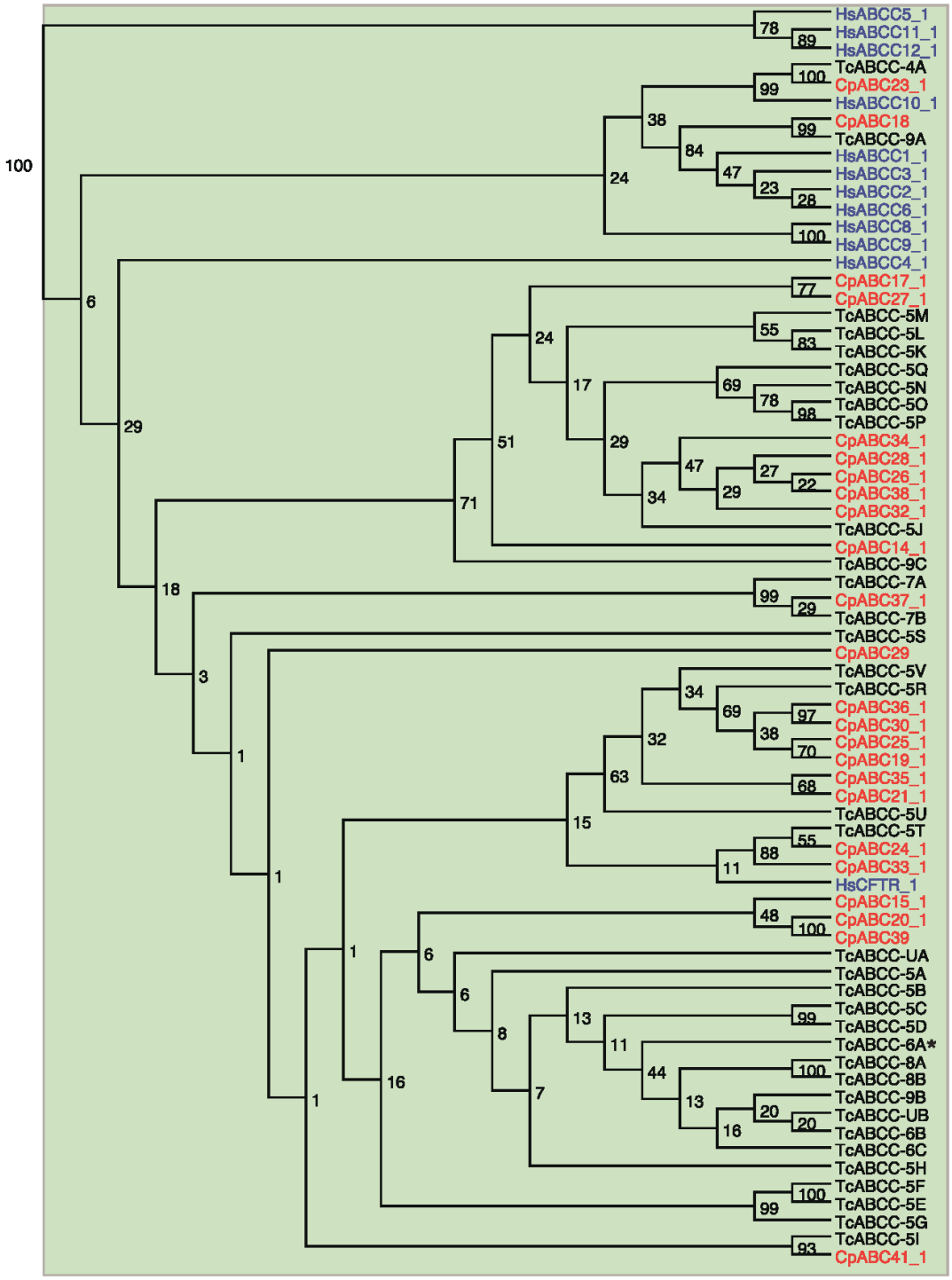

Figure 4. Phylogenetic tree of NBDs1 from subfamily ABCC. Red, *C. populi* (Cp); blue, *H. sapiens* (Hs); black, *T. castaneum* (Tc). *Hs*ABC8 and 9 are sulfonylurea receptors (SUR); *, *T. castaneum* with phenotype after RNAi. Numbers at nodes represent bootstrap values.
doi:10.1371/journal.pone.0098637.g004

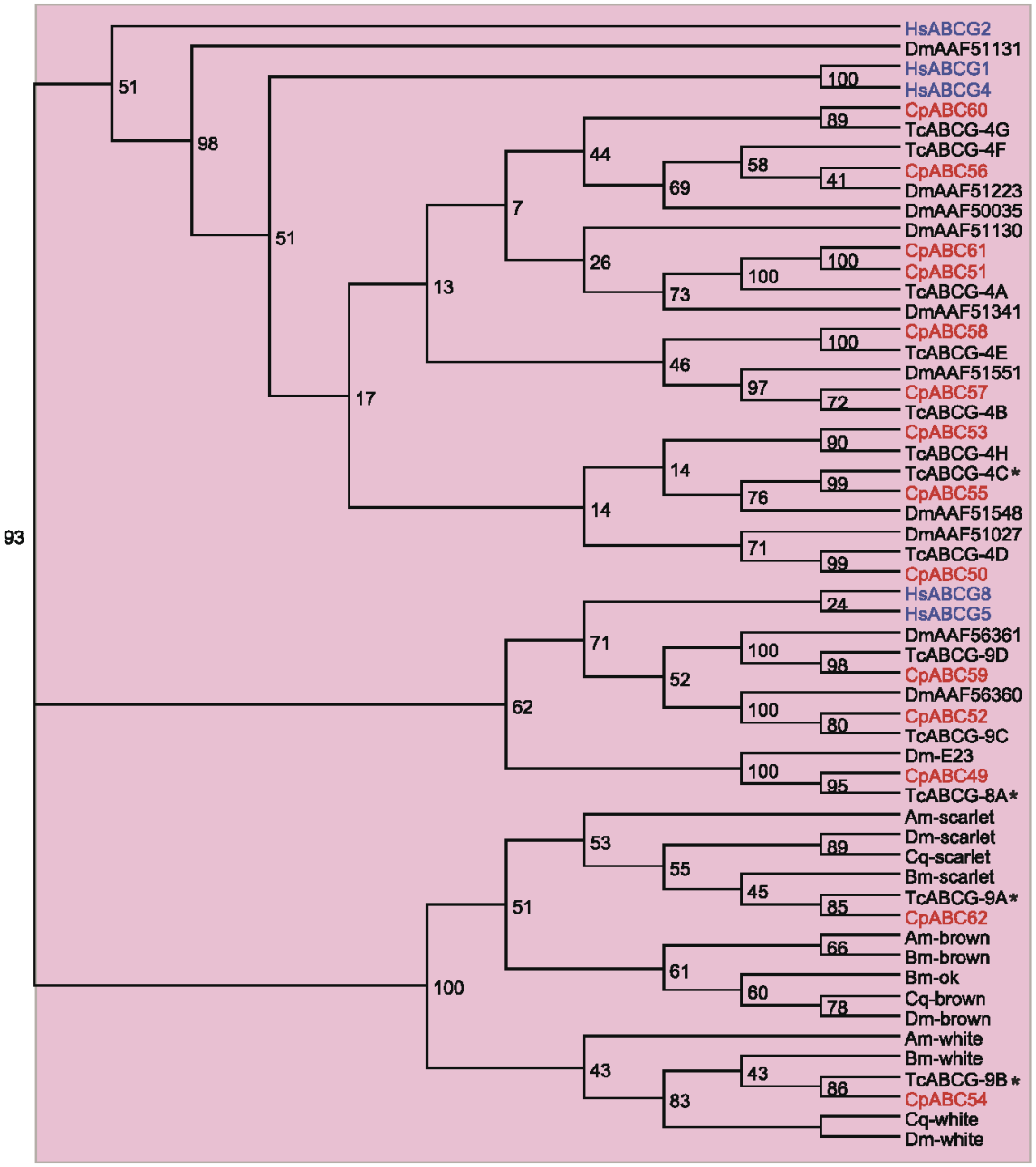

Figure 5. Phylogenetic tree of subfamily ABCG. Red, *C. populi* (Cp); blue, *H. sapiens* (Hs); black, *T. castaneum* (Tc), *D. melanogaster* (Dm), *B. mori* (Bm), *A. mellifera* (Ap), *C. quinquefasciatus* (Cq). *, *T. castanaeum* with phenotype after RNAi. Numbers at nodes represent bootstrap values.
doi:10.1371/journal.pone.0098637.g005

essential for cellular functions in all insect tissues including initiation of translation [104,105] and ribosome biogenesis [120].

In the following, we describe differential expression of putative ABC transporters in the different larval tissues:

Gut tissue. We found 17 transcripts abundant in the intestinal tissue of *C. populi* encoding members of the following subfamilies: one sequence of ABCA, two of ABCB, eight of ABCC, five of ABCG (Figure 6). The existence of ABC transporters in the gut influences the absorption and bioavailability of nutrients, water, ions and plant derived compounds. The predicted ABCA subfamily sequence *Cp*ABC5 exhibits a high mRNA level only in the gut tissue. Its deduced protein clusters

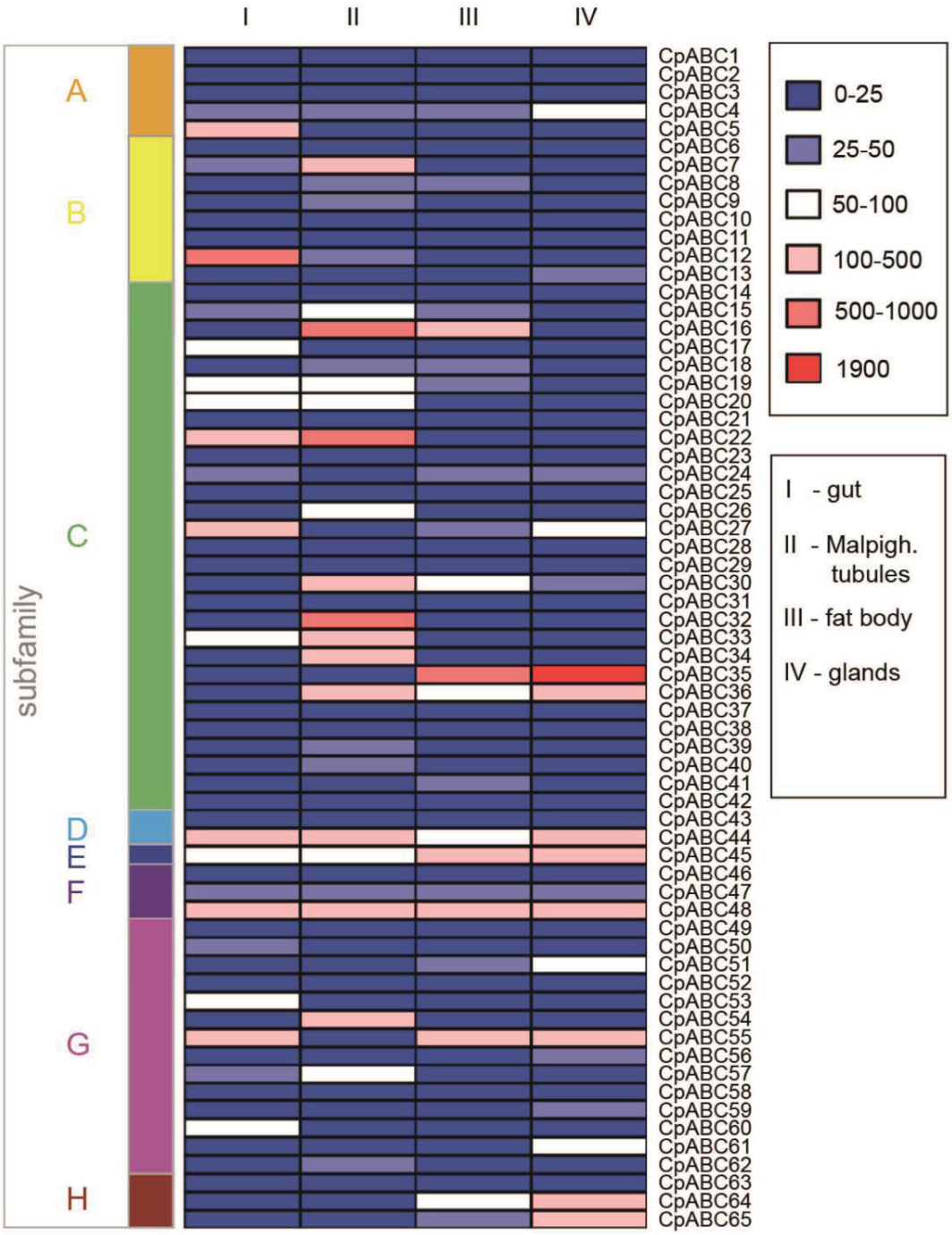

Figure 6. Colormap of the expression profiles of the 65 putative ABC genes of *C. populi*. Values are shown for four different tissues of larvae: gut, Malpighian tubules, fat body and glands. Counts of RNA-seq reads (derived from three replicates for each tissue) normalized to the effective library size and to the length of the corresponding sequences (see Table S2 for data). Expression levels are illustrated by a six grade color scale representing the sequence coverage for each transcript for each tissue, respectively.
doi:10.1371/journal.pone.0098637.g006

together with *Tc*ABCA-9A/B of *T. castaneum* (Figure 2). The silencing of these two red flour beetle genes resulted in high mortality and severe defects in wing and elytra development, depending on the developmental stage of treatment. This indicates an essential function for cell physiology, but a ligand has not been identified for these proteins to date [52]. The closest homolog in humans is ABCA3 which is related to phospholipid transfer but also to the modulation of cell susceptibility to chemotherapy of tumors [11,121,122]. Thus, *Cp*ABC5 may have a special function in this tissue, in addition to a role in lipid trafficking.

The highest transcript level of ABC transporters in the intestinal tissue was detected for *Cpabc12* which was classified into the subfamily B. It is also expressed in Malpighian tubules but ten times less. *Cp*ABC12 is a full transporter, and possesses most likely homology to human ABCB1 (MDR1, P-glycoprotein), 4, 5 and 11 as well as the *D. melanogaster* Mdr50 (bootstrap value of 63 percent) (Figure 3). Though ABCB4 acts in humans as a transporter for phospholipids in the liver [75], it is involved in the zebrafish's cellular resistance to noxious chemicals [123]. Except for ABCB11, which is a bile salt transporter [76], all the homologous vertebrate ABCB members can confer multidrug resistance [72–74]. We hypothesize a function in the translocation of phytochemicals for *Cp*ABC12 in the gut of *C. populi*. *Cpabc7* is the second ABCB candidate with a high expression level in the gut, albeit not as high as *Cpabc12*. Moreover, *Cpabc7* is 3 times more highly expressed in the Malpighian tubules than in the gut. *Cp*ABC7 is most homologous to the human ABCB6 (Figure 3) which is reported to be localized in the Golgi apparatus [124], mitochondria, plasma membranes [125], vesicular structures [126], or endolysosomes and lysosomes of cells [79]. ABCB6 is discussed to play a role in the heme metabolism [78,127–129], in the drug and arsenite resistance of cells [130–132].

All eight ABCC candidates highly expressed in the gut tissue cluster together with human CFTR, SURs, 'long' MRPs and ABCC4. This implies a broad substrate spectrum for these insect transporters which, however, cannot be specified further from our phylogenetic analysis (Figure 4, S1).

All five ABCG candidates highly expressed in the larval gut tissue cluster together with the human ABCG1 and ABCG4. These proteins are involved in sterol homeostasis. Among these *C. populi* ABCG sequences, *Cpabc55* showed the most elevated transcript level. Its deduced protein is homologous to *Tc*ABCG-4C (Figure 5) whose involvement in the transport of lipids to the cuticle has been suggested and, thus, that it is required for the formation of a waterproof barrier in the epicuticle [52]. *Cpabc55* is also highly expressed in glands and fat body tissue but not in the Malpighian tubules. The expression of *Tcabcg-4c* was higher in intestinal/excretory tissues than in carcass tissue [52]. The function of the other four ABCG transporters cannot be predicted from our analyses. However, it has been demonstrated recently that an ABCG1-homolog in the fungus *Grossmannis clavigera* confers tolerance to monoterpenes which contributes to the fungus' ability to cope with the chemical defence of its host plant [133]. Therefore, the ABCG proteins' specificity in insects may not be limited to sterols or lipids but may have a broader substrate spectrum - that is not known to date. Besides trafficking of physiological substrates, the identified ABC transporters (particularly the ABCC candidates together with the ABCB *Cp*ABC12, and the ABCGs *Cp*ABC50, 53, 60) in the gut tissue may also play a critical role in regulating the absorption of plant secondary metabolites or influence the effectiveness of pesticides in the phytophagous *C. populi*.

Malpighian tubules. Insect Malpighian tubules are critical for osmoregulation. Moreover, the tubules have the capability to excrete actively a broad range of organic solutes and xenobiotics, such as insecticides. Recently, we have shown a role of the excretion system in the homeostasis of phytochemicals in the larval body of leaf beetles [39]. Additionally, the tubules play a significant role in immunity by sensing bacterial infections and mounting an effective killing response by secretion of antimicrobial peptides [134]. We found 21 transcripts abundant in the Malpighian tubules of *C. populi* encoding members of the following subfamilies known to contain multidrug resistance proteins: four of ABCB, 14 of ABCC, three of ABCG (Figure 6).

Among the four predicted ABCB members displaying a high mRNA level in the Malpighian tubules, two, *Cp*ABC7 and 12, were already described in the gut section above. The third candidate, *Cp*ABC8, is most similar to human mitochondrial ABCB10 (Figure 3). For ABCB10 different roles have been suggested, including protection against toxic reactive oxygen species, heme synthesis, or peptide transport [77,135,136]. For this tissue, we speculate that it is involved in antimicrobial peptide transfer. The forth ABCB protein, *Cp*ABC9, clusters together with the human mitochondrial ABCB7 which is involved in the iron-sulfur cluster assembly essential for multiple metabolic pathways throughout the cell (Figure 3) [80,81]. RNAi of the homologous *Tc*ABCB-5A demonstrated the pivotal function of this gene in the red flour beetle: its down-regulation resulted in severe morphological defects and high mortality depending on the developmental stage treated [52]. Hence, the three most likely mitochondrial localized ABCB candidates, namely *Cp*ABC7-9, are proposed to be of vital importance in the cells. However, for *Cp*ABC12, which is a full transporter and clusters to human proteins related to xenobiotic resistance, we can predict a similar function in the larval excretion system.

Most putative ABC transporter transcripts identified in *C. populi* are present at a high level in the excretion system of the juvenile beetles compared to the other tissues. Particularly, the 14 candidates belonging to the ABCC subfamily are the most highly transcribed compared to other subfamilies in this tissue. Remarkably, one of the highly expressed candidates, *Cp*ABC16, clusters in our phylogeny together with *Cp*ABC35 (*Cp*MRP) which is involved in the accumulation of plant-derived metabolites (Figure S1) [40]. Therefore, it is tempting to speculate a role for *Cp*ABC16 in the excretion of phytochemicals in *C. populi* larvae.

Among the three candidates of the G-subfamily, two are highly transcribed only in the Malpighian tubules: *Cp*ABC54 is a homolog of *Tc*ABCG-9B from the White group and *Cp*ABC62 is homologous to *Tc*ABCG-9A from the Scarlet group (Figure 5). RNAi targeting *Tcabcg-9a* or *b* resulted in both cases not only in white eyes but also in a whitish appearance of the Malpighian tubules due to the absence of tryptophan metabolites/kynurenine and pteridines. These eye pigment precursors are stored and processed in the larval tubules before being released for further conversion into pigments in the developing adult eyes [137–139]. In addition, in *D. melanogaster* White is expressed in intracellular vesicles in tubule principal cells, suggesting that White participates in vesicular transepithelial transport of cGMP [140]. *Cp*ABC57 is the only ABCG candidate that is also expressed in the intestine and belongs to the human ABCG1 and ABCG4 branch (Figure 5). Taken together, the conspicuous overrepresentation of drug-resistance associated proteins, including the ABCC proteins together with the members of the subfamily B (*Cp*ABC12) and G (*Cp*ABC57), in the excretion system suggests a role for these candidates in the extrusion of xenobiotics or phytochemicals from the larval body.

Fat body. The fat body of insects is a polymorphic tissue. It performs a vast array of fundamental activities in the intermediary

metabolism and is involved in maintaining the homeostasis of hemolymph proteins, lipids, and carbohydrates [141]. Predominantly, the storage of lipid reserves in the form of glycogen and triglycerides is essential in the life of holometabolous insects, primarily in their survival of metamorphosis [142]. In humans, members of the subfamilies A, B, D and G are known to be involved in lipid transport [11,143]. In principal, we found the expression of ABC transporters in the larval fat body of *C. populi* to be low compared to the other tested tissues (Figure 6). From the ABCB subfamily, we identified in the fat body only *Cpabc8* exhibiting a low transcript level comparable to that of the Malpighian tubules. As described above, it clusters with the human mitochondrial ABCB10 which is associated with different functions, also described above, but not particularly with lipid transfer.

From ABCG we found *Cpabc51* and *Cpabc55* with high expression in the fat body. Both deduced proteins cluster to the human ABCG1 and ABCG4 (Figure 5). Only one sequence was exclusively expressed in this body part, namely *Cpabc41*, a member of the subfamily C (Figure 4). Other ABCC members which are highly expressed in this tissue are the homologous *Cpabc16* and *Cpabc35* (Figure S1). *Cp*ABC35 is known to translocate phytochemicals [40].

Noticeably, we found high expression of putative ABCH genes (*Cpabc64, Cpabc65)* in the fat body tissue. Up to now the function of this insect specific subfamily has been unclear. However, RNAi targeting *Tcabch-9c* in the flour beetle revealed a lethal, desiccated phenotype similar to the silencing of *Tcabcg-4c* mentioned above. This ABCH member also seems to be involved directly or indirectly in the transport of lipids from epidermal cells to the cuticle [52]. Based on our data we can hypothesize a role for ABC transporters in phytochemical translocation (by members of the subfamily C and the ABCG candidate), in cuticle formation (by members of the ABCH subfamily) in the fat body, but not particularly in the lipid storage of this tissue. Transporters which are important for this function might be lowly expressed and therefore not detected in our analyses.

Defensive glands. The nine pairs of defensive glands enable larvae of *C. populi* to chemically defend themselves *via* deterrent secretions. Each of these dorsal glands is composed of several secretory cells which are attached to a large reservoir. The anti-predatory effect of the secretions can be attributed to salicylaldehyde synthesized within the reservoir by a few enzymatic reactions from the pre-toxin salicin, which is sequestered from the host plant [35,37]. Recent studies have identified *Cp*ABC35 (*Cp*MRP) which is essential for the sequestration of salicin [40]. It is associated with the accumulation of the plant-derived metabolite in intracellular storage vesicles. Intriguingly, *Cpabc35* is the only predominant transcript in the defensive glands of *C. populi* (Figure 6). Its expression level lies far beyond all other ABC transporters in all tissues. There are four additional predicted ABCC proteins with high expression clustering to the human CFTR, SURs, 'long' MRPs and ABCC4, but not particularly to *Cp*ABC35 (Figure 4, S1). In *T. castaneum* another member of this group (not homologues to *Cp*ABC35; Figure 4) has been identified as playing a role in the production of secretions in odiferous stink glands [144]. The silencing of *Tcabcc-6a* (TC015346) in *T. castaneum* resulted in a strong reduction of alkenes in the secretions produced by abdominal and prothoracic glands. Although a substrate for *Tc*ABCC-6A has not been described as yet, the hypothesis can be advanced that ABC transporters functioning in the formation of secretions seem to be a widespread phenomenon in insects.

Besides ABCC proteins, members of the subfamilies B, G and H also have elevated mRNA levels in the defensive glands. *Cp*ABC13 is a member of the B-subfamily exclusively expressed in the defensive glands. It clusters particularly with the human mitochondrial ABCB8 (Figure 3). ABCB8 is known to be responsible for iron transport and doxorubicin resistance in melanoma cells *via* the protection of mitochondrial DNA from doxorubicin-induced DNA damage [145].

Among the five candidates of the ABCG also possessing a high mRNA level in the defensive glands, *Cp*ABC56, 59 and 61 are expressed only in this tissue. *Cp*ABC59 clusters to the human ABCG5:ABCG8 that pump cholesterol and other sterol derivatives, and all of the four other proteins cluster to human ABCG1 and 4, which may have a broader substrate spectrum including xenobiotics (Figure 5).

Remarkably, the expression of putative ABCH genes (*Cpabc64, Cpabc65)* was almost 3 times higher in the glandular tissue compared to the fat body tissue. Owing to this, the two ABCH proteins may have a special function as yet unknown in the defensive glands, but they may also be associated with the formation of the cuticle reservoir for storage of secretions. Furthermore, in the defensive glands there are also ABC candidates potentially associated with the translocation of phytochemicals or other xenobiotics.

RNAi with Predominant ABC Transporter – *Cpabc35* (*CpMRP*)

Conspicuously, only one ABC sequence, namely *Cpabc35,* displays an exceedingly high transcript level in the defensive glands of *C. populi.* As recently described [40], its function and key role in the sequestration of defensive compound precursors has been demonstrated. In order to test cooperative or compensation effects of other ABC genes, we performed RNAi silencing experiments for *Cpabc35.* Ten days after the injection of *Cpabc35*-dsRNA and *gfp*-dsRNA, glandular tissues were dissected and two biological replicates for each treatment were sequenced. The normalized counts of all transcripts of all samples were calculated. Thereafter, the $\log_2$ fold-changes of the silenced ABC transporter (*gfp*-injected samples as control) and adjusted p-values were determined using the DESeq package. In all samples (either in RNA-seq or quantitative real-time PCR experiments), we observed varying transcript levels corresponding to the individual biological variance and diversity despite similar developmental stage or living conditions during sample preparation.

The silencing of *Cpabc35* resulted in a significant decrease of its own transcript level (adjusted p-value (padj) = 7.31E-15). One additional ABC transporter, *Cpabc50*, belonging to subfamily G, was determined as differentially expressed (slight upregulation). In non-treated larvae, *Cpabc50* is expressed only in the gut tissue (Figure 6). Its deduced protein clusters together with the human ABCG1 and ABCG4 (Figure 5). However, *Cp*ABC50 could not compensate the function of the salicin translocation into storage vesicles, and, hence, its function remains unclear. Overall, *Cp*ABC35 is an exclusive and highly specific transporter used in the sequestration process, which explains its extraordinarily high transcript level in the defensive glands.

Conclusion

Phytophagous beetles are adapted to cope with the chemical defense of their host plant. The larvae of the poplar leaf beetle, *C. populi,* evolved the ability to sequester the plant-derived compound salicin and to use it for their own defense against their enemies. The sequestration process proceeds *via* barriers with different selectivity. While the uptake from the gut lumen into the hemolymph together with the excretion by Malpighian tubules is

non-selective, the translocation into the defensive glands is selective. In these glands two barriers must be passed: a selective membrane on the hemolymph side and a non-selective membrane on the side towards the cuticle reservoir containing the defensive secretions. Based on our analyses, we predicted specific ABC proteins that are related to the translocation of plant-derived compounds in the larvae. In the gut of *C. populi*, genes of the subfamilies A, B, C and G are predominantly expressed. Almost all of these ABC candidates have been linked in our phylogenetic trees with proteins known to be associated with xenobiotic or drug resistance and which may, therefore, contribute to the non-selective translocation into the larval hemocoel. But depending on the localization of the proteins in the intestinal cells, they may also take part in the detoxification of plant metabolites or pesticides by back-exporting them into the gut lumen. The Malpighian tubules are dominated by candidates of subfamilies B, C and G. In particular, members of the multidrug-related ABCC-group are present in great numbers in this tissue, which suggests a role in the previously observed non-selective phytochemical extrusion by the excretion system. In the defensive glands the salicin-transporting ABCC protein *Cp*ABC35 (*Cp*MRP) is extraordinarily highly expressed in comparison to the other tested tissues. It is localized intracellularly in storage compartments of the gland cells and accumulates salicin in these vesicles for further exocytosis into the glandular reservoir. *Cp*ABC35 has a broad substrate spectrum of phytochemicals and controls the non-selective barrier into the reservoir. The differential expression analysis of *Cp*ABC35-silenced defensive glands in comparison to control samples corroborated the observation that the function cannot be compensated by any other ABC transporter with overlapping substrate selectivity in this particular compartment of the glandular cells. The occurrence of other drug-resistant related ABC transporters in the defensive glands may contribute to the selectivity in the membrane of the hemolymph side of the glandular cells by extruding unused plant-derived compounds from these cells. Thus, ABC transporters are key components in the homeostasis control of phytochemicals in the sequestering poplar leaf beetle larvae.

Supporting Information

Figure S1 Phylogenetic tree of subfamilies ABCC (green) and ABCB (magenta). Some transporters contain two NBDs (NBD1 as *Cp*ABCX_1 and NBD2 as *Cp*ABCX_2), others contain only one NBD. Numbers at nodes represent bootstrap values. *C. populi* (Cp); *H. sapiens* (Hs); *T. castaneum* (Tc), *D. ponderosae* (Dp), *D. melanogaster* (Dm), *B. mori* (Bm); *Microsporida* (Msp).
(EPS)

Figure S2 Phylogenetic tree of subfamilies D (light blue), E (blue), and F (purple). Red, *C. populi* (Cp); blue, *H. sapiens* (Hs); black, *T. castaneum* (Tc), *D. melanogaster* (Dm). Star, *T. castanaeum* with phenotype after RNAi. Numbers at nodes represent bootstrap values.
(EPS)

Figure S3 Relative mRNA levels of selected putative ABC transporters in the different tissues of juvenile *C. populi*. Expression data were determined by carrying out RNA-seq (A) and quantitative real-time PCR (B) experiments (n = 3–4, mean ± SD). *CpEF1alpha* and *CpeIF4a* were used for normalization of transcript quantities. ABC gene subfamilies of *C. populi* are color-coded and grouped by their tissue-specific expression level (orange, ABCA; yellow, ABCB; green, ABCC; purple; ABCF; pink, ABCH).
(EPS)

Table S1 *De novo* assembly of the transcript catalogue of *C. populi*. The numbers of assembled transcripts and average length after assembly and reassembly show the usefulness of reassembling.
(XLSX)

Table S2 Accession numbers of homologous sequences added to *C. populi*'s ABC transporter sequences to calculate phylogenetic trees.
(XLSX)

Table S3 Overview of the raw sequence data. The table exhibits the RNA derived specimens, number of reads, sequencing technology and sequencing mode.
(XLSX)

Table S4 List of 65 predicted ABC transporters of *C. populi*. Accession numbers and sequence length of cDNAs encoding putative ABC transporters of *C. populi* and their corresponding read counts normalized to the effective library size as well as to the sequence length of all ABC transporters in the different larval tissues are presented.
(XLSX)

Table S5 Primer sets used in quantitative real-time PCR and RNAi experiments.
(XLSX)

Table S6 Predicted domain distribution in the deduced protein sequences of all identified ABC transporters of *C. populi*.
(XLSX)

Acknowledgments

We sincerely thank Angelika Berg, Ivonne Goerlich, and Christin Meißner for technical assistance. The authors would like to express their gratitude to Prof. Dr. Hans Merzendorfer from the University of Osnabrueck, Germany, for providing ABC protein sequences from *Tribolium castaneum*. We thank Manuela Baerwolff from the Thueringer Landesanstalt für Landwirtschaft, Referat Nachwachsende Rohstoffe, Dornburg-Camburg, Germany for the support of the beetles' collection. Furthermore, we wish to thank Melinda Palmer Kolb for the critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: AB MS DW ASS. Performed the experiments: RRG MG AB ASS. Analyzed the data: DW MS ASS MG AB. Contributed reagents/materials/analysis tools: WB. Wrote the paper: AB MS ASS DW WB.

References

1. Dassa E (2011) Natural history of ABC systems: not only transporters. In: Sharom FJ, editor. Essays in Biochemistry: ABC Transporters. 19 42.
2. Holland IB (2011) ABC transporters, mechanisms and biology: an overview. In: Sharom FJ, editor. Essays in Biochemistry: ABC Transporters. 1 17.
3. Jin MS, Oldham ML, Zhang QJ, Chen J (2012) Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans*. Nature 490: 566 569.
4. Al-Shawi MK (2011) Catalytic and transport cycles of ABC exporters. In: Sharom FJ, editor. Essays in Biochemistry: ABC Transporters. 63 83.
5. Dassa E, Bouige P (2001) The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. Res Microbiol 152: 211 229.
6. Dean M, Annilo T (2005) Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genom Hum Genet 6: 123 142.

7. Wang B, Dukarevich M, Sun EI, Yen MR, Saier MH (2009) Membrane porters of ATP-binding cassette transport systems are polyphyletic. J Membr Biol 231: 1 10.
8. Annilo T, Chen ZQ, Shulenin S, Costantino J, Thomas L, et al. (2006) Evolution of the vertebrate ABC gene family: analysis of gene birth and death. Genomics 88: 1 11.
9. Dean M, Rzhetsky A, Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. Genome Res 11: 1156 1166.
10. Kretzschmar T, Burla B, Lee Y, Martinoia E, Nagy R (2011) Functions of ABC transporters in plants. In: Sharom FJ, editor. Essays in Biochemistry: ABC Transporters. 145 160.
11. Quazi F, Molday RS (2011) Lipid transport by mammalian ABC proteins. In: Sharom FJ, editor. Essays in Biochemistry: ABC Transporters. 265 290.
12. Sharom FJ (2011) The P-glycoprotein multidrug transporter. In: Sharom FJ, editor. Essays in Biochemistry: ABC Transporters. 161 178.
13. Slot AJ, Molinski SV, Cole SPC (2011) Mammalian multidrug-resistance proteins (MRPs). In: Sharom FJ, editor. Essays in Biochemistry: ABC Transporters. 179 207.
14. Tamaki A, Ierano C, Szakacs G, Robey RW, Bates SE (2011) The controversial role of ABC transporters in clinical oncology. In: Sharom FJ, editor. Essays in Biochemistry: ABC Transporters. 209 232.
15. Ewart GD, Howells AJ (1998) ABC transporters involved in transport of eye pigment precursors in Drosophila melanogaster. In: Ambudkar SV, Gottesman MM, editors. Methods Enzymol. 213 224.
16. Mackenzie SM, Brooker MR, Gill TR, Cox GB, Howells AJ, et al. (1999) Mutations in the white gene of *Drosophila melanogaster* affecting ABC transporters that determine eye colouration. Biochim Biophys Acta 1419: 173 185.
17. Ohare K, Murphy C, Levis R, Rubin GM (1984) DNA-sequence of the white locus of *Drosophila melanogaster*. J Mol Biol 180: 437 455.
18. Quan GX, Kanda T, Tamura T (2002) Induction of the white egg 3 mutant phenotype by injection of the double-stranded RNA of the silkworm white gene. Insect Mol Biol 11: 217 222.
19. Sumitani M, Yamamoto DS, Lee JM, Hatakeyama M (2005) Isolation of white gene orthologue of the sawfly, *Athalia rosae* (Hymenoptera) and its functional analysis using RNA interference. Insect Biochem Mol Biol 35: 231 240.
20. Tatematsu K-i, Yamamoto K, Uchino K, Narukawa J, Iizuka T, et al. (2011) Positional cloning of silkworm white egg 2 (w-2) locus shows functional conservation and diversification of ABC transporters for pigmentation in insects. Genes Cells 16: 331 342.
21. Ricardo S, Lehmann R (2009) An ABC transporter controls export of a *Drosophila* germ cell attractant. Science 323: 943 946.
22. Hock T, Cottrill T, Keegan J, Garza D (2000) The E23 early gene of *Drosophila* encodes an ecdysone-inducible ATP-binding cassette transporter capable of repressing ecdysone-mediated gene activation. Proc Natl Acad Sci U S A 97: 9519 9524.
23. Buss DS, Callaghan A (2008) Interaction of pesticides with p-glycoprotein and other ABC proteins: A survey of the possible importance to insecticide, herbicide and fungicide resistance. Pestic Biochem Physiol 90: 141 153.
24. Labbe R, Caveney S, Donly C (2011) Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera. Insect Mol Biol 20: 243 256.
25. Jones CM, Toe HK, Sanou A, Namountougou M, Hughes A, et al. (2012) Additional selection for insecticide resistance in urban malaria vectors: DDT resistance in *Anopheles arabiensis* from Bobo-Dioulasso, Burkina Faso. PloS One 7.
26. Gahan LJ, Pauchet Y, Vogel H, Heckel DG (2010) An ABC transporter mutation is correlated with insect resistance to *Bacillus thuringiensis* Cry1Ac toxin. PLoS Genet 6.
27. Atsumi S, Miyamoto K, Yamamoto K, Narukawa J, Kawai S, et al. (2012) Single amino acid mutation in an ATP-binding cassette transporter gene causes resistance to Bt toxin Cry1Ab in the silkworm, *Bombyx mori*. Proc Natl Acad Sci U S A 109: E1591 E1598.
28. Liu SM, Zhou S, Tian L, Guo EN, Luan YX, et al. (2011) Genome-wide identification and characterization of ATP-binding cassette transporters in the silkworm, *Bombyx mori*. BMC Genomics 12.
29. Roth CW, Holm I, Graille M, Dehoux P, Rzhetsky A, et al. (2003) Identification of the *Anopheles gambiae* ATP-binding cassette transporter superfamily genes. Mol Cells 15: 150 158.
30. Gomez-Zurita J, Hunt T, Kopliku F, Vogler AP (2007) Recalibrated tree of leaf beetles (Chrysomelidae) indicates independent diversification of angiosperms and their insect herbivores. PloS One 2.
31. Termonia A, Hsiao TH, Pasteels JM, Milinkovitch MC (2001) Feeding specialization and host-derived chemical defense in Chrysomeline leaf beetles did not lead to an evolutionary dead end. Proc Natl Acad Sci U S A 98: 3909 3914.
32. Opitz SEW, Mueller C (2009) Plant chemistry and insect sequestration. Chemoecology 19: 117 154.
33. Pasteels JM, Duffey S, Rowell-Rahier M (1990) Toxins in chrysomelid beetles possible evolutionary sequence from *de novo* synthesis to derivation from food-plant chemicals. J Chem Ecol 16: 211 222.
34. Pasteels JM, Rowell-Rahier M, Braekman JC, Dupont A (1983) Salicin from host plant as precursor of salicyl aldehyde in defensive secretion of chrysomeline larvae. Physiol Entomol 8: 307 314.
35. Kuhn J, Pettersson EM, Feld BK, Burse A, Termonia A, et al. (2004) Selective transport systems mediate sequestration of plant glucosides in leaf beetles: A molecular basis for adaptation and evolution. Proc Natl Acad Sci U S A 101: 13808 13813.
36. Smiley JT, Horn JM, Rank NE (1985) Ecological effects of salicin at 3 trophic levels - new problems from old adaptations. Science 229: 649 651.
37. Michalski C, Mohagheghi H, Nimtz M, Pasteels JM, Ober D (2008) Salicyl alcohol oxidase of the chemical defense secretion of two chrysomelid leaf beetles - Molecular and functional characterization of two new members of the glucose-methanol-choline oxidoreductase gene family. J Biol Chem 283: 19219 11928.
38. Gross J, Podsiadlowski L, Hilker M (2002) Antimicrobial activity of exocrine glandular secretion of *Chrysomela* larvae. J Chem Ecol 28: 317 331.
39. Discher S, Burse A, Tolzin-Banasch K, Heinemann SH, Pasteels JM, et al. (2009) A versatile transport network for sequestering and excreting plant glycosides in leaf beetles provides an evolutionary flexible defense strategy. ChemBioChem 10: 2223 2229.
40. Strauss AS, Peters S, Boland W, Burse A (2013) ABC transporter functions as a pacemaker for sequestration of plant glucosides in leaf beetles. eLife 2: e01096.
41. Bodemann RR, Rahfeld P, Stock M, Kunert M, Wielsch N, et al. (2012) Precise RNAi-mediated silencing of metabolically active proteins in the defence secretions of juvenile leaf beetles. Proc R Soc Lond, Ser B: Biol Sci 279: 4126 4134.
42. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53 59.
43. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644 652.
44. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, et al. (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19: 651 652.
45. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, et al. (2008) The Pfam protein families database. Nucleic Acids Res 36: D281 D288.
46. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: D290 D301.
47. Keeling PJ, Fast NM (2002) Microsporidia: Biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol 56: 93 116.
48. Toguebaye BS, Marchand B, Bouix G (1988) Microsporidia of the Chrysomelidae. In: Jolivet P, Petitpierre E, Hsiao TH, editors. Biology of Chrysomelidae: Springer Netherlands. 399 416.
49. Bauer LS, Pankratz HS (1993) *Nosema scripta* N. Sp. (Microsporida, Nosematidae), a microsporidian parasite of the cottonwood leaf beetle, *Chrysomela scripta* (Coleoptera, Chrysomelidae). J Eukaryot Microbiol 40: 135 141.
50. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30: 3059 3066.
51. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688 2690.
52. Broehan G, Kroeger T, Lorenzen M, Merzendorfer H (2013) Functional analysis of the ATP-binding cassette (ABC) transporter gene family of *Tribolium castaneum*. BMC Genomics 14.
53. Langmead B (2010) Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics Chapter 11: Unit 11.17.
54. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106.
55. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, et al. (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nature protocols 8: 1765 1786.
56. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.
57. Lander ES, Waterman MS (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics 2: 231 239.
58. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2 $^{\Delta\Delta CT}$ method. Methods 25: 402 408.
59. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin Chem 55: 611 622.
60. Albrecht C, Viturro E (2007) The ABCA subfamily - gene and protein structures, functions and associated hereditary diseases. Pflugers Arch - Eur J Physiol 453: 581 589.
61. Broccardo C, Luciani MF, Chimini G (1999) The ABCA subclass of mammalian transporters. Biochim Biophys Acta 1461: 395 404.
62. Zarubica A, Trompier D, Chimini G (2007) ABCA1, from pathology to membrane function. Pflugers Arch - Eur J Physiol 453: 569 579.
63. Tsybovsky Y, Orban T, Molday RS, Taylor D, Palczewski K (2013) Molecular organization and ATP-induced conformational changes of ABCA4, the photoreceptor-specific ABC transporter. Structure 21: 854 860.
64. Juliano RL, Ling V (1976) A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta 455: 152 162.

65. Ueda K, Cardarelli C, Gottesman MM, Pastan I (1987) Expression of a full-length cDNA for the human "MDR1" gene confers resistance to colchicine, doxorubicin, and vinblastine. Proc Natl Acad Sci U S A 84: 3004 3008.
66. Gros P, Croop J, Housman D (1986) Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. Cell 47: 371 380.
67. Gottesman MM, Fojo T, Bates SE (2002) Multidrug resistance in cancer: Role of ATP-dependent transporters. Nature Reviews Cancer 2: 48 58.
68. Yu M, Ocana A, Tannock IF (2013) Reversal of ATP-binding cassette drug transporter activity to modulate chemoresistance: why has it failed to provide clinical benefit? Cancer Metastasis Rev 32: 211 227.
69. Zhou SF (2008) Structure, function and regulation of P-glycoprotein and its clinical relevance in drug disposition. Xenobiotica 38: 802 832.
70. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718 1722.
71. Seeger MA, van Veen HW (2009) Molecular basis of multidrug transport by ABC transporters. Biochim Biophys Acta 1794: 725 737.
72. Sarkadi B, Homolya L, Szakacs G, Varadi A (2006) Human multidrug resistance ABCB and ABCG transporters: Participation in a chemoimmunity defense system. Physiol Rev 86: 1179 1236.
73. Chen KG, Valencia JC, Gillet JP, Hearing VJ, Gottesman MM (2009) Involvement of ABC transporters in melanogenesis and the development of multidrug resistance of melanoma. Pigment Cell Melanoma Res 22: 740 749.
74. Kawanobe T, Kogure S, Nakamura S, Sato M, Katayama K, et al. (2012) Expression of human ABCB5 confers resistance to taxanes and anthracyclines. Biochem Biophys Res Commun 418: 736 741.
75. Oude Elferink RP, Paulusma CC (2007) Function and pathophysiological importance of ABCB4 (MDR3 P-glycoprotein). Pflugers Arch - Eur J Physiol 453: 601 610.
76. Stieger B, Meier Y, Meier PJ (2007) The bile salt export pump. Pflugers Arch - Eur J Physiol 453: 611 620.
77. Herget M, Tampe R (2007) Intracellular peptide transporters in human - compartmentalization of the "peptidome". Pflugers Arch - Eur J Physiol 453: 591 600.
78. Chavan H, Khan MMT, Tegos G, Krishnamurthy P (2013) Efficient purification and reconstitution of ATP Binding Cassette Transporter B6 (ABCB6) for functional and structural studies. J Biol Chem 288: 22658 22669.
79. Kiss K, Brozik A, Kucsma N, Toth A, Gera M, et al. (2012) Shifting the paradigm: The putative mitochondrial protein ABCB6 resides in the lysosomes of cells and in the plasma membrane of erythrocytes. Plos One 7.
80. Pondarre C, Antiochos BB, Campagna DR, Greer EL, Deck KM, et al. (2006) The mitochondrial ATP-binding cassette transporter ABCB7 is essential in mice and participates in cytosolic iron-sulfur cluster biogenesis. Hum Mol Genet 15: 953 964.
81. Ye H, Rouault TA (2010) Human iron-sulfur cluster assembly, cellular iron homeostasis, and disease. Biochemistry 49: 4945 4956.
82. Ichikawa Y, Bayeva M, Ghanefar M, Potini V, Sun L, et al. (2012) Disruption of ATP-binding cassette B8 in mice leads to cardiomyopathy through a decrease in mitochondrial iron export. Proc Natl Acad Sci U S A 109: 4152 4157.
83. Podsiadlowski L, Matha V, Vilcinskas A (1998) Detection of a P-glycoprotein related pump in Chironomus larvae and its inhibition by verapamil and cyclosporin A. Comp Biochem Physiol B: Biochem Mol Biol 121: 443 450.
84. Aurade RM, Jayalakshmi SK, Sreeramulu K (2010) P-glycoprotein ATPase from the resistant pest, *Helicoverpa armigera*: Purification, characterization and effect of various insecticides on its transport function. Biochim Biophys Acta 1798: 1135 1143.
85. Aurade RM, Jayalakshmi SK, Udikeri SS, Sreeramulu K (2012) Modulation of P-glycoprotein ATPase of *Helicoverpa armigera* by cholesterol: Effects on ATPase activity and interaction of insecticides. Arch Insect Biochem Physiol 79: 47 60.
86. Luo L, Sun YJ, Wu YJ (2013) Abamectin resistance in *Drosophila* is related to increased expression of P-glycoprotein via the dEGFR and dAkt pathways. Insect Biochem Mol Biol 43: 627 634.
87. Porretta D, Gargani M, Bellini R, Medici A, Punelli F, et al. (2008) Defence mechanisms against insecticides temephos and diflubenzuron in the mosquito *Aedes caspius*: the P-glycoprotein efflux pumps. Med Vet Entomol 22: 48 54.
88. Srinivas R, Udikeri SS, Jayalakshmi SK, Sreeramulu K (2004) Identification of factors responsible for insecticide resistance in Helicoverpa armigera. Comp Biochem Physiol, C: Toxicol Pharmacol 137: 261 269.
89. Lanning CL, Fine RL, Corcoran JJ, Ayad HM, Rose RL, et al. (1996) Tobacco budworm P-glycoprotein: Biochemical characterization and its involvement in pesticide resistance. Biochim Biophys Acta Gen Subj 1291: 155 162.
90. Murray CL, Quaglia M, Arnason JT, Morris CE (1994) A putative nicotine pump at the metabolic blood-brain-barrier of the tobacco hornworm. J Neurobiol 25: 23 34.
91. Sorensen JS, Dearing MD (2006) Efflux transporters as a novel herbivore countermechanism to plant chemical defenses. J Chem Ecol 32: 1181 1196.
92. Begun DJ, Whitley P (2000) Genetics of alpha-amanitin resistance in a natural population of *Drosophila melanogaster*. Heredity 85: 184 190.
93. Tapadia MG, Lakhotia SC (2005) Expression of mdr49 and mdr65 multidrug resistance genes in larval tissues of *Drosophila melanogaster* under normal and stress conditions. Cell Stress Chaperones 10: 7 11.
94. Petschenka G, Pick C, Wagschal V, Dobler S (2013) Functional evidence for physiological mechanisms to circumvent neurotoxicity of cardenolides in an adapted and a non-adapted hawk-moth species. Proc R Soc Lond, Ser B: Biol Sci 280: 20123089.
95. Bangert I, Tumulka F, Abele R (2011) The lysosomal polypeptide transporter TAPL: more than a housekeeping factor? Biol Chem 392: 61 66.
96. Zhou SF, Wang LL, Di YM, Xue CC, Duan W, et al. (2008) Substrates and inhibitors of human multidrug resistance associated proteins and the implications in drug development. Curr Med Chem 15: 1981 2039.
97. Chen Z-S, Tiwari AK (2011) Multidrug resistance proteins (MRPs/ABCCs) in cancer chemotherapy and genetic diseases. FEBS J 278: 3226 3245.
98. Aleksandrov AA, Aleksandrov LA, Riordan JR (2007) CFTR (ABCC7) is a hydrolyzable-ligand-gated channel. Pflugers Arch - Eur J Physiol 453: 693 702.
99. Hunt JF, Wang C, Ford RC (2013) Cystic Fibrosis Transmembrane Conductance Regulator (ABCC7) Structure. Cold Spring Harb Perspect Med 3.
100. Bryan J, Munoz A, Zhang X, Dufer M, Drews G, et al. (2007) ABCC8 and ABCC9: ABC transporters that regulate K(+) channels. Pflugers Arch - Eur J Physiol 453: 703 718.
101. Heckel DG (2012) Learning the ABCs of Bt: ABC transporters and insect resistance to *Bacillus thuringiensis* provide clues to a crucial step in toxin mode of action. Pestic Biochem Physiol 104: 103 110.
102. Morita M, Imanaka T (2012) Peroxisomal ABC transporters: Structure, function and role in disease. Biochim Biophys Acta 1822: 1387 1396.
103. Kashiwayama Y, Seki M, Yasui A, Murasaki Y, Morita M, et al. (2009) 70-kDa peroxisomal membrane protein related protein (P70R/ABCD4) localizes to endoplasmic reticulum not peroxisomes, and NH2-terminal hydrophobic property determines the subcellular localization of ABC subfamily D proteins. Exp Cell Res 315: 190 205.
104. Tian Y, Han X, Tian D-l (2012) The biological regulation of ABCE1. IUBMB Life 64: 795 800.
105. Andersen DS, Leevers SJ (2007) The essential *Drosophila* ATP-binding cassette domain protein, pixie, binds the 40 S ribosome in an ATP-dependent manner and is required for translation initiation. J Biol Chem 282: 14752 14760.
106. Kerr ID (2004) Sequence analysis of twin ATP binding cassette proteins involved in translational control, antibiotic resistance, and ribonuclease L inhibition. Biochem Biophys Res Commun 315: 166 173.
107. Kerr ID, Haider AJ, Gelissen IC (2011) The ABCG family of membrane-associated transporters: you don't have to be big to be mighty. Br J Pharmacol 164: 1767 1779.
108. Moitra K, Silverton L, Limpert K, Im K, Dean M (2011) Moving out: from sterol transport to drug resistance - the ABCG subfamily of efflux pumps. Drug Metabol Drug Interact 26: 105 111.
109. Tarr PT, Tarling EJ, Bojanic DD, Edwards PA, Baldan A (2009) Emerging new paradigms for ABCG transporters. Biochim Biophys Acta 1791: 584 593.
110. Woodward OM, Koettgen A, Koettgen M (2011) ABCG transporters and disease. FEBS J 278: 3215 3225.
111. Wang L, Kiuchi T, Fujii T, Daimon T, Li M, et al. (2013) Mutation of a novel ABC transporter gene is responsible for the failure to incorporate uric acid in the epidermis of ok mutants of the silkworm, *Bombyx mori*. Insect Biochem Mol Biol 43: 562 571.
112. Itoh TQ, Tanimura T, Matsumoto A (2011) Membrane-bound transporter controls the circadian transcription of clock genes in *Drosophila*. Genes Cells 16: 1159 1167.
113. Liu S, Li Q, Liu Z (2013) Genome-wide identification, characterization and phylogenetic analysis of 50 catfish ATP-binding cassette (ABC) transporter genes. Plos One 8.
114. Popovic M, Zaja R, Loncar J, Smital T (2010) A novel ABC transporter: The first insight into zebrafish (*Danio rerio*) ABCH1. Mar Environ Res 69: S11 S13.
115. Dermauw W, Osborne EJ, Clark RM, Grbic M, Tirry L, et al. (2013) A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*. BMC Genomics 14.
116. Sturm A, Cunningham P, Dean M (2009) The ABC transporter gene family of *Daphnia pulex*. BMC Genomics 10: 170.
117. Xie X, Cheng T, Wang G, Duan J, Niu W, et al. (2012) Genome-wide analysis of the ATP-binding cassette (ABC) transporter gene family in the silkworm, *Bombyx mori*. Mol Biol Rep 39: 7281 7291.
118. Petzold A, Reichwald K, Groth M, Taudien S, Hartmann N, et al. (2013) The transcript catalogue of the short-lived fish *Nothobranchius furzeri* provides insights into age-dependent changes of mRNA levels. BMC Genomics 14: 185.
119. Stock M, Gretscher RR, Groth M, Eiserloh S, Boland W, et al. (2013) Putative sugar transporters of the mustard leaf beetle *Phaedon cochleariae*: Their phylogeny and role for nutrient supply in larval defensive glands. PLoS One 8: e84461.
120. Dong JS, Lai R, Jennings JL, Link AJ, Hinnebusch AG (2005) The novel ATP-binding cassette protein ARB1 is a shuttling factor that stimulates 40S and 60S ribosome biogenesis. Mol Cell Biol 25: 9859 9873.
121. Overbeck TR, Hupfeld T, Krause D, Waldmann-Beushausen R, Chapuy B, et al. (2013) Intracellular ATP-Binding Cassette Transporter A3 is expressed in lung cancer cells and modulates susceptibility to cisplatin and paclitaxel. Oncology 84: 362 370.
122. Aung T, Chapuy B, Vogel D, Wenzel D, Oppermann M, et al. (2011) Exosomal evasion of humoral immunotherapy in aggressive B-cell lymphoma modulated by ATP-binding cassette transporter A3. Proc Natl Acad Sci U S A 108: 15336 15341.

123. Fischer S, Kluver N, Burkhardt-Medicke K, Pietsch M, Schmidt AM, et al. (2013) ABCB4 acts as multixenobiotic transporter and active barrier against chemical uptake in zebrafish (*Danio rerio*) embryos. BMC Biol 11.
124. Tsuchida M, Emi Y, Kida Y, Sakaguchi M (2008) Human ABC transporter isoform B6 (ABCB6) localizes primarily in the Golgi apparatus. Biochem Biophys Res Commun 369: 369 375.
125. Paterson JK, Shukla S, Black CM, Tachiwada T, Garfield S, et al. (2007) Human ABCB6 localizes to both the outer mitochondrial membrane and the plasma membrane. Biochemistry 46: 9443 9452.
126. Jalil YA, Ritz V, Jakimenko A, Schmitz-Salue C, Siebert H, et al. (2008) Vesicular localization of the rat ATP-binding cassette half-transporter rAbcb6. Am J Physiol Cell Physiol 294: C579 C590.
127. Krishnamurthy P, Xie T, Schuetz JD (2007) The role of transporters in cellular heme and porphyrin homeostasis. Pharmacol Ther 114: 345 358.
128. Ulrich DL, Lynch J, Wang Y, Fukuda Y, Nachagari D, et al. (2012) ATP-dependent mitochondrial porphyrin importer ABCB6 protects against phenylhydrazine toxicity. J Biol Chem 287: 12679 12690.
129. Krishnamurthy P, Schuetz JD (2011) The role of ABCG2 and ABCB6 in porphyrin metabolism and cell survival. Curr Pharm Biotechnol 12: 647 655.
130. Szakacs G, Annereau JP, Lababidi S, Shankavaram U, Arciello A, et al. (2004) Predicting drug sensitivity and resistance: Profiling ABC transporter genes in cancer cells. Cancer Cell 6: 129 137.
131. Chavan H, Oruganti M, Krishnamurthy P (2011) The ATP-Binding Cassette transporter ABCB6 is induced by arsenic and protects against arsenic cytotoxicity. Toxicol Sci 120: 519 528.
132. Yasui K, Mihara S, Zhao C, Okamoto H, Saito-Ohara F, et al. (2004) Alteration in copy numbers of genes as a mechanism for acquired drug resistance. Cancer Res 64: 1403 1410.
133. Wang Y, Lim L, DiGuistini S, Robertson G, Bohlmann J, et al. (2013) A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. New Phytol 197: 886 898.
134. Dow JAT, Davies SA (2006) The Malpighian tubule: Rapid insights from post-genomic biology. J Insect Physiol 52: 365 378.
135. Bayeva M, Khechaduri A, Wu RX, Burke MA, Wasserstrom JA, et al. (2013) ATP-binding cassette B10 regulates early steps of heme synthesis. Circul Res 113: 279 287.
136. Chloupkova M, LeBard LS, Koeller DM (2003) MDL1 is a high copy suppressor of ATM1: Evidence for a role in resistance to oxidative stress. J Mol Biol 331: 155 165.
137. Sullivan DT, Sullivan MC (1975) Transport defects as physiological basis for eye color mutants of *Drosophila melanogaster*. Biochem Genet 13: 603 613.
138. Ryall RL, Howells AJ (1974) Ommochrome biosynthetic pathway of *Drosophila melanogaster* - variations in levels of enzyme activities and intermediates during adult development. Insect Biochem 4: 47 61.
139. Sullivan DT, Grillo SL, Kitos RJ (1974) Subcellular-localization of first 3 enzymes of ommochrome synthetic pathway in *Drosophila melanogaster*. J Exp Zool 188: 225 233.
140. Evans JM, Day JP, Cabrero P, Dow JAT, Davies S-A (2008) A new role for a classical gene: White transports cyclic GMP. J Exp Biol 211: 890 899.
141. Haunerland NH, Shirk PD (1995) Regional and functional differentiation in the insect fat body. Annu Rev Entomol 40: 121 145.
142. Arrese EL, Soulages JL (2010) Insect fat body: Energy, metabolism, and regulation. Annu Rev Entomol 55: 207 225.
143. Tarling EJ, Vallim TQD, Edwards PA (2013) Role of ABC transporters in lipid transport and human disease. Trends Endocrinol Metab 24: 342 350.
144. Li JW, Lehmann S, Weissbecker B, Naharros IO, Schutz S, et al. (2013) Odoriferous defensive stink gland transcriptome to identify novel genes necessary for quinone synthesis in the Red Flour Beetle, *Tribolium castaneum*. PLoS Genet 9.
145. Elliott AM, Al-Hajj MA (2009) ABCB8 mediates doxorubicin resistance in melanoma cells by protecting the mitochondrial genome. Mol Cancer Res 7: 79 87.

4.2 Manuscript 2: "Glandular beta-glucosidases in juvenile *Chrysomelina* leaf beetles support the evolution of a host-plant-dependent chemical defence"

Insect Biochemistry and Molecular Biology 58 (2015) 28–38

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Glandular β-glucosidases in juvenile Chrysomelina leaf beetles support the evolution of a host-plant-dependent chemical defense

Peter Rahfeld [a], Wiebke Haeger [a, b], Roy Kirsch [b], Gerhard Pauls [a], Tobias Becker [a], Eva Schulze [d], Natalie Wielsch [c], Ding Wang [a], Marco Groth [e], Wolfgang Brandt [d], Wilhelm Boland [a], Antje Burse [a, *]

[a] Department of Bioorganic Chemistry, Max Planck Institute for Chemical Ecology, Jena, Germany
[b] Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany
[c] Research Group Mass Spectrometry/Proteomics, Max Planck Institute for Chemical Ecology, Jena, Germany
[d] Department of Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany
[e] Genome Analysis Group, Leibniz Institute for Age Research, Fritz Lipmann Institute, Jena, Germany

ARTICLE INFO

Article history:
Received 14 November 2014
Received in revised form
5 January 2015
Accepted 6 January 2015
Available online 14 January 2015

Keywords:
Insect
β-Glucosidase
Enzyme kinetics
RNA interference
Homology modeling
Evolution

ABSTRACT

Plant-feeding insects are spread across the entire plant kingdom. Because they chew externally on leaves, leaf beetle of the subtribe Chrysomelina *sensu stricto* are constantly exposed to life-threatening predators and parasitoids. To counter these pressures, the juveniles repel their enemies by displaying glandular secretions that contain defensive compounds. These repellents can be produced either *de novo* (iridoids) or by using plant-derived precursors. The autonomous production of iridoids pre-dates the evolution of phytochemical-based defense strategies. Both strategies include hydrolysis of the secreted non-toxic glycosides in the defensive exudates. By combining *in vitro* as well as *in vivo* experiments, we show that iridoid *de novo* producing as well as sequestering species rely on secreted β-glucosidases to cleave the pre-toxins. Our phylogenetic analyses support a common origin of chrysomeline β-glucosidases. The kinetic parameters of these β-glucosidases demonstrated substrate selectivity which reflects the adaptation of Chrysomelina *sensu stricto* to the chemistry of their hosts during the course of evolution. However, the functional studies also showed that the broad substrate selectivity allows building a chemical defense, which is dependent on the host plant, but does not lead to an "evolutionary dead end".

1. Introduction

Insects represent the most species-rich lineage of animals on earth and they are, because of their pervasive abundance, a food source of a multitude of predators, pathogens and parasitoids. These numerous insect–enemy interactions led to the evolution of sophisticated defense strategies, which contributed substantially to the success of the insect class. One of those strategies is based on noxious compounds. For example, to repel their natural enemies, leaf beetles (family Chrysomelidae) of the subtribe Chrysomelina *sensu stricto* developed exocrine glands for releasing repellents. The larvae, in particular, have a remarkable way of delivering these toxins: when disturbed, they display droplets of defensive secretions on their backs by contracting specific muscles to evert nine pairs of glandular reservoirs (Pasteels et al., 1982). These droplets contain chemically diverse deterrents which are synthesized from non-toxic glucosidic precursors in the glandular reservoirs; these reservoirs function as small bioreactors (Fig. 1) (Blum et al., 1978; Meinwald et al., 1977; Pasteels et al., 1982, 1989; Sugawara et al., 1979; Termonia and Pasteels, 1999). The precursors can be of endogenous (*de novo* production) or exogenous (sequestration) origin.

Phylogenetic analyses of Chrysomelina *sensu stricto* species revealed that the composition of their larval secretions reflects a step-wise scenario of host-plant adaptation (Termonia et al., 2001). Accordingly, the evolutionary history started with the *de novo* production of deterrent iridoids (cyclopentanoid monoterpenoids) not relying on the secondary metabolites from host plants (Fig. 1) (Oldham et al., 1996; Veith et al., 1994). Derived from this autonomous biosynthesis, some Chrysomelina species evolved a defensive

* Corresponding author. Tel.: +49 (0) 3641 571265; fax: +49 (0) 3641 571202.
E-mail address: aburse@ice.mpg.de (A. Burse).

http://dx.doi.org/10.1016/j.ibmb.2015.01.003
0965-1748/

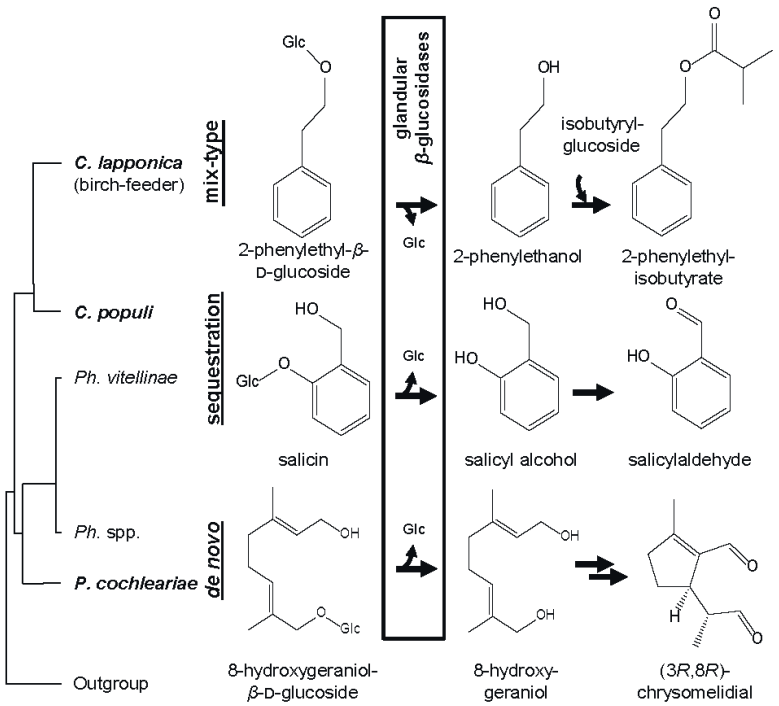

Fig. 1. Steps of the deterrent biosynthesis found in the glandular reservoirs of *Chrysomelina* juveniles. For the *C. lapponica*, only one example reaction is represented. The common enzymatic activity is highlighted with a box. The occurrence of those biosynthetic pathways in selected chrysomeline leaf beetle genera is plotted onto their phylogeny according to Termonia et al. (2001). Glc, glucose. *Ph.* spp., *Phratora* species.
The figure is adapted from Kirsch et al. (2011) and Rahfeld et al. (2014).

strategy that depends on the sequestration of salicin, a plant-derived precursor from Salicaceae used to produce the volatile deterrent salicylaldehyde (Pasteels et al., 1983; Termonia et al., 2001). Juveniles of the monophyletic *interrupta* group that evolved from sequestration-dependent species have acquired the ability to import several glucosidically bound alcohols. These alcohols are esterified after release, resulting in a diversity of esters in the exudate of the larvae (Blum et al., 1978; Hilker and Schulz, 1994). Some members of this group have shifted their host plants from Salicaceae to Betulaceae. Thus, they no longer depend solely on salicin-containing host plants and at the same time they have avoided detection by the predators and parasitoids that seek out sources of salicylaldehyde odor (Gross et al., 2004).

In the juveniles of Chrysomelina *sensu stricto*, all compounds reaching the glandular reservoir *via* the hemolymph are glucosides (*de novo* produced or sequestered) (Pasteels et al., 1989, 1990). As indicated by physiological studies, a complex influx–efflux transport network of varying selectivity guides the glucosides through the insect body (Discher et al., 2009; Kuhn et al., 2004; Kunert et al., 2008; Strauss et al., 2013). As soon as the glucosides have been secreted into the reservoirs, they are metabolized in the same way in all chrysomeline larvae. The first activation step, i.e. the hydrolysis of the glucose moiety from the precursors, is predicted to be facilitated in all species by β-glucosidases (Laurent et al., 2005; Soetens et al., 1993) (Fig. 1). Subsequently, the released alcohol is oxidized by glucose–methanol–choline (GMC) oxidoreductases in iridoid-as well as salicylaldehyde-producing species (Brueckmann et al., 2002; Kirsch et al., 2011; Rahfeld et al., 2014; Veith et al., 1996).

Whereas the use of sequestered or *de novo*-produced glycosides and β-glucosidase for defense purposes is very well known and has been functionally characterized for plants (Mithoefer and Boland, 2012; Pentzold et al., 2014), few reports of such two-component systems for defense are available from insects (Berenbaum, 1995; Dobler, 2001). The caterpillars of the six-spot burnet moth (*Zygaena filipendulae*), for example, are able to produce *de novo* or to sequester cyanogenic glycosides and then to store them in the hemolymph and defense cavities together with a β-glucosidase named linamarase (Zagrobelny et al., 2008; Zagrobelny and Moeller, 2011). The cyanogenesis is triggered when insects are injured (Franzl et al., 1989). The cabbage aphids (*Brevicoryne brassicae*) sequester glucosinolates from their brassicaceous host plants and store them in the hemolymph. After injuring the aphids, the ascorbate-dependent β-thioglucosidase (called myrosinase), which is stored separately in the non-flight muscles, hydrolyzes the glucosinolates and releases the toxins (Jones et al., 2001, 2002; Kazana et al., 2007). This type of "mustard oil bomb" defense is also utilized by flea beetles (*Phyllotreta striolata*). The adults sequester and store glucosinolates from their host plants which can be hydrolyzed by a myrosinase (Beran et al., 2014).

Unlike the defense strategies described above, the toxic deterrent produced by Chrysomelina larvae is stored separately from the larval body in the glandular reservoir. The Chrysomelina defense strategy is a fascinating adaption of the classic two-component system, which does not rely on the sacrifice of whole insects as is the case with cabbage aphids. Independent of the mode of action, all these defense strategies depend on glycoside hydrolases.

The β-glucosidases involved in the binary defense systems of plants and insects are mostly members of the glycoside hydrolases family 1 (GH1). The GH superfamily comprises over 100 families, classified through structural or amino acid sequence similarities (Henrissat et al., 1995). GH1s are able to hydrolyze the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (Henrissat and Davies, 1997; Lombard et al., 2014). With a few exceptions, such as the dhurrin hydrolyzing dhurrinase 1 from *Sorghum bicolor* (Verdoucq et al., 2004), β-glucosidases have a broad range of substrates.

Using Chrysomelina *sensu strico* as a model system, we have focused on elucidating the glandular defense metabolism and processes contributing to metabolic diversity in defensive secretions. For this purpose we have identified the glandular-specific β-glucosidases from iridoid-producing (*Phaedon cochleariea*), salicin-sequestering (*Chrysomela populi*) and ester-utilizing (*Chrysomela lapponica*) larvae. Their importance for the formation of deterrents in defensive secretions from glucosidically-bound pre-toxins has been verified by RNA interference (RNAi) *in vivo*. Further, the determination of the kinetic parameters of the glucosidases allows conclusions about the mechanisms that play a role in adaptation to host plants or in host plant shifts.

2. Material & methods

See electronic supplementary material (ESM) for the complete proteome analyses of the secretions by data-independent liquid chromatography–mass spectrometry detection (LC–MS^E), transcriptome generation, cloning procedures, detailed quantitative PCR (qPCR) procedure, phylogenic reconstructions, gas chromatography–mass spectrometry (GC–MS) and liquid chromatography–mass spectrometry (LC–MS) analysis, enzymatic synthesis of 2-phenylethyl-β-D-glucoside, homology modeling, protein assays, glucose measurement in the secretions, all primer sequences (Table S1) and accession numbers (Table S2).

All chemicals were purchased from Sigma–Aldrich (St. Louis, MO, USA), Carl Roth (Karlsruhe, Germany) or Serva (Heidelberg, Germany), if not stated otherwise.

2.1. Leaf beetle rearing and sample processing

The mustard leaf beetle *P. cochleariae* (F.) was lab-reared on *Brassica oleracea* convar. *capitata* var. *alba* (Gloria F1) in 16 L:8 D cycle conditions and 15 ± 2 °C. *C. populi* (L.) was collected near Dornburg, Germany (latitude 51.015, longitude 11.64), on *Populus maximowiczii* × *Populus nigra* and then lab-reared on its host plant under conditions comparable to *P. cochleariae*. Birch-feeding *C. lapponica* (L.) was collected from *Betula rotundifolia* in the Altai Mountains in East Kazakhstan, close to Uryl, near the Burhat Pass (2130 m altitude, 49°07,438′ N, 86°01, 365′ E). *C. lapponica* was reared under field conditions on its host plant.

Larval secretions were collected in glass capillaries (i.d.: 0.28 mm, o.d.: 0.78 mm, length 100 mm; Hirschmann, Eberstadt, Germany) and subsequently depleted in ethyl acetate for analysis of low molecular weight compounds or in 1× SDS-sample buffer (50 mM Tris–HCl pH 6.8, 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 12.5 mM EDTA, 0.02% bromophenol blue) for protein analysis. To isolate total RNA, the bodies of third-instar larvae were dissected into head, gut, Malpighian tubules, fat body, glands. The tissue samples were stored in 100 µl lysis-buffer (Life Technologies, Carlsbad, CA, USA) with the addition of 1 µl ExpressArt Nucleo-Guard (Amp Tec GmbH, Hamburg, Germany) at −80 °C until needed.

2.2. Identification and cloning of the P. cochleariae glandular β-glucosidase (PcgbGlc), C. populi glandular β-glucosidase (CpgbGlc) and the C. lapponica (birch-feeder) glandular β-glucosidase (ClgbGlc_B)

The data from the proteome analyses of *P. cochleariae*, *C. populi* and *C. lapponica* secretions were screened for proteins belonging to the glycosyl hydrolase family 1 (Glyco_hydro_1 (PF00232)) by using Pfam-search (http://pfam.sanger.ac.uk/) (Punta et al., 2012). Three transcripts (one for each species) encoding proteins of the glycosyl hydrolase family 1 were identified and subsequently amplified *via* PCR. In the reactions, the Phusion High-Fidelity DNA Polymerase (Fisher Scientific – Germany GmbH, Schwerte, Germany) was used to amplify full-length sequence of *Pc*gbGlc, *Cp*gbGlc and *Cl*gbGlc_B from glandular cDNA templates (Table S1). After purification with a PCR-purification kit (Roche, Basel, Switzerland), the resulting DNA-fragments were cloned into pIB/V5-HIS-TOPO vector (Life Technologies, Carlsbad, CA, USA). The sequences were confirmed by custom Sanger sequencing and compared to the assembled sequences available in our transcriptome libraries (Rahfeld et al., 2014; Stock et al., 2013). The sequences of *Pc*gbGlc, *Cp*gbGlc and *Cl*gbGlc_B were registered at GenBank (Table S2).

2.3. Heterologous expression of PcgbGlc, CpgbGlc and ClgbGlc_B in High Five insect cells and protein purification

Heterologous expression was carried out in the insect cell line High Five (Life Technologies, Carlsbad, CA, USA). The constructs pIB-*Pc*gbGlc, pIB-*Cp*gbGlc and pIB-*Cl*gbGlc_B were transfected with the FuGeneHD–Kit (Promega GmbH, MA, USA) and MA Lipofection Enhancer (IBA GmbH, Göttingen, Germany) according to the manufacturer's instructions. After one day of incubation at 27 °C, the cultures were supplied with 60 µg/ml blasticidin (Life Technologies, Carlsbad, CA, USA) to initiate the selection of stable cell lines. Afterwards, the insect cells were selected over three passages. The cultivation of the stable cell lines for protein expression was carried out in six 75 cm^2 cell culture flasks, each containing 15 ml culture media (expressfiveTM (Life Technologies, Carlsbad, CA, USA), 20 µg/ml blasticidin, 1× Protease Inhibitor HP Mix (SERVA Electrophoresis GmbH, Heidelberg, Germany)). After 3 days of growth, the supernatants were collected and the cells were discarded (4000 × g, 10 min, 4 °C). The supernatants containing *Pc*gbGlc, *Cp*gbGlc or *Cl*gbGlc_B were dialyzed overnight at 4 °C against 50 mM NaH_2PO_4, 10 mM imidazole (Pufferan), 5% (v/v) glycerol, pH 7.5.

The subsequent purification was done with HisPureCobalt (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions with alterations in the composition of the elution buffer (50 mM NaH_2PO_4, 150 mM imidazole (Pufferan), 5% (v/v) glycerol, pH 7.5). The elution fractions were then desalted through PD midiTrap G-25 (GE Healthcare, Munich, Germany) in storage buffer (50 mM NaH_2PO_4, 2% (v/v) glycerol, pH 7). The last step, the removal of the remaining glucose, was a dialysis overnight at 4 °C against the storage buffer. To calculate the yield of dialyzed proteins, the Pierce BCA Protein Assay Kit (Life Technologies, Carlsbad, CA, USA) was used according to the manufacturer's instructions.

To confirm their identity, the purified proteins were separated by any-kD gradient gels (Bio-Rad Laboratories, Munich, Germany) in 1D-SDS-PAGE and then analyzed *via* Nano-UPLC–MSE as described in Rahfeld et al. (2014).

2.4. Enzymatic assays

The activities of the glandular β-glucosidases were monitored through the release of the glucose from the supplemented

substrates. Therefore, the Amplite Glucose Quantitation Kit *Red Fluorescence* was used according to the manufacturer's instructions with alteration in the assay buffer (100 mM sodium acetate/acetic acid, pH 5.5). The assays were monitored in an Infinite M200 Microplate Reader (Tecan Group Ltd., Maennedorf, Switzerland) at 37 °C. *Pc*gbGlc, *Cp*gbGlc and *Cl*gbGlc_B were incubated with 8-hydroxgeraniol-glucoside (0.1–50 mM), salicin (0.1–100 mM) and 2-phenylethyl-β-D-glucoside (0.1–50 mM). The replicates are performed from single purified protein fraction.

2.5. Kinetic data analysis

The obtained primary fluorescence data were assigned to a corresponding glucose concentration based on a simultaneously measured glucose standard. The slope of every primary plot was then used for a calculation of the kinetic parameters with GraphPad Prism version 5.04 (GraphPad Software, La Jolla, CA, USA) using the built-in enzyme kinetics module.

2.6. Silencing the glandular β-glucosidases of P. cochleariae (PcgbGlc), C. populi (CpgbGlc) and C. lapponica (birch-feeder) (ClgbGlc_B) by RNAi

The coding sequences of *Pc*gbGlc, *Cp*gbGlc and *Cl*gbGlc_B were analyzed for possible down-regulation of unintended targets (off-targets) according to a computational method described by Bodemann et al. (2012). The analyses did not show evidence of off-targets in the transcript libraries, for a critical value of at least 21 continuous nucleotides.

For the dsRNA constructs, 450-bp fragments from the coding sequences of *PcgbGlc* and *CpgbGlc* and 310-bp fragment from the coding sequences of *ClgbGlc_B* were amplified by the Phusion High-Fidelity DNA Polymerase (Fisher Scientific – Germany GmbH, Schwerte, Germany) (Table S1). After purification with a PCR-purification kit (Roche, Basel, Switzerland), the resulting fragments were cloned into T7-promotor site free pIB/V5-HIS-TOPO vectors (Life Technologies, Carlsbad, CA, USA). For dsRNA synthesis, templates with opposing T7-promotor sites were amplified out of sequenced pIB-*450bpPcgbGlc* as well as pIB-*450bpCpgbGlc* and further processed as described in Bodemann et al. (2012). The concentration of dsRNA was adjusted to 1 μg/μl. Early second-instar larvae of *P. cochleariae* or *C. populi* were used for injections. They were collected 7 d after hatching and treated with 100 nl (100 ng) dsRNA of *450bpPcgbGlc* or *450bpCpgbGlc*. In the case of *C. lapponica*, the injection took place under field conditions. The available juveniles (late second-instar and early third-instar larvae) were used for the injection of 100 nl (100 ng) dsRNA of *450bpCpgbGlc*. The dsRNA of *720bpeGFP* was used as described in Bodemann et al. (2012) for control treatments.

3. Results

3.1. Identification, sequence analysis and localization of β-glucosidases in juvenile Chrysomelina sensu stricto

To identify the enzymes responsible for the activation of deterrent precursors, larval secretions from *P. cochleariae*, *C. populi* and *C. lapponica* (birch-feeder) were at first separated by 1D-SDS-PAGE. After recovering, the protein bands were then analyzed by LC–MS^E. For each of the three secretions, the peptides derived from proteins of a molecular weight of 60–75 kDa were matched to data base proteins that belonged to the GH1 (Glyco_hydro_1 (PF00232/E.C. 3.2.1.21)) (Table S3). The full-length amplification and sequencing of the corresponding transcripts led to coding sequences (cds) of 1605 bp (535 amino acids (aa)) for *Pc*gbGlc from *P. cochleariae*, 1611 bp (537 aa) for *Cp*gbGlc from *C. populi*, and 1611 bp (537 aa) for *Cl*gbGlc_B from *C. lapponica*. On the amino acid level, the sequences possess at least 71% identity with each other. For all three proteins, N-terminal signal peptides with a length of 16 aa were indicated by cleavage site predictions (SignalP 4.1: http://www.cbs.dtu.dk/services/SignalP/) (Petersen et al., 2011).

Sequence comparison of the three Chrysomelina β-glucosidases to other β-glucosidases – for example, from the termite *Neotermes koshunensis* (PDB: 3VIL_A) (Jeng et al., 2011), the red flour beetle *Tribolium castaneum* (XP_972437.1) or a human cytosolic β-glucosidase (PDB: 2JFE_X) (Tribolo et al., 2007) – showed only a low degree of identity, about 36% on aa level. The protein alignment, however, illustrated conserved aa residues known to be responsible for substrate binding (Fig. S1) (Jeng et al., 2011, 2012). Furthermore, all three chrysomeline β-glucosidases have conserved glutamates: E182 acting as a catalytic nucleophile and E391 as an acid-base residue (Fig. S1), this is descriptive for members of the glycosyl hydrolase family 1 (PF00232/E.C. 3.2.1.21), according to the classification of Carbohydrate-Active enZymes (CAZy: http://www.cazy.org/) (Henrissat and Davies, 1997; Levasseur et al., 2013).

We used qPCR to test if the β-glucosidase abundance in the defensive glands is also reflected in the mRNA levels of the corresponding transcripts. All genes were predominantly expressed in the defensive glands, in comparison to the analyzed gut, Malpighian tubules and fat body tissues from juvenile *P. cochleariae*, *C. populi* and *C. lapponica* (birch-feeder) (Fig. S2). Based on these results and those from the proteome analyses of the glandular exudates, gland-specific functions of the β-glucosidases are indicated.

3.2. Purification of the glandular β-glucosidases and their enzyme kinetics

In order to study their functional properties and substrate preferences, the three glandular β-glucosidases were heterologously expressed in a eukaryotic system and subsequently purified by affinity chromatography. The plasmids carrying the cds for the β-glucosidases (pIB-*PcgbGlc*, pIB-*CpgbGlc* and pIB-*ClgbGlc_B*) were used for transfection of High Five cells. The yield of recombinant protein purified from 6 × 10 ml (each 75 cm^2) cell culture supernatant was 25.9 μg for *Pc*gbGlc, 49.7 μg for *Cp*gbGlc and 13.6 μg for *Cl*gbGlc_B (Fig. S3). The identities of the proteins were validated by LC–MS^E analysis (Table S4). The catalytic β-glucosidase activity of each purified enzyme and of the un-transfected cells was tested at first by assays in the presence of 8-hydroxygeraniol-β-D-glucoside (Fig. S4). The appearance of 8-hydroxygeraniol in the reactions verified the estimated hydrolase activity. The elution fraction of un-transfected cells did not show any activity (Fig. S4). The activity of the β-glucosidases from *P. cochleariae* and *C. populi* is optimal at a pH range of 4.5–6 (Fig. S5) fitting to the pH value of 4–5 observed in the glandular secretions of Chrysomelina larvae, their natural catalytic environment (Kirsch, 2011). Therefore, all following enzymatic assays were performed at pH 5.5.

The kinetic parameters of each glandular β-glucosidase were determined in assays containing the physiological precursors of the three deterrent pathways. Hence, we used 8-hydroxygeraniol-β-D-glucoside (8HGG) the intermediate in the iridoid metabolism of *P. cochleariae*, salicin the intermediate of the salicylaldehyde synthesis of *C. populi* and 2-phenylethyl-β-D-glucoside (2PhEG) the intermediate of ester production found in *C. lapponica* (birch-feeder) (Fig. 1). The activity of the enzymes was measured by determining the release of the glucose molecule in relation to a glucose standard, and the results are summarized in Table 1. Briefly, all three enzymes, *Pc*gbGlc, *Cp*gbGlc and *Cl*gbGlc_B, exhibited the highest catalytic efficiency for 8HGG, the intermediate of the

Table 1
Kinetic constants for the glandular β-glucosidases. Activities of purified recombinant enzymes from *Chrysomela lapponica* (birch-feeder), *Chrysomela populi* and *Phaedon cochleariae* resulted from the incubation with different substrates. Standard deviation is presented in brackets.

Glandular β-glucosidase	Substrate	K_m [mM]	V_{max} [μmol/min/mg]	k_{cat} [min^{-1}]	k_{cat}/K_m [mM^{-1} min^{-1}]
Phaedon cochleariae	8-Hydroxygeraniol-β-D-glucoside	8.98 (±2.57)	1.92	118.7 (±12.1)	13.21
	Salicin	33.66 (±7.82)	2.82	174.6 (±19.9)	5.19
	2-Phenylethyl-β-D-glucoside	8.34 (±0.68)	0.61	37.6 (±0.9)	4.52
Chrysomela populi	8-Hydroxygeraniol-β-D-glucoside	4.05 (±1.03)	10.87	676.4 (±55.8)	166.93
	Salicin	17.35 (±1.88)	6.44	400.2 (±15.1)	23.07
	2-Phenylethyl-β-D-glucoside	31.36 (±5.66)	1.84	114.2 (±11.1)	3.64
Chrysomela lapponica (birch-feeder)	8-Hydroxygeraniol-β-D-glucoside	6.32 (±0.67)	2.62	161.6 (±4.6)	25.59
	Salicin	8.04 (±0.74)	2.40	147.9 (±3.5)	18.40
	2-Phenylethyl-β-D-glucoside	13.69 (±2.92)	1.55	95.6 (±9.1)	6.98

iridoid pathway. *Cp*gbGlc had higher catalytic activity than *Pc*gbGlc or *Cl*gbGlc_B in the presence of 8HGG. Salicin was most efficiently converted by *Cp*gbGlc, followed by *Cl*gbGlc_B and *Pc*gbGlc. The intermediate for ester synthesis, 2PhEG, was more efficiently hydrolyzed by *Cl*gbGlc_B than by the two other enzymes.

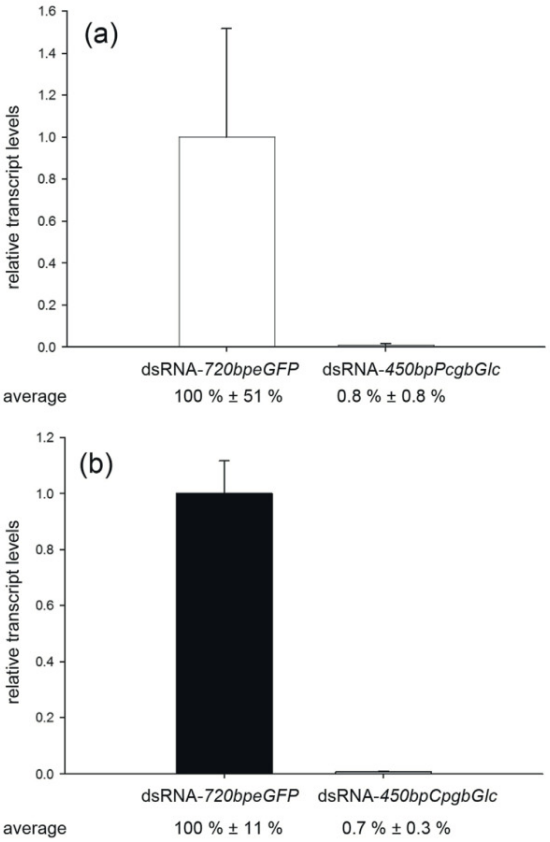

Fig. 2. Transcript level of β-glucosidases in larvae of *P. cochleariae* and *C. populi* after gene silencing by RNAi. (a) *PcgbGlc* (white) and (b) *CpgbGlc* (black) in defensive gland cells seven days after treating larvae with dsRNA-*720bpeGFP* or dsRNA-*450bpPcgbGlc*/dsRNA-*450CpgbGlc*. For the normalization of transcript quantities, *PcRPS3*, *PcRPS18* and *CpeIF4A*, *CpEF1a* were used and the transcript level of the dsRNA-*eGFP* was set to 1. Average transcript levels are shown below the graph as percentage in comparison to RNAi-*eGFP*. Error bars indicate the standard error. The data were normalized by taking the logarithm (to base 10) before statistical analyses to pass the Normality Test (Shapiro–Wilk); (a) $p = 0.046$; (b) $p = 0.040$.

3.3. Silencing glandular β-glucosidases in Chrysomelina sensu stricto larvae

To analyze the *in vivo* relevance of each characterized β-glucosidase from *P. cochleariae*, *C. populi* and *C. lapponica* in the biosynthetic pathway of the deterrents in the larval glandular secretions, RNAi was used. The down-regulation of the targeted transcripts in the glandular tissue was surveyed by qPCR for *P. cochleariae* and *C. populi* (Fig. 2). Seven days after injecting the dsRNA targeting *PcgbGlc* in *P. cochleariae* or *CpgbGlc* in *C. populi* larvae a down-regulation by ~99% of the respective transcripts was detected, compared to the *eGFP* treatment control in both species.

To identify alterations in the deterrent composition, we analyzed glandular secretions by GC–MS and/or LC–MS analyses after silencing *PcgbGlc*, *CpgbGlc* or *ClgbGlc_B*. In the juvenile exudates of *P. cochleariae*, *PcgbGlc* knock-down led to a reduction of chrysomelidial (9.7 min) (Fig. S6), the final product of the iridoid pathway (Veith et al., 1994). At the same time, the intermediate 8HGG (20.8 min) (Veith et al., 1994) appeared (Fig. 3) due to the reduced hydrolase activity and the continuous secretion process of the glucosides into the glandular reservoirs. The secretions had a clear-transparent yellow color, unlike the control treatments (Fig. 3).

As in the silenced *P. cochleariae* larvae, *C. populi* larval secretions with a down-regulated β-glucosidase contained only a small amount of salicylaldehyde (4.2 min), the end product of the hydroxybenzaldehyde pathway (Pasteels et al., 1983) (Fig. S6). Moreover, the substrate of the hydrolysis, namely salicin (13.1 min), accumulated in the secretions; there it finally resulted in the formation of salicin crystals in the gland reservoirs (Fig. 3).

The GC–MS analysis of *C. lapponica* secretions after "in field" silencing revealed that the butyrate esters (*cis*-3-hexenylisobutyrate 8.50 min; *cis*-3-hexenyl-2-methylbutyrate 10.47 min; benzylisoburyrate 11.85 min; 2-phenylethyl-2-methylbutyrate 13.73 min; 2-phenylethylisobutyrate 13.91 min) occurring in the control treatment were missing in the *ClgbGlc_B* knock-down secretions, which indicated the glandular β-glucosidase had been successfully silenced (Fig. S6). Taken together, the RNAi experiments verified the glucosidically-bound intermediates assayed as substrates for the β-glucosidases (*Pc*gbGlc, *Cp*gbGlc, *Cl*gbGlc_B) and the importance of the enzymes for producing deterrents in the glandular system of chrysomeline larvae.

To confirm the source of the free glucose in the secretions, the concentration of glucose was measured before and after RNAi knock-down in *P. cochleariae* and *C. populi* secretions. In the secretion of non-treated larvae 5.24 ± 1.23% (291.1 ± 68.4 μmo/l) glucose for *P. cochleariae* and 6.58 ± 1.63% (365.3 ± 90.9 μmol/l) glucose for *C. populi* were detected (Fig. 4). The amount of free glucose decreased after silencing the glandular β-glucosidases to 0.87 ± 0.42% (48.3 ± 23.3 μmo/l) glucose for *P. cochleariae* and 0.33 ± 0.15% (18.5 ± 8.3 μmo/l) glucose for *C. populi* (Fig. 4). The

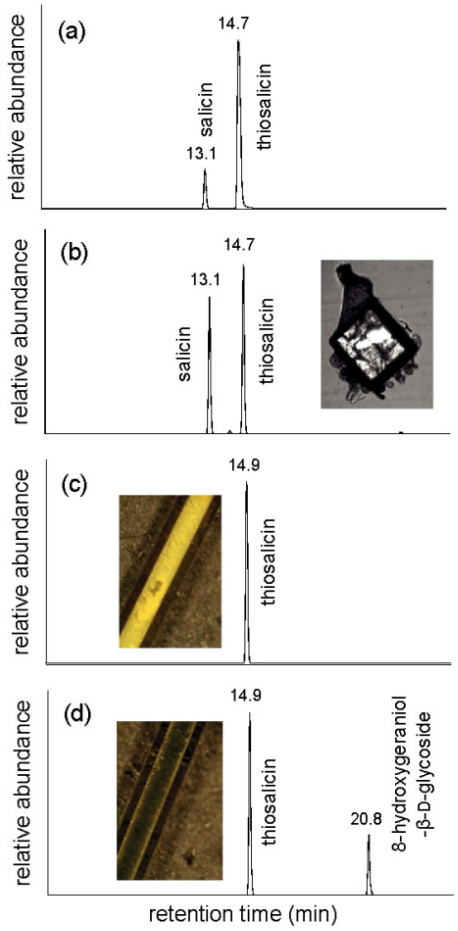

Fig. 3. LC–MS analysis of larval secretions seven days after β-glucosidases' silencing. Treatment of *C. populi* (mass range (+/−1): 331 + 347) with dsRNA-*720bpeGFP* (a), dsRNA-*450bpCpgbGlc* (b) and of *P. cochleariae* (mass range (+/−1): 347 + 377) with dsRNA-*720bpeGFP* (c), dsRNA-*450bpPcgbGlc* (d). The pictures show a salicin-crystal dissected out of a defensive gland from a treated (dsRNA-*450bpCpgbGlc*) larva (b); changes in the color of the secretions collected in capillaries after RNAi treatment (dsRNA-*450bpPcgbGlc*) (d) in comparison to the control (c). 1 mM thiosalicin (14.7 min/14.9 min) was used as an internal standard.

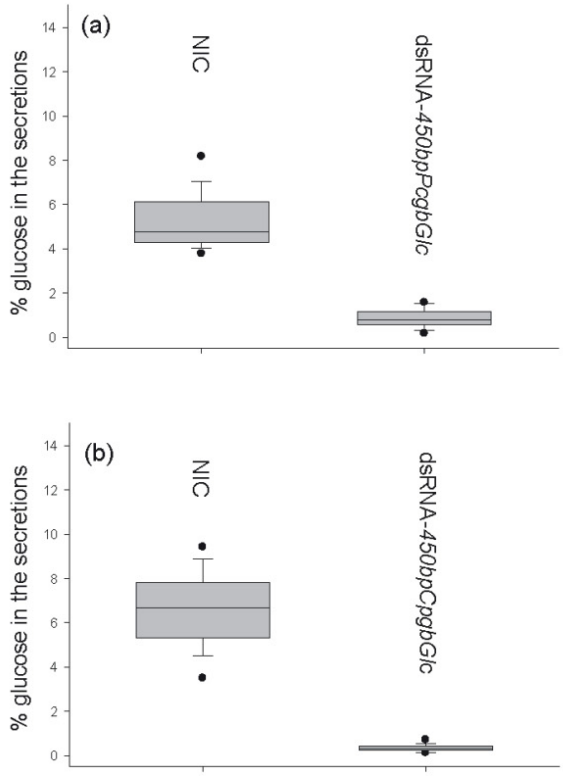

Fig. 4. Analysis of glucose concentration (%) in the secretions seven days after β-glucosidases' silencing. *P. cochleariae* (a) and *C. populi* (b) larvae were injected with dsRNA (dsRNA-*450bpPcgbGlc*/dsRNA-*450bpCpgbGlc*) and compared to non-injected control (NIC) larvae. The data were normalized by taking the logarithm (to base 10) before statistical analyses to pass the Normality Test (Shapiro–Wilk); (a) $p > 0.001$; (b) $p > 0.001$.

results supports that the hydrolysis of the glycosides is the only source of glucose in the secretions.

3.4. Homology modeling of the glandular β-glucosidases

Modeling the three β-glucosidases in the presence of the different substrates tested in the kinetic assays showed that the active centers (where the glucose moiety is bound) with the conserved glutamates in positions 183 and 391 (which catalyzed the hydrolysis) are identical in all models (Fig. 5, Figs. S7–S9). The only detectable variations in the binding site are found in positions 190, 186 and 328.

Two binding models could be generated for *Pc*gbGlc and *Cl*gbGlc_B in the presence of 8HGG. In the case of *Cp*gbGlc, only one model for the binding of 8HGG was generated (Fig. 5, Figs. S7–S9). Because 2PhEG is only one of several glucosidically bound alcohols (Hilker and Schulz, 1994; Tolzin-Banasch et al., 2011) sequestered into the secretions of *C. lapponica* (birch-feeder), the computer-based modeling of additional ester precursors onto the theoretical structure of *Cl*gbGlc_B was performed and revealed that these precursors were also suitable substrates for this β-glucosidase (Table S5).

3.5. Phylogenetic analysis of glucosyl hydrolases

In order to determine if the glandular β-glucosidases of chrysomeline larvae evolved once or arose independently, we performed phylogenetic analyses. A set of β-glucosidases from different insect orders was used. This set includes representative sequences from three different beetle families: the red flour beetle *T. castaneum* (Tenebrionidae), the mountain pine beetle *Dendroctonus ponderosae* (Curculionidae) and a comprehensive GH1 dataset from the striped flea beetle *P. striolata* (Chrysomelidae) (Beran et al., 2014; Keeling et al., 2013; Richards et al., 2008). In addition to the glandular β-glucosidases characterized here, further GH1 sequences were retrieved from the transcriptomes of *P. cochleariae*, *C. populi* and *C. lapponica*, and added to the analyses (Table S2). In general, no orthology of GH1 sequences exists between different beetle families. This is in accordance with previous GH1 phylogenetic

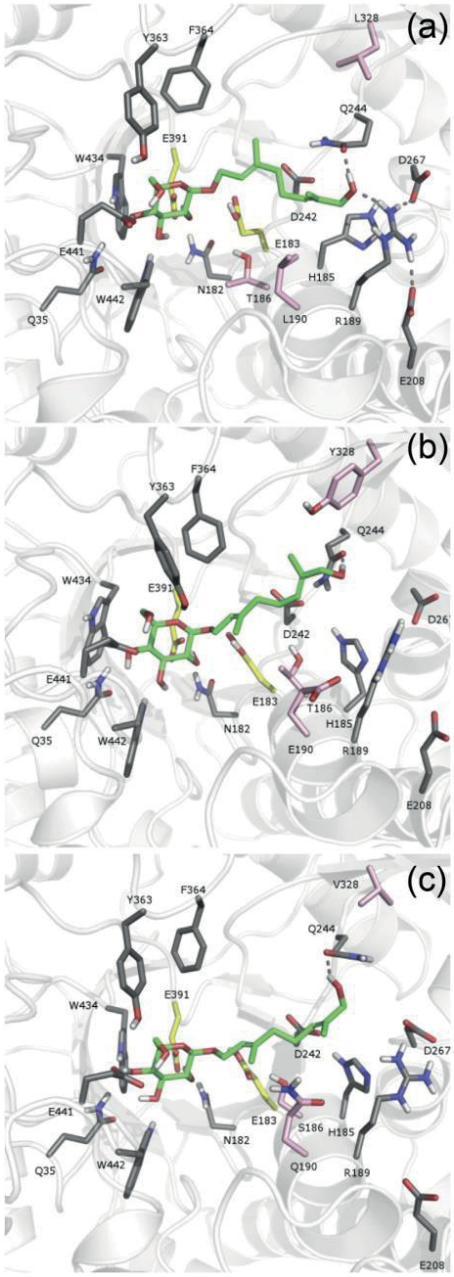

Fig. 5. Homology modeling of β-glucosidases. The active site in *Cp*gbGlc (a), *Pc*gbGlc (b), *Cl*gbGlc_B (c) are shown with the bound substrate 8-hydroxygeraniol-β-D-glycoside (8HGG), presented in green. For *Pc*gbGlc (a) and *Cl*gbGlc_B (c), only one of two putative binding positions of 8HGG is presented (Figs. S7 and S9). The picture was created with Pymol. All presented and labeled amino acids (aa) are involved in the formation of the binding site; the yellow aa are responsible for the hydrolysis of the glycosides, the pink ones are the aa which are varied in the different species. Oxygen is red, nitrogen is blue and hydrogen is white.

reconstructions (Beran et al., 2014; Eyun et al., 2014). But both the Bayesian inferred phylogeny and the Maximum Likelihood reconstruction supported a clade harboring the larval glandular β-glucosidases (Fig. 6). This strongly indicates the orthology of these β-glucosidases and, thus, a common ancestry. Moreover, members of a group of *P. striolata* GH1 sequences, including a functional myrosinase identified earlier (Beran et al., 2014), form a distinct clade with the chrysomeline glandular β-glucosidases, supported by a high posterior probability and bootstrap value (1/96).

4. Discussion

Hydrolysis of glycosides occurs in the metabolism of all organisms. Enzymes that catalyze such reactions often belong to the glycosyl hydrolase family 1 (Henrissat and Davies, 1997). In plants, GH1s play an important role in defense, particularly in the metabolism of various glycosides, but also in symbiosis, cell wall catabolism, lignification, signaling and secondary metabolic processes (Cairns and Esen, 2010). Corresponding to this functional diversity, plants have a huge number of β-glucosidase genes. For example, in the dicotyledonous *Arabidopsis thaliana* genome, 48 putative β-glucosidase genes (Xu et al., 2004) and in the monocotyledonous *Oryza sativa* genome 40 putative β-glucosidase genes (Opassiri et al., 2006) are known. In contrast, insect genomes (such as those from *Apis mellifera* or *Drosophila melanogaster*) and transcriptomes (such as that from *P. striolata*) contain considerably fewer (5–25) GH1s (Beran et al., 2014; Pentzold et al., 2014), probably due to the fact that insects use β-glucosidases almost exclusively for digestive purposes, either in the gut (Terra, 2012) or in salivary glands (Tokuda et al., 2002). Only a few insect β-glucosidases are known to be utilized in defense mechanisms (Beran et al., 2014; Zagrobelny and Moeller, 2011). Here we provide additional examples from the juvenile Chrysomelina *sensu stricto*.

In the defensive secretion of Chrysomelina *sensu stricto* larvae, the hydrolysis of secreted glycosides leads to the aglucons, priming them as deterrents through the following oxidation (Fig. 1) (Brueckmann et al., 2002; Veith et al., 1996). We identified the β-glucosidases in three different species: *Pc*gbGlc in *P. cochleariae*, *Cp*gbGlc in *C. populi*, and *Cl*gbGlc_B in *C. lapponica* (birch-feeder).

The heterologously expressed and purified enzymes revealed Michaelis Menten constants (K_M) for the hydrolysis of salicin of 33.66 mM for *Pc*gbGlc, of 17.35 mM for *Cp*gbGlc and of 8.04 mM for *Cl*gbGlc_B; these constants are in the same range as the already characterized intestinal β-glucosidase from *Reticulitermes flavipes* (RfBGluc-1: K_M for salicin 34.3 mM) (Scharf et al., 2010) and the linamarase from *Zygaena transalpina* (K_M for linamarin 7.8 mM) (Franzl et al., 1989). Despite their different functions in insects, the β-glucosidases have similar kinetic properties.

Our enzyme assays demonstrated that all three chrysomeline glandular β-glucosidases were able to catalyze the hydrolysis of the physiologically relevant intermediates 8HGG, salicin and 2PhEG. In general, all three enzymes were able to hydrolyze the "ancestral" substrate 8HGG with high efficiency, which is in accordance to the hypothesis of a single β-glucosidase ancestor. However, in determining K_M and catalytic efficiency (k_{cat}/K_M), we discovered that the enzymes show a preference for their physiological substrates. Salicin, for example, was converted more efficiently by the glandular β-glucosidases of *C. populi* followed by those of *C. lapponica* (birch-feeder) compared to *Pc*gbGlc, most likely an adaptation to the sequestered substrate. The glandular β-glucosidase from *C. lapponica* had the highest efficiency for 2PhEG, one of its physiological substrates. Overall, the kinetic parameters of the β-glucosidases show a tendency following the evolutionary history of chemical defense strategies in Chrysomelina *sensu stricto* (Fig. 1).

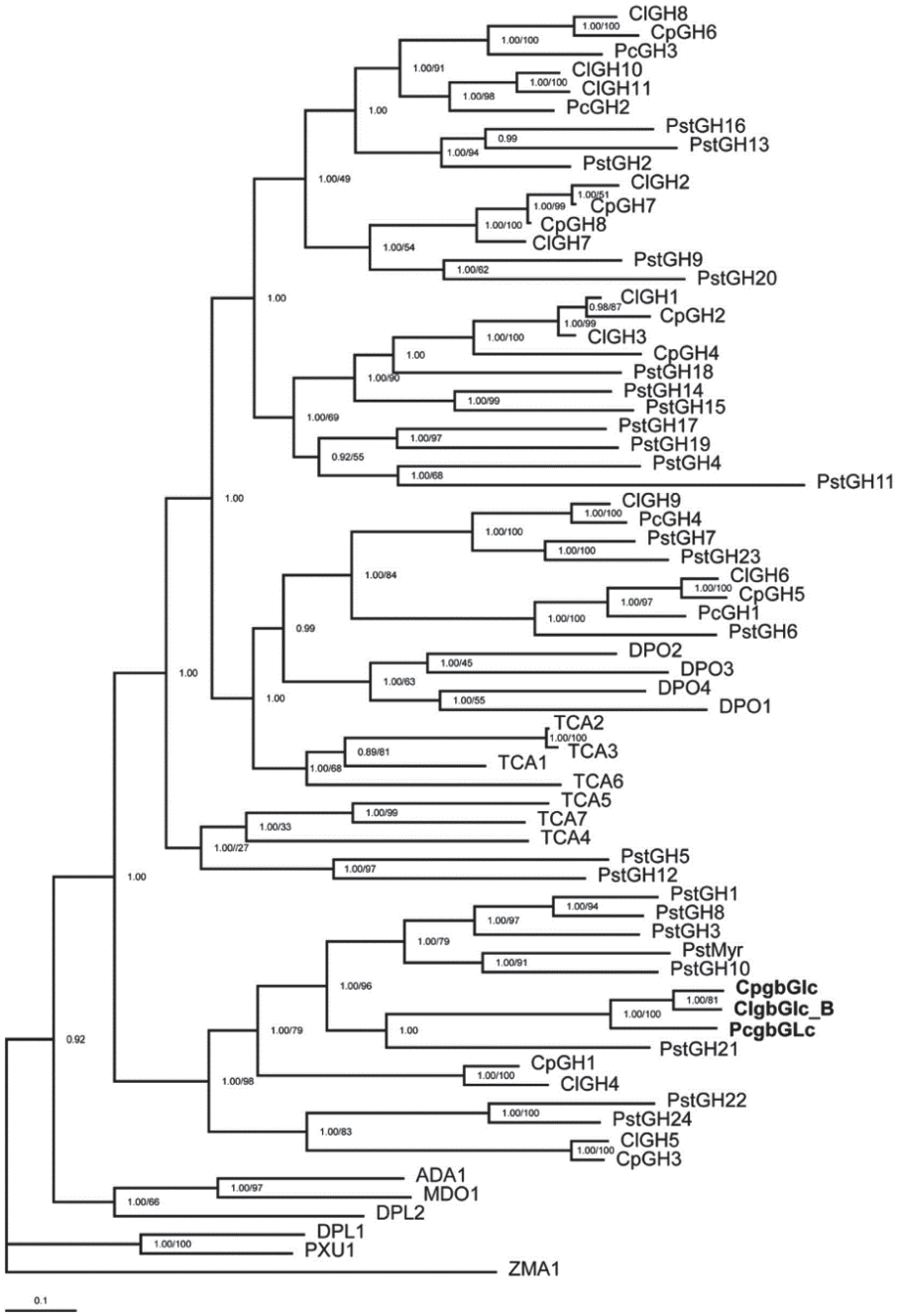

Fig. 6. Phylogeny of Chrysomelina *sensu stricto* glandular β-glucosidases and related glycosyl hydrolases including protein sequences of other insects. The phylogenetic tree was generated using a Bayesian inference method. Posterior probability values are shown next to each node. The second numbers, exemplarily indicated, represent bootstrap values based on a neighbor-joining algorithm. Abbreviations: *Cp* (*C. populi*), *Cl* (*C. lapponica*), *Pc* (*Phaedon cochleariae*), *TCA* (*Tribolium castaneum*), *Pst* (*Phyllotreta striolata*), *DPO* (*Dendroctonus ponderosae*), *DPL* (*Danaus plexippus*), *ADA* (*Anopheles darlingi*), *MDO* (*Musca domestica*), *ZMA* (*Zea mays*), *PXU* (*Papilio xuthus*), GH (glycosyl hydrolase), gbGlc (glandular β-glucosidase), Myr (myrosinase).

The computer-based homology modeling of the substrate binding in the active center of the glandular β-glucosidases explains the high catalytic efficiency of *Cp*gbGlc for 8HGG. *Cp*gbGlc has, in comparison to *Pc*gbGlc or *Cl*gbGlc_B, an aa exchange in position 190 (Fig. 5). The substitution of hydrophilic aa's (E190 and Q190, respectively) by a hydrophobic leucine helps 8HGG to bind efficiently to *Cp*gbGlc. The binding is supported by the hydrophobic effect, which is not present in the other two enzymes. Guided and fixed by the hydrophobic leucine, the hydroxyl group of 8HGG forms hydrogen bonds with R189 and Q244 in *Cp*gbGlc. In contrast, the hydrophilic aa's at position 190 in *Pc*gbGlc and *Cl*gbGlc_B (E190 and Q190, respectively) can serve as alternative hydrogen bond donor/acceptor for the hydroxyl group of 8HGG (Fig. 5, Figs. S7–S9). Hence, the probably more efficient binding of 8HGG also might be the reason why the calculation of the 8HGG binding to *Cp*gbGlc yielded only one potential model while the calculation of the 8HGG binding to *Pc*gbGlc and *Cl*gbGlc_B yielded in each case two potential model. Apart from this effect in the position 190, no additional impact on the enzyme activity is supported for the aa's in the variable positions 186, 190 and 328 of our models. They may have an influence on the mode of substrate binding or of product releasing, which is not covered by the modeling.

Juvenile Chrysomelina evolved barriers with various degrees of selectivity for the translocation of glucosides (be they of exo- or endogenous origin) through the larval body and for the synthesis of deterrents from the glucosides in the defensive glands. For example, uptake from gut lumen into hemolymph and excretion from hemolymph into feces by Malpighian tubules are considered as non-selective barriers (Discher et al., 2009). In the defensive glands non-selective and selective barriers control the metabolic diversity in the secretions, and these, in turn, have an effect on different trophic levels: the composition of secretions can be either beneficial for the insect by affiliating it with its enemy and competitor-free space or it can be harmful, increasing insects' attractiveness to parasites or predators. A recently functionally characterized ABC transporter is a component of the non-selective barriers in the defensive glands. It has been demonstrated to be essential for the vesicular shuttling of glucosides from the inside of the gland associated secretory cells into the reservoir (Strauss et al., 2013). Upstream of this ABC transporter, a selective uptake mechanism of glucosides from hemolymph into the secretory cells was indicated by previous glucoside feeding and injection experiments (Discher et al., 2009; Kuhn et al., 2004; Kunert et al., 2008). The detailed analysis of this barrier is in progress.

Examples of glandular enzymes involved in the deterrent synthesis which contribute to a selective barrier with a tight substrate range are the GMC oxidoreductases; these catalyze the reaction after the deglucosylation (Fig. 1). Here the responsible enzymes, *Pc*8HGO and *Cpop*SAO, were adapted from different clades of the multigene GMC superfamily, probably due to selection pressure from altered deterrent precursors (Rahfeld et al., 2014). In contrast, the activity of the β-glucosidases identified herein adds another non-selective barrier in the processes by which deterrents are produced in Chrysomelina *sensu stricto*. In a nutshell, to counter a variation in their environment or composition of their deterrent metabolism the larvae only have to adjust a few proteins without exchanging the whole set. The three glandular β-glucosidases *Pc*gbGlc, *Cp*gbGlc and *Cl*gbGlc are able to hydrolyze most of the potentially secreted glucosides, and thus perfectly fit to this demand of plasticity during e.g. host plant shifts occurring in the Chrysomelina taxon.

Our *in vitro* experiments suggested that the glandular β-glucosidases are indispensable for deterrent production. The larvae also showed clear phenotypes after β-glucosidase silencing: the secretions of *P. cochleariae* lost the yellow color, as happened when the oxidase was silenced (Rahfeld et al., 2014), whereas the secretions of *C. populi* were occasionally crystallized. The latter effect was due to the high accumulation of salicin, which could not be cleaved in the gland reservoir, up to the point of crystallization. Therefore we propose that the level of available glucosides in the secretions doesn't impede the process of glucoside secretion by the secretory cell into the gland reservoir. Based on our results from *in vitro* and *in vivo* as well as modeling experiments, we assume that generally only one β-glucosidase is located in the secretions, and this β-glucosidase is able to hydrolyze very different glycosides.

The measured concentration of free glucose in the secretions was similar to already described measurements performed for *C. populi* (210 μmol/l) (Pasteels et al., 1983) and decreased after the knock-down of the glandular β-glucosidases. This reveals that the cleavage of imported glycosides is the only source of free glucose in the secretions. The glucose left over after silencing was only due to residual β-glucosidase activity in the secretions.

The observations of a low surface tension of the RNAi secretions (no drop formation after the everting of the glands) led us to conclude that one potential function of the glucose in the secretions, which reached a concentration of ~5–6%, could be droplet formation, a well-known characteristic of glucose (Hoorfar et al., 2006).

Phylogenetic analysis of chosen insect GH family members showed a species-specific diversification commonly observed in this enzyme family (Beran et al., 2014). Remarkably, the chrysomeline *Pc*gbGlc, *Cp*gbGlc and *Cl*gbGlc_B are true orthologs and cluster together with the myrosinase from *P. striolata*, a β-thioglucosidase involved in the chemical defense of this flea beetle, and other GH homologs from this species (Fig. 6) (Beran et al., 2014). Here a hypothesis can be advanced that this cluster might be related to GH1s with a function outside of digestion. To validate this particular clade, in the future more sequences of GHs have to be functionally analyzed. Our data indicate that β-glucosidases are important components in the secondary metabolism for essential signals in insect defense or communication.

Acknowledgments

We sincerely thank Dr. Maritta Kunert and Dr. Stefan von Reuß for their support in chemicals analyses, Dr. Magdalena Stock for her support in bioinformatics, and Prof. Jacques M. Pasteels for the fruitful discussions and his everlasting interest in our projects. The authors would like to express their gratitude to Angelika Berg, Ivonne Goerlich, and Yvonne Hupfer for technical assistance. Very special thanks are due to Alija Gabdullina, Katon-Karagay State National Park, Kazakhstan, for her support during sample collection of *C. lapponica*. We also like to thank Emily Wheeler for her critical reading of the manuscript. This work was financially supported by the Max Planck Society and the German Research Foundation (BU 1862/2-1).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ibmb.2015.01.003.

References

Beran, F., Pauchet, Y., Kunert, G., Reichelt, M., Wielsch, N., Vogel, H., Reinecke, A., Svatos, A., Mewis, I., Schmid, D., Ramasamy, S., Ulrichs, C., Hansson, B.S., Gershenzon, J., Heckel, D.G., 2014. *Phyllotreta striolata* flea beetles use host plant defense compounds to create their own glucosinolate–myrosinase system. Proc. Natl. Acad. Sci. U. S. A. 111, 7349–7354.

Berenbaum, M.R., 1995. The chemistry of defense – theory and practice. Proc. Natl. Acad. Sci. U. S. A. 92, 2–8.

Blum, M.S., Wallace, J.B., Duffield, R.M., Brand, J.M., Fales, H.M., Sokoloski, E.A., 1978. Chrysomelidial in defensive secretion of leaf beetle *Gastrophysa cyanea* Melsheimer. J. Chem. Ecol. 4, 47–53.

Bodemann, R.R., Rahfeld, P., Stock, M., Kunert, M., Wielsch, N., Groth, M., Frick, S., Boland, W., Burse, A., 2012. Precise RNAi-mediated silencing of metabolically active proteins in the defence secretions of juvenile leaf beetles. Proc. R. Soc. Biol. Sci. 279, 4126–4134.

Brueckmann, M., Termonia, A., Pasteels, J.M., Hartmann, T., 2002. Characterization of an extracellular salicyl alcohol oxidase from larval defensive secretions of *Chrysomela populi* and *Phratora vitellinae* (Chrysomelina). Insect Biochem. Mol. Biol. 32, 1517–1523.

Cairns, J.R.K., Esen, A., 2010. beta-Glucosidases. Cell. Mol. Life Sci. 67, 3389–3405.

Discher, S., Burse, A., Tolzin-Banasch, K., Heinemann, S.H., Pasteels, J.M., Boland, W., 2009. A versatile transport network for sequestering and excreting plant glycosides in leaf beetles provides an evolutionary flexible defense strategy. ChemBioChem 10, 2223–2229.

Dobler, S., 2001. Evolutionary aspects of defense by recycled plant compounds in herbivorous insects. Basic Appl. Ecol. 2, 15–26.

Eyun, S.I., Wang, H.C., Pauchet, Y., Ffrench-Constant, R.H., Benson, A.K., Valencia-Jimenez, A., Moriyama, E.N., Siegfried, B.D., 2014. Molecular evolution of glycoside hydrolase genes in the western corn rootworm (*Diabrotica virgifera virgifera*). PLoS One 9.

Franzl, S., Ackermann, I., Nahrstedt, A., 1989. Purification and characterization of a beta-glucosidase (Linamarase) from the hemolymph of Zygaena-Trifolii Esper, 1783 (Insecta, Lepidoptera). Experientia 45, 712–718.

Gross, J., Fatouros, N.E., Neuvonen, S., Hilker, M., 2004. The importance of specialist natural enemies for *Chrysomela lapponica* in pioneering a new host plant. Ecol. Entomol. 29, 584–593.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., Davies, G., 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc. Natl. Acad. Sci. U. S. A. 92, 7090–7094.

Henrissat, B., Davies, G., 1997. Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. 7, 637–644.

Hilker, M., Schulz, S., 1994. Composition of larval secretion of *Chrysomela lapponica* (Coleoptera, Chrysomelidae) and its dependence on host-plant. J. Chem. Ecol. 20, 1075–1093.

Hoorfar, M., Kurz, M.A., Policova, Z., Hair, M.L., Neumann, A.W., 2006. Do polysaccharides such as dextran and their monomers really increase the surface tension of water? Langmuir 22, 52–56.

Jeng, W.Y., Wang, N.C., Lin, C.T., Chang, W.J., Liu, C.I., Wang, A.H.J., 2012. High-resolution structures of *Neotermes koshunensis* beta-glucosidase mutants provide insights into the catalytic mechanism and the synthesis of glucoconjugates. Acta Crystallogr. D. 68, 829–838.

Jeng, W.Y., Wang, N.C., Lin, M.H., Lin, C.T., Liaw, Y.C., Chang, W.J., Liu, C.I., Liang, P.H., Wang, A.H.J., 2011. Structural and functional analysis of three beta-glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. J. Struct. Biol. 173, 46–56.

Jones, A.M.E., Bridges, M., Bones, A.M., Cole, R., Rossiter, J.T., 2001. Purification and characterisation of a non-plant myrosinase from the cabbage aphid *Brevicoryne brassicae* (L.). Insect Biochem. Mol. Biol. 31, 1–5.

Jones, A.M.E., Winge, P., Bones, A.M., Cole, R., Rossiter, J.T., 2002. Characterization and evolution of a myrosinase from the cabbage aphid *Brevicoryne brassicae*. Insect Biochem. Mol. Biol. 32, 275–284.

Kazana, E., Pope, T.W., Tibbles, L., Bridges, M., Pickett, J.A., Bones, A.M., Powell, G., Rossiter, J.T., 2007. The cabbage aphid: a walking mustard oil bomb. Proc. R. Soc. B 274, 2271–2277.

Keeling, C.I., Yuen, M.M.S., Liao, N.Y., Docking, T.R., Chan, S.K., Taylor, G.A., Palmquist, D.L., Jackman, S.D., Nguyen, A., Li, M., Henderson, H., Janes, J.K., Zhao, Y.J., Pandoh, P., Moore, R., Sperling, F.A.H., Huber, D.P.W., Birol, I., Jones, S.J.M., Bohlmann, J., 2013. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. Genome Biol. 14.

Kirsch, R., 2011. Identification and Characterization of Defense Related Enzymes in Chrysomelina Larvae (Coleoptera: Chrysomelidae): Contributions to Understand the Evolutionary and Molecular Dynamics of Chemical Defense in Leaf Beetles. Friedrich-Schiller-Universität, Jena, Germany.

Kirsch, R., Vogel, H., Muck, A., Vilcinskas, A., Pasteels, J.M., Boland, W., 2011. To be or not to be convergent in salicin-based defence in chrysomeline leaf beetle larvae: evidence from *Phratora vitellinae* salicyl alcohol oxidase. Proc. R. Soc. Biol. Sci. 278, 3225–3232.

Kuhn, J., Pettersson, E.M., Feld, B.K., Burse, A., Termonia, A., Pasteels, J.M., Boland, W., 2004. Selective transport systems mediate sequestration of plant glucosides in leaf beetles: a molecular basis for adaptation and evolution. Proc. Natl. Acad. Sci. U. S. A. 101, 13808–13813.

Kunert, M., Soe, A., Bartram, S., Discher, S., Tolzin-Banasch, K., Nie, L., David, A., Pasteels, J., Boland, W., 2008. De novo biosynthesis versus sequestration: a network of transport systems supports in iridoid producing leaf beetle larvae both modes of defense. Insect Biochem. Mol. Biol. 38, 895–904.

Laurent, P., Braekman, J.C., Daloze, S., 2005. Insect chemical defense. Top. Curr. Chem. 240, 167–229.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., Henrissat, B., 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol. Biofuels 6.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495.

Meinwald, J., Jones, T.H., Eisner, T., Hicks, K., 1977. Defense-mechanisms of arthropods. 56. New methylcyclopentanoid terpenes from larval defensive secretion of a chrysomelid beetle (Plagiodera-Versicolora). Proc. Natl. Acad. Sci. U. S. A. 74, 2189–2193.

Mithoefer, A., Boland, W., 2012. Plant defense against herbivores: chemical aspects. Annu. Rev. Plant Biol. 63, 431–450.

Oldham, N.J., Veith, M., Boland, W., Dettner, K., 1996. Iridoid monoterpene biosynthesis in insects: evidence for a de novo pathway occurring in the defensive glands of *Phaedon armoraciae* (Chrysomelidae) leaf beetle larvae. Naturwissenschaften 83, 470–473.

Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A., Cairns, J.R.K., 2006. Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 beta-glucosidase. BMC Plant Biol. 6.

Pasteels, J.M., Braekman, J.C., Daloze, D., Ottinger, R., 1982. Chemical defense in chrysomelid larvae and adults. Tetrahedron 38, 1891–1897.

Pasteels, J.M., Duffey, S., Rowell-Rahier, M., 1990. Toxins in chrysomelid beetles – possible evolutionary sequence from de novo synthesis to derivation from food-plant chemicals. J. Chem. Ecol. 16, 211–222.

Pasteels, J.M., Rowell-Rahier, M., Braekman, J.C., Daloze, D., Duffey, S., 1989. Evolution of exocrine chemical defense in leaf beetles (Coleoptera, Chrysomelidae). Experientia 45, 295–300.

Pasteels, J.M., Rowell-Rahier, M., Braekman, J.C., Dupont, A., 1983. Salicin from host plant as precursor of salicylaldehyde in defensive secretion of chrysomeline larvae. Physiol. Entomol. 8, 307–314.

Pentzold, S., Zagrobelny, M., Roelsgaard, P.S., Moeller, B.L., Bak, S., 2014. The multiple strategies of an insect herbivore to overcome plant cyanogenic glucoside defence. PLoS One 9.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. Nucleic Acids Res. 40, D290–D301.

Rahfeld, P., Kirsch, R., Kugel, S., Wielsch, N., Stock, M., Groth, M., Boland, W., Burse, A., 2014. Independently recruited oxidases from the glucose–methanol–choline oxidoreductase family enabled chemical defences in leaf beetle larvae (subtribe Chrysomelina) to evolve. Proc. R. Soc. B 281.

Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W., Gibbs, R., Bucher, G., Friedrich, M., Grimmelikhuijzen, C.J.P., Klingler, M., Lorenzen, M.D., Roth, S., Schroder, R., Tautz, D., Zdobnov, E.M., Muzny, D., Attaway, T., Bell, S., Buhay, C.J., Chandrabose, M.N., Chavez, D., Clerk-Blankenburg, K.P., Cree, A., Dao, M., Davis, C., Chacko, J., Dinh, H., Dugan-Rocha, S., Fowler, G., Garner, T.T., Garnes, J., Gnirke, A., Hawes, A., Hernandez, J., Hines, S., Holder, M., Hume, J., Jhangiani, S.N., Joshi, V., Khan, Z.M., Jackson, L., Kovar, C., Kowis, A., Lee, S., Lewis, L.R., Margolis, J., Morgan, M., Nazareth, L.V., Nguyen, N., Okwuonu, G., Parker, D., Ruiz, S.J., Santibanez, J., Savard, J., Scherer, S.E., Schneider, B., Sodergren, E., Vattahil, S., Villasana, D., White, C.S., Wright, R., Park, Y., Lord, J., Oppert, B., Brown, S., Wang, L.J., Savard, J., Liu, Y., Worley, K., Elsik, C.G., Reese, J.T., Elhaik, E., Landan, G., Graur, D., Arensburger, P., Atkinson, P., Beidler, J., Demuth, J.P., Drury, D.W., Du, Y.Z., Fujiwara, H., Maselli, V., Osanai, M., Robertson, H.M., Tu, Z., Wang, J.J., Wang, S.Z., Song, H., Zhang, L., Sodergren, E., Werner, D., Stanke, M., Morgenstern, B., Solovyev, V., Kosarev, P., Brown, G., Chen, H.C., Ermolaeva, O., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Maglott, D., Pruitt, K., Sapojnikov, V., Souvorov, A., Mackey, A.J., Waterhouse, R.M., Wyder, S., Zdobnov, E.M., Kriventseva, E.V., Kadowaki, T., Bork, P., Aranda, M., Bao, R.Y., Beermann, A., Berns, N., Bolognesi, R., Bonneton, F., Bopp, D., Butts, T., Chaumot, A., Denell, R.E., Ferrier, D.E.K., Gordon, C.M., Jindra, M., Klingler, M., Lan, Q., Lattorff, H.M.G., Laudet, V., von Levetsow, C., Liu, Z.Y., Lutz, R., Lynch, J.A., da Fonseca, R.N., Posnien, N., Reuter, R., Roth, S., Schinko, J.B., Schmitt, C., Schoppmeier, M., Shippy, T.D., Simonnet, F., Marques-Souza, H., Tomoyasu, Y., Trauner, J., Van der Zee, M., Vervoort, M., Wittkopp, N., Wimmer, E.A., Yang, X.Y., Jones, A.K., Sattelle, D.B., Ebert, P.R., Nelson, D., Scott, J.G., Muthukrishnan, S., Kramer, K.J., Arakane, Y., Zhu, Q.S., Hogenkamp, D., Dixit, R., Jiang, H.B., Zou, Z., Marshall, J., Elpidina, E., Vinokurov, K., Oppert, C., Evans, J., Lu, Z.Q., Zhao, P.C., Sumathipala, N., Altincicek, B., Vilcinskas, A., Williams, M., Hultmark, D., Hetru, C., Hauser, F., Cazzamali, G., Williamson, M., Li, B., Tanaka, Y., Predel, R., Neupert, S., Schachtner, J., Verleyen, P., Raible, F., Walden, K.K.O., Robertson, H.M., Angeli, S., Foret, S., Schuetz, S., Maleszka, R., Miller, S.C., Grossmann, D., Consortium, T.G.S., 2008. The genome of the model beetle and pest *Tribolium castaneum*. Nature 452, 949–955.

Scharf, M.E., Kovaleva, E.S., Jadhao, S., Campbell, J.H., Buchman, G.W., Boucias, D.G., 2010. Functional and translational analyses of a beta-glucosidase gene (glycosyl hydrolase family 1) isolated from the gut of the lower termite *Reticulitermes flavipes*. Insect Biochem. Mol. Biol. 40, 611–620.

Soetens, P., Pasteels, J.M., Daloze, D., 1993. A simple method for in-vivo testing of glandular enzymatic-activity on potential precursors of larval defensive compounds in *Phratora*-species (Coleoptera, Chrysomelinae). Experientia 49, 1024–1026.

Stock, M., Gretscher, R., Groth, M., Eiserloh, S., Boland, W., Burse, A., 2013. Transcriptome-wide analysis of the SLC2 sugar transporter family in the mustard leaf beetle *Phaedon cochleariae*. PLoS One 8, e84461.

Strauss, A.S., Peters, S., Boland, W., Burse, A., 2013. ABC transporter functions as a pacemaker for sequestration of plant glucosides in leaf beetles. Elife 2, e01096.

Sugawara, F., Matsuda, K., Kobayashi, A., Yamashita, K., 1979. Defensive secretion of chrysomelid larvae *Gastrophysa atrocyanea* Motschulsky and *Phaedon brassicae* Baly. J. Chem. Ecol. 5, 635–641.

Termonia, A., Hsiao, T.H., Pasteels, J.M., Milinkovitch, M.C., 2001. Feeding specialization and host-derived chemical defense in chrysomeline leaf beetles did not lead to an evolutionary dead end. Proc. Natl. Acad. Sci. U. S. A. 98, 3909–3914.

Termonia, A., Pasteels, J.M., 1999. Larval chemical defence and evolution of host shifts in *Chrysomela* leaf beetles. Chemoecology 9, 13–23.

Terra, R.W., 2012. Biochemistry and molecular biology of digestion. Insect Mol. Biol. Biochem. 365–418.

Tokuda, G., Saito, H., Watanabe, H., 2002. A digestive beta-glucosidase from the salivary glands of the termite, *Neotermes koshunensis* (Shiraki): distribution, characterization and isolation of its precursor cDNA by 5′- and 3′-RACE amplifications with degenerate primers. Insect Biochem. Mol. Biol. 32, 1681–1689.

Tolzin-Banasch, K., Dagvadorj, E., Sammer, U., Kunert, M., Kirsch, R., Ploss, K., Pasteels, J.M., Boland, W., 2011. Glucose and glucose esters in the larval secretion of *Chrysomela lapponica*; selectivity of the glucoside import system from host plant leaves. J. Chem. Ecol. 37, 195–204.

Tribolo, S., Berrin, J.G., Kroon, P.A., Czjzek, M., Juge, N., 2007. The crystal structure of human cytosolic beta-glucosidase unravels the substrate aglycone specificity of a family 1 glycoside hydrolase. J. Mol. Biol. 370, 964–975.

Veith, M., Dettner, K., Boland, W., 1996. Stereochemistry of an alcohol oxidase from the defensive secretion of larvae of the leaf beetle *Phaedon armoraciae* (Coleoptera: Chrysomelidae). Tetrahedron 52, 6601–6612.

Veith, M., Lorenz, M., Boland, W., Simon, H., Dettner, K., 1994. Biosynthesis of iridoid monoterpenes in insects – defensive secretions from larvae of leaf beetles (Coleoptera, Chrysomelidae). Tetrahedron 50, 6859–6874.

Verdoucq, L., Moriniere, J., Bevan, D.R., Esen, A., Vasella, A., Henrissat, B., Czjzek, M., 2004. Structural determinants of substrate specificity in family 1 beta-glucosidases – novel insights from the crystal structure of sorghum dhurrinase-1, a plant beta-glucosidase with strict specificity, in complex with its natural substrate. J. Biol. Chem. 279, 31796–31803.

Xu, Z.W., Escamilla-Trevino, L.L., Zeng, L.H., Lalgondar, M., Bevan, D.R., Winkel, B.S.J., Mohamed, A., Cheng, C.L., Shih, M.C., Poulton, J.E., Esen, A., 2004. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. Plant Mol. Biol. 55, 343–367.

Zagrobelny, M., Bak, S., Moeller, B.L., 2008. Cyanogenesis in plants and arthropods. Phytochemistry 69, 1457–1468.

Zagrobelny, M., Moeller, B.L., 2011. Cyanogenic glucosides in the biological warfare between plants and insects: the Burnet moth-Birdsfoot trefoil model system. Phytochemistry 72, 1585–1592.

4.3 Manuscript 3: "Selective adaptation within the chemosensory system of the leaf beetle, *Chrysomela lapponica*, following host plant shift"

Selective adaptation within the chemosensory system of the leaf beetle, *Chrysomela lapponica*, following host plant shift

Ding Wang[1], Stefan Pentzold[1], Maritta Kunert[1], Marco Groth[2], Wolfgang Brandt[3], Jacques M. Pasteels[4], Wilhelm Boland[1], Antje Burse[1*]

[1] Max Planck Institute for Chemical Ecology, Department of Bioorganic Chemistry, Hans-Knöll-Str. 8, D-07745 Jena, Germany

[2] Leibniz Institute on Aging – Fritz Lipmann Institute, Beutenbergstraße 11, D-07745 Jena, Germany

[3] Leibniz Institute of Plant Biochemistry, Weinberg 3, D-06120 Halle (Saale), Germany

[4] Université Libre de Bruxelles, Department of Biology, Av. F.D. Roosevelt, 50, B-1050 Brussels, Belgium

*Corresponding author. Phone: ++49(0) 3641-571265, Fax: ++49(0) 3641-571202, E-mail: aburse@ice.mpg.de

ABSTRACT

Due to its fundamental relevance to shape host selection behavior, we have analyzed the chemosensory repertoire of the specialized leaf beetle, *Chrysomela lapponica,* which evolved distinct populations that shifted host from willow (*Salix* sp., Salicaceae) to birch (*Betula rotundifolia*, Betulaceae). Our analysis of volatiles showed that the two plant species differed regarding their ratio of mainly terpenoids, such as (*E,E*)-α-farnesene and 4,8-dimethylnona-1,3,7-triene, chemical cues potentially used by the beetles to discriminate their hosts. We identified 113 chemosensory candidate genes in *C. lapponica*: 31 odor binding proteins (OBPs), 12 chemo sensory proteins, 4 sensory neuron membrane proteins, 41 olfactory receptors (ORs), 8 gustatory receptors, and 17 ionotropic receptors by RNA-seq. Differential expression analyses of the antennal chemosensory genes of adult beetles revealed upregulation of the minus-C OBP *Clap*OBP27 in the willow-feeders. In comparison, the two minus-C OBPs, *Clap*OBP20 and *Clap*OBP02, and the OR *Clap*OR17 were upregulated in the birch-feeders. These differences have also been mirrored in the transcript levels of the adult legs from both populations. Homology modeling of the three minus-C OBPs showed a varying surface polarity of the ligand binding pockets. Accordingly, affinity calculations indicated that the non-polar *Clap*OBP27 binds preferentially hydrophobic plant-derived terpenoids (e.g. (*E,E*)-α-farnesene) while *Clap*OBP20 can accommodate also more hydrophilic phytochemicals (e.g. cis-3-hexenyl acetate). We suggest that the host plant shift of *C. lapponica* through a loss of host preference and a gain of tolerance to the phytochemicals of birch as novel host plant is accompanied by the evolution of minus-C OBPs and ORs.

INTRODUCTION

Phytophagous beetles have undergone a 140 million years lasting period of coevolution and coadaptation with their host plants (Labandeira & Currano 2013; Wang *et al.* 2013). Nowadays, it is debated how such an interaction itself diversifies and how the interaction affects the modes and rates of the speciation of beetles and plants (Barrett & Heil 2012; Futuyma & Agrawal 2009; Tilmon 2008). The most successful herbivorous beetle lineages (Curculionoidea and Chrysomeloidea), forming the clade "Phytophaga", have developed different patterns of life history strategies to exploit their plant diet effectively (Farrell 1998; Farrell & Sequeira 2004;

Fernandez & Hilker 2007; Gómez-Zurita *et al.* 2007). Plant secondary metabolites that are often used in anti-herbivore defense are key players in this relationship. As insects have to adapt to their noxious effect, plant compounds can contribute to the specialization of phytophagous beetles to their host plants (Mithöfer & Boland 2012). Although the adaptation to plant metabolites promotes diet conservatism, it does not lead inevitably to evolutionary dead ends (Day *et al.* 2016; Termonia *et al.* 2001). Over ecological and evolutionary time scales, both plant and insect herbivores may change their geographic range generating novel plant-herbivore interactions including host plant shifts.

Host plant shifts have also occurred during the evolutionary history of the leaf beetle subtribe Chrysomelina (Chrysomelidae, Chrysomeloidea). Some species of the monophyletic *interrupta*-group escaped the plant constraints by shifting host plant families (Termonia *et al.* 2001). In the species *Chrysomela lapponica*, for example, geographically separated populations have evolved that colonize either willow (*Salix* spp.; Salicaceae) or birch (*Betula* spp.; Betulaceae) in the Eurasian Palearctic (Geiselhardt *et al.* 2015; Zvereva *et al.* 2016). The recent reconstructions of the host-shift history of allopatric *C. lapponica* populations disentangled that willow is the ancestral feeding niche and that the transition to Betulaceae occurred several times independently, possibly after the last glacial episode, i.e. during the last 10,000 years (Mardulyn *et al.* 2011). Besides resistance traits of host plants, also enemy effectors emerge as central players determining host affiliation. Ecological studies carried out on the different *C. lapponica* populations revealed that the juveniles of the willow inhabitants are frequently exposed to specialized parasitoids and predators while the grazers on birch escaped this top-down pressure and occupied thus an enemy-free space (Gross *et al.* 2004b; Zvereva & Rank 2003, 2004).

In the affiliations of novel host plants the insect chemosensory system represents the first barrier to be overcome (del Campo *et al.* 2001). Based on nutritional and secondary metabolites this system discriminates among a mosaic of different plant species and triggers physiological processes and an appropriate feeding behavior (Dahanukar *et al.* 2005). The evolution of the sensory repertoire could provoke and reinforce adaptations of insects to new hosts. The populations of *C. lapponica* selecting different host plant families represent an excellent model system to investigate the role of the chemosensory system during and after host plant shifts by herbivorous insects.

In all insects two major chemosensory mechanisms, the sense of smell and the sense of taste, largely orchestrate host plant selection. While olfaction is a sensor enabling insects to

89 detect and discriminate between numerous volatile molecules, taste is the sensory impression of
90 mainly nonvolatile substances (Van Naters & Carlson 2006). The reception of chemical cues
91 from the environment is mediated by peripheral receptor neurons residing in peripheral organs
92 such as antennae, maxillary palps, or legs which are covered by hair-like sensilla (Hallem *et al.*
93 2006). The sensilla house the dendrites of a varying number of these neurons which project into
94 the central nervous system (Yarmolinsky *et al.* 2009). In both senses, the receptor proteins of
95 peripheral neurons play a pivotal role as biological transducers that convert external chemical
96 signals into a sensory input. These receptors can signal through G-protein dependent second
97 messenger cascades or operate as ligand-gated ion-channels (Sato *et al.* 2008; Sato *et al.* 2011;
98 Wicher *et al.* 2008; Zhang *et al.* 2011).
99 The members of the olfactory receptor (OR) family are known to function as
100 heteromultimeres composed of a ligand-specific OR and an OR co-receptor (Orco) (Leal 2013).
101 ORs act typically in concert with the soluble odorant binding proteins (OBPs) or chemosensory
102 proteins (CSPs) (Pelosi *et al.* 2014). These proteins capture hydrophobic odorants or pheromones
103 in the sensillar lymph and lead subsequently to the activation of the membrane integrated
104 receptors. Other classes of receptors involved in chemosensation comprise the ionotropic
105 receptors (IRs) (Rytz *et al.* 2013) and sensory neuron membrane proteins (SNMPs) (Vogt *et al.*
106 2009). Insects sense CO_2 and hydrophilic tastants via gustatory receptors (GRs). GRs can form
107 hetero-multimeres functioning without accessory soluble proteins (Zhang *et al.* 2011). GRs share
108 motives with ORs in their transmembrane domains, and phylogenetically they are suggested to
109 predate the expansion of the insect ORs (Missbach *et al.* 2014). As taste neurons express a
110 manifold subset of GRs, the design principles of taste are much less understood to date than the
111 one receptor-one neuron system facilitating olfaction.
112 In the present study, we test the hypothesis if a host plant shift of the specialist
113 phytophagous beetle species, *C. lapponica*, is accompanied by the modulation of the
114 chemosensory repertoire. For this reason, we present a comparative inventory of the
115 chemosensory systems based on transcriptome sequences from two Kazakh *C. lapponica*
116 populations that differ in their host plant preference for either willow or birch. By using RNA-
117 sequencing, we studied the expression profiles of the chemosensory components in antennae and
118 legs in males and females of the two populations and propose a function mainly of OBPs and
119 ORs in host plant adaptation.
120

RESULTS

Volatile composition of willow and birch

Host selection behavior in phytophagous insects is triggered in the first place by volatile plant compounds before nutritional and toxic plant factors become effective. To initiate host plant shifts, these volatiles may therefore be of critical importance. To elucidate these chemical cues likely manipulating *C. lapponica* host finding, we have analyzed the volatile bouquet of *Betula rotundifolia* and *Salix* sp. host plants. As insects face in nature usually plants that are challenged by biotic or abiotic stressors we compared the scents of untreated leaves with those treated with coronalon or with those wounded mechanically. Coronalon is a known synthetic elicitor of the plant response against herbivory, microbial infection and of stress physiology (Lauchli *et al.* 2002; Schüler *et al.* 2004).

Our analyses revealed that in general the volatile emission was much lower in the untreated than in the coronalon-treated or wounded leaf material of both plant species (Table S1, Supporting information). With one exception, both plant species surprisingly shared a qualitative similar volatile pattern. However, they appeared to differ in the quantity of their emitted compounds. The willow species released as the major component of its volatile bouquet more (*E,E*)-α-farnesene (17), and as minor components more sabinene (2), *E*-myroxide (10), germacrene D (15), and (*Z,E*)-α-farnesene (16) in comparison to the birch leaf samples (Fig. 1). A minor amount of salicylaldehyde has been detected exclusively in willow and not in birch samples. Further, volatiles emitted by the birch samples could all be detected in the willow species. Quantitatively, however, we found an increased release of DMNT (4,8-dimethylnona-1,3,7-triene) (18) as major component, as well as of eucalyptol (5), (*Z*)-ß-ocimene (6), ß-caryophyllene (13), α-humulene (14), and linalool (9) as minor components in comparison to the willow volatile emission.

Green leaf volatiles (C_6 aldehydes, alcohols and esters derived from fatty acids) are emitted in trace amounts from undamaged plant tissue, while they are emitted promptly after cell disruption (D'Auria *et al.* 2007; Turlings *et al.* 1995). As expected, artificially wounded leaves of the two plant species produced more green leaf volatiles than the coronalon-treated or untreated leaf samples. Both species had the compound pattern in common, but produced different amounts of the green leaf volatiles. In general, willow released higher amounts than birch (Table S1, Supporting information).

153 *Construction of transcriptome libraries of birch- and willow-adapted C. lapponica*

154 Given the quantitative differences in the volatile pattern of willow and birch host plants, we
155 asked whether these differences are mirrored in the chemosensory system of the two *C. lapponica*
156 populations. In detail, we have structured our study around the classical organ for the perception
157 of smell, the antennae. From the external morphology antennae typically house porous
158 chemosensory sensilla as it is the case in *C. lapponica* (Fig. 2Aa). According to the number of
159 pores in their cuticle we could distinguish two sensilla types on the antennae: (*i*) uniporous
160 sensilla (considered as gustatory chemosensilla or chemo-mechanosensilla, Fig. 2Ab), and (*ii*)
161 multiporous sensilla (considered as olfactory chemosensilla, Fig. 2c) (Zacharuk & Shields 1991).
162 In order to define the identified chemosensory components as characteristic for the principal
163 olfactory organ of *C. lapponica*, we have contrasted our results from the antennae with those of
164 the legs. Though the legs of *C. lapponica* are also covered by various types of sensilla, they seem
165 to lack the multiporous types (Fig. 2B, C). Instead we found sensilla that contain a single opening
166 on their tips (Fig. 2Cd). The observed external morphology of sensilla gives us the hint that smell
167 and taste may possibly be perceived differently in antenna and legs of *C. lapponica.*

168 For creating a catalogue of chemosensory genes, we have sequenced cDNA derived from
169 pooled individuals from different developmental stages as well as from two biological replicates
170 each of *C. lapponica's* separated population- and sex-specific antennae and legs. The resulting
171 raw sequence data are listed in Table S2, Supporting information. For our transcriptome reference
172 libraries we obtained 31,612 assembled cDNAs (contigs) with an average length of approx. 1,260
173 bp and an N50 length of 2,048 bp in willow-feeding specialists; 34,154 contigs with an average
174 length of approx. 1,166 bp and an N50 length of 1,904 bp in birch-feeding specialists.

175

176 *Identification of putative chemosensory proteins in birch- and willow-adapted C. lapponica*

177 The created transcriptome reference libraries from both populations were separately translated
178 into all possible amino acid sequences and subject to motif- and sequence alignment-based
179 database searches. Further, we integrated our annotated sequences and homologous sequences
180 from other insects into our phylogenetic trees to group the putative *C. lapponica* proteins, to
181 characterize their relationships and also to propose a function for these beetle proteins. The
182 identified putative binding proteins and receptors of the chemosensory system are listed
183 separately in Table 1 for each of the *C. lapponica* populations.

OBPs represent small (10 to 30 kDa), globular and soluble proteins that contribute to the sensitivity of the olfactory system by transporting odorants through the sensillar lymph. Based on our analysis we identified a total of 31 OBPs in the sequence library of each of the *C. lapponica* populations. Sequence comparison by using blastp showed that the putative OBPs from the two populations share a sequence homology of more than 96%. Except *Clap*OBP22, all of the OBPs represent full-length proteins. Despite conserved protein features, including the signal peptide, the six α-helix domains and the cysteine motives, the *C. lapponica* OBP family members were divergent in terms of protein length (131-263 amino acids) and cysteine profiles.

On the basis of distinctive structural features and phylogenetic relationships we identified four main subgroups of OBPs: classic, antenna binding protein II (ABPII), plus-C, and minus-C (Fig. 3). In accordance with previous phylogenetic analyses (Andersson *et al.* 2013; Dippel *et al.* 2014) we could show that the basal OBP group seems to be the classic, whereas all other groups are internal clades of this subfamily. In our tree the subgroups ABPIIs and minus-C OPBs appeared to have had independent origins. Further, we see mostly lineage specific expansions, particularly in minus-C and plus-C subgroups. Only two classic OBP genes were found with clear orthology relationships across the insects tested in our study: *Obp29* and *Obp10*; a finding which may indicate a conserved function for these genes.

In each of the two *C. lapponica* populations we found six classic OBPs and five ABPIIs. The characteristic hallmark of these proteins are six cysteine residues at conserved positions with a C-pattern of $C1X_{23\text{-}40}C2X_3C3X_{38\text{-}44}C4X_{8\text{-}21}C5X_8C6$ (Fig. 4A) (Xu *et al.* 2009). Unlike in *T. castaneum* or in *D. melanogaster*, we could not find an expansion of *C. lapponica* classic OBPs. Among the classic OBPs, the 243 amino acid long *Clap*OBP29 featured a modified C-pattern that had three additional cysteine residues instead of C1 in the above mentioned C-pattern. Its orthologues possess also unusually long sequence and a unique cysteine pattern. A particular function, however, has not been identified for these unusual OBPs as yet.

The subgroup of ABPIIs forms three clades. In one clade, the characterized OBP Lush of *D. melanogaster*, crucial for detecting pheromones (Xu *et al.* 2005), is localized together with *Clap*OBP01 and *Clap*OBP30 (Fig. 3, bootstrap value of 87%) (Vieira & Rozas 2011). Remarkably, another clade is formed not only by ABPIIs but contains also four *C. lapponica* minus-C OBPs.

In contrast to the classic OBPs, minus-C OBPs lack the second and the fifth conserved cysteine residues (Fan *et al.* 2011). Twenty of all the predicted OBPs in *C. lapponica* comprised a

motif of the minus-C OBPs, $C1X_{28-34}C2X_{35-39}C3X_{16-22}C_4$ (Fig. 4B). Two of the minus-C OBPs (*Clap*OBP06 262 amino acids, *Clap*OBP19 263 amino acids) contained a dimer minus-C pattern. Most of the minus-C OBPs are localized in two distinct clusters and only a few are scattered across the phylogeny. The minus-C OBPs from *D. melanogaster* were all clustered in a branch separated from the beetles' sequences. Plus-C OBPs have not been identified from *C. lapponica* in our study.

Insect **CSPs** are also small, globular and soluble in the sensillar lymph, but smaller (about 100-120 amino acids) and more conserved compared to OBPs (Pelosi *et al.* 2014). In each of the two *C. lapponica* populations, we identified 12 CSPs. The sequence comparison of both CSP sets revealed that the CSP pairs shared at least 93% amino acid identity. All CSP candidates represented full-length proteins showing the conserved C-pattern, $C1X_6C2X_{18}C3X_2C4$ (Fig. 4C) (Xu *et al.* 2009). Among all the *C. lapponica* CSPs, the candidate *Clap*CSP11 contained with 283 amino acids the longest amino acid sequence. Bootstrapping (Fig. 5) revealed a clade of CSPs with a value of 100% that included only the longest CSPs, *Clap*CSP11, *Tcas*CSP6 (251 amino acids) and *Ityp*CSP4 (214 amino acids).

SNMPs are membrane proteins representing essential components of the inventory of neurons that respond to pheromone (Benton *et al.* 2007; Nichols & Vogt 2008). We identified four SNMPs in each of the two *C. lapponica* populations. The amino acid identity of each matched candidate is very high (99%) between both populations. Except the candidate *Clap*SNMP03, which consists of two fragments (*Clap*SNMP03p01 and *Clap*SNMP03p02) in both populations, all other *Clap*SNMPs are represented by full-length proteins with 515-534 amino acids. Our phylogenetic analysis revealed that the four *Clap*SNMPs from both populations are divided into two subgroups, SNMP group 1 and SNMP group 2 (Fig. 6). The candidates *Clap*SNMP01 and *Clap*SNMP02 from *C. lapponica* populations clustered together into SNMP group 1 while the other candidates (*Clap*SNMP03p01/03p02 and *Clap*SNMP04) were located in the SNMP group 2.

We identified 38 **ORs** of *C. lapponica* in the willow population and 34 ORs in the birch population. Comparing the OR sequences of both populations, 31 ORs share high amino acid identities (30 sequences share ≥93% identity; due to a gap *Clap*OR14 shares 82% identity) with a counterpart in the other population. Among them is also the universal odorant co-receptor, *Clap*COR1 with 480 amino acids. Three ORs appeared to be birch-specific and seven ORs were

willow-specific with ≤55% identities (Table 1). Among all the identified ORs, 10 ORs are represented by complete proteins composed of 363 to 480 amino acids with 5-7 transmembrane domains.

By sequence alignments we observed that the region at the C-terminus of *C. lapponica* ORs is more conserved than that at the N-terminus. This conserved region is a loop of roughly 50 amino acids between the sixth and seventh alpha-helix and contains three conspicuous motives (Fig. S1, Supporting information). In our alignments it was found that tryptophan, arginine and serine are highly prevalent in these motives. These motives are also known from other insect ORs presumably involved in protein-protein interactions (Benton *et al.* 2006; Miller & Tu 2008).

We identified eight **GRs** including one trehalose receptor, *Clap*TR. All of them were detected in both *C. lapponica* populations with high amino acid similarities (7 sequences share ≥96% identity; due to a gap *Clap*GR05 shares 82% identity). Three candidates are represented as full-length coding sequence: *Clap*GR01 with 440 amino acids and 8 TMDs, *Clap*GR02 and *Clap*TR possess 385 (8 TMDs) and 299 (7 TMDs) amino acids, respectively. Due to the fact that insect ORs and GRs belong to one chemoreceptor superfamily (Robertson *et al.* 2003), the *C. lapponica* ORs and GRs were combined in our phylogenetic analysis (Fig. 7). Except *Clap*GR06, *Clap*TR and *Ityp*GR6, all of other GRs are grouped together with a bootstrap value of 45%. *Clap*GR01, 05 and 07 clustered into a CO_2 clade characterized by *Dmel*GR21a and *Dmel*GR63a (Kwon *et al.* 2007). *Clap*GR03 clustered with *Dmel*GR43a group. *Clap*TR and *Ityp*GR6 clustered together in one clade. *Clap*GR06 clustered next to the GR group, but in the OR group. As described in previous studies (Andersson *et al.* 2013; Engsontia *et al.* 2008), seven subgroups (from 1 to 7) of ORs could be found.

IRs represent also membrane proteins, but unlike ORs and GRs they include only three transmembrane domains (Silbering & Benton 2010; Wicher 2015). They are more closely related to ionotropic glutamate receptors (iGluRs) (Croset *et al.* 2010; Rytz *et al.* 2013). In order to distinguish IRs from iGluRs, we carried out phylogenetic analyses. Twenty putative receptors of *C. lapponica* clustered distinctly into the family of iGluRs and 17 into the family of IRs with a bootstrap value of 92% (Fig. 8). Twelve of the 17 IRs share high amino acid identities (≥96%) in the coding sequence in both *C. lapponica* populations. The remaining IRs possess low amino acid identities with ≤39% among the two populations and are therefore considered as population-specific IRs (Table 1). *Clap*IR01 (924 amino acids) and *Clap*IR12 (626 amino acids) are full-

length coding sequences. Our phylogenetic analysis revealed that the IRs from *C. lapponica* could be divided into two general subgroups: co-receptor IRs and antennal IRs (Abuin *et al.* 2011; Croset *et al.* 2010). The candidates *Clap*IR01 and *Clap*IR02 clustered into co-receptor *Dmel*IR25a orthologues and *Dmel*IR8a orthologues, respectively, that are located in the clade of iGluRs. The remaining 15 IRs of *C. lapponica* formed ten orthologous groups with other insect species in the subgroup of antennal IRs. Gene duplication events were found in the species-specific divergent IR subgroup of *T. castaneum* and *D. melanogaster*, but not in the IRs of *C. lapponica*.

Expression profiling of chemosensory genes in birch- and willow-adapted C. lapponica

In total we have identified 113 unique sequences encoding putative members of six chemosensory protein families from both populations of *C. lapponica*. To understand the differences in the chemosensory system after host plant shift on the molecular level, we have carried out RNA-seq of selected organs of female and male adults each collected from *C. lapponica* willow and birch populations. We have calculated the transcript levels in counts per million mapped reads (CPM) for each gene in each of the 16 RNA-seq libraries (Table S3, Supporting information). Based on the results we have filtered out 79 from the total of 113 sequences that were expressed in least one library with a CPM≥1 for following analyses (Fig. 9). Among the 34 discarded sequences were also those putative ORs and IRs that were initially identified as population-specific. As they exhibited CPMs≤1, they may not play a role in the adult, but in other developmental stages. Regardless of host plant adaptation, genes encoding OBPs, CSPs and SNMPs were in general higher expressed than the receptor genes for ORs, GRs and IRs. According to our statistic approach ($\log_2$fold≥2, P-value≤0.05 and FDR≤0.05), males and females did not show significant differences in their expression levels and were therefore treated as biological replicates in all subsequent calculations.

Differential expression of chemosensory genes in the antennae and legs of birch- and willow-adapted C. lapponica

The comparative analysis of the expression levels of the 79 genes identified in the willow- and birch-feeding beetles unveiled differences in all the six defined chemosensory gene families between the two populations except in the SNMP family (Fig. 9). The four SNMP genes

exhibited similar CPMs in the corresponding tissue samples of each population. Our statistical evaluation of the observed expression differences indicated genes encoding OBPs and ORs as significant differentially expressed. These candidates are further described below.

In the antennae of the willow-adapted population only one OBP, *Clap*OBP27, was four times higher expressed than in the antennae of the birch-adapted population (Fig. 10A). As a result of host plant shift, in the antennae of the birch-adapted population *Clap*OBP02 was eight times, *Clap*OBP20 18-times and *Clap*OR17 75-times higher expressed compared to the antennae of the willow-adapted population. The comparative analysis of legs revealed a similar expression pattern of these chemosensory genes between the two populations. Further, we observed an at least five times higher expression of five more OBPs, including *Clap*OBP05, 07, 08, 12, and 28 in the legs of birch-feeders compared to the samples of the willow-feeding population. In the legs of the willow-feeding beetles additionally *Clap*OR02 was six times higher expressed compared to the legs of the birch-feeders (Fig. 10B).

Functional consideration of differentially expressed OBPs

To further characterize the OBPs which displayed a differential expression associated with the scenario of host plant shift, we have inspected their phylogenetic relationships. All these OBPs with a significant differential expression in the antennal tissue (*Clap*OBP02, 20, 27) were classified as minus-C OBPs (Fig. 10). In our phylogenetic tree, *Clap*OBP02 and 20, which were highly expressed in the birch-adapted population, fall together into a branch that comprises a *C. lapponica*-specific expansion of this minus-C OBP subgroup (Fig. 3). By further analyzing the expression levels of these two minus-C OBPs, we found that they are not only differentially expressed among the two populations, but also within each population. They were least six times higher expressed in the legs than in the antennae of the beetles from both populations (Fig. 9; Fig. S2, Supporting information).

The remaining minus-C OBP *Clap*OBP27 diverged from the above mentioned expression pattern in a way as it exhibited an almost equal expression level in both antennae and legs in the willow-affiliated beetles. Within the birch-feeding beetles, *ClapOBP27* was significantly ten times higher expressed in antennae than in the legs of the birch-feeders (Fig. 9; Fig. S2, Supporting information). By comparing the populations, however, it became clear that *Clap*OBP27 is typically upregulated in the willow-adapted *C. lapponica*. On inspection of the phylogenetic tree this minus-C OBP clusters into a branch together with members of the ABPIIs,

predominantly chemosensory in function in insect antennae, which include OBP83a and OBP83b from *D. melanogaster* in their immediate vicinity (Fig. 3). Based on the homology it is conceivable that *Clap*OBP27 differs in its function from the other two minus-C OBPs from *C. lapponica*.

The OBPs particularly upregulated in the legs of the birch-adapted population belong, with the exception of *Clap*OBP28, to the minus-C OBPs. Phylogenetically, the minus-C OBPs *Clap*OBP05, 08, and 12 cluster together, while *Clap*OBP07 falls together with *Clap*OBP27 upregulated in willow-feeders. The identified *Clap*OBP28 has been classified as a candidate of the subfamily of ABPIIs. Accordingly, though *ClapOBP28* showed a differential expression in the legs of the two populations, its highest expression has been detected in the antennae samples with CPM values of 127 and 166 in birch- and willow-feeders, respectively (Fig. 9; Fig. S2, Supporting information). From our phylogenetic analysis, a relationship of *Clap*OBP28 to the *D. melanogaster* ABPIIs, OBP69a, 83a and 83b can be inferred (Fig. 3). In summary we have identified, with one exception, minus-C OBPs that have switched their abundance during the host plant shift of *C. lapponica*. Compared to classic OBPs and ABPIIs, the genes of the minus-C OBPs have undergone an expansion in *C. lapponica*.

Functional consideration of differentially expressed ORs

The receptor candidate gene for *Clap*OR17 was significantly higher expressed in the birch-feeding population, in both antennae and legs, compared to the samples from the willow-feeding population (Fig. 10). *Clap*OR17 had about the same CPM value in the antennae and legs of the birch-feeders and exhibited the second highest expression level among all identified ORs after the co-receptor (Fig. 9; Table S3, Supporting information). Our phylogenetic analysis revealed the clustering of *Clap*OR17 into the subgroup 4, 5, 6 with relationship to several *Tcas*ORs (Fig. 7). *Clap*OR02 displayed a higher expression in the legs of the willow population than in those of the birch population. Phylogenetically, *Clap*OR02 is a member of the OR group 2. After blast searches the homology to functionally characterized insect receptors of *Clap*OR17 as well as *Clap*OR02 was very low (<30% sequence identity).

Homology modeling of OBP02, 20 and 27 and ligand docking

As antennae represent the classical odor-sensory organs triggering host selection behavior from a distance, we have analyzed the ligand binding abilities of the three OBPs, *Clap*OBP27 from

373 willow-feeders and *Clap*OBP02 together with *Clap*OBP20 from birch-feeders, which have been
374 differentially expressed in the antennae. For this purpose, homology modeling has been carried
375 out in the presence of selected host plant odors that were identified in our analysis (Table S1,
376 Supporting information).
377 The three protein models are of excellent quality indicated by a Ramachandran Plot with
378 more than 90% of all amino acid residues in the most favored regions not containing outliers (and
379 other stereo-chemical parameters are also inside or even better than required) after PROCHECK
380 analyses. The overall z-score values of Prosa II analysis were also in the expected range for the
381 native folding of protein models.
382 All three binding proteins form at least seven α-helices which define the internal ligand
383 binding pocket of each protein (Fig. 11A, B; Fig. S3, Supporting information). The N-terminal α-
384 helices are connected to the second helix by a rather flexible loop which allows different
385 orientations. Furthermore, the last seven amino acid residues of the C-termini may form flexible
386 lids to close or open up the binding sites (particularly seen for OBP02 in Fig. 11B - red tail).
387 In order to evaluate the ligand binding affinities, we have calculated fitness score values
388 for selected willow and birch volatiles which are listed for the most favored docking poses of
389 each ligand in all three proteins (Table S4, Supporting information). In general, the more positive
390 these values are, the higher should be the affinity. From all three binding proteins *Clap*OBP27
391 seems to be favored to bind all the odorants tested, especially the hydrophobic terpenoids.
392 Though the three proteins have a folding pattern in the central core in common, they
393 differ from each other regarding size and lipophilic/hydrophilic surface potential of the ligand
394 binding cavities (Fig. 11C, D; Fig. S3, Supporting information). *Clap*OBP27 possess the most
395 distinct hydrophobic ligand binding pocket, mainly formed by the side chains of F30, L73, I90,
396 L91, L94, L106, L131, and F142. This is also reflected by the preferred binding of the
397 hydrophobic ligands, such as (*E,E*)-α-farnesene or (*Z,E*)-α-farnesene, with highest fitness scores
398 (Table S4, Supporting information). In comparison, the two other OBPs contain more amino
399 residues whose side chains are capable of contributing to hydrogen bonds in the ligand cavity,
400 with consequences for the ligand binding abilities of the proteins. In the structure of *Clap*OBP02
401 the carbonyl group of *cis*-3-hexenyl acetat, for example, forms hydrogen bonds to the hydroxyl
402 groups of the tyrosine side chains of Y70 and 126 and in the structure of *Clap*OBP20 with H32
403 (Fig. 11G, H; Fig. S3, Supporting information). Similar results were obtained for other ligands
404 with hydrophilic moieties like salicylaldehyde or methyl-salicylate.

Also the hydrophobic DMNT, as main component of the birch volatile pattern, was predicted as possible ligand for the three modeled OBPs. We did, however, not observe a binding preference for this compound by the two OBPs higher expressed in the birch-feeding beetles (*Clap*OBP02 and *Clap*OBP20). In summary, the 3D-protein models and related docking studies demonstrated that all the investigated compounds may bind to the odorant binding proteins with varying affinity due to individual differences in polarity and the architecture of the ligand binding cavities.

DISCUSSION

Due to its physiological and ecological relevance, we have studied the chemosensory repertoire of *C. lapponica*, a beetle species that forms allopatric populations adapted to feed on either *Salix* spp. or *Betula* spp. plants. Phylogenetically, willow species are considered as the ancient host plant while birch species represent the novel host (Termonia *et al.* 2001). According to Gross et al. (Gross *et al.* 2004a) no-choice feeding experiments demonstrated that willow-adapted beetles refused to feed on birch whereas birch-feeder could be reared on willow. In dual choice experiments, however, birch-feeder preferred the natural host over willow. These studies have been carried out on populations from the regions of Finland and Czech Republic. We have made similar observation in feeding experiments using Kazakh *C. lapponica* from the willow (*Salix* sp.) and birch (*B. rotundifolia*) populations. Hence, during the separation of the populations the birch-feeders not only expanded their host plant spectrum, but developed already a preference for the members of another plant family. *C. lapponica* represents thus a model system to take a unique snapshot of the divergence in the chemosensory repertoire underlying the transition from one host plant species to another – a critical process often preceding speciation (Fitzpatrick *et al.* 2008; Futuyma 2008).

In order to identify potential ligands for the chemosensory candidates of adult *C. lapponica*, we have analyzed volatiles emitted from either *Salix* sp. or *B. rotundifolia* leaves. Although both species have most components of the measured volatile bouquet in common (with the exception of salicylaldehyde), the quantitative ratio between the components was characteristic for each host plant. Among all the identified volatiles, willow produced, for example, higher amounts of (*E,E*)-α-farnesene, while birch released more DMNT. These are likely ligands that can be used by *C. lapponica* beetles to orientate and discriminate between the two different host plant species. The similarity in volatile quality and the fact that both birches

and willows occur frequently together in the same habitat (Fatouros *et al.* 2006; Gross *et al.* 2004a), may have favored the host plant switch from Salicaceae to Betulaceae. Taken together, when comparing the volatile composition of the host plants it seems reasonable to assume that *C. lapponica* shifted host to a chemically similar plant species – via an "olfactory bridge".

In our transcriptome reference library of *C. lapponica* we have identified members of the gene families comprising OBPs, SNMPs, CSPs, ORs, GRs, and IRs. Though probably not all of the 113 chemosensory genes identified here are responsible for host plant choice, we have found candidates that differ significantly between birch- and willow-feeders, and that can be considered as key players in identification of the natural host plant. In the antennae of willow-feeders, the minus-C *Clap*OBP27 was higher expressed than in birch-feeders, while the minus-C OBPs *Clap*OBP02 and *Clap*OBP20 and the *Clap*OR17 were higher expressed in the birch- than in the willow-feeders. Phylogenetically, *Clap*OBP27 clusters within the group of ABPII together with *Dmel83a* and *Dmel83*b, whose ligands have not been identified yet. However, a homologous protein, *Ccap*OBP83a-2 from the Mediterranean fruit fly, *Ceratitis capitate*, displays a high affinity towards (*E,E*)-α-farnesene (Siciliano *et al.* 2014). Hence, together with the calculated high binding affinity inferred from our modeling, (*E,E*)-α-farnesene can be anticipated as ligand for *Clap*OBP27. This would be in accordance with the high production of this compound by the natural host plant of the willow-feeders. Given its higher expression in antennae compared to legs, *Clap*OBP27 seems to convey primarily volatile cues.

In comparison, the other two differentially expressed minus-C OBPs are abundant in both antennae and legs. From the homology modeling we could infer differences in the ligand binding abilities among the binding proteins; in general *Clap*OBP20 and *Clap*OBP02 seem to bind all the ligands tested weaker than *Clap*OBP27. However, due to the higher polarity of the cavities the according amino acid residues of *Clap*OBP20 and *Clap*OBP02 could contribute to hydrogen-bonds and therefore, interact also with more hydrophilic compounds than in case of *Clap*OBP27. *Clap*OBP20, for example, is predicted to possess a higher affinity towards methyl-salicylate, salicylaldehyde and cis-3-hexenyl acetate than *Clap*OBP27. Hence, the differences in the ligand binding properties of both proteins may indicate a role not only limited to olfaction. Instead, they may serve additional purposes and may contribute to host plant affiliation such as tolerance of plant toxins which has been suggested for *Drosophila sechellia* (Hungate *et al.* 2013). This hypothesis is supported by the differential expression of six more minus-C OBPs in the legs between the two populations.

469 Strikingly, minus-C OBPs have been found to be differentially expressed during host
470 plant shift. One of the reasons might be that among all binding proteins in particular the minus-C
471 OBP genes in *C. lapponica* have experienced an expansion, presumably due to gene duplication
472 events, compared to the four minus-C OBP genes in *D. melanogaster*. This expansion of minus-C
473 OBP genes has been described also from other herbivorous beetles including *Pyrrhalta*
474 *maculicollis* and *P. aenescens* (Zhang *et al.* 2016), *Phyllotreta striolata* (Wu *et al.* 2016) or
475 *Colaphellus bowringi* (Li *et al.* 2015). Furthermore, due to the lack of a third disulfide bridge
476 stabilizing the 3D-structure, the minus-C OBP may have the capability to bind different odorants
477 with various functional groups (Schwaighofer *et al.* 2014). The manifold primary sequence as
478 well as the structural properties of minus-C OBPs may consequently contribute to the adaptation
479 of herbivorous insects in short periods of time to changing environmental conditions including
480 affiliation of novel host plants.
481 Furthermore, the parallel upregulation of the olfactory receptor *Clap*OR17 together with
482 two OBPs exclusively found in the birch-feeding populations of *C. lapponica* may indicate that
483 both types of olfaction-related proteins act in combination. Volatile ligands from birch could be
484 bound by *Clap*OBP02 and/or *Clap*OBP20 to be transferred through the sensilla lymph to finally
485 interact with *Clap*OR17 situated on the olfactory neuron membrane. Whether the ligand activates
486 the OR while still bound to the OBP or whether the ligand is released from the OBP before it
487 activates the OR to generate downstream signaling pathways (Leal 2013) requires further
488 investigation. Similar studies using inter-population variations contributing to phenotypic
489 plasticity in host plant use are limited to date to *Drosophila* flies; thus, the understanding of
490 general mechanisms underlying olfactory processes and shifts within species is still in its
491 infancies (Pan & Volkan 2014). RNA-seq analyses comparing different cacti-adapted populations
492 of *D. mojavensis* demonstrated, for example, that changes in host use were accompanied by
493 changes in the olfactory system including the expression profile of ORs in adult heads including
494 antennae and maxillary palps (Crowley-Gall *et al.* 2016). With our pioneering endeavor we hence
495 add an example from the most diversified animal lineage on earth, the beetles (Coleoptera).
496 Given the feeding choice behavior and our differential analyses we suggest that the host plant
497 shift of *C. lapponica* populations occurred through a loss of host preference and gain of tolerance
498 to the phytochemicals of betulacaeous plants accompanied by the evolution of specific OBPs and
499 ORs.

The significant differences in the set-up of OBPs and ORs between birch- and willow-feeders were not found when comparing male and female beetles; none of the chemosensory genes differed between the sexes. This is similar to the situation found in *T. castaneum* males and females in which expression levels of both sexes show only minor dimorphism with respect to expression levels of OBPs, IRs, GRs, ORs, and SNMPs (Dippel *et al.* 2016; Dippel *et al.* 2014). Furthermore, the number and distribution of the different chemoreceptive sensilla does not show any difference between *T. castaneum* males and females (Dippel *et al.* 2016). These findings are in line with the situation found in flea beetles *P. striolata* (Wu *et al.* 2016), but in contrast to Lepidopterans such as *Manduca sexta* (Große-Wilde *et al.* 2010) or *Cydia pomonella* (Walker *et al.* 2016). Thus, our results on *C. lapponica* support the notion that there are no apparent differences in the chemosensory transcript levels between male and female beetles.

Taken together, our study highlights the complexity of smell and taste sensation underlying adaptation to new ecological environments in a plant-feeding beetle species. It provides novel insights into the chemosensory system that donates to phenotypic divergence and the evolution of host plant preference in insect herbivores.

METHODS

Collection and rearing of C. lapponica

Birch-feeding *C. lapponica* (L.) was collected from *B. rotundifolia* in the Altai Mountains in East Kazakhstan, close to Uryl, near the Burkhat Pass (2130 m altitude, 49°07.438' N 86°01.365' E). Willow-feeding *C. lapponica* was collected also in the Altai Mountains in East Kazakhstan, near Katon-Karagay, from *Salix* sp. (2207 m altitude, 49°02.573' N 85° 39.209' E). *C. lapponica* was reared under field conditions on its corresponding host plant. Individuals from *C. lapponica* were dissected and the desired tissues and organs stored in the RNA stabilization reagent, RNAlater, according to the manufacturer's recommendations until needed.

RNA preparing, RNA library construction and sequencing

RNA sequencing (RNA-seq) of both populations (birch-specific and willow-specific *C. lapponica*) was carried out using the next-generation sequencing technique of Illumina (Illumina, Inc., San Diego, California USA) (Bentley *et al.* 2008). For creating transcriptome reference libraries total RNA was pooled from entire individuals collected from all developmental stages including male and female adults, pupae, and first- to third-instar larvae each of a population. For

differential expression analyses entire legs and entire antennae dissected from 20 adult females and 20 adult males per biological replicate from each of a population was used for total RNA isolation (two biological replicates per prepared tissue, sex and host plant specificity). All total RNA samples were prepared according to (Bodemann *et al.* 2012). Around 2.5 µg total RNA of each sample was used for the library preparation with the TruSeq RNA Sample Prep Kit v2 (Illumina Inc., San Diego, USA) according to the manufacturer's description. In order to obtain longer fragments for the transcriptome reference libraries, the fragmentation step during the preparation procedure was reduced to four minutes.

The libraries of pooled samples each from birch-specific and willow-specific *C. lapponica* for the reference transcriptome was sequenced using a GenomeAnalyzer IIx (GAIIx, Illumina Inc., San Diego, USA) in 100-bp paired-end mode. The two libraries were pooled in one lane. The eight libraries of the tissue samples were sequenced using a HiSeq2000 in a 50-bp single-end mode by pooling of four libraries per lanes. All reads were extracted in FastQ format using CASAVA v1.8 (GAIIx) or v1.8.2 (HiSeq) (Illumina Inc., San Diego, USA).

De Novo assembly of transcriptomes from C. lapponica

Transcriptome reference libraries were created from each population separately. The raw RNA-seq reads were subject to adapter removal and to the trimming of the low quality regions from 3'-end with a minimum Phred score threshold of 20 by using the tool cutadapt v1.8.1 (Martin 2011). Afterwards, the trimmed paired-end reads of pooled samples and the trimmed single-end reads of tissue samples were *de novo* assembled together by applying the open source tool Trinity v2012–03–17 (Grabherr *et al.* 2011) with a minimal contig length of 300 bp. In order to reconstruct full-length transcriptomes, the above *de novo* assembled transcripts were reassembled using the TGI Clustering tools (vJan. 2009) (Pertea *et al.* 2003) with a minimum overlap length of 100 bp and sequence similarity of 90%.

Annotation of de novo assembled transcript libraries and identification of proteins implicated in chemosensation of C. lapponica

The assembled transcripts were translated into all six possible open reading frames (ORFs) by using tool "transeq" from package EMBOSS v6.3.1. To obtain their annotation, the derived protein sequences were searched against the Pfam database (update, Jan. 2013) with an e-value cutoff of 1e-4. All hits classified into the Pfam protein families PBP_GOBP (PF01395), OS-D

564 (PF03392) and CD36 (PF01130), were preliminarily identified as OBPs, CSPs and SNMPs,
565 respectively.
566 To identify IRs, ORs and GRs, we created custom reference databases of receptors
567 described from other insect species including *T. castaneum*, *M. sexta*, *B. mori*, two bark beetles
568 (*D. ponderosae* and *I. typographus*), and *D. melanogaster*, whose sequences were deposited in
569 GenBank (NCBI). We performed blastp v2.2.29+ searches against the custom databases with an
570 e-value cutoff of 1e-1 to identify chemosensory genes from the *C. lapponica* transcriptome
571 libraries.
572 To verify the chemosensory proteins identified by Pfam or by comparison with our
573 custom databases, all the sequences were subsequently searched via blastp v2.2.29+ (e-value 1e-
574 3) approach against the NCBI non-redundant database (update, Oct. 2014) (Camacho *et al.* 2009).
575 The top ten hits were inspected manually regarding a putative function in chemosensory genes of
576 *C. lapponica*. The species-specific sequences of *C. lapponica* were given temporary designations
577 as numbered series in the form of ClapXXyy, XX: chemosensory transcript; yy: number. In
578 addition, the population specific sequences of *C. lapponica* are named ClapXX-Wyy and
579 ClapXX-Byy for willow-specific and birch-specific, respectively.
580 To allocate the longest ORFs in all the identified transcripts, the derived protein
581 sequences were aligned with their corresponding custom reference databases by using MAFFT
582 version7 (option E-INS-I with default parameters) (Katoh & Standley 2013). The full-length
583 ORFs and the incomplete sequences with more than 100 amino acids were selected for further
584 analyses.
585
586 *Calculation of Phylogenetic Trees*
587 The population-specific and the longest identical chemosensory protein sequences between *C.*
588 *lapponica* feeding on willow and feeding on birch were aligned with the homologous protein
589 sequences derived from other insect species (Fasta data set, Supporting information) (Andersson
590 *et al.* 2013; Attrill *et al.* 2016; Croset *et al.* 2010; Dippel *et al.* 2014; Engsontia *et al.* 2008;
591 Grosse-Wilde *et al.* 2011) by applying the E-INS-i methods from MAFFT with default
592 parameters. To calculate the phylogenetic tree, RAxML v7.2.8 (Stamatakis 2006), a program
593 based on maximum-likelihood inference was used. For phylogenetic analysis of the olfactory
594 transcripts of *C. lapponica*, the best fitted model of protein evolution was chosen by using Perl
595 script ProteinModelSelection.pl, which was downloaded from http://sco.h-

its.org/exelixis/web/software/raxml/). The maximum-likelihood phylogenetic tree was reconstructed with a bootstrap test of 100 replicates in RAxML.

Analysis of differentially expressed chemosensory genes in the two C. lapponica populations

To determine identical chemosensory transcripts in both populations, we compared the transcripts of birch-specific and willow-specific populations by using blastp with an e-value cut-off of 1e-3.

To compare the transcript expression levels of the antennae and legs (male and female) from both populations of *C. lapponica*, we mapped the tissue RNA-seq reads onto the *de novo* assembled transcriptome of *C. lapponica* feeding on willow with the open source tool Bowtie2 v2.2.9 (Langmead & Salzberg 2012) using default parameters. Afterwards, the R package edgeR (Robinson *et al.* 2010), which is part of the Bioconductor package, was used to estimate abundance and detect differentially expressed transcripts in the two different tissues. To remove very low counts across all libraries, we selected the transcripts that expressed in two or more libraries with counts per million mapped reads >1. Trimmed mean of M-values normalization (edgeR default normalization method) was applied to remove technical variability (accounting for compositional difference between the libraries). Using the Cox-Reid (CR) profile-adjusted likelihood method to estimate dispersions, the generalized linear model (GLM) was selected to testing for different expression (DE) of transcripts with a P-value cutoff 0.05 and false discovery rate (FDR) cutoff 0.05. To avoid the DE caused by low expression level between samples, we focused on transcripts that had at least 10 CPM in one or both comparable samples.

Homology modeling and calculations

Protein homology modeling of three odorant binding proteins were automatically performed with YASARA (Krieger *et al.* 2009). The resulting models were evaluated by YASARA, and if appropriate a final model was created by merging best folded fragments from different models followed by energy minimization. The quality of all models was checked for native folding by energy calculations with PROSA II (Sippl 1990) and for stereo-chemical quality by PROCHECK (Laskowski *et al.* 1993).

For ClapOBP27 YASARA created 16 homology models based on alignments with several already crystallized odorant binding proteins (pdb-codes: 2ERB (Wogulis *et al.* 2006), 4PT1 (Zheng *et al.* 2015), 3OGN (Mao *et al.* 2010), 3CZ0 and 3D73 (Pesenti *et al.* 2009), 3R72, 3Q8I,

627 3K1E). A final hybrid model was formed mainly based on 3R72 (sequence identity 19.8%,
628 sequence similarity 47.2%) with inclusion of short template fragments from 3K1E and 4PT1.
629 For ClapOBP02 YASARA created 20 homology models (pdb-codes: 2JPO (Damberger *et*
630 *al.* 2007); 3D78, 3D73and 3CZ0 (Pesenti *et al.* 2009); 2QEB (Mans *et al.* 2007); 3VB1 and
631 3V2L (Ziemba *et al.* 2013); 4PT1, 3OGN, 3K1E). A final hybrid model was formed mainly
632 based on 3VB1 (sequence identity 14.4%, sequence similarity 35.1%) with inclusion of short
633 template fragments from 2JPO and 4PT1.
634 ClapOBP20 YASARA created 22 homology models (pdb-codes: 3S0D and 3S0G
635 (Spinelli *et al.* 2012), 3R1P (Lagarde *et al.* 2011), 1C3Z (Rothemund *et al.* 1999), 3DYE (Calvo
636 *et al.* 2009), 1DQE (Sandler *et al.* 2000), 3D78, 3R72, 3V2L, 3CZ0). A final hybrid model was
637 formed mainly based on 3V2L (sequence identity 20.0%, sequence similarity 38.0%) with
638 inclusion of short template fragments from 3R1P and 3DYE.
639 The 3D-structures of all the ligands were constructed with MOE (Molecular Operating
640 Environment, 2013.08 (2016); Chemical Computing Group Inc., Montreal, QC, Canada). The
641 putative binding sites were identified based on the structure of *Anopheles gambiae* odorant
642 binding protein 20 with bound polyethylene glycol (PDB 3V2L (Ziemba *et al.* 2013)). In all three
643 cases a radius of 15 Å was applied to define the active site for docking using the coordinates of
644 following atoms as origin: ClapOBP20: L73-CD1, ClapOBP02: L125-CG, ClapOBP27: L131-
645 CD1. Two side chains of each protein were considered to be flexible (ClapOBP27: F142, M34,
646 ClapOBP02: Y70, Y126, ClapOBP20: F124, Y133). Docking studies were performed with
647 GOLD using the ChemPLP scoring functions. For all other options standard settings were
648 applied.
649
650 *Volatile analysis using GC-MS*
651 For GC-MS a TRACE MS (Thermo-Finnigan) device equipped with a ZB5 column (15 m, 0.25
652 mm I.D, 0.25 µm film thickness) was used with a 10 m guard column (Phenomenex,
653 Aschaffenburg, Germany). Mass spectra were measured in electron impact (EI) mode at 70 eV,
654 33-450 m/z. Volatiles were eluted under programmed conditions: 40 °C (2 min isotherm),
655 followed by heating at 10 °C min^{-1} to 220 °C and at 30 °C min^{-1} to 280 °C, using helium (1.5 ml
656 min^{-1}) as the carrier gas. The GC injector (split ratio 1:7), transfer line and ion source were set at
657 220, 280 and 200 °C, respectively. 25 cm long branches of birch or willow were enclosed with
658 polyethylene terephthalate (PET) foil (Toppits Bratschlauch, Minden, Germany). The volatile

collection time was 6 hours. For volatile collection push pull systems were used. One system is equipped with two rotary vane pumps (model G 12/02 EB, Gardner Denver Thomas GmbH, Fürstenfeldbruck, Germany), one for providing fresh charcoal cleaned air (flow 1.0 l/min) and one for volatile collection (flow 0.9 l/min). The volatiles were collected on PorapackQ (20 mg, Supelco, Bellefonte, PA, USA) and eluted with 90 µl dichloromethane containing an internal standard (1-bromodecane, 50 ng/µl, Fluka, Germany). Freshly collected tree branches without any treatment, coronalon (0.1 mmol/l) treated plants (sprayed two times the evening before collection and let the leaves dry) and mechanical wounded branches (scratched by a pattern wheel) were sampled. Coronalon is a synthetic 6-ethyl indanoyl isoleucine conjugate that induces various plant stress responses including the induction of volatiles against herbivore attack (Schüler *et al.* 2004). The compounds were identified by using standard compounds or Kováts indices (Kováts 1958) and MassFinder 4 software (D. Hochmuth (scientific consulting), 1999-2010, Hamburg, Germany, www.massfinder.com) ((*E*)-Myroxide, β-Bourbonene, Germacrene D, (*Z,E*)-α-Farnesene, (*E,E*)-α-Farnesene).

Field emission scanning electron microscopy

The animals were fixed with 2.5 % (v/v) glutaraldehyde in cacodylate buffer for 60 min. Afterwards the samples were washed three times for 10 min with cacodylate buffer and dehydrated in ascending ethanol concentrations (30, 50, 70, 90 and 100%) for 20 min each. Subsequently, the samples were critical-point dried using liquid CO_2 and sputter coated with gold (thickness approx. 4 nm) using a SCD005 sputter coater (BAL-TEC, Liechtenstein) to avoid surface charging. Finally the specimens were investigated with a field emission scanning electron microscope LEO-1530 Gemini (Carl Zeiss NTS GmbH, Oberkochen, Germany).

ACKNOWLEDGMENTS

We sincerely thank Dr. Sandor Nietzsche and Susanne Linde for their support in electron microscopy. The authors would like to express their gratitude to Kerstin Ploß, Ivonne Goerlich and Ivonne Heinze for technical assistance. Very special thanks are due to Alija Gabdullina, Katon-Karagay State National Park, Kazakhstan, for her support during sample collection of *C. lapponica*. This work was financially supported by the Max Planck Society.

690 **REFERENCES**

691

692 Abuin L, Bargeton B, Ulbrich MH*, et al.* (2011) Functional architecture of olfactory ionotropic glutamate
693 receptors. *Neuron* **69**, 44-60.
694 Andersson MN, Grosse-Wilde E, Keeling CI*, et al.* (2013) Antennal transcriptome analysis of the
695 chemosensory gene families in the tree killing bark beetles, Ips typographus and Dendroctonus
696 ponderosae (Coleoptera: Curculionidae: Scolytinae). *BMC Genomics* **14**, 198.
697 Attrill H, Falls K, Goodman JL*, et al.* (2016) FlyBase: establishing a Gene Group resource for Drosophila
698 melanogaster. *Nucleic Acids Res* **44**, D786-792.
699 Barrett LG, Heil M (2012) Unifying concepts and mechanisms in the specificity of plant-enemy
700 interactions. *Trends in Plant Science* **17**, 282-292.
701 Bentley DR, Balasubramanian S, Swerdlow HP*, et al.* (2008) Accurate whole human genome sequencing
702 using reversible terminator chemistry. *Nature* **456**, 53-59.
703 Benton R, Sachse S, Michnick SW, Vosshall LB (2006) Atypical membrane topology and heteromeric
704 function of Drosophila odorant receptors in vivo. *PLoS Biology* **4**, e20.
705 Benton R, Vannice KS, Vosshall LB (2007) An essential role for a CD36-related receptor in pheromone
706 detection in Drosophila. *Nature* **450**, 289-293.
707 Bodemann RR, Rahfeld P, Stock M*, et al.* (2012) Precise RNAi-mediated silencing of metabolically active
708 proteins in the defence secretions of juvenile leaf beetles. *Proceedings of the Royal Society of*
709 *London, Series B: Biological Sciences* **279**, 4126-4134.
710 Calvo E, Mans BJ, Ribeiro JM, Andersen JF (2009) Multifunctionality and mechanism of ligand binding in a
711 mosquito antiinflammatory protein. *Proceedings of the National Academy of Sciences of the USA*
712 **106**, 3728-3733.
713 Camacho C, Coulouris G, Avagyan V*, et al.* (2009) BLAST+: architecture and applications. *BMC*
714 *Bioinformatics* **10**, 421.
715 Croset V, Rytz R, Cummins SF*, et al.* (2010) Ancient protostome origin of chemosensory ionotropic
716 glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genetics* **6**, e1001064.
717 Crowley-Gall A, Date P, Han C*, et al.* (2016) Population differences in olfaction accompany host shift in
718 *Drosophila mojavensis*. *Proceedings of the Royal Society of London, Series B: Biological Sciences*
719 **283**.
720 D'Auria JC, Pichersky E, Schaub A, Hansel A, Gershenzon J (2007) Characterization of a BAHD
721 acyltransferase responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in
722 Arabidopsis thaliana. *The Plant Journal* **49**, 194-207.
723 Dahanukar A, Hallem EA, Carlson JR (2005) Insect chemoreception. *Current Opinion in Neurobiology* **15**,
724 423-430.
725 Damberger FF, Ishida Y, Leal WS, Wuthrich K (2007) Structural basis of ligand binding and release in
726 insect pheromone-binding proteins: NMR structure of Antheraea polyphemus PBP1 at pH 4.5.
727 *Journal of Molecular Biology* **373**, 811-819.
728 Day EH, Hua X, Bromham L (2016) Is specialization an evolutionary dead end? Testing for differences in
729 speciation, extinction and trait transition rates across diverse phylogenies of specialists and
730 generalists. *Journal of Evolutionary Biology* **29**, 1257-1267.
731 del Campo ML, Miles CI, Schroeder FC*, et al.* (2001) Host recognition by the tobacco hornworm is
732 mediated by a host plant compound. *Nature* **411**, 186-189.
733 Dippel S, Kollmann M, Oberhofer G*, et al.* (2016) Morphological and transcriptomic analysis of a beetle
734 chemosensory system reveals a gnathal olfactory center. *BMC Biology* **14**, 90.

735 Dippel S, Oberhofer G, Kahnt J, *et al.* (2014) Tissue-specific transcriptomics, chromosomal localization,
736 and phylogeny of chemosensory and odorant binding proteins from the red flour beetle
737 *Tribolium castaneum* reveal subgroup specificities for olfaction or more general functions. *BMC*
738 *Genomics* **15**, 1141.
739 Engsontia P, Sanderson AP, Cobb M, *et al.* (2008) The red flour beetle's large nose: an expanded odorant
740 receptor gene family in *Tribolium castaneum*. *Insect Biochemistry and Molecular Biology* **38**, 387-
741 397.
742 Fan J, Francis F, Liu Y, Chen JL, Cheng DF (2011) An overview of odorant-binding protein functions in
743 insect peripheral olfactory reception. *Genetics and Molecular Research* **10**, 3056-3069.
744 Farrell BD (1998) Inordinate fondness explained - why are there so many beetles. *Science* **281**, 555-559.
745 Farrell BD, Sequeira AS (2004) Evolutionary rates in the adaptive radiation of beetles on plants. *Evolution*
746 **58**, 1984-2001.
747 Fatouros NE, Hilker M, Gross J (2006) Reproductive isolation between populations from Northern and
748 Central Europe of the leaf beetle *Chrysomela lapponica* L. *Chemoecology* **16**, 241-251.
749 Fernandez P, Hilker M (2007) Host plant location by Chrysomelidae. *Basic and Applied Ecology* **8**, 97-116.
750 Fitzpatrick BM, Fordyce JA, Gavrilets S (2008) What, if anything, is sympatric speciation? *Journal of*
751 *Evolutionary Biology* **21**, 1452-1459.
752 Futuyma D (2008) Sympatric speciation: norm or exception. In: *Specialization, Speciation, and Radiation:*
753 *the Evolutionary Biology of Herbivorous Insects*, pp. 136-148. University of California Press.
754 Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores.
755 *Proceedings of the National Academy of Sciences of the USA* **106**, 18054-18061.
756 Geiselhardt S, Hilker M, Muller F, Kozlov MV, Zvereva EL (2015) Inter- and intrapopulation variability in
757 the composition of larval defensive secretions of willow-feeding populations of the leaf beetle
758 *Chrysomela lapponica*. *Journal of Chemical Ecology* **41**, 276-286.
759 Gómez-Zurita J, Hunt T, Kopliku F, Vogler AP (2007) Recalibrated tree of leaf beetles (Chrysomelidae)
760 indicates independent diversification of angiosperms and their insect herbivores. *PLoS ONE* **2**.
761 Grabherr MG, Haas BJ, Yassour M, *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data
762 without a reference genome. *Nature Biotechnology* **29**, 644-652.
763 Gross J, Fatouros NE, Hilker M (2004a) The significance of bottom-up effects for host plant specialization
764 in *Chrysomela* leaf beetles. *Oikos* **105**, 368-376.
765 Gross J, Fatouros NE, Neuvonen S, Hilker M (2004b) The importance of specialist natural enemies for
766 *Chrysomela lapponica* in pioneering a new host plant. *Ecological Entomology* **29**, 584-593.
767 Grosse-Wilde E, Kuebler LS, Bucks S, *et al.* (2011) Antennal transcriptome of Manduca sexta. *Proceedings*
768 *of the National Academy of Sciences of the United States of America* **108**, 7449-7454.
769 Große-Wilde E, Stieber R, Forstner M, *et al.* (2010) Sex-specific odorant receptors of the tobacco
770 hornworm *Manduca sexta*. *Frontiers in cellular neuroscience* **4**, 22.
771 Hallem EA, Dahanukar A, Carlson JR (2006) Insect odor and taste receptors. *Annual Review of*
772 *Entomology* **51**, 113-135.
773 Hungate EA, Earley EJ, Boussy IA, *et al.* (2013) A locus in Drosophila sechellia affecting tolerance of a host
774 plant toxin. *Genetics* **195**, 1063-1075.
775 Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in
776 performance and usability. *Molecular Biology and Evolution* **30**, 772-780.
777 Kováts E (1958) Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1:
778 Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helvetica Chimica*
779 *Acta* **41**, 1915-1932.
780 Krieger E, Joo K, Lee J, *et al.* (2009) Improving physical realism, stereochemistry and side-chain accuracy
781 in homology modeling: four approaches that performed well in CASP8. *Proteins* **77**, 114-122.
782 Kwon JY, Dahanukar A, Weiss LA, Carlson JR (2007) The molecular basis of CO2 reception in Drosophila.
783 *Proceedings of the National Academy of Sciences of the USA* **104**, 3574-3578.

784 Labandeira CC, Currano ED (2013) The fossil record of plant-insect dynamics. *Annual Review of Earth and*
785 *Planetary Sciences* **41**, 287-311.
786 Lagarde A, Spinelli S, Tegoni M, *et al.* (2011) The crystal structure of odorant binding protein 7 from
787 *Anopheles gambiae* exhibits an outstanding adaptability of its binding site. *Journal of Molecular*
788 *Biology* **414**, 401-412.
789 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359.
790 Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK - A program to check the
791 stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291.
792 Lauchli R, Schueler G, Boland W (2002) Selective induction of secondary metabolism in *Phaseolus lunatus*
793 by 6-substituted indanoyl isoleucine conjugates. *Phytochemistry* **61**, 807-817.
794 Leal WS (2013) Odorant reception in insects: roles of receptors, binding proteins, and degrading
795 enzymes. *Annual Review of Entomology* **58**, 373-391.
796 Li X-M, Zhu X-Y, Wang Z-Q, *et al.* (2015) Candidate chemosensory genes identified in *Colaphellus*
797 *bowringi* by antennal transcriptome analysis. *BMC Genomics* **16**, 1028.
798 Mans BJ, Calvo E, Ribeiro JM, Andersen JF (2007) The crystal structure of D7r4, a salivary biogenic amine-
799 binding protein from the malaria mosquito *Anopheles gambiae*. *Journal of Biological Chemistry*
800 **282**, 36626-36633.
801 Mao Y, Xu X, Xu W, *et al.* (2010) Crystal and solution structures of an odorant-binding protein from the
802 southern house mosquito complexed with an oviposition pheromone. *Proceedings of the*
803 *National Academy of Sciences of the United States of America* **107**, 19102-19107.
804 Mardulyn P, Othmezouri N, Mikhailov YE, Pasteels JM (2011) Conflicting mitochondrial and nuclear
805 phylogeographic signals and evolution of host-plant shifts in the boreo-montane leaf beetle
806 *Chrysomela lapponica*. *Molecular Phylogenetics and Evolution* **61**, 686-696.
807 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
808 *EMBnet.journal* **17**, 10--12.
809 Miller R, Tu Z (2008) Odorant receptor C-terminal motifs in divergent insect species. *Journal of Insect*
810 *Science* **8**.
811 Missbach C, Dweck HK, Vogel H, *et al.* (2014) Evolution of insect olfactory receptors. *eLife* **3**, e02115.
812 Mithöfer A, Boland W (2012) Plant defense against herbivores: Chemical aspects. *Annual Review of Plant*
813 *Biology* **63**, 431-450.
814 Nichols Z, Vogt RG (2008) The SNMP/CD36 gene family in Diptera, Hymenoptera and Coleoptera:
815 *Drosophila melanogaster, D. pseudoobscura, Anopheles gambiae, Aedes aegypti, Apis mellifera,*
816 *and Tribolium castaneum*. *Insect Biochemistry and Molecular Biology* **38**, 398-415.
817 Pan JW, Volkan PC (2014) Mechanisms of development and evolution of the insect olfactory system. *Cell*
818 *& Developmental Biology* **2013**.
819 Pelosi P, Iovinella I, Felicioli A, Dani FR (2014) Soluble proteins of chemical communication: an overview
820 across arthropods. *Frontiers in physiology* **5**.
821 Pertea G, Huang X, Liang F, *et al.* (2003) TIGR Gene Indices clustering tools (TGICL): a software system for
822 fast clustering of large EST datasets. *Bioinformatics* **19**, 651-652.
823 Pesenti ME, Spinelli S, Bezirard V, *et al.* (2009) Queen bee pheromone binding protein pH-induced
824 domain swapping favors pheromone release. *Journal of Molecular Biology* **390**, 981-990.
825 Robertson HM, Warr CG, Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene
826 superfamily in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the*
827 *USA* **100**, 14537-14542.
828 Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression
829 analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
830 Rothemund S, Liou YC, Davies PL, Krause E, Sonnichsen FD (1999) A new class of hexahelical insect
831 proteins revealed as putative carriers of small hydrophobic ligands. *Structure* **7**, 1325-1332.

Rytz R, Croset V, Benton R (2013) Ionotropic Receptors (IRs): Chemosensory ionotropic glutamate receptors in Drosophila and beyond. *Insect Biochemistry and Molecular Biology* **43**, 888-897.

Sandler BH, Nikonova L, Leal WS, Clardy J (2000) Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. *Chemistry and Biology* **7**, 143-151.

Sato K, Pellegrino M, Nakagawa T*, et al.* (2008) Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **452**, 1002-1006.

Sato K, Tanaka K, Touhara K (2011) Sugar-regulated cation channel formed by an insect gustatory receptor. *Proceedings of the National Academy of Sciences of the USA* **108**, 11680-11685.

Schüler G, Mithöfer A, Baldwin IT*, et al.* (2004) Coronalon: a powerful tool in plant stress physiology. *FEBS Letters* **563**, 17-22.

Schwaighofer A, Pechlaner M, Oostenbrink C*, et al.* (2014) Insights into structural features determining odorant affinities to honey bee odorant binding protein 14. *Biochemical and Biophysical Research Communications* **446**, 1042-1046.

Siciliano P, He XL, Woodcock C*, et al.* (2014) Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, Ceratitis capitata. *Insect Biochemistry and Molecular Biology* **48**, 51-62.

Silbering AF, Benton R (2010) Ionotropic and metabotropic mechanisms in chemoreception: 'chance or design'? *Embo Reports* **11**, 173-179.

Sippl MJ (1990) Calculation of conformational ensembles from potentials of mean force - an approach to the knowledge-based prediction of local structures in globular-proteins. *Journal of Molecular Biology* **213**, 859-883.

Spinelli S, Lagarde A, Iovinella I*, et al.* (2012) Crystal structure of *Apis mellifera* OBP14, a C-minus odorant-binding protein, and its complexes with odorant molecules. *Insect Biochemistry and Molecular Biology* **42**, 41-50.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

Termonia A, Hsiao TH, Pasteels JM, Milinkovitch MC (2001) Feeding specialization and host-derived chemical defense in Chrysomeline leaf beetles did not lead to an evolutionary dead end. *Proceedings of the National Academy of Sciences of the USA* **98**, 3909-3914.

Tilmon KJ (2008) *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* Univ California Press, 2120 Berkeley Way, Berkeley, Ca 94720 USA.

Turlings TC, Loughrin JH, McCall PJ*, et al.* (1995) How caterpillar-damaged plants protect themselves by attracting parasitic wasps. *Proceedings of the National Academy of Sciences* **92**, 4169-4174.

Van Naters WvdG, Carlson JR (2006) Insects as chemosensors of humans and crops. *Nature* **444**, 302-307.

Vieira FG, Rozas J (2011) Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biol Evol* **3**, 476-490.

Vogt RG, Miller NE, Litvack R*, et al.* (2009) The insect SNMP gene family. *Insect Biochemistry and Molecular Biology* **39**, 448-456.

Walker WB, Gonzalez F, Garczynski SF, Witzgall P (2016) The chemosensory receptors of codling moth *Cydia pomonella*–expression in larvae and adults. *Scientific Reports* **6**, 23518.

Wang B, Zhang HC, Jarzembowski EA (2013) Early Cretaceous angiosperms and beetle evolution. *Frontiers in Plant Science* **4**.

Wicher D (2015) Chapter Two - Olfactory Signaling in Insects. In: *Progress in Molecular Biology and Translational Science* (ed. Richard G), pp. 37-54. Academic Press.

Wicher D, Schäfer R, Bauernfeind R*, et al.* (2008) Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* **452**, 1007-1011.

879 Wogulis M, Morgan T, Ishida Y, Leal WS, Wilson DK (2006) The crystal structure of an odorant binding
880 protein from Anopheles gambiae: evidence for a common ligand release mechanism.
881 *Biochemical and Biophysical Research Communications* **339**, 157-164.
882 Wu Z, Bin S, He H, *et al.* (2016) Differential Expression Analysis of Chemoreception Genes in the Striped
883 Flea Beetle Phyllotreta striolata Using a Transcriptomic Approach. *PLoS ONE* **11**, e0153067.
884 Xu P, Atkinson R, Jones DN, Smith DP (2005) Drosophila OBP LUSH is required for activity of pheromone-
885 sensitive neurons. *Neuron* **45**, 193-200.
886 Xu YL, He P, Zhang L, *et al.* (2009) Large-scale identification of odorant-binding proteins and
887 chemosensory proteins from expressed sequence tags in insects. *BMC Genomics* **10**, 632.
888 Yarmolinsky DA, Zuker CS, Ryba NJP (2009) Common sense about taste: from mammals to insects. *Cell*
889 **139**, 234-244.
890 Zacharuk RY, Shields VD (1991) Sensilla of immature insects. *Annual Review of Entomology* **36**, 331-354.
891 Zhang B, Zhang W, Nie RE, *et al.* (2016) Comparative transcriptome analysis of chemosensory genes in
892 two sister leaf beetles provides insights into chemosensory speciation. *Insect Biochemistry and*
893 *Molecular Biology* **79**, 108-118.
894 Zhang H-J, Anderson AR, Trowell SC, *et al.* (2011) Topological and functional characterization of an insect
895 gustatory receptor. *PLoS ONE* **6**, e24111.
896 Zheng J, Li J, Han L, *et al.* (2015) Crystal structure of the Locusta migratoria odorant binding protein.
897 *Biochemical and Biophysical Research Communications* **456**, 737-742.
898 Ziemba BP, Murphy EJ, Edlin HT, Jones DNM (2013) A novel mechanism of ligand binding and release in
899 the odorant binding protein 20 from the malaria mosquito *Anopheles gambiae*. *Protein Science*
900 **22**, 11-21.
901 Zvereva EL, Hunter MD, Zverev V, Kozlov MV (2016) Factors affecting population dynamics of leaf beetles
902 in a subarctic region: The interplay between climate warming and pollution decline. *Science of*
903 *the Total Environment* **566**, 1277-1288.
904 Zvereva EL, Rank NE (2003) Host plant effects on parasitoid attack on the leaf beetle *Chrysomela*
905 *lapponica*. *Oecologia* **135**, 258-267.
906 Zvereva EL, Rank NE (2004) Fly parasitoid Megaselia opacicornis uses defensive secretions of the leaf
907 beetle *Chrysomela lapponica* to locate its host. *Oecologia* **140**, 516-522.

908

909 **DATA ACCESSIBILITY**

910 The raw sequence data are stored in the Sequence Read Archive (SRA) of the National Center for

911 Biotechnology Information (NCBI) with the accession number . The corresponding BioProject

912 is .

913

914 **AUTHOR CONTRIBUTIONS**

915 DW, JMP, WBo, AB designed research; DW, MK, MG, WBr, JMP, AB performed research;

916 DW, SP, WBr analyzed data; DW, SP, WBo, AB wrote the paper.

917

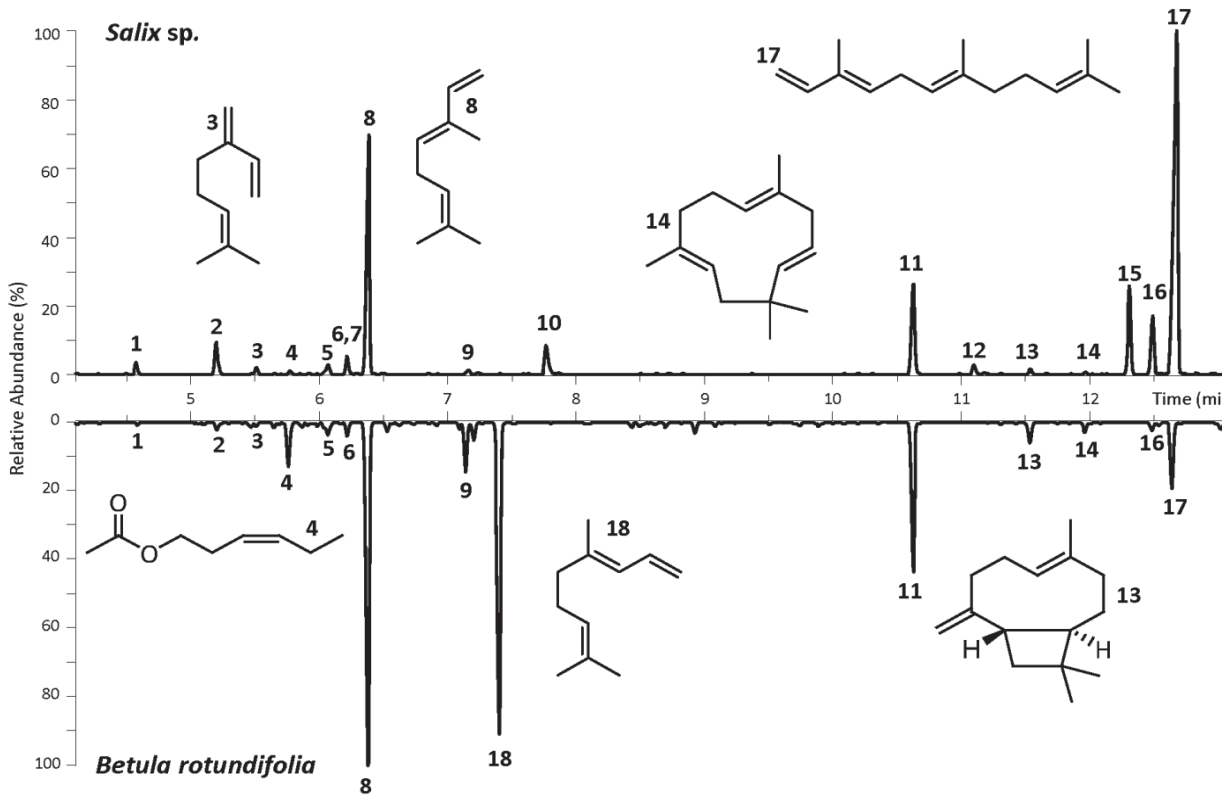

Figure 1. Gas chromatograms of the volatile composition of the host plants colonized by two *C. lapponica* populations. The plants (*Salix* sp. and *Betula rotundifolia*) were treated with coronalon (0.1 mmol/l). The volatiles were collected for 6 hours on PorapackQ using a push pull system. **1,** α-pinene; **2,** sabinene; **3,** myrcene; **4,** cis-3-hexenyl acetate; **5,** eucalyptol; **6,** (*Z*)-β-ocimene; **7,** salicylaldehyde; **8,** (*E*)-β-ocimene; **9,** linalool; **10,** (*E*)-myroxide; **11,** 1-bromodecane (internal standard); **12,** β-bourbonene; **13,** β-caryophyllene; **14,** α-humulene; **15,** germacrene D; **16,** (*Z,E*)-α-farnesene; **17,** (*E,E*)-α-farnesene; **18,** DMNT (4,8-dimethylnona-1,3,7-triene).

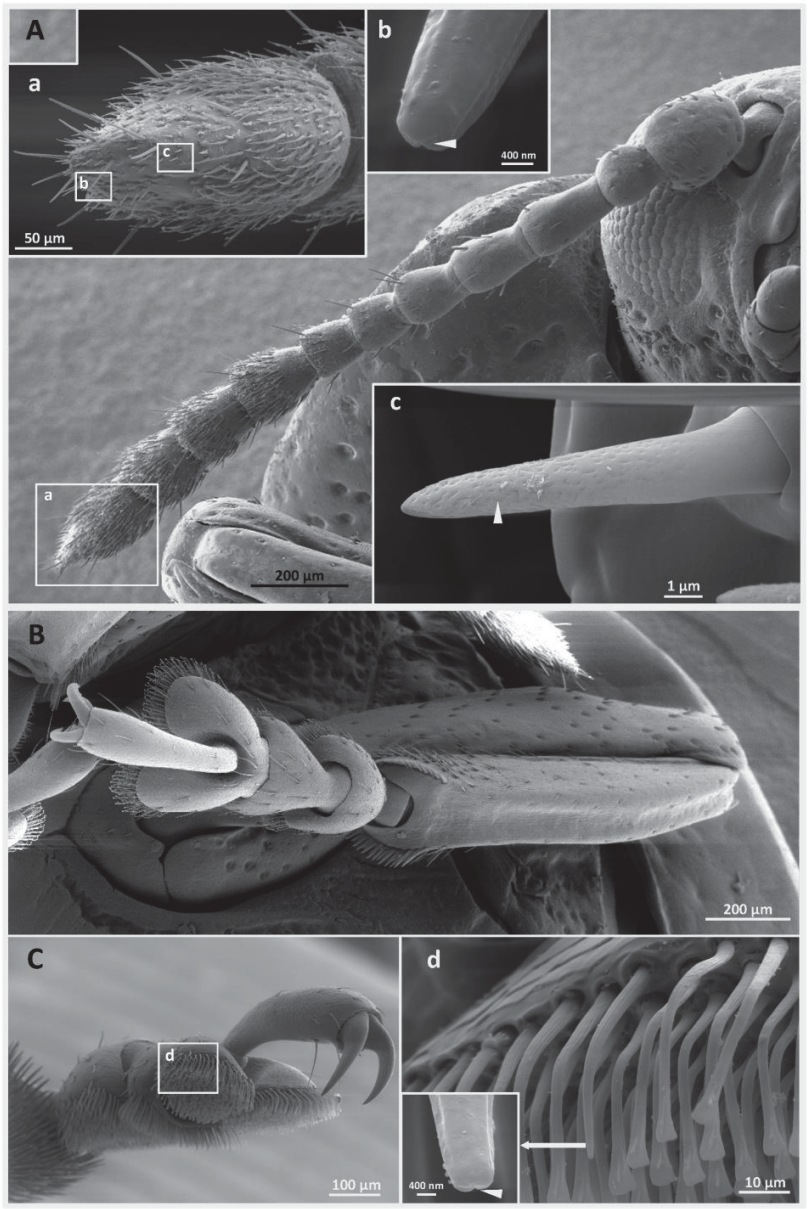

Figure 2. Scanning electron micrographs of the external morphology of the chemosensory organs of a birch-feeding *C. lapponica* female, three days after eclosion. **A**, Filiforme antenna; **a**, tip of the ninth flagellomer showing diversity of sensilla; **b**, tip of an antennal sensillum showing terminal pore (triangle); **c**, multiporous sensillum basiconicum; **B**, dorsal view of a leg; **C**, ventral view of the tarsus, d, tip of a tarsal sensillum showing terminal pore (triangle)..

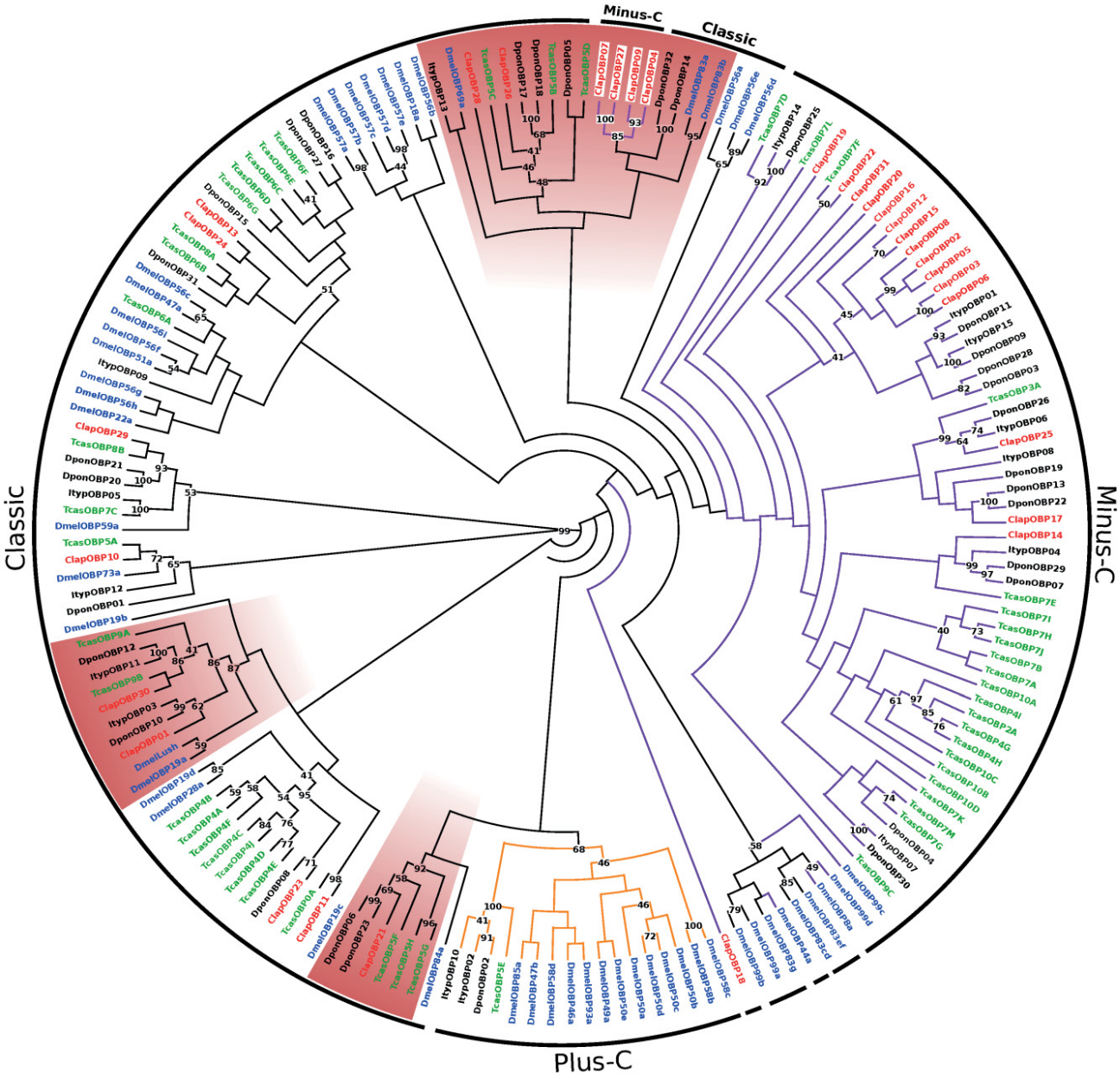

Figure 3. Phylogenetic tree of OBPs. Blue: *D. melanogaster* (Dmel); Green: *T. castaneum* (Tcas); Black: *D. ponderosae* (Dpon) and *I. typographus* (Ityp); Red: *C. lapponica* (Clap). Four subgroups of OBPs: Classic (black edges), Minus-C (purple edges), Plus-C (orange edges) and ABPII (shaded in brown) are identified. Numbers at nodes represent bootstrap values based on 100 replicates, which are shown when the value ≥ 40%.

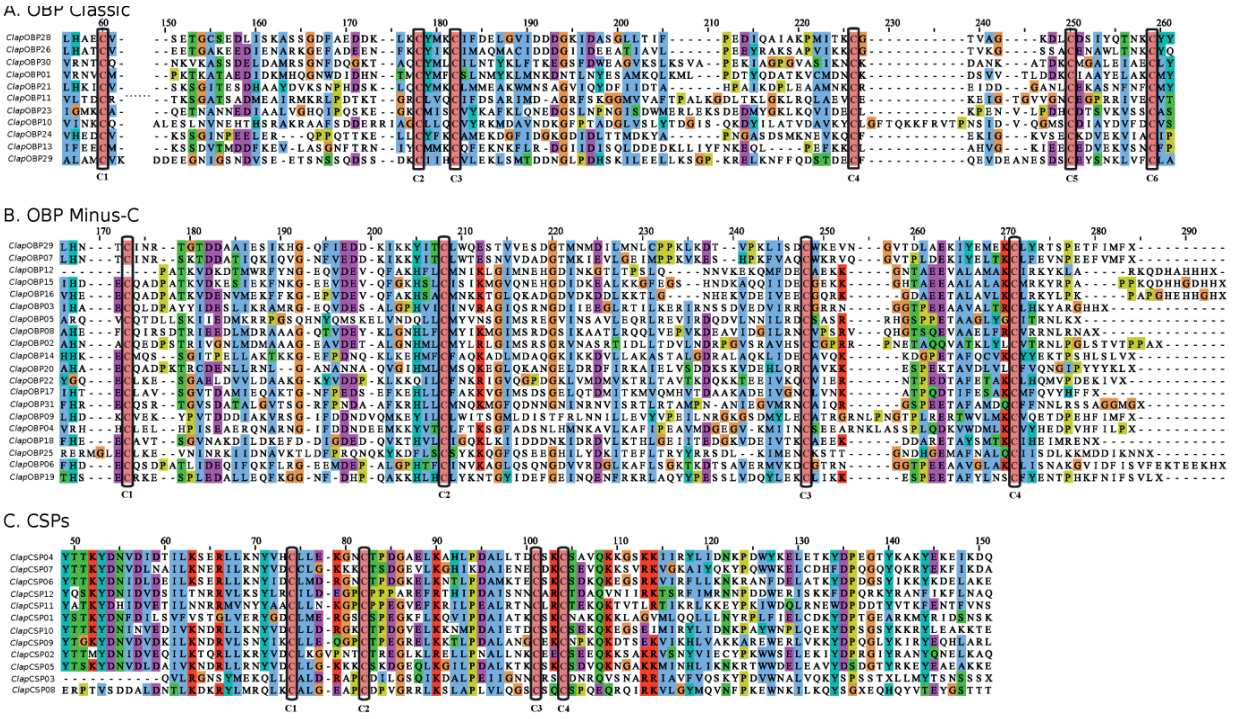

Figure 4. Multiple protein sequence alignments of classic OBPs, minus-C OBPs and CSPs from *C. lapponica* (Clap) using the program MAFFT. Only the longest identical OBPs and CSPs between *C. lapponica* feeding on willow and on birch were chosen. Conserved cysteines are framed in black.

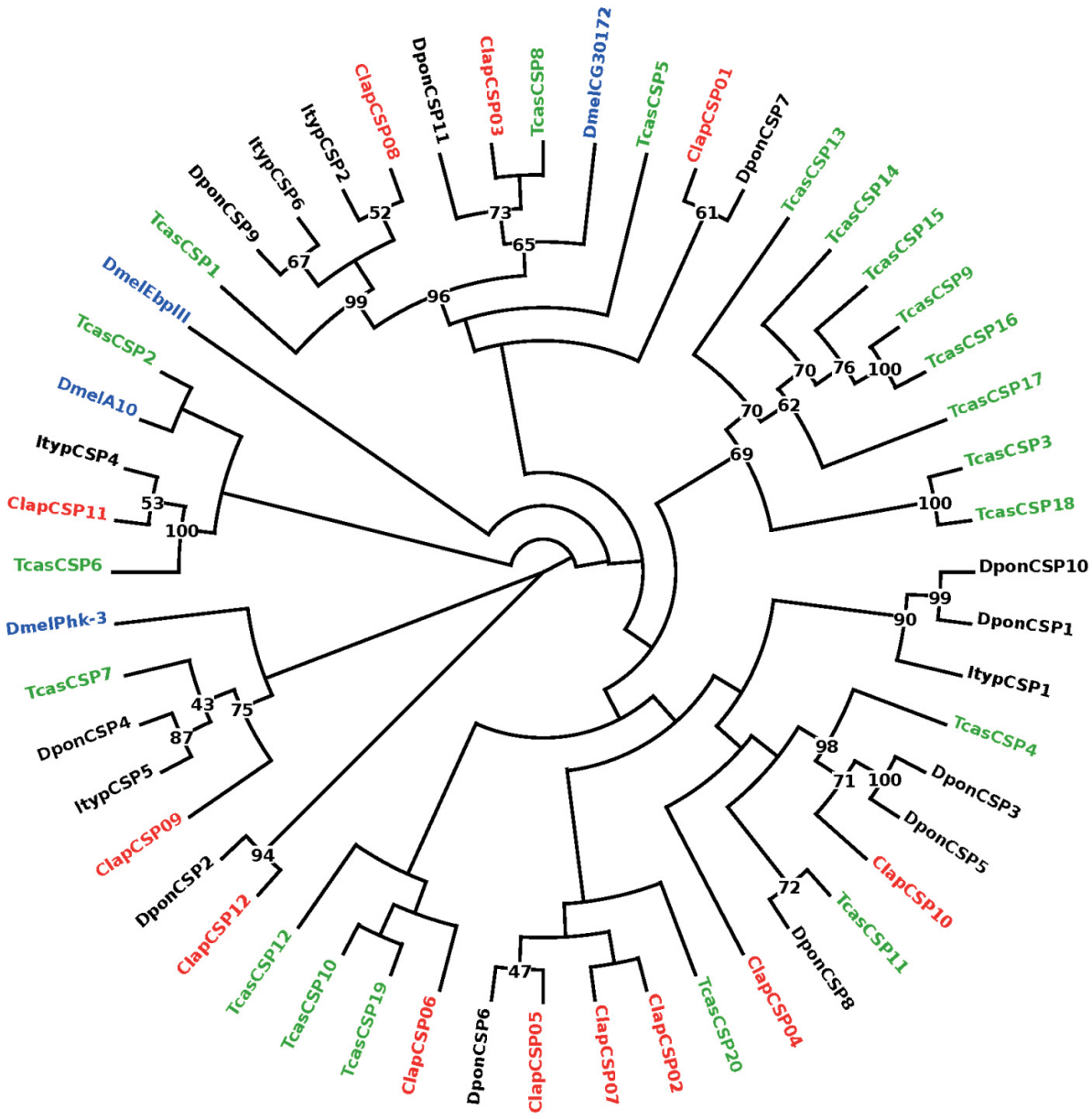

Figure 5. Phylogenetic tree of CSPs. Blue: *D. melanogaster* (Dmel); Green: *T. castaneum* (Tcas); Black: *D. ponderosae* (Dpon) and *I. typographus* (Ityp); Red: *C. lapponica* (Clap). Numbers at nodes represent bootstrap values based on 100 replicates, which are shown when the value ≥ 40%.

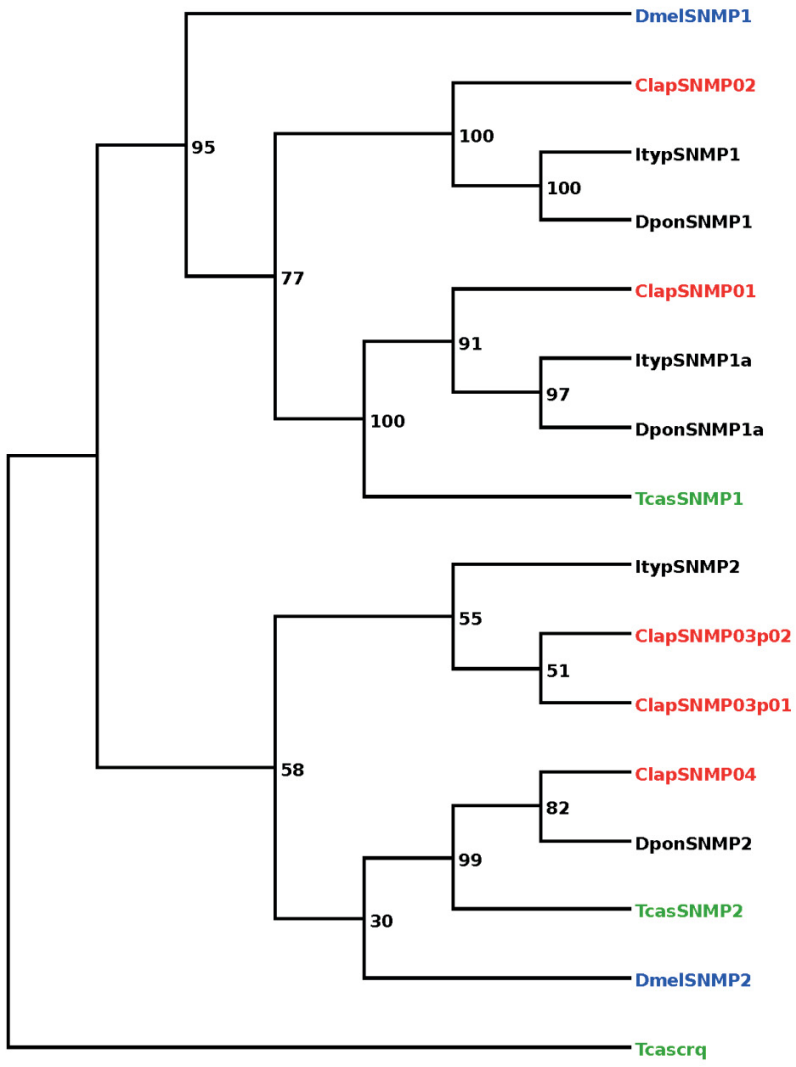

Figure 6. Phylogenetic tree of SNMPs. Blue: *D. melanogaster* (Dmel); Green: *T. castaneum* (Tcas); Black: *D. ponderosae* (Dpon) and *I. typographus* (Ityp); Red: *C. lapponica* (Clap). Two subgroups of SNMPs are identified. Numbers at nodes represent bootstrap values based on 100 replicates, which are shown when the value ≥ 40%.

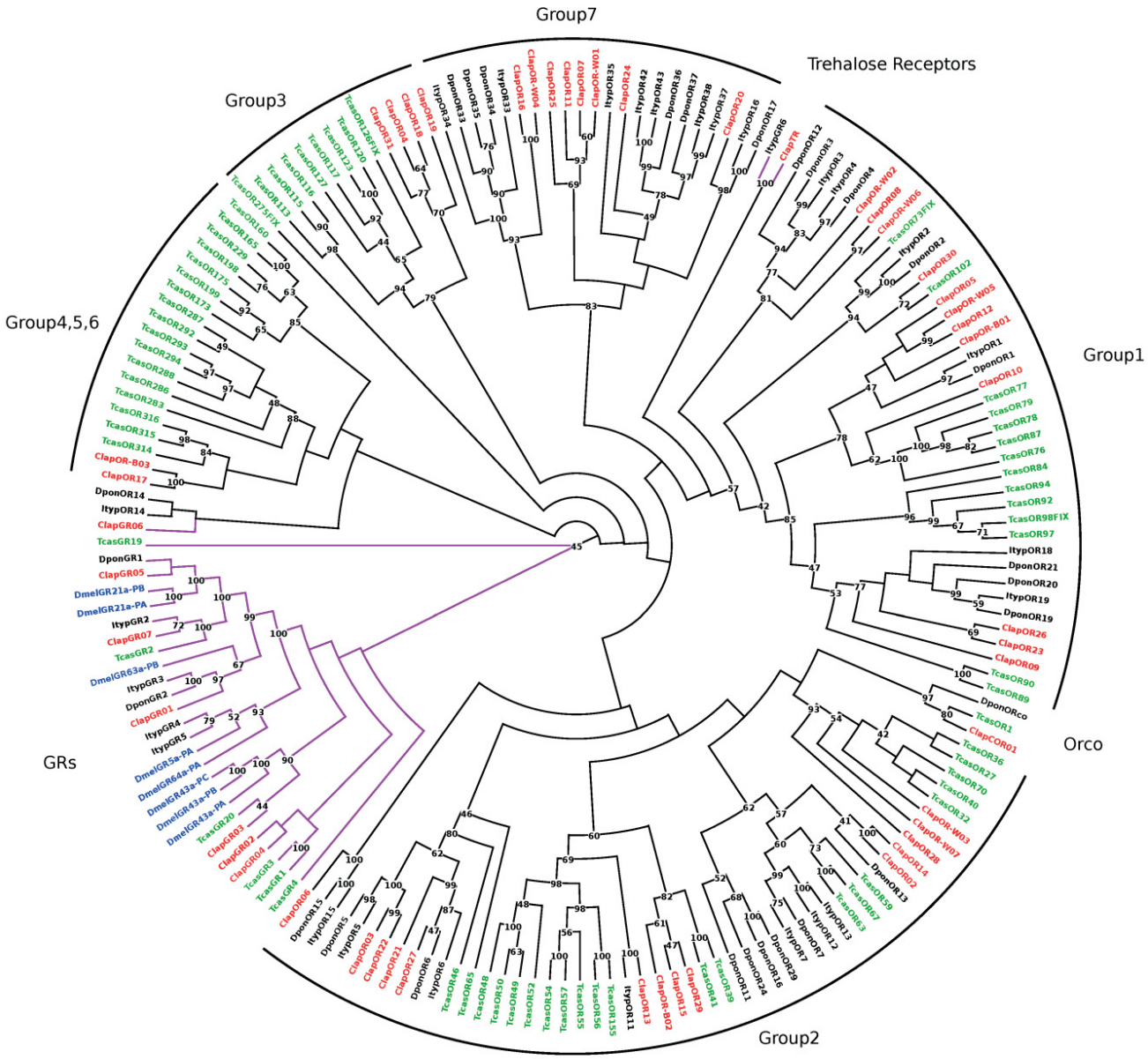

Figure 7. Phylogenetic tree of ORs and GRs. Blue: *D. melanogaster* (Dmel); Green: *T. castaneum* (Tcas); Black: *D. ponderosae* (Dpon) and *I. typographus* (Ityp); Red: *C. lapponica* (Clap). Seven subgroups 1-7 of ORs are identified. Numbers at nodes represent bootstrap values based on 100 replicates, which are shown when the value ≥ 40%.

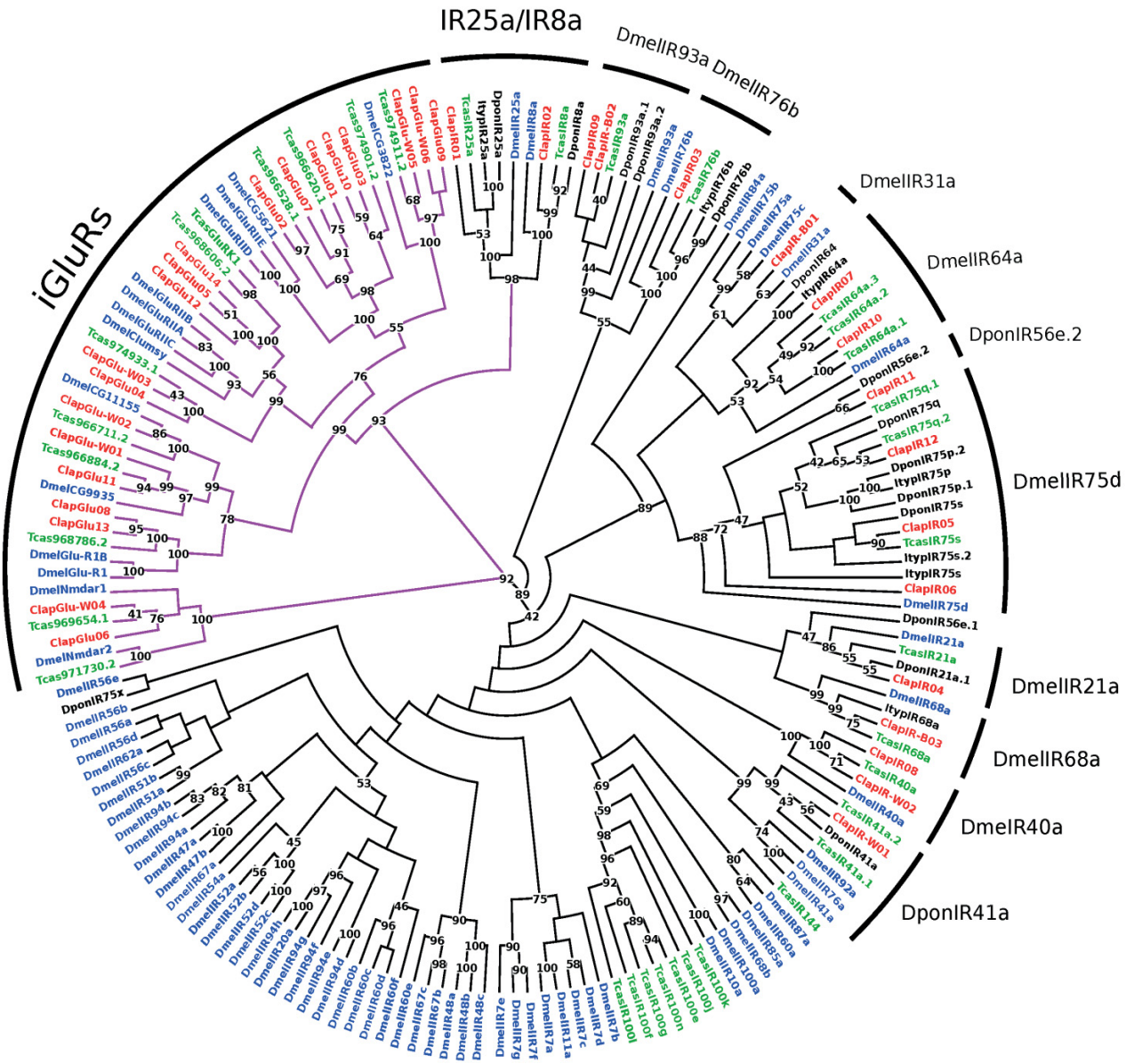

Figure 8. Phylogenetic tree of IRs and iGluRs. Blue: *D melanogaster* (Dmel); Green: *T. castaneum* (Tcas); Black: *D. ponderosae* (Dpon) and *I. typographus* (Ityp); Red: *lapponica* (Clap). Purple edges: iGluRs subgroup. Numbers at nodes represent bootstrap values based on 100 replicates, which are shown when the value ≥ 40%.

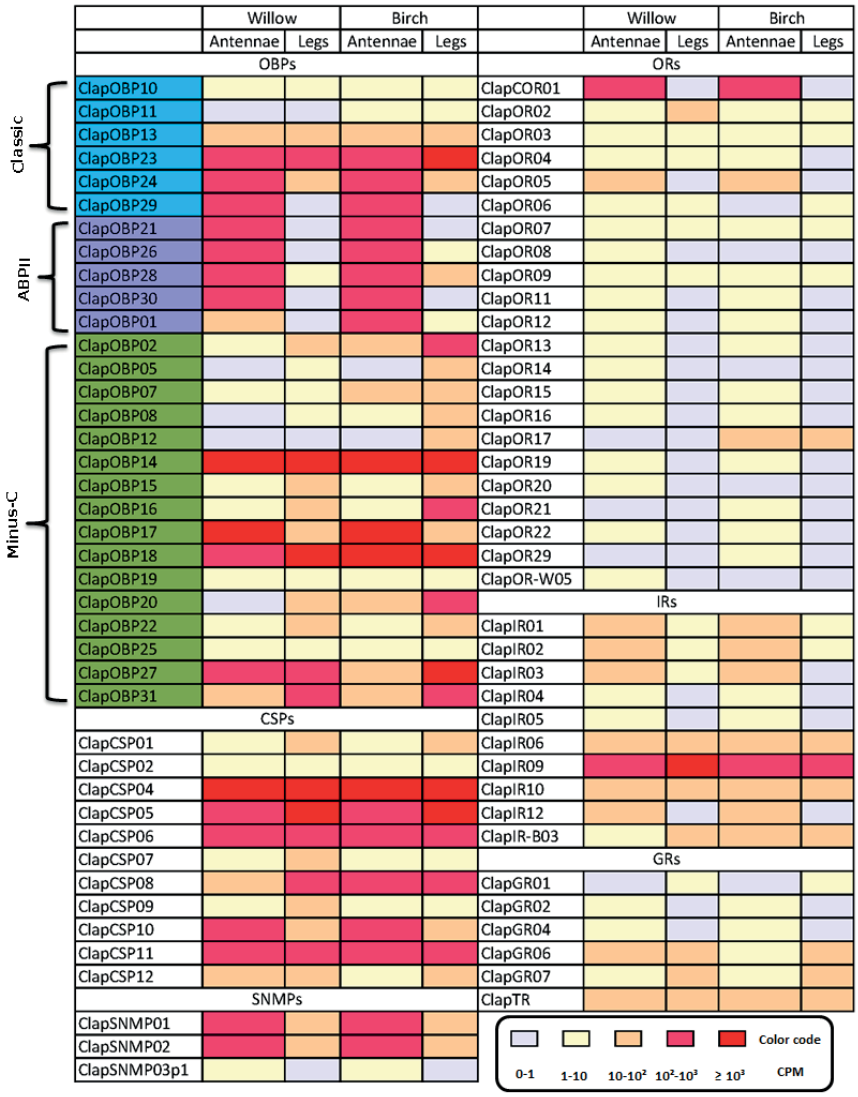

Figure 9. Expression profiles of 80 unique genes from six chemoreception families: OBPs, SNMPs, CSPs, ORs, IRs and GRs from willow-feeding or birch-feeding *C. lapponica* in antennae and legs based on CPM values. RNA-seq reads were normalized to the effective library size. The CPM value of each tissue is derived from four replicates: two in male and female, respectively. Candidates were chosen according to their CPM values of ≤ 1 in at least one of the examined tissues. OBPs are divided into three subgroups classic OBPs (blue), ABPIIs (purple) and minus-C OBPs (green).

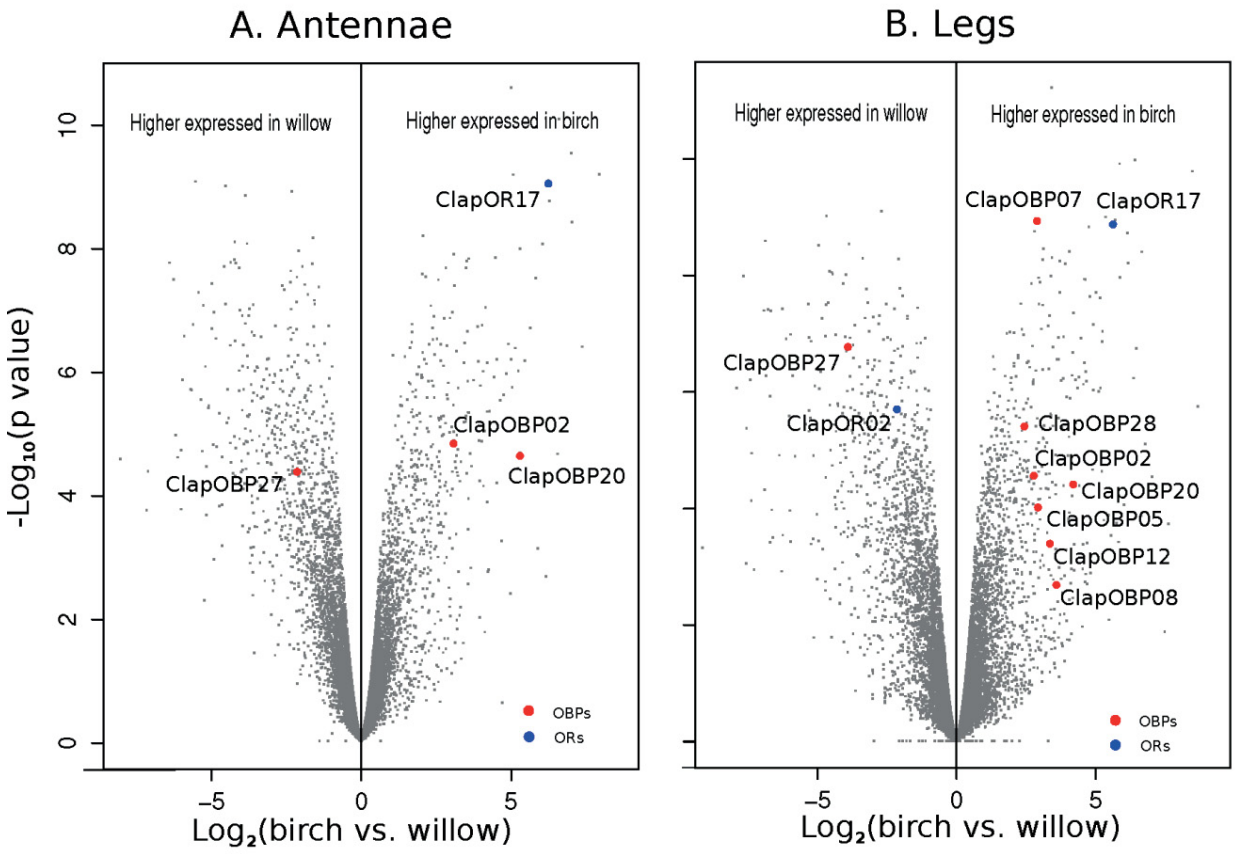

Figure 10. Volcano Plot showing significant differences in the expression level of all chemoreception genes of *C. lapponica* when comparing willow and birch populations. Grey points: differently expressed genes between two populations of *C. lapponica*; red and blue points: significantly different expressed OBPs and ORs between two populations. Significantly different: log2fold≥2, P-value≤0.05 and FDR≤0.05

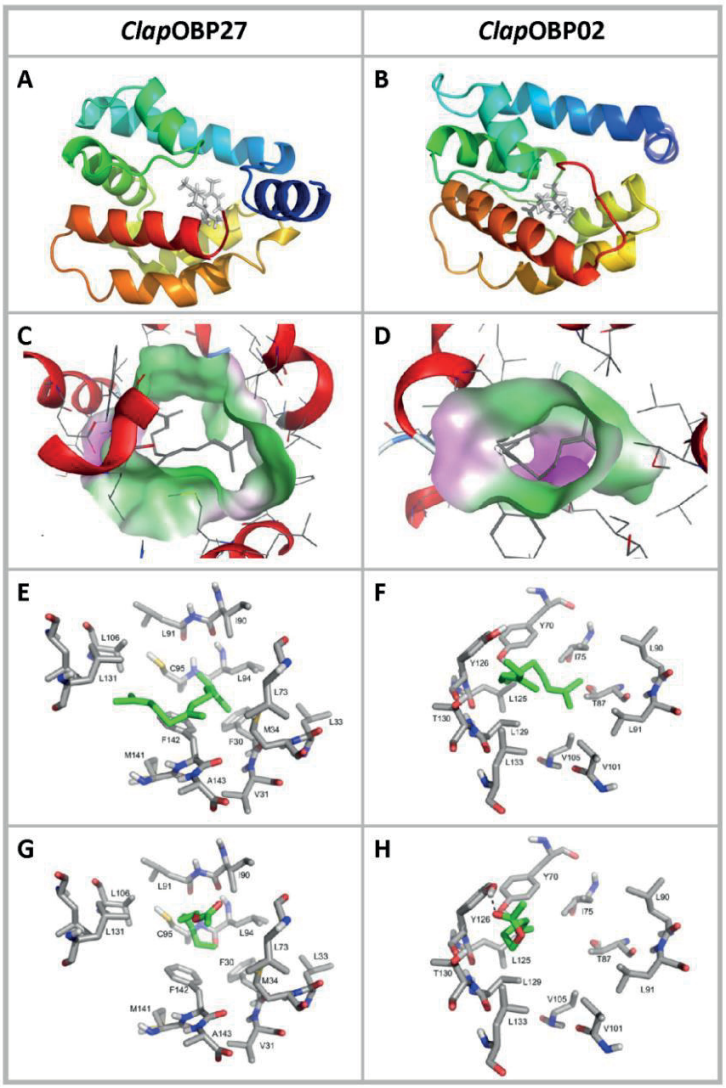

Figure 11. Comparison of tertiary structure models and docking studies of the minus-C OBPs, *Clap*OBP27 (upregulated in willow-feeders) and *Clap*OBP02 (upregulated in birch-feeders). A & B, rainbow representation of the 3-D models (N-terminus dark blue, C-terminus red); C & D, graphical representation of the lipophilic (green) and hydrophilic (red) potential of the binding site of the ligands with docked (*E,E*)-α-farnesene; E & F, details of the interactions of (*E,E*)-α-farnesene in the binding site for each protein; G & H, details of the interactions of cis-3-hexenylacetat in the binding site for each protein. Ligands are highlighted by green carbon atoms.

Table 1. Number of identified chemosensory proteins from willow or birch-feeding populations of *C. lapponica*

	OBPs	CSPs	SNMPs	ORs	GRs	IRs
species-specific	31	12	4	31	8	12
willow-specific	-	-	-	7	-	2
birch-specific	-	-	-	3	-	3

5 General discussion

My dissertation focuses on the specific application and optimization of computationally intensive RNA-seq in ecologically relevant non-model insects. The main analyses including *de novo* assembly of transcriptomes from *P. cochleariae*, *C. populi* and *C. lapponica*, annotation of genes and determination of differential gene expression are described in the manuscripts 3.1-3.3. Three molecular biological topics that centered on the defensive and host plant selection strategies of leaf beetles are explained in the published manuscripts. Aiming to complement the discussion of my publications, I will address the entire workflow of the computational analyses in the following sections and the biological relevance of the results.

5.1 Optimizing RNA-seq analysis

5.1.1 Designing RNA-seq experiments

Sequencing depth. Without genome or/and genetic information, the *de novo* assembly of transcriptomes provides a cost-effective method to study non-model organisms by applying RNA-seq technology (Oppenheim *et al.* 2015). The first challenge during the process of the experimental design was to optimize the sequencing depth prior to the start of the analysis. According to the research from Francis (Francis *et al.* 2013), I chose a sequencing depth of roughly 30 million reads for whole-animals and at least 20 million reads for single-tissue in RNA-seq analysis from the beetles. Increasing the total number of RNA-seq reads (up 30 million reads) improves the detection of the low abundance transcripts but at the same time it creates more noises in different expression analyses (Tarazona *et al.* 2011). A balance between transcript coverage and noise is hence critical for the analysis of RNA-seq data.

Biological replicates. Another important parameter to design an efficient and optimal RNA-seq experiment is to decide the number of biological replicates (Liu *et al.* 2014). Zhang (Zhang *et al.* 2014) observed that the number of detected differentially expressed gene increased continuously as the number of replicates increased. At least three biological replicates per condition are recommended to detect gene expression differences based on statistical analysis (Conesa *et al.* 2016; Lin *et al.* 2016). However an economical limitation would not allow for a

large number of biological replicates, such as in my thesis, only the RNA-seq data for *C. populi* had three biological replicates, the remaining two species had two biological replicates per tissue. In comparison with three biological replicates, the statistical power to detect differential gene expression variations with two biological samples is decreased. Only if the expression differences were greater than 2.5-fold between conditions, the genes can be detected as significantly different for the samples that had two biological replicates (Figure 5). RNA-seq samples with three biological replicates would detect 1.8-fold or greater changes (Lin *et al.* 2016). To gain reliable results I used the housekeeping genes (such as *EF-1 alpha* and *eIF4a*) for normalization to estimate variability across all samples in the experiment (Anders *et al.* 2013).

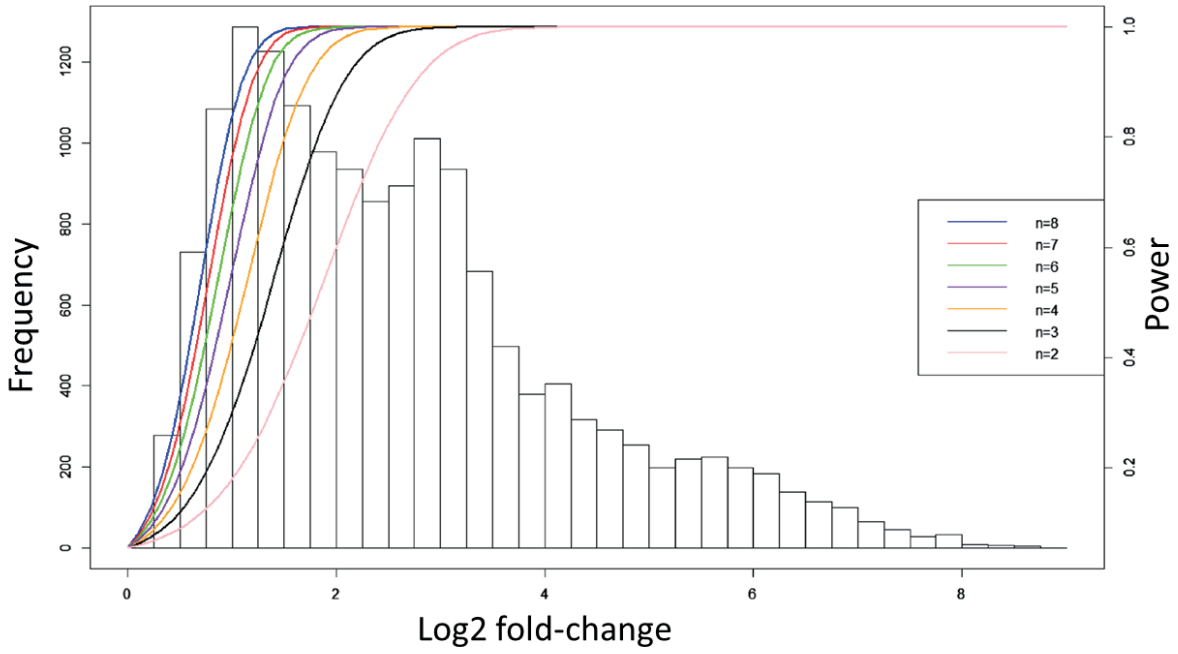

Figure 5. Statistical power analysis. Detectable fold-change versus statistical power for n=2, 3, 4, 5, 6, 7, and 8 biological replicates per sample. Adapted from Lin et al. (2016).

5.1.2 *De novo* assembly

In order to reconstruct a *de novo* transcript library, I selected the most suitable bioinformatic tools for the RNA-seq analysis. Owing to its higher performance in almost all categories using single k-mer method for both small and large data sets, in particular the software Trinity is one of the most convenient tools (Zhao *et al.* 2011). A comparative study of *de novo* assembled libraries between mere pooled whole-animals RNA-seq data and the RNA-

seq data of pooled whole-animals together with tissues in *C. populi* showed that the detection of GRs from eight to 26 increased (analysis not shown). To identify rare transcripts, such as those of odorant or gustatory receptors, from *C. lapponica* the RNA-seq reads from pooled whole-animals of different life stages (100bp paired-end) and different tissues (50bp single-end) were *de novo* assembled together to increase the sequencing depth (see details in the manuscript 3.3).

5.1.3 RNA-seq statistical analysis

Besides the reconstruction of transcriptomes without any information of the genome, another important approach of RNA-seq is the comparative analysis of gene expression levels of two or more conditions such as different species, treatments, life stages, tissues, sexes and so on (Anders & Huber 2010). RNA-seq allows determining the expression of specific genes in any tissue at any time in non-model organisms. To obtain an optimal analysis result, the most critical points of RNA-seq analysis to be considered are the normalization method, the underlying data distribution assumption and the above mentioned biological replicates (Lin *et al.* 2016).

Normalization. To estimate the statistical relevance of differences observed in RNA-seq data sets, the measures of RNA abundance should be normalized in each sample set. Due to the technical bias, such as the length of transcripts and the sequencing depth of samples, the raw read counts are not directly comparable in the research (Robinson & Oshlack 2010). R package DESeq and edgeR are two widely used tools for statistical tests in RNA-seq analysis from the Bioconductor project. Both normalization methods are based on the hypothesis that most genes are not differently expressed. Lin (Lin *et al.* 2016) conducted a comparative study of DESeq and edgeR with other tools such as total count, upper quatile (Bullard *et al.* 2010), median (Dillies *et al.* 2013), RPKM. The result was that the DESeq normalization and TMM from edgeR performed best in the analysis, because both normalization methods not only properly aligned the distribution of the tested data across samples but they also could compensate effectively for RNA-Seq data with a large dynamic range. They have been used for the analysis of the beetles RNA-seq data in my study. DESeq and edgeR possess similar stochastic power but edgeR is more sensitive to outliers than DESeq (Anders *et al.* 2013). Therefore I preferred to use

the R package edgeR in the manuscript 3.3 to balance the extreme values of read counts, because only two biological replicates of each tissue were available due to economical and environmental limitations. DESeq was used in the study of *C. lapponica* in the manuscript 3.1. I combined the normalized read counts with the Lander/Waterman equation (Lander & Waterman 1988) to calculate the average coverage per base in each transcript of each biological replicate. Comparison with RPKM, this combination is better able to describe the biological insight.

Different expression analysis. Based on the read counts, RNA-seq quantification for each gene varied among the samples. This is mainly due to technical and biological variation found in read count data. After normalization of raw read counts, choosing the most fitting statistical model plays an important role for estimating differentially expressed genes. Usually, the variance is assumed to equal the mean value of the technical variation (Bullard *et al.* 2010). However, the variance among biological replicates, particular among small-scale biological replicates, is more variable than in technical replicates (Langmead *et al.* 2010; Robinson & Smyth 2007). This is known as over-dispersion. To account for over-dispersion, a negative binomial distribution is more useful than a Poisson distribution being more tolerant with high variance for random sampling of RNA-seq data. It was observed that the best analysis scheme for the multifactor RNA-seq experiments was to apply either DESeq or edgeR software to generalize a linear model assuming a negative binomial distribution (Anders & Huber 2010; Lin *et al.* 2016). Both were used in manuscript 3.1 and 3.3, respectively.

5.1.4 Outlook and perspective in next-generation sequencing

With cost reduction of RNA-seq, it has become the standard method for transcriptome analysis. Besides the numerous applications of RNA-seq, there are two emerging active fields of RNA-seq, these are: single-cell RNA-seq (Saliba *et al.* 2014) and long-read sequencing (Cho *et al.* 2014).

Single-cell RNA-seq is using very small amounts of starting mRNA from just one single cell. The protocol from Smart-seq (Ramskold *et al.* 2012) allows researchers to study an individual cell and its biological process, allowing to identify new cell types. In biomedical research such as cancer studies, it allows to illuminate the genetic properties of tumor cells and resolves solid

tumor heterogeneity (Schmidt & Efferth 2016). Fluidigm C1 (single-cell mRNA sequencing method) (Xin *et al.* 2016) and DROP-seq (Macosko *et al.*) perform the analysis of 100 to 10,000 cells at a time, and is carried out to identify cell subgroups within populations.

Long-read sequencing is a new sequencing technology also called third-generation sequencing, such as Pacific-Biosciences SMRT (van Dijk *et al.* 2014) and Oxford Nanopore (Drmanac *et al.* 2010), which overcome the limitations from second-generation sequencing especially the limitation of short-read length that negatively affect the reconstruction of full-length transcripts, gene isoform identification, determination of complex genomic regions, and methylation detection (Rhoads & Au 2015). This technology offers amplification-free, single-molecule sequencing of cDNA. The length of long-read sequencing is with an N50 (Ekblom & Wolf 2014) of more than 20kb and maximum read lengths over 60kb. However, the long-read sequencing technology has its own limitations such as high error rate and high costs. A combined utility of short-read and long-read sequencing helps researcher to determinate genome-free isoform-resolved transcriptomes with low error rate.

5.2 Phylogenetic analysis

There are numerous programs to construct phylogenetic trees. In these programs the most common methods are NJ, parsimony, ML and Bayesian. Indeed, there is no best method, but there are best fitting methods. Choosing a method depends on the data. For example, NJ algorithm is very fast but it loses information by compressing the sequences into a distance matrix. It is suitable for very closely related data. The parsimony method is fast enough for the analysis of hundreds of sequences. The best tree from parsimony is always the shortest possible tree that requires the fewest evolutionary changes. One disadvantage is, obviously, that it might not reflect the true evolutionary relationships between data. Another disadvantage is: it could be inconsistent under ‘long branch attraction’. I compared the ML and Bayesian methods that allowed using more information from the sequence multi-alignment and models of evolution. They are hence well suited to analyze distantly related sequences. One drawback is, however, that they are computationally intensive because of the confidence assessing of the trees using the statistical technique ‘bootstrapping’ and posterior probability.

However, the Bayesian method is more sensitive to long branch attraction biases (Kolaczkowski & Thornton 2009; Susko 2008), therefore I chose ML in my thesis.

5.3 Key-role of ABC transporter in defensive system in *C. populi*

Toxic or repellent compounds which are stored in defensive glands protect larvae of leaf beetles against attackers/predators or parasites. To reach the defensive gland, the glucosidic precursors of the repellents have to pass through several membranous barriers. In *C. populi*, only one ABC transporter (*Cp*MRP) has been unambiguously identified until now to be essential for the sequestration of phytochemicals (Strauss *et al.* 2013).

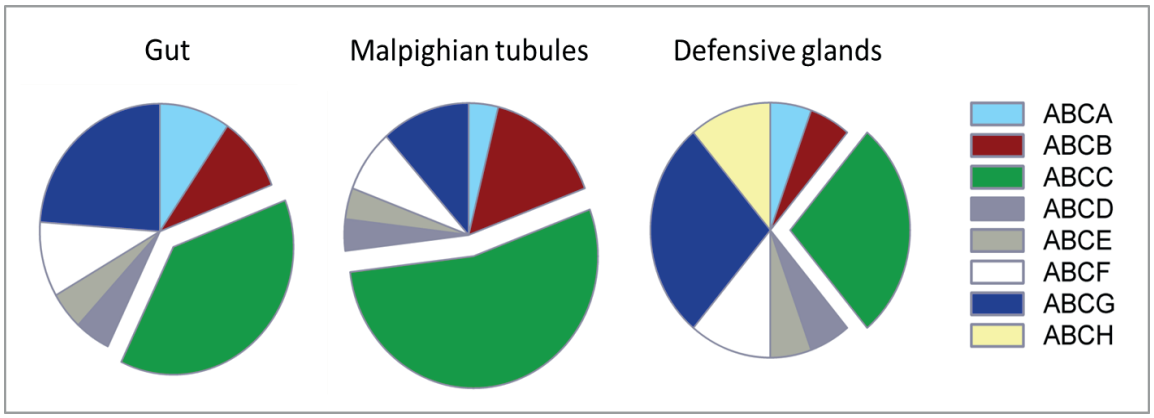

Figure 6. Distribution of ABC proteins from different subfamilies which are highly expressed in three different tissues of *C. populi* larvae. Only sequences with more than 25 counts were considered as being highly expressed.

5.3.1 Tissue expression profiling analysis

I identified 65 putative ABC transporters classified into eight subfamilies: ABCA to ABCH in *C. populi* based on RNA-seq and phylogenetic analysis. Based on RNA-seq analysis in tissues, the different expression levels of ABC transporters showed that the transcripts of subfamilies A, B, C and G were predominately expressed in the gut tissue of *C. populi* (Figure 6). In the phylogentic analysis, almost all of these ABC transporter sequences have been linked with proteins which are associated with xenobiotic or drug resistance (see details in manuscript 3.1). In the Malpighian tubules the ABC transporter candidates from subfamilies B, C and G are dominant. It was observed that the ABCC subfamily has undergone an expansion in silkworm

(Xie *et al.* 2012), red flour beetle (Broehan *et al.* 2013) and also in the *C. populi* with 29 putative members. The multidrug-related ABCC subfamily may play an important role in response to the excretion of waste compounds including phytochemicals (O'Donnell & Rheault 2005). One ABCC transporter displayed an exceptionally high expression level, in particular in the defensive glands. This ABC transporter *Cp*ABC35 (*Cp*MRP) possessed an exceptionally high transcript level, 7000 times higher than in the gut tissue. Silencing of this transporter resulted in the lack of transport of the precursor glucosides for the prodution of salicylaldehyde in the reservoir. It suggests that *Cp*ABC35 may play a key role in the sequestration of salicin from the larvae of *C. populi.*

5.3.2 Comparative analysis of *Cp*ABC35 (*Cp*MRP) in leaf beetles

The larvae of all Chrysomelina share a uniform defensive system, a measure of functional similarity was identified through comparing of similarities of sequences. One highly similar protein sequence from ancestral species *P. cochlearia* was identified with 86% amino acid identity of *Cp*ABC35. In *C. lapponica* a highly similar sequence with at least 96% at amino acid level was also identified in the larval glands of willow- and birch-feeding populations, being almost identical with 99% amino acid identity between the two populations. These identical protein sequences to *Cp*ABC35 from *P. cochlearia* and *C. lapponica* were also highly expressed in the defensive gland tissue of their larvae. Using ABC transporter to overcome the glandular membrane barrier seems to be a common mechanism in the taxon Chrysomelina. Outside of Chrysomelina, an ABCC transporter with high similarity of *Cp*ABC35 from other insects such as red flour beetles, *T. castanaeum,* showed high expression levels in the larval glands which store a stinky substrate (*p*-quinones) for chemical defense (Li *et al.* 2013). Silencing by RNAi of these sequences led to the decrease of defensive coumpounds in the gland tissue. It suggests that not only in the Chrysomelina but also in other insect species, ABC transporters are more widespread for translocation processes in defensive glands.

Taken together, ABC transporters play a key role in the functional model of sequestration of secondary metabolite of plants in insects. In particular, molecular sequence analysis of ABC transporter candidate *Cp*ABC35 from three Chrysomelina species showed that the more advanced specialists (e.g *C. lapponica* feed on willow) may have a homolog of

*Cp*ABC35 that is already pre-adapted to overcome the chemical constraints of shifting to a new host plant (such as birch). In accordance with the opinion that the phytophagous insects often shift host plants with high chemical similarity, the host shift from willow to birch may be achieved through the existence of a "phytochemical bridge"(Gross *et al.* 2004b; Kirsch *et al.* 2011; Menken & Roessingh 1998).

5.4 Adaptation of defensive system in leaf beetles

The defensive system of phytophagous leaf beetles is adapted to the secondary metabolites of host plants. They possess a common mechanism: Nutrition including glucosides pass through the gut membrane into hemolymph; Plant glucosides are tranferred into the glandular system; In the glandular reservoir the glucosides are catalyzed into defensive compounds. In this process two membrane barriers have to be overcome. In comparison with the non-selective uptake of plant-derived glucosides through gut membrane, a selective transfer is observed from hemolymph into the glandular system. A broad substrate spectrum ABCC transporter (*Cp*MRP) acts as a pacemarker (Strauss *et al.* 2013). The high similarity of this transporter among leaf beatles strongly implies that it may allow the affiliation of novel host plants. Similarly, in the glandular tissues the beta-glucosidase represents also an unselective action (seen manuscript 3.2). Beta-glucosidase shares at least 71% identity with each other in *P. cochlearia*, *C. populi* and *C. lapponica* feeding on birch. The functional analysis shows that selectivety of physiological preferred substrates depands on the host plant. However, they are able to hydrolyze a broad spectrum of substrates including glucosidases which are not encountered in the respective larvae in nature. It allows to build a chemical defense that is dependent on the host plant in the larval gland. The high level of specialization doesn't necessarily lead to an "evolutionary dead end" (Termonia *et al.* 2001).

5.5 The chemosensory system in *C. lapponica*

The leaf beetles *C. lapponica* have developed a successful strategy, shifting the host plants from salicin-rich willows to salicin-free birches. For the larvae, the benefits of a salicin-free diet translate to a lowered risk of predation and parasitism that are attracted by

salicylaldehyde. Adaptation in smell and taste is the first step to recognize the different food sources.

5.5.1 Comparative analysis between populations

Of the 113 chemosensory candidate genes which I identified in both populations of *C. lapponica*, 15 transcripts were "population specific" in ORs and IRs. Population specific means there were no sequences with high amino acid level (higher than 80%) in the other population. All of them were expressed very lowly in antennae and legs of adults. These population specific chemosensory transcripts may play a role in the other life stages. On the other hand it was observed in *D. melanogaster* that alternatively spliced ORs, Or46aA and Or46aB share only 35% amino acid identity (Ray *et al.* 2014). Therefore, to identify population specific chemosensory transcripts it not only depends on the protein sequence similarity but also on a detailed analysis of genetic locus. Due to the lack of genome information of *C. lapponica* in this study, the unique sequences that share low amino acid identity between both populations were identified as population specific chemosensory proteins. I also identified one receptor *Clap*GR05 in both populations that share 82% amino acid identity. A gap of 21 amino acids were observed between both populations of *C. lapponica* in the alignment. It may indicate the occurrence of alternative splicing events. However either truncation or extension or skipping of intron (extron) sequences are not clearly studied. The two alternatively spliced gustatory receptors of *Clap*GR05 probably possessed different functions between willow-feeder and birch-feeder populations. This aspect was also represented in alternatively spliced *Or46a* from fruit flies. However, the very low expression levels of both spliced *Glap*GR05 in the tissue expression profiling didn't allow us to draw any further conclustions with the given data.

5.5.2 Comparison analysis focusing on OBPs between antennae and legs

Besides the differences between the two populations, I also have identified significant differences among the two organs – antennae (olfactory organ) and legs (non-olfactory organ) (Figure 7). We observed significant high expression of predominantly classic OBPs and ABPII in antennae – seven out of 11 OBPs (Figure 8); a fact that is consistent with findings in other insects including *T. castaneum, B. mori, A. meliffera* (Dippel *et al.* 2014; Foret & Maleszka 2006; Qiao *et al.* 2013). Only the classic OBP23 was higher expressed in legs than in antennae. In contrast to the classic OBPs and ABPII, the Minus-C OBPs were mainly expressed in legs with two exceptions, OBP17 and OBP27 which exhibited a higher expression in antennae. Unlike OBPs, most of CSPs from both populations of *C. lapponica* display no different expression levels between antennae and legs. They are highly expressed in both tissues. CSPs contain the second most highly expressed candidates in all chemosensorty families in *C.lapponica*. Antennae and legs are the most cuticle-rich tissues. Some CSPs may function in cuticle synthesis (Foret *et al.* 2007). Several OBPs were also observed to be highly expressed in cuticle but not specifically high in the chemosensory hairs (Galindo & Smith 2001; Park *et al.* 2000). It is observed also in SNMPs, ORs and IRs, the highly or very highly expressed candidates show upregulated expression in legs or antenna. As expected (Gu *et al.* 2014; Vogt *et al.* 2009), SNMP01, SNMP02 and coOR01 expressed higher in antennae. The co-receptor of ORs is the single gene with high expression level in both *C.lapponica* populations. One highly expressed IR from our data (*Clap*IR09) formed a group with DmIR93a (bootstrapping value 99), which is coexpressed in neurons surrounding the sacculus (Benton *et*

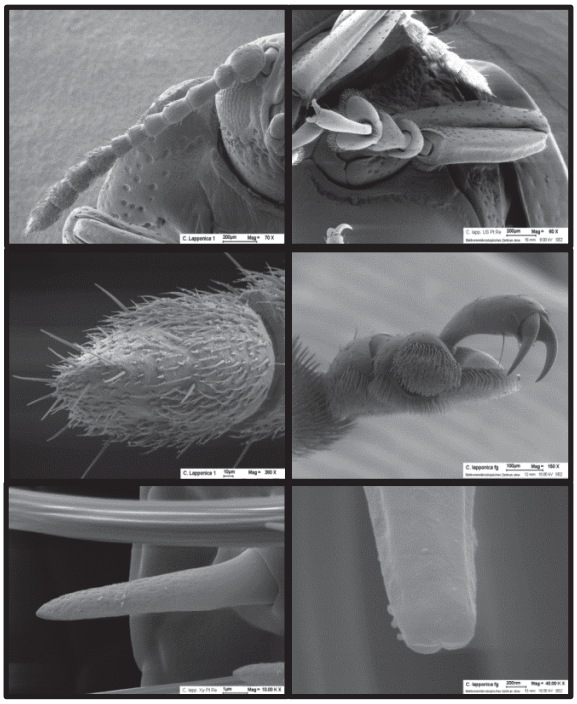

Figure 7. Scanning electron micrographs of the external morphology of a birch-feeding *C. lapponica* female. A. antennae; B. tip of an antenna; C. tip of an antenna sensillum; D. Leg; E. tarsus; F. tip of a tarsal sensillum.

al. 2009). IR09 expressed higher in legs in willow population.

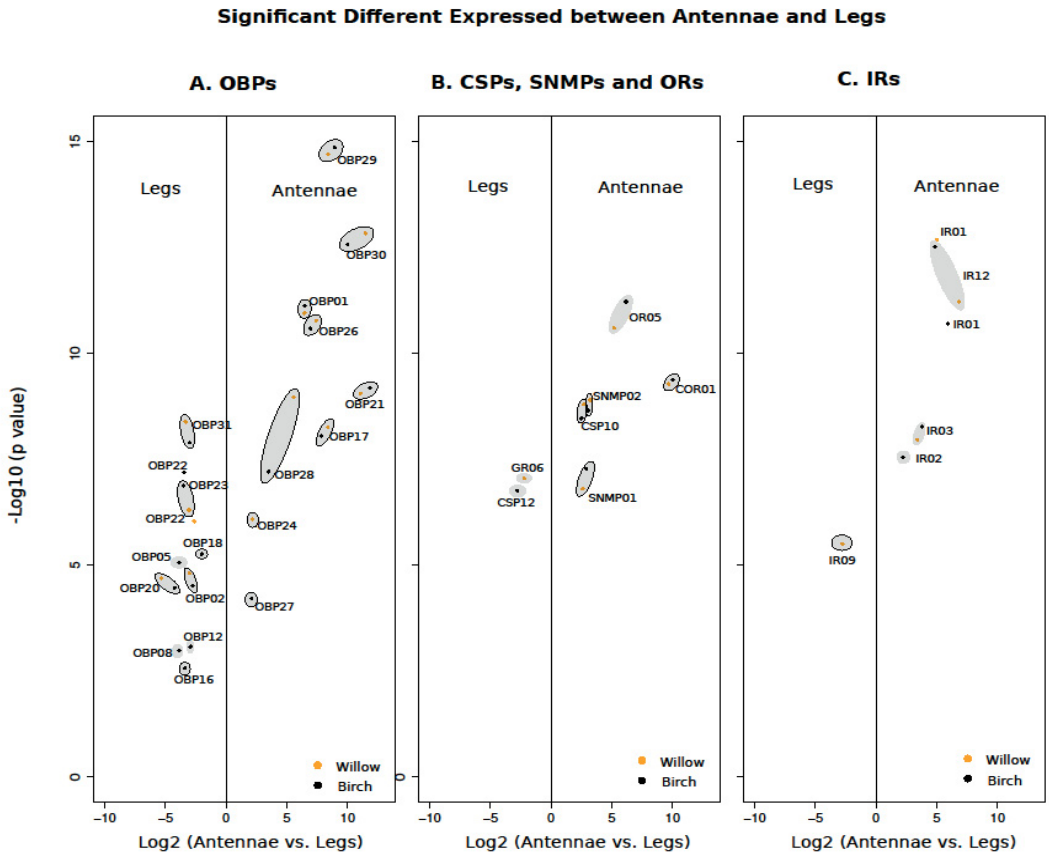

Figure 8. Significant different expressed chemosensory genes between antennae and legs in both populations of *C. lapponca*. Yellow and black points: significant different expressed chemosensory genes between antennae and legs from willow and birch. Grey cloud locates the identified genes from both populations. The cloud is circled in black that shows the chemosensory genes highly expressed in at least one population (count per million ≥ 100). Significantly different: log2fold≥2, P-value≤0.05 and PDR≤0.05

The only one classic OBP28 and other six minus-C OBPs are upregulated in the legs of birch-feeder population. This result in combination with comparable analysis between antennae and legs in each population showed that five OBPs 02, 05, 08, 12 and 20 of the above upregulated minus-C OBPs are also higher expressed in legs than in antennae in at least one birch-feeder population Fig. OBPs highly expressed in legs have been reported in species like *Adelphocoris lineolatus*, *Bactrocera dorsalis*, *Sitodiplosis mosellana* and *Apolygus lucorum*. In legs OBPs may be carriers of ligands other than odorants. Hai-Bin Yuan suggested that OBPs may play a role in gustation rather than in olfaction in legs from *Apolygus lucorum* (Yuan *et al.*

2015). OBPs occur also in saliva, and proteomics studies of defensive secretions indicated that these proteins are also components of such extracellular fluids. In *Helicoverpa armigera*, an OBP which is abundant in male seminal fluid and transferred to females, is able to carry oviposition deterrents to label fertilized eggs, thus prompting the female moth away from the location where the first egg was laid (Sun *et al.* 2012). Hence, despite the function in olfaction, OBPs may have more roles in different physiological processes than known today.

Taken together, in *C. lapponica*, there are in total 10 minus-C OBPs that are significantly higher expressed in legs than in antennae in the both populations. Five of them are higher expressed in birch-feeder than in willow-feeder population. Due to the lack of the third disulfide bridge of protein 3D structure of OBPs, it allows the minus-C OBPs to bind to different molecules (Schwaighofer *et al.* 2014). Higher expressed OBPs in the legs of birch-feeder than in willow-feeder are involved in gustatory chemosensation such as in honey bees, fruit flies and ants (de Brito Sanchez *et al.* 2008; Ling *et al.* 2014; McKenzie *et al.* 2014). When moving on the leaf surface, the adults might rasp it and taste the released molecules with their legs in order to identify a possible food source.

6 Summary

RNA-seq provides a way for discovery and quantification of transcriptional elements as an alternative to hybridization based technologies like microarrays. It offers measurement of expression independent of any reference sequence with/without prior knowledge about genome sequence or gene information. However, finding the most appropriate in numerous various methods of RNA sequencing is a challenge in this study.

This thesis focuses on the RNA-seq analysis for three non-model leaf beetles without genome sequences. Three manuscripts are presented that comprise transcript library *de novo* assembly, transcript identification/quantification, normalization and differential expression analysis by applying RNA-seq analysis. Depending on the specific question, a complete R- and perl-based computational pipeline for the analyses of RNA-seq data was implemented. To identify and annotate the investigated genes the transcript libraries were aligned to public databases and in-house data sets. Phylogenetic analysis was used to characterize the functions of gene. The investigated transcripts could be grouped with sequence homologies from well-studied species for additional information. Tissue expression profiling was analyzed by applying R package DESeq and edgeR. In the manuscript 3.1 qPCR and RNAi were used for verification of RNA-seq. RNAi targeting the most abundant ABC transporter in *C. populi* resulted in a strong decrease in the amount of deterrent in the defensive gland in larvae.

Based on the RNA-seq data, the adaptation mechanisms of *P. cochleariae*, *C. populi* and *C. lapponica* to the host plants in defensive system and olfactory system were studied.

Plants and insects have evolved a variety of beneficial and deleterious interactions. Host plants developed toxic compounds as a chemical defense against herbivores. Meanwhile the insects take advantage of the secondary toxic metabolites of the host plants to detect food source or oviposition sites. Insects evolved sophisticated strategies to circumvent the noxious effects of plant toxins including uptake avoidance, metabolic conversion, target alteration or sequestration (Hartmann 2004; Speed *et al.* 2015). Sequestration involves uptake, storage and occasionally conversion of plant-derived compounds. Often these compounds, which are either *de novo* produced or sequestrated from plant-derived glucosides, are used for insect's own

chemical defense against their enemies. According to the phylogenetic analysis of Chrysomelina species from Termonia 2001, a more advanced species *C. lapponica* has undergone a host plant shift from willow to birch. The larvae possess a mixed strategy of *de novo* and sequestration.

The transfer of non-toxic precursors through gut membranes is a non-selective uptake of phenolglucosides, while transfer between hemolymph and the glandular system is selective. An ABCC transporter *Cp*ABC35 (*Cp*MRP) from *C. populi* has a pacemaker function in the transport of glucosides. Knockdown of this transporter led to the lack of defensive compounds in the reservoirs of larvae. The broad substrate spectrum of this ABCC transporter and beta-glucosidase among *P. cochlearia*, *C. populi* and *C. lapponica* represent the non-selective barriers in the defensive system for host plant shifts. A high sequence similarity among them indicates that a pre-adaptation of these sequences increases the adaptation to a new host plant.

The population of *C. lapponica* feeding on birch benefits from the enemy-free niche where the larvae suffer less from the specialist predation and parasitism that are attracted by salicylaldehyde. The chemosensory system of beetle adults has adapted to the odor of novel plants. A comparison analysis of chemosensory genes between two populations (feeder willow and feeder birch) showed that minus-C OBPs and ORs were differentially expressed in olfactory organs (antennae) and non-olfactory organs (legs). The affinity calculation of OBP homology modeling and mostly leaf-released chemical cues, indicated that non-polar OBP27 preferred to bind hydrophobic plant-derived terpenoids (such as (*E*,*E*)-alpha-farnesene from willow) while OBP20 bound more hydrophilic phytochemicals (*cis*-3-hexenyl acetate from birch) (see manuscript 3.3). The lack of the third disulfide bridge in the 3D-structure enable minus-C OBPs to bind various different chemostimuli. In combination with comparative analysis of chemosensory genes between antennae and legs, it showed that minus-C OBPs not only in antennae but also in legs may play an important role to adapt to the chemostimuli of novel plants.

In conclusion, the combination of RNA-seq, phylogentic analysis and qPCR as well as RNAi can be very powerful in molecular biology of non-model organisms. Although this study was conducted in leaf beetles, the analysis could be generalized to study other non-model organisms.

7 Zusammenfassung

Alternativ zur auf Hybridisierung basierenden Microarray-Technologie bietet RNA-seq eine Methode zur Entdeckung und Quantifizierung von den transkriptionellen Elementen. Es kann unabhängig von Referenzsequenzen und dem Vorhandensein von Vorkenntnissen über die Genomsequenz oder Geninformationen zur Genexpressionsmessung verwendet werden. Allerdings ist das Finden der am besten geeigneten unter zahlreichen möglichen Methoden der RNA-Sequenzierung eine der Herausforderungen in dieser Arbeit.

Diese Dissertation legt den Fokus auf die RNA-seq-Analyse von drei Nicht-Modell-Blattkäfern ohne gegebene Genomsequenzen. In den drei Manuskripten werden durch Anwendung der RNA-seq-Analyse Transkriptbibliotheken durch *De-Novo*-Assemblierung erzeugt, deren Transkripte identifiziert und quantifiziert sowie Normalisierung und differentielle Expressionsanalyse durchgeführt. Eine vollständige R- und Perl-basierte Verarbeitungspipeline wurde unter Berücksichtigung der jeweiligen Fragestellung für die Analyse der RNA-seq-Daten umgesetzt. Zur Identifizierung und Annotation der untersuchten Gene wurde die Transkript-Bibliothek mit öffentlichen und eigenen Datenbanken aligniert. Zur Charakterisierung der Funktion der Gene wurde phylogenetische Analyse eingesetzt. Die untersuchten Transkripte konnten zur Gewinnung zusätzlicher Information mit entsprechenden Sequenzhomologien aus gut erforschten Spezies gruppiert werden. Unter Anwendung der R-Pakete DESeq und edgeR wurden die Gewebeexpressionsprofile erstellt. Im Manuskript 3.1 wurden qPCR und RNAi zur Überprüfung der Ergebnisse aus RNA-seq verwendet. Der Knockout mittels RNAi des häufigsten ABC-Transporter von *C. populi* führte zur starken Reduzierung der Abschreckungsmittel in den Verteidigungsdrüsen der Larven.

Auf Basis der RNA-seq-Daten wurden die Anpassungsmechanismen *von P. cochleariae*, *C. populi* und *C. lapponica* an die Wirtspflanzen bezüglich des olfaktorischen und des Verteidigungssystems untersucht.

Pflanzen und Insekten haben eine Vielzahl von nützlichen und schädlichen Wechselwirkungen hervorgebracht. Wirtspflanzen entwickelten toxische Verbindungen zur chemischen Abwehr gegen Herbivoren. Gleichzeitig verwenden Insekten die sekundären toxischen Metaboliten der Wirtspflanzen, um Futterquellen oder Eiablagestellen zu erkennen.

Insekten haben ausgereifte Strategien entwickelt, um die schädlichen Wirkungen der Pflanzengifte zu umgehen, z.B. Verhinderung der Aufnahme, metabolische Umwandlung, Veränderung der Zielproteine oder Sequestrierung . Sequestrierung beinhaltet die Aufnahme, Speicherung und gelegentliche die Umwandlung der pflanzlichen Verbindungen. Oft werden Verbindungen, *de novo* produzierte oder aus pflanzlichen Glukoside sequestrierte, von den Insekten für ihre eigene Verteidigung gegen Feinde eingesetzt. Nach der phylogenetischen Analyse der Chrysomelina-Arten von Termonia 2001 gab es bei der weitentwickelte Spezies *C. lapponica* einen Wechsel der Wirtspflanze von Weide zu Birke. Deren Larven verwenden eine Mischstrategie aus *de novo*-Erzeugung und Sequestration.

Der Transport von nichttoxischen Vorprodukten durch die Darmmembran passiert als nicht-selektive Aufnahme von Phenolglukosiden, während der Transport zwischen Hämolymphe und dem Drüsensystem selektiv ist. Ein ABCC-Transporter *Cp*ABC35 (*Cp*MRP) fungiert als Schrittmacher beim Transport der Glukoside. Ein Knockdown dieses Transporters führte zu einem Mangel an Verteidigungsverbindungen im Reservoir der Larven. Das breite Substratspektrum dieses ABCC-Transporters und Beta-Glukosidase bei *P. cochlearia*, *C. populi* und *C. lapponica* deutet auf nichtselektive Barrieren im Verteidigungssystem bei Wechsel der Wirtspflanzen hin. Eine hohe Sequenzähnlichkeit zwischen den Arten weist darauf hin, dass eine Präadaptation dieser Sequenzen die evolutionäre Anpassung neuer Wirtpflanzen erhöht.

Die birkenfressende *C. lapponica*-Population profitiert von einer Nische ohne Feinde, in der die Larven weniger durch spezialisierte Parasiten und Räubern, die durch Salicylaldehyd angelockt werden, bedroht sind. Das chemosensorische System der erwachsenen Käfer hat sich dem Geruch der neuen Pflanzen angepasst. Eine Vergleichsanalyse der chemosensorischen Gene zwischen zwei Populationen (birkenfressend und weidefressend) zeigte, dass minus-C OBPs und ORs in Geruchsorganen (Antennen) und Organen ohne Geruchssinn (Beine) signifikant unterschiedlich exprimiert sind. Die Affinitätsberechnung von OBP Homologie-Modellierung und hauptsächlich durch Blätter freigesetzter chemischer Signale deutet darauf hin, dass unpolare OBP27 pflanzliche hydrophober Terpenoide-Bind (z.B. (*E*,*E*)-alpha-Farnesen aus Weide) bevorzugt. Gleichzeitig band OBP20 mehr mit hydrophilen Phytochemikalien (*cis*-3-Hexenylacetat aus Birken) (Manuskript 3.3). Das Fehlen der dritten Disulfidbrücke in der 3D-

Struktur erlaubt minus-C OBPs, an verschiedene chemische Orientierungsstoffe zu binden. In Verbindung mit der Vergleichsanalyse der chemosensorischen Gene zwischen Antennen und Beinen zeigt dies, dass minus-C OBPs sowohl in Antenne wie in den Beinen eine wichtige Rolle spielen könnten bei der Anpassung an die Orientierungsstoffe neuer Wirtspflanzen.

Für die Molekularbiologie von Nicht-Modell-Organismen bietet die Kombination von RNA-seq, phylogenetischer Analyse und qPCR sowie RNAi enormes Potential. Zwar wurde in dieser Arbeit mit Blattkäfern gearbeitet, aber die Analyse ließe sich zur Untersuchung anderer Nicht-Modell-Organismen verallgemeinern.

8 Danksagung

An dieser Stelle möchte ich meinen besonderen Dank nachstehenden Personen entgegen bringen, ohne deren Mithilfe die Anfertigung dieser Promotionsschrift niemals zustande gekommen wäre:

Zuerst möchte ich Dr. Antje Burse danken, die mir die Käferwelt vorgestellt hat. Ich danke ihr herzlich für die Betreuung meiner Arbeit. Antje unterstützte mein Studium mit einem offenen Ohr und Geduld für Fragen, Diskussion und neuen Ideen. Außerdem bedanke ich mich sehr dafür, dass Sie mein Englisch in der Vorbereitung der Vorträge, Poster und Paper ausführlich verbessert hat.

Prof. Dr. Wilhelm Boland möchte ich für die freundliche Überlassung des hochinteressanten Themas und die Bereitstellung des Arbeitsplatzes herzlich danken.

Ich danke auch alle Mitgliedern der Käfergruppe und allen Kollegen meiner Abteilung, die meine Arbeit unterstützt haben. Insbesondere möchte ich Dr. Magdalena Stock, Dr. Marco Groth, Dr. Andreas Petzold, Thomas Brockmöller, Martin Bens, ZhiHao Lin für die hilfreichen Diskussionen über aktuelle Bioinformatik-Analysen danken. Hier möchte ich mich auch Prof. Dr. Reinhard Guthke bedanken, dass er mir erlaubte an den Seminaren seiner Bioinformatik-Forschungsgruppe im Leibniz-Institute in Jena teilzunehmen.

Dr. Natalie Wielsch möchte ich auch für die Proteom-Analysen danken.

Dr. Marrita Kunert, Kerstin Ploss und Angelika Berg danke ich für das Käfersammeln, die Haltung sowie die Blattduftstoffanalyse.

Ein großer Dank gilt auch meiner Familie und Freunden, die mich in jeder Hinsicht unterstützt haben. Danke für die Zeit die wir miteinander verbracht haben! Der allergrößte Dank aber gilt Bernhard, der mich immer begleitet hat, auch wenn ich ihn gestresst habe.

Vielen Dank für Alles!!

9 Eigenständigkeitserklärung

Hiermit erkläre ich, dass mir die geltende Promotionsordnung (§ 5 PromO) der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist. Ich habe die vorliegende Arbeit selbständig und nur mit angegebenen Hilfesmitteln, Literatur und Daten angefertigt. Die übernommenen Inhalte aus anderen Quellen und von anderen Personen sind eindeutig kenntlich gemacht. Alle Personen, die sich an dieser Arbeit und meiner Publikationen teilgenommen haben, wurden in der Co-Autoren List und der Danksagung ausgeführt.

Ich versichere, dass weder ein Promotionsberater in Anspruch genommen wurde, noch das Dritte unmittelbar oder mittelbar von mir geldwerte Leistung für Arbeiten erhalten haben, die im Zusammenhang mit den Inhalt der vorgelegten Dissertation stehen. Die vorliegende Arbeit wurde zu keinem früheren Zeitpunkt, weder im Inland noch im Ausland in gleicher oder ähnlicher Form bei einer anderen Hochschule oder Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens eingereicht.

Ding Wang
Jena, den

10 References

Agrawal AA, Petschenka G, Bingham RA, Weber MG, Rasmann S (2012) Toxic cardenolides: chemical ecology and coevolution of specialized plant–herbivore interactions. *New Phytologist* **194**, 28-45.

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology* **11**, R106.

Anders S, McCarthy DJ, Chen Y*, et al.* (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc* **8**, 1765-1786.

Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169.

Andrews S (2016) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ashburner M, Ball CA, Blake JA*, et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29.

Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB (2009) Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. *Cell* **136**, 149-162.

Blum MS, Wallace JB, Duffield RM*, et al.* (1978) Chrysomelidial in the defensive secretion of the leaf beetleGastrophysa cyanea Melsheimer. *J Chem Ecol* **4**, 47-53.

Bodemann RR, Rahfeld P, Stock M*, et al.* (2012) Precise RNAi-mediated silencing of metabolically active proteins in the defence secretions of juvenile leaf beetles. *Proc Biol Sci* **279**, 4126-4134.

Boland W (2015) Sequestration of plant-derived glycosides by leaf beetles: A model system for evolution and adaptation. In: *Proceedings of the Beilstein Bozen Symposium 2014 – Chemistry and Time*, pp. 38-48.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

Brenner S, Johnson M, Bridgham J*, et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*

18, 630-634.

Broehan G, Kroeger T, Lorenzen M, Merzendorfer H (2013) Functional analysis of the ATP-binding cassette (ABC) transporter gene family of *Tribolium castaneum*. *Bmc Genomics* **14**, 6.

Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *Bmc Bioinformatics* **11**, 94.

Buss DS, Callaghan A (2008) Interaction of pesticides with p-glycoprotein and other ABC proteins: A survey of the possible importance to insecticide, herbicide and fungicide resistance. *Pesticide Biochemistry and Physiology* **90**, 141-153.

Celaj A, Markle J, Danska J, Parkinson J (2014) Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. *Microbiome* **2**, 39.

Cho H, Davis J, Li X*, et al.* (2014) High-resolution transcriptome analysis with long-read RNA sequencing. *PLoS One* **9**, e108095.

Chopra R, Burow G, Farmer A*, et al.* (2014) Comparisons of *de novo* transcriptome assemblers in diploid and polyploid species using peanut (Arachis spp.) RNA-Seq data. *PLoS One* **9**, e115055.

Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res* **38**, 1767-1771.

Conesa A, Madrigal P, Tarazona S*, et al.* (2016) A survey of best practices for RNA-seq data analysis. *Genome Biology* **17**, 13.

de Brito Sanchez MG, Chen C, Li J*, et al.* (2008) Behavioral studies on tarsal gustation in honeybees: sucrose responsiveness and sucrose-mediated olfactory conditioning. *Journal of comparative physiology, A* **194**, 861-869.

de Magalhaes JP, Finch CE, Janssens G (2010) Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Res Rev* **9**, 315-323.

del Campo ML, Miles CI, Schroeder FC*, et al.* (2001) Host recognition by the tobacco hornworm is mediated by a host plant compound. *Nature* **411**, 186-189.

Dillies MA, Rau A, Aubert J*, et al.* (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* **14**, 671-683.

Dippel S, Oberhofer G, Kahnt J*, et al.* (2014) Tissue-specific transcriptomics, chromosomal localization, and phylogeny of chemosensory and odorant binding proteins from the red flour beetle *Tribolium*

castaneum reveal subgroup specificities for olfaction or more general functions. *Bmc Genomics* **15**, 1141.

Djegbe I, Agossa FR, Jones CM*, et al.* (2014) Molecular characterization of DDT resistance in *Anopheles gambiae* from Benin. *Parasit Vectors* **7**, 409.

Drmanac R, Sparks AB, Callow MJ*, et al.* (2010) Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science* **327**, 78-81.

Duan JL, Xia C, Zhao GY, Jia JZ, Kong XY (2012) Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. *Bmc Genomics* **13**.

Eben A, Barbercheck ME, Martín AS (1997) Mexican diabroticite beetles: II. Test for preference of cucurbit hosts by Acalymma and Diabrotica spp. *Entomologia Experimentalis et Applicata* **82**, 63-72.

Ekblom R, Wolf JB (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl* **7**, 1026-1042.

Feld BK, Pasteels JM, Boland W (2001) *Phaedon cochleariae* and *Gastrophysa viridula* (Coleoptera : Chrysomelidae) produce defensive iridoid monoterpenes *de novo* and are able to sequester glycosidically bound terpenoid precursors. *Chemoecology* **11**, 191-198.

Felsenstein J (1993) *PHYLIP (Phylogeny Inference Package)* Department of Genetics, University of Washington, Seattle. http://evolution.genetics.washington.edu/phylip.html

Finn RD, Coggill P, Eberhardt RY*, et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-285.

Foret S, Maleszka R (2006) Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res* **16**, 1404-1413.

Foret S, Wanner KW, Maleszka R (2007) Chemosensory proteins in the honey bee: Insights from the annotated genome, comparative analyses and expressional profiling. *Insect Biochem Mol Biol* **37**, 19-28.

Fossog Tene B, Poupardin R, Costantini C*, et al.* (2013) Resistance to DDT in an urban setting: common mechanisms implicated in both M and S forms of *Anopheles gambiae* in the city of Yaounde Cameroon. *PLoS One* **8**, e61408.

Francis WR, Christianson LM, Kiko R*, et al.* (2013) A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. *Bmc Genomics* **14**, 167.

Galindo K, Smith DP (2001) A large family of divergent *Drosophila* odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetics* **159**, 1059-1072.

Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* **14**, 685-695.

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews: Genetics* **17**, 333-351.

Grabherr MG, Haas BJ, Yassour M*, et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652.

Gross J, Fatouros NE, Hilker M (2004a) The significance of bottom-up effects for host plant specialization in Chrysomela leaf beetles. *Oikos* **105**, 368-376.

Gross J, Fatouros NE, Neuvonen S, Hilker M (2004b) The importance of specialist natural enemies for *Chrysomela lapponica* in pioneering a new host plant. *Ecological Entomology* **29**, 584-593.

Gross J, Podsiadlowski L, Hilker M (2002) Antimicrobial activity of exocrine glandular secretion of Chrysomela larvae. *J Chem Ecol* **28**, 317-331.

Gu SH, Sun L, Yang RN*, et al.* (2014) Molecular characterization and differential expression of olfactory genes in the antennae of the black cutworm moth *Agrotis ipsilon*. *PLoS One* **9**, e103420.

Guindon S, Dufayard JF, Lefort V*, et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

Haas BJ, Papanicolaou A, Yassour M*, et al.* (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512.

Hallem EA, Dahanukar A, Carlson JR (2006) Insect odor and taste receptors. *Annu Rev Entomol* **51**, 113-135.

Han Y, Gao S, Muegge K, Zhang W, Zhou B (2015) Advanced Applications of RNA Sequencing and Challenges. *Bioinform Biol Insights* **9**, 29-46.

Hartmann T (2004) Plant-derived secondary metabolites as defensive chemicals in herbivorous insects: a case study in chemical ecology. *Planta* **219**, 1-4.

Hilker M, Schulz S (1994) Composition of larval secretion of *Chrysomela lapponica* (Coleoptera, Chrysomelidae) and its dependence on host plant. *J Chem Ecol* **20**, 1075-1093.

Hornett EA, Wheat CW (2012) Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *Bmc Genomics* **13**, 361.

Hou Z, Jiang P, Swanson SA*, et al.* (2015) A cost-effective RNA sequencing protocol for large-scale gene expression studies. *Sci Rep* **5**, 9570.

J.M. Pasteels MR-R, J.C. Braekman, A. Dupont (1983) Salicin from host plant as precursor of salicylaldehyde in defensive secretion of chrysomeline larvae. *Physiological Entomology* **8**, 307-314.

Khang TF, Lau CY (2015) Getting the most out of RNA-seq data analysis. *Peerj* **3**.

Kirsch R, Vogel H, Muck A*, et al.* (2011) To be or not to be convergent in salicin-based defence in chrysomeline leaf beetle larvae: evidence from *Phratora vitellinae* salicyl alcohol oxidase. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **278**, 3225-3232.

Kogenaru S, Qing Y, Guo Y, Wang N (2012) RNA-seq and microarray complement each other in transcriptome profiling. *Bmc Genomics* **13**, 629.

Kolaczkowski B, Thornton JW (2009) Long-branch attraction bias and inconsistency in Bayesian phylogenetics. *PLoS One* **4**, e7891.

Kuhn J, Pettersson EM, Feld BK*, et al.* (2004) Selective transport systems mediate sequestration of plant glucosides in leaf beetles: A molecular basis for adaptation and evolution. *Proc Natl Acad Sci U S A* **101**, 13808-13813.

Lander ES, Waterman MS (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231-239.

Langmead B, Hansen KD, Leek JT (2010) Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biology* **11**, R83.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25.

Laurent P, Braekman J-C, Daloze S (2005) Insect chemical defense. In: *Chemistry of Pheromones and Other Semiochemicals II* (ed. Schulz S), pp. 167-229. Springer, Berlin / Heidelberg.

Leal WS (2013) Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annu Rev Entomol* **58**, 373-391.

Li J, Lehmann S, Weissbecker B*, et al.* (2013) Odoriferous Defensive stink gland transcriptome to identify novel genes necessary for quinone synthesis in the red flour beetle, *Tribolium castaneum*. *PLoS Genet* **9**, e1003596.

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930.

Lin Y, Golovnina K, Chen ZX*, et al.* (2016) Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*. *Bmc Genomics* **17**, 28.

Ling F, Dahanukar A, Weiss LA, Kwon JY, Carlson JR (2014) The molecular and cellular basis of taste coding in the legs of *Drosophila*. *Journal of Neuroscience* **34**, 7148-7164.

Liu Y, Zhou J, White KP (2014) RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301-304.

Macosko Evan Z, Basu A, Satija R*, et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214.

Marchant A, Mougel F, Almeida C*, et al.* (2015) *De novo* transcriptome assembly for a non-model species, the blood-sucking bug *Triatoma brasiliensis*, a vector of Chagas disease. *Genetica* **143**, 225-239.

Martin M (2016) Cutadapt removes adapter sequences from high-throughput sequencing reads.

McKenzie SK, Oxley PR, Kronauer DJC (2014) Comparative genomics and transcriptomics in ants provide new insights into the evolution and function of odorant binding and chemosensory proteins. *Bmc Genomics* **15**.

Meinwald J, Jones TH, Eisner T, Hicks K (1977) New Methylcyclopentanoid Terpenes from the Larval Defensive Secretion of a Chrysomelid Beetle (Plagiodera versicolora). *Proc Natl Acad Sci U S A* **74**, 2189-2193.

Menken SBJ, Roessingh P (1998) Evolution of Insect-Plant Associations: Sensory Perception and Receptor Modifications Direct Food Specialization and Host Shifts in Phytophagous Insects. In: *Endless Forms: Species and Speciation* (eds. Howard DJ, Berlocher SH), pp. 145–156. Oxford Univ. Press, New York, NY.

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255-264.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628.

Nichols Z, Vogt RG (2008) The SNMP/CD36 gene family in Diptera, Hymenoptera and Coleoptera: *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles gambiae*, *Aedes aegypti*, *Apis mellifera,* and *Tribolium castaneum*. *Insect Biochem Mol Biol* **38**, 398-415.

O'Donnell MJ, Rheault MR (2005) Ion-selective microelectrode analysis of salicylate transport by the Malpighian tubules and gut of *Drosophila melanogaster*. *J Exp Biol* **208**, 93-104.

Oldham NJ, Veith M, Boland W, Dettner K (1996) Iridoid monoterpene biosynthesis in insects: Evidence for a *de novo* pathway occurring in the defensive glands of *Phaedon armoraciae* (Chrysomelidae) leaf beetle larvae. *Naturwissenschaften* **83**, 470-473.

Oppenheim SJ, Baker RH, Simon S, DeSalle R (2015) We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. *Insect Molecular Biology* **24**, 139-154.

Pabinger S, Rodiger S, Kriegner A, Vierlinger K, Weinhausel A (2014) A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomol Detect Quantif* **1**, 23-33.

Park SK, Shanbhag SR, Wang Q*, et al.* (2000) Expression patterns of two putative odorant-binding proteins in the olfactory organs of *Drosophila melanogaster* have different implications for their functions. *Cell Tissue Res* **300**, 181-192.

Pasteels JM, Braekman JC, Daloze D, Ottinger R (1982) Chemical defence in chrysomelid larvae and adults. *Tetrahedron* **38**, 1891-1897.

Pelosi P, Iovinella I, Felicioli A, Dani FR (2014) Soluble proteins of chemical communication: an overview across arthropods. *Frontiers in Physiology* **5**.

Pertea G, Huang X, Liang F*, et al.* (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* **19**, 651-652.

Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61-65.

Qiao HL, Deng PY, Li DD*, et al.* (2013) Expression analysis and binding experiments of chemosensory proteins indicate multiple roles in *Bombyx mori*. *Journal of Insect Physiology* **59**, 667-675.

Ramskold D, Luo S, Wang YC*, et al.* (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology* **30**, 777-782.

Rau A, Maugis-Rabusseau C, Martin-Magniette ML, Celeux G (2015) Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models. *Bioinformatics* **31**, 1420-1427.

Ray A, van Naters WG, Carlson JR (2014) Molecular determinants of odorant receptor function in insects. *J Biosci* **39**, 555-563.

Rhoads A, Au KF (2015) PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* **13**, 278-289.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**, R25.

Robinson MD, Smyth GK (2007) Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881-2887.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Saliba AE, Westermann AJ, Gorski SA, Vogel J (2014) Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res* **42**, 8845-8860.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467.

Sato K, Pellegrino M, Nakagawa T*, et al.* (2008) Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **452**, 1002-1006.

Sato K, Tanaka K, Touhara K (2011) Sugar-regulated cation channel formed by an insect gustatory receptor. *Proceedings of the National Academy of Sciences, USA* **108**, 11680-11685.

Schmidt F, Efferth T (2016) Tumor Heterogeneity, Single-Cell Sequencing, and Drug Resistance. *Pharmaceuticals (Basel)* **9**.

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086-1092.

Schwaighofer A, Pechlaner M, Oostenbrink C*, et al.* (2014) Insights into structural features determining odorant affinities to honey bee

odorant binding protein 14. *Biochemical and Biophysical Research Communications* **446**, 1042-1046.
Soetens P, Pasteels JM, Daloze D (1993) A simple method for in vivo testing of glandular enzymatic activity on potential precursors of larval defensive compounds in Phratora species (Coleoptera: Chrysomelinae). *Experientia* **49**, 1024-1026.
Speed MP, Fenton A, Jones MG, Ruxton GD, Brockhurst MA (2015) Coevolution can explain defensive secondary metabolite diversity in plants. *New Phytologist* **208**, 1251-1263.
Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313.
Stamp N (2001) Enemy-free space via host plant chemistry and dispersion: assessing the influence of tri-trophic interactions. *Oecologia* **128**, 153-163.
Strauss AS, Peters S, Boland W, Burse A (2013) ABC transporter functions as a pacemaker for sequestration of plant glucosides in leaf beetles. *Elife* **2**, e01096.
Sugawara F, Matsuda K, Kobayashi A, Yamashita K (1979) Defensive secretion of chrysomelid larvaeGastrophysa atrocyanea Motschulsky andPhaedon brassicae Baly. *J Chem Ecol* **5**, 635-641.
Sun YL, Huang LQ, Pelosi P, Wang CZ (2012) Expression in antennae and reproductive organs suggests a dual role of an odorant-binding protein in two sibling Helicoverpa species. *PLoS One* **7**, e30040.
Susko E (2008) On the distributions of bootstrap support and posterior distributions for a star tree. *Syst Biol* **57**, 602-612.
Swofford D (2003) *PAUP*: phylogenetic analysis using parsimon (* and other methods).* https://www.scienceopen.com/document?vid=a59ba3b2-9972-437a-ae3b-3ad61c20edd0
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729.
Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A (2011) Differential expression in RNA-seq: a matter of depth. *Genome Res* **21**, 2213-2223.
Termonia A, Hsiao TH, Pasteels JM, Milinkovitch MC (2001) Feeding specialization and host-derived chemical defense in Chrysomeline leaf beetles did not lead to an evolutionary dead end. *Proceedings of the National Academy of Sciences* **98**, 3909-3914.

Termonia A, Pasteels JM (1999) Larval chemical defence and evolution of host shifts in Chrysomela leaf beetles. *Chemoecology* **9**, 13-23.

Trapnell C, Roberts A, Goff L, *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578.

UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204-212.

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. *Trends in Genetics* **30**, 418-426.

Van Verk MC, Hickman R, Pieterse CM, Van Wees SC (2013) RNA-Seq: revelation of the messengers. *Trends Plant Sci* **18**, 175-179.

Veith M, Lorenz M, Boland W, Simon H, Dettner K (1994) Biosynthesis of iridoid monoterpenes in insects: defensive secretions from larvae of leaf beetles (Coleoptera: Chrysomelidae). *Tetrahedron* **50**, 6859-6874.

Veith M, Oldham NJ, Dettner K, Pasteels JM, Boland W (1997) Biosynthesis of defensive allomones in leaf beetle larvae - stereochemistry of salicylalcohol oxidation in phratora vitellinae and comparison of enzyme substrate and stereospecificity with alcohol oxidases from several iridoid producing leaf beetles. *J Chem Ecol* **23**, 429-443.

Vogt RG, Miller NE, Litvack R, *et al.* (2009) The insect SNMP gene family. *Insect Biochem Mol Biol* **39**, 448-456.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews: Genetics* **10**, 57-63.

Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.

Wicher D, Schäfer R, Bauernfeind R, *et al.* (2008) *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* **452**, 1007-1011.

Wilkens S (2015) Structure and mechanism of ABC transporters. *F1000Prime Rep* **7**, 14.

Xiao Y, Liu K, Zhang D, *et al.* (2016) Resistance to *Bacillus thuringiensis* Mediated by an ABC Transporter Mutation Increases Susceptibility to Toxins from Other Bacteria in an Invasive Insect. *PLoS Pathog* **12**, e1005450.

Xie X, Cheng T, Wang G, *et al.* (2012) Genome-wide analysis of the ATP-binding cassette (ABC) transporter gene family in the silkworm, *Bombyx mori*. *Mol Biol Rep* **39**, 7281-7291.

Xie Y, Wu G, Tang J, *et al.* (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **30**, 1660-1666.

Xin Y, Kim J, Ni M, *et al.* (2016) Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. *Proc Natl Acad Sci U S A* **113**, 3293-3298.

Xu DL, Long H, Liang JJ, *et al.* (2012) *De novo* assembly and characterization of the root transcriptome of *Aegilops variabilis* during an interaction with the cereal cyst nematode. *Bmc Genomics* **13**.

Yuan HB, Ding YX, Gu SH, *et al.* (2015) Molecular Characterization and Expression Profiling of Odorant-Binding Proteins in *Apolygus lucorum*. *PLoS One* **10**, e0140562.

Zhang H-J, Anderson AR, Trowell SC, *et al.* (2011) Topological and functional characterization of an insect gustatory receptor. *PLoS One* **6**, e24111.

Zhang ZH, Jhaveri DJ, Marshall VM, *et al.* (2014) A comparative study of techniques for differential expression analysis on RNA-Seq data. *PLoS One* **9**, e103207.

Zhao QY, Wang Y, Kong YM, *et al.* (2011) Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *Bmc Bioinformatics* **12 Suppl 14**, S2.

11 Curriculum Vitae

Personal

Name:	Ding Wang
Current address:	Schröding Str. 36
Date of Birth:	05. 09. 1979
City of Birth:	Tian Jin (天津)
Nationality:	China

Academic Education

03/2012 – current	Doctoral student, Max Planck Institute for Chemical Ecology, Jena Department of Bioorganic Chemistry First supervisor: Prof. Dr. Wilhelm Boland
08/2011 – 01/2012	Diploma thesis, Leibniz Institute for Natural Product Research and Infection Biology, Jena Department of Cell and Molecular Biology
10/2004 – 07/2011	Study of Bioinformatic, Friedrich-Schiller-University, Jena

Publications

Wang, D., Pentzold, S., Kunert, M., Groth, M., Brandt, W., Pasteels, J. M., Boland, W., Burse, A. (2016). Selective adaptation within the chemosensory system of the leaf beetle, *Chrysomela lapponica*, following host plant shift. Molecular Ecology. (In preparation)

Gretscher, R., Streicher, P. E., Strauß, A., Wielsch, N., Stock, M., **Wang, D.,** Boland, W., Burse, A. (2016). A common theme in extracellular fluids of beetles: extracellular superoxide dismutases crucial for balancing ROS in response to microbial challenge. Scientific Reports, 6: 24082. doi:10.1038/srep24082.

Rahfeld, P., Häger, W., Kirsch, R., Pauls, G., Becker, T., Schulze, E., Wielsch, N., **Wang, D.,** Groth, M., Brandt, W., Boland, W., Burse, A. (2015). Glandular beta-glucosidases in juvenile *Chrysomelina* leaf beetles support the evolution of a host-plant-dependent chemical defence. Insect Biochemistry and Molecular Biology, 58, 28-38. doi:10.1016/j.ibmb.2015.01.003.

Strauß, A., **Wang, D.,** Stock, M., Gretscher, R., Groth, M., Boland, W., Burse, A. (2014). Tissue-specific transcript profiling for ABC transporters in the sequestering larvae of the phytophagous leaf beetle *Chrysomela populi*. PLoS One, 9(6): e98637. doi:10.1371/journal.pone.0098637.