

**Integrative statistical methods for decoding molecular responses
to insect herbivory in *Nicotiana attenuata***

Dissertation

To fulfil the requirements for the degree of
Doctor of Philosophy (Ph. D.)

Submitted to the Council of Faculty of Biology and Pharmacy,
Friedrich Schiller University Jena

By

Jyotasana Gulati, M.S.

Born on 16.03.1985 in India



Max-Planck Institute for
Chemical Ecology



Friedrich-Schiller
Universität jena

Gutachter:

Prof. Dr. Ian Thomas Baldwin, Max-Planck Institute for Chemical Ecology, Jena, Germany

Prof. Dr. Christoph Kaleta, Friedrich-Schiller-Universität Jena, Jena, Germany

Prof. Dr. Georg Jander, Boyce Thompson Institute for Plant Research, New York, U.S.A

Beginn der Promotion: 07/01/2010

Eingereicht am: 21/08/2013

Tag der Verteidigung: 16/12/2013

TABLE OF CONTENT

Chapter1. General Introduction.....	1
Chapter2. Overview of Manuscripts.....	10
Chapter3. Tissue Specific Diurnal Rhythms of Metabolites and Their Regulation during Herbivore Attack in a Native Tobacco, <i>Nicotiana attenuata</i>	14
Chapter4. Deciphering Herbivory-Induced Gene-to-Metabolite Dynamics in <i>Nicotiana attenuata</i> Tissues Using a Multifactorial Approach.....	37
Chapter5. An integrative statistical method to explore herbivory-specific responses in plants...	67
Chapter6. The roots of plant defenses: Integrative multivariate analyses uncover dynamic behaviors of roots' gene and metabolic networks elicited by leaf herbivory.....	72
Chapter7. Discussion.....	109
Summary.....	118
References.....	122
Eigenständigkeitserklärung.....	127
Curriculum Vitae.....	128
Acknowledgements.....	130

Chapter 1

General Introduction

***Nicotiana attenuata*: a model system in Chemical ecology**

Chemical communication occurs at all the levels of biological organization, including regulation of cells and organs within an organism, as well as ecological interactions among organisms. The production of chemical signals in plants and animals is regulated through hormones and signal transduction pathways. The field of science that addresses the role of these chemical signals in the interaction of organisms with their biotic or abiotic environment as well as the evolutionary and behavioral consequences of these interactions is termed as chemical ecology. This discipline provides the understanding of molecular and genetic mechanisms of biological signal transduction controlling several interactions among organisms such as the attraction of mates, defense against enemies, competition for resources, etc. Among these interactions, the study of chemistry of plant defense strategies is an excellent example that is benefitted from the advancement of molecular and chemical biology. Advances in molecular biology, analytical chemistry and genetics coupled with the increasing availability of high throughput “omics” technologies have recently been applied using theoretical approaches of systems biology to study the functional basis of plant-insect interactions.

Nicotiana attenuata is a wild tobacco species native to the Great Basin Dessert in United States. It germinates in the post fire habitats from long-lived seed banks to form monocultures in nitrogen-rich soils (Baldwin and Morse, 1994; Baldwin et al., 1994). After germination, the newly emerged seedlings struggle against intense intraspecific competition within population and highly variable biotic and abiotic challenges. The perception of fatty acid-amino acid conjugates (FACs) in the oral secretions of *Manduca sexta*, a specialist herbivore of *N. attenuata*, triggers the activation of protein kinases and jasmonic acid (JA) biosynthesis and signaling (Figure 1). JA and its metabolites, collectively called as jasmonates, mediate profound changes in the expression of regulatory and structural genes which in turn elicit both direct and indirect defenses in *N. attenuata*. In direct defense to herbivore attack, plants produce toxic, anti-nutritive and/or anti-digestive compounds such as alkaloids, phenolamids and terpenoids (Bennett and Wallsgrove, 1994; Kessler and Baldwin, 2002; Mithofer and Boland, 2012). In *N. attenuata*, insect attack or its simulation by the

application of its OS results in significant reconfigurations in metabolic and growth processes, in the de novo production of defense compounds such as nicotine (Steppuhn et al., 2004), diterpene glycosides (Heiling et al., 2010) and phenopropanoid polyamine conjugates (Kaur et al., 2010; Onkokesung et al., 2012). Systemically induced defenses - the accumulation of defensive compounds in distal intact systemic leaves during insect herbivory - is also well-characterized in *N. attenuata* (Baldwin, 2001; Wu and Baldwin, 2010; Onkokesung et al., 2012). In addition, the release of volatile organic compounds to attract predators of herbivores constitute indirect defense traits in *N. attenuata* (Kessler, 2004; Schuman, 2012). Since the production of plant defense traits incurs significant fitness cost, plants have evolved mechanisms to balance tolerance and defense. *N. attenuata* has developed a key tolerance strategy in which roots act as a sink tissue for sequestration of partitioned assimilates to facilitate re-growth after herbivore attack (Schwachtje et al., 2006).

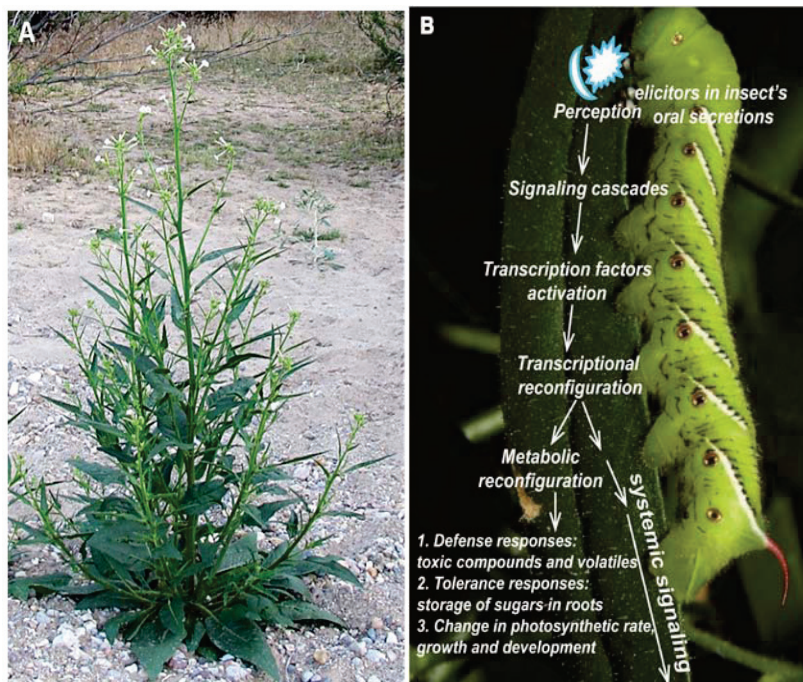


Figure 1: (A) *Nicotiana attenuata* is a wild tobacco species native to Southwestern North America. (B) A schematic summarizing the information flow that results in the activation of defense mechanisms in an herbivore attacked leaf: The perception of herbivore-derived elicitors triggers the activation of signaling cascades which translates into large scale transcriptional and metabolic reconfigurations that induces direct and indirect defense mechanisms. Systemic signaling from the attacked leaf results in the activation of defense responses in distal intact (systemic) leaves. (Photos were taken by Danny Kessler)

N. attenuata is among the few model plants for which “omics” technologies have been applied to understand chemically mediated heterotrophic interactions (Halitschke et al., 2003; Giri et al., 2006; Kang et al., 2006; Gulati et al., 2013). Vast amounts of data generated require the development of techniques such as multivariate analysis (MVA) that can correctly frame the extremely complex biological designs into statistical models.

Functional-genomics studies using conjunctive reasoning

Biology has always been a science of complex systems, mainly studied for three inherent properties (a) robustness: the ability to adapt to environmental changes and the availability of components with redundant functionality (Csete and Doyle, 2002; Kitano, 2002), (b) modularity: physical and functional insulation of subsystems (Alm and Arkin, 2003), and (c) emergence: the behavior that cannot be predicted by simply analyzing the structure of individual components (Van Regenmortel, 2004, 2004). With advances in molecular genetics, research in last three decades has majorly been benefited from the application of reductionist approach which is based on the concept that complex biological systems are simply a sum of their constituent parts. This approach has also established a niche in chemical ecology by identifying regulatory elements that play key roles in chemical traits behind physiological responses (Kessler and Baldwin, 2002; Wu and Baldwin, 2010). For example, when a gene is disrupted using genetic manipulations, its function can be inferred from the phenotype of the organism. However, reductionist approach cannot adequately account for the above mentioned properties of a complex biological network (Van Regenmortel, 2004; Van Norman and Benfey, 2009).

In the last decade, the emergence of high-throughput technologies has enabled the simultaneous detection of a large number of alterations in molecular components of a system and has caused a major shift from a reductionist to a more holistic or systems biology based studies. Systems biology is an inter-disciplinary field of research which studies complex biological systems by examining all of the components and their interactions in the context of the whole system (Ideker et al., 2001; Ideker, 2004). The implementation of systems biology framework starts by defining the components of a biological system, profiling of relevant biochemical and genetic data (transcripts, metabolites, proteins, lipids, etc) on a global scale and the formulation of an initial model of interactions among elements that regulate the observed traits. The next step is to perturb the components of the system and study the results of the perturbation. Models are further refined based on the inferences which then leads to a

set of new testable hypotheses (Joyce and Palsson, 2006). Auffray et al (Auffray et al., 2003) proposed following four axioms that should be considered while studying systems biology:

1. Contextualization: characterization of objects in context of the associated environment in which they function. It also accounts for the associated complexity in biological systems when gene function depends on the context such as developmental stage or environment, which reductionism fails to incorporate.
2. Relatedness: studying dynamic behavior of interacting objects, that is achievable using high throughput multivariate “omics” studies.
3. Conditionality: identifying the rules that determine the behavior of interacting objects, which is specifically dependent on axiom 2 since the combination of multiple data sets would ease the identification of regulatory mechanism of the given trait.
4. Pertinence: identifying few relevant modules from the large set of interactions, this in other terms called as dimension-reduction.

These axioms define a framework for large scale functional genomics studies with the main goal of developing system-level understanding of biological systems using conjunctive reasoning which implies deducing results from collective “omics” studies. Methods of integration for meta-analyses of “omics” data can simply be summarized in three steps:

1. Identification of network scaffolds which represents the connections between cellular components.
2. Decomposition of scaffolds into modules which are the portions of the network that are most active under a given condition and therefore best explain the observed behavior.
3. Predicting network behavior using previously developed system models.

Several efforts have been made in integrating different “omics” data to facilitate the functional identification of networks controlling growth and development, and response to biotic and abiotic stresses in plants (Joyce and Palsson, 2006; Long et al., 2008; Moreno-Risueno et al., 2010; Liberman et al., 2012). Earlier work of gene-to-metabolite associations studied dynamic response in *Arabidopsis* during sulphur and nitrogen depletion at global level (Hirai et al., 2005; Hirai et al., 2007; Sawada et al., 2009). Similar efforts have been made to integrate transcripts and proteins to identify potential biomarkers for roots, flowers, leaves, and seeds (Baerenfaller et al., 2008). The power of systems approach to understand biological

networks is the identification of emergent network properties that is revealed while studying interactions among components in different “omics” data sets. For example, A recent large scale study of root transcriptional activity identified an emergent property which represents root-specific short rhythms that determine the periodicity in lateral root development in *Arabidopsis* (Moreno-Risueno et al., 2010).

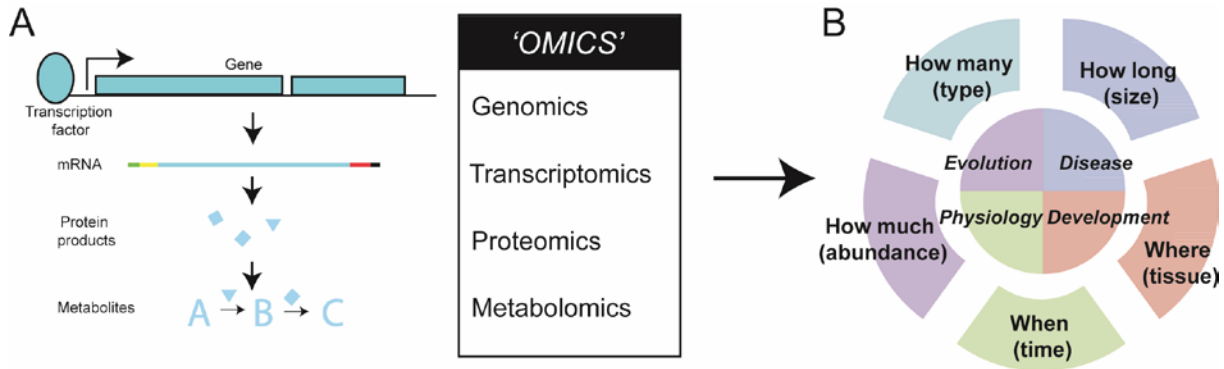


Figure 2: (A) Components of “omics” data sets: DNA (genomics) is first transcribed to mRNA (transcriptomics) and translated into protein (proteomics), which can catalyze chemical reactions that give rise to metabolites (metabolomics). (B) These components are profiled for different factorial designed experiments to simultaneously measure their abundance, size, type, tissue localization and time of action, for studying evolution, disease, development or physiology of an organism.

Central problem of dimensionality in “omics”

With the advancement in data acquisition techniques and computing facilities, practical environment of “omics” world has gained several layers of dimension. The first increment is the manifold increase in the simultaneous sampling of different experimental designs. Figure 2B represents the 5 dimensions of different components (transcriptomics, proteomics, metabolomics, etc.) which are commonly incorporated in a complex experiment to measure parameters such as their abundance, length, localization, time of expression, different types, influence of genomic variation, etc. while studying development, physiology, disease and evolution. At the same time, these “omics” measurements, with attributes – features (genes, proteins, metabolites etc) and individuals (samples, subjects etc), represent an expression scenario with the number of features p being much larger than the number of samples n , often described as “large p small n ” problem (Johnstone and Titterington, 2009). However from a statistical point of view, it is desirable that the number of experimental units n should significantly exceed the number of features p if inference is to be made about the

data. Therefore, in order to improve the efficiency and accuracy of data analysis, it is essential to reduce the dimension of the data. Dimension Reduction (DR) techniques are often applied as a preprocessing step or as part of data analysis to simplify the data model. The strategies of identifying low-dimensional representations for the original high dimensional data set are majorly of two types:

1. Feature extraction: It involves transferring original features of a data set to a more compact set of dimensions while retaining as much information as possible, through the application of some mapping functions. Well known extraction methods include Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)
2. Feature selection: It involves a family of techniques that finds a minimum subset of the original features set based on some criteria, rather than transforming the data to an entirely new set of dimensions. The commonly used methods to identify set of interesting features in “omics” data are clustering and statistical modeling.

Presently, developing effective methods to interpret these high dimensional data sets and thereby deriving fundamental and applied biological information about whole systems is one of the major challenges of computational biology. The most effective method to derive information from complex biological experiments without introducing noise is statistical modeling, which is an intuitive method of describing the influence of known variables on the observed data.

Statistical modeling of high dimensional data

Clustering analysis identifies predominant patterns within the data by grouping features (genes, metabolites etc) that have comparable patterns of variation across samples. However, statistical modeling identifies features affected by a given stimulus and provides specific information regarding individual response patterns. This involves modeling the dependence of feature's response on a specific environmental variable such as tissue type, timing or treatments, using a set of interpretable parameters. If this representation adequately accounts the response pattern of a feature, then the corresponding model parameters may describe the timing, magnitude or duration of the response. This strategy is the basis of two group comparison for identifying differentially expressed genes/metabolites between normal and perturbed conditions. As previously mentioned, high-throughput experiments are now designed to study perturbations of biological networks by combination of various factors. Therefore, these multifactorial experimental designs require biologically interpretable models

that could identify the effect of different treatments on a give feature. While the parameter estimation of these models is similar to those of classical factorial experiments, the interpretation of parameters account for the duration and magnitude of effects, interactions of different treatments and the interactions among features. This extracted information can effectively be used in designing “low cost, high impact” experiments.

Additionally, time course molecular profiling has attracted increasing attention in biological research to identify dynamic patterns of features. Time course experiments can be broadly classified into two types: periodic and developmental. Periodic time course experiments are designed to study cell cycles and circadian/diurnal regulation mechanism, the processes that are usually observed showing regular temporal patterns. They are often studied as one-sample problem where the aim is to identify genes which change over time. On the other hand, developmental time courses measure expression levels in a growth and developmental processes or after the application of treatments and are often observed with arbitrary patterns. They are studied as two or more sample problems, with no assumption of the pattern, to identify genes which change differently over time between two or more conditions. However, most temporal biological “omics” data sets contain a limited number of time points with very few replicates and measured on biological processes which are often shifted in time. Classical methods for the study of static “omics” datasets are of limited application as they cannot consider specific ordering of time points in the series. Also, traditional methods developed primarily to analyze time series data, eg. Auto regressive moving average model (ARMA) or wavelet methods are not the preferred choices because of the short sampled time series data sets.

Several methods have been specially designed for biological dynamic “omics” that are plagued by above mentioned limitations and few broad ideas behind these methods are briefly introduced here. An intuitive way of selecting the genes of interests in one-sample case is the use of one-way ANOVA that models the gene expression on the time variable as a factor and the example is the famous study by Wang et al. that identified genes with different temporal profiles in a developmental stage in *Caenorhabditis elegans* (Wang and Kim, 2003). Similarly, two or more sample case problems can be analyzed using ANOVA structures that include time and biological conditions and their interaction terms in the model (Park et al., 2003). Several variations of ANOVA approaches have been utilized to model developmental time course microarray data sets. (mixed-effect ANOVA (Romagnolo et al., 2002; Wang and Kim, 2003), modified ANOVA (Park et al., 2003)). However, a common problem underlying

all these attempts was the assumption that samples measured at different time points are independently distributed multivariate normal vectors. The commonly used alternatives are the permutation based methods that simulate the null distribution of test statistic to estimate the false discovery rate (FDR). A method recently developed based on non-parametric ANOVA structure to simultaneously analyze more than one factor measured across time series using boot-strap is a part of the statistical framework proposed in this thesis (Zhou et al., 2010).

Second widely applied class of methods is based on regression models and the most commonly used are those fitting continuous curves using B-splines. For example, Breeze et al. used splines based clustering to identify temporal profiles of genes sampled over 22 time points, in order to understand regulation of leaf senescence in *Arabidopsis* (Breeze et al., 2011). However, these methods have been shown unsuitable for short time courses (Bar-Joseph et al., 2003).

Another class of method was developed primarily to correct the problem associated with the inference of large number of false positives (FP) and false negatives (FN) arising from the above described analyses. This problem arises from the measurements of large number of genes compared to the small number of replicates (large p , small n problem) which then leads to the poor estimation of “within” variances or co-variance matrices. The idea of moderation in which the gene-specific variances are corrected by moving it towards a common value, estimated from the entire gene set, has been applied extensively to analyze microarray data (Tai and Speed, 2006) and it is the basis for the well known tool used in microarrays studies – SAM (significance analysis of microarrays) which adjusted t-statistic by changing the standard deviation in the data sets (Tusher et al., 2001).

Objective of the thesis

The main objective of my Ph.D. work was to explore regulatory mechanisms mediating plant defense against herbivores in *N. attenuata* using statistical models. I have reported and discussed the findings in following six chapters:

Chapter 2: Overview of the manuscripts

Chapter 3: Manuscript I: Tissue Specific Diurnal Rhythms of Metabolites and Their Regulation during Herbivore Attack in a Native Tobacco, *Nicotiana attenuata*.

Chapter 4: Manuscript II: Deciphering Herbivory-Induced Gene-to-Metabolite Dynamics in *Nicotiana attenuata* Tissues Using a Multifactorial Approach.

Chapter 5: Manuscript III: An integrative statistical method to explore herbivory-specific responses in plants.

Chapter 6: Manuscript IV: The roots of plant defenses: Integrative multivariate analyses uncover dynamic behaviors of roots' gene and metabolic networks elicited by leaf herbivory.

Chapter 7: Discussion

CHAPTER 2

Overview of Manuscripts

MANUSCRIPT I

Tissue Specific Diurnal Rhythms of Metabolites and Their Regulation during Herbivore Attack in a Native Tobacco, *Nicotiana attenuata*

Sang-Gyu Kim, Felipe Yon, Emmanuel Gaquerel, **Jyotasana Gulati** and Ian T. Baldwin

Published in Plos One 2011, **6**: e26214

In this manuscript, we studied tissue-specific diurnal rhythms and their regulation in the generalized and specialized metabolism of *Nicotiana attenuata*. We developed a liquid chromatography-mass spectrometry procedure to profile metabolites of simulated herbivory-elicited source leaves and unelicited sink leaves and roots in a time course experiment. We identified metabolites with statistically significant diurnal patterns in sink/source leaves and roots. We found that roots and leaves had distinct set of oscillating metabolites which mainly peaked at dusk or night in roots while leaf metabolites peaked during the day. Many of these oscillating metabolites showed more pronounced systemic responses to simulated herbivory in un-attacked tissues. With these results, we demonstrated the significance of diurnal regulation in metabolic reconfiguration during plant's responses to herbivore attack.

Sang-Gyu Kim, Ian T. Baldwin conceived and designed the experiments. Sang-Gyu Kim, Felipe Yon, Emmanuel Gaquerel performed the experiments. Sang-Gyu Kim, Emmanuel Gaquerel, Jyotasana Gulati analyzed the data. Felipe Yon contributed reagents/materials/analysis tools. Sang-Gyu Kim, Emmanuel Gaquerel, Ian T. Baldwin wrote the manuscript.

MANUSCRIPT II

Deciphering Herbivory-Induced Gene-to-Metabolite Dynamics in *Nicotiana attenuata* Tissues Using a Multifactorial Approach

Jyotasana Gulati, Sang-Gyu Kim, Ian T. Baldwin and Emmanuel Gaquerel

Published in Plant Physiology 2013, **162**: 1042-1059

In this manuscript, we designed a method to explore spatio-temporal dynamics of activation of herbivory-induced changes in gene-to-metabolite networks in *Nicotiana attenuata*. We conducted time-series transcriptome and metabolome profiling of simulated herbivory-elicited source leaves and unelicited sink leaves and roots. To explore differential expression patterns of genes and metabolites which are activated during shoot systemic signaling, we analyzed dynamic response patterns obtained by comparing treated versus untreated systemic leaves using bootstrap-based nonparametric ANOVA models on the projected time-vector space. Using these response patterns, we highlighted branch-specific functional organization and transition points in the oxylipin gene network. We also studied nonlinearities in gene-metabolite associations involved in the acyclic diterpene glucoside pathway, which is highly activated in systemic tissues during herbivory, after selectively extracting modules obtained by self-organizing maps based spatio-temporal resolution of induced molecular processes.

Jyotasana Gulati designed the research and data analysis, performed research, interpreted results and wrote the manuscript. Sang-Gyu Kim designed and performed the experiment. Ian T. Baldwin designed the experiment and wrote the manuscript. Emmanuel Gaquerel designed the research and data analysis, performed the research, interpreted results and wrote the manuscript.

MANUSCRIPT III

An integrative statistical method to explore herbivory-specific responses in plants

Jyotasana Gulati, Ian T. Baldwin and Emmanuel Gaquerel

Accepted for Publication in Plant Signaling & Behavior 2013, 8:e25638; PMID: 23857359

In this manuscript, we reviewed our newly developed approach to analyze whole-plant molecular responses in a time course multivariate data-set by combining an extended self-organizing map (SOM) based dimensionality reduction method with bootstrap-based non-parametric ANOVA models. We extended the application of this method to extract genes showing simulated-herbivory specific elicitation in systemic (distal from the treatment sites) tissues using motif analysis for different combinations of treatment applied to *Nicotiana attenuata*. We analyzed dynamic response patterns obtained by comparing transcriptomic data of treated and untreated systemic leaves collected for W+OS (diluted oral secretion from larvae of the specialist herbivore *Manduca sexta* applied into mechanically produced puncture wounds) and W+W (mechanically produced puncture wounds treated with water) treatment type to differentiate OS-specific systemic response from those inherent to mechanical wounding. As a proof of principal, we compared time response behavior of genes involved in the phenylpropanoid pathway, which is known to be activated in systemic tissues during herbivory, for 2 comparisons: (a) Control versus W+OS (responses to combined herbivory and mechanical wounding), (b) W+OS versus W+W (OS-specific responses).

Jyotasana Gulati, Ian T. Baldwin and Emmanuel Gaquerel designed and performed the research, interpreted results and wrote the manuscript.

MANUSCRIPT IV

The roots of plant defenses: Integrative multivariate analyses uncover dynamic behaviors of roots' gene and metabolic networks elicited by leaf herbivory

Jyotasana Gulati, Ian T. Baldwin and Emmanuel Gaquerel

In Review: Genome Biology 2013

In this manuscript, we explored the important role played by roots in above ground stress responses in *Nicotiana attenuata* by identifying multivariate descriptors from a time-course factorial experiment. The data set consists of previously published time-series transcriptome and metabolome of simulated herbivory-elicited source leaves and unelicited sink leaves and roots. From single time point analyses, we observed an interesting pattern of co-linearity in the up- and down-regulation of genes and metabolites across the entire time series in treated (source) and untreated (sink) leaves and large transcriptomic and metabolic changes in roots but without co-linearity across the time series. Instead, we observed a unique pattern in roots that manifest itself in a similar number of induced and suppressed genes separated by a short time lag which when analyzed using multivariate time series analysis resulted in two principal trends characterized by: (a) an inversion of root-specific semidiurnal (12h) gene oscillations and (b) transcriptional changes with major amplitude effects that translated into a distinct suite of root-specific secondary metabolites.

Jyotasana Gulati, Ian T. Baldwin and Emmanuel Gaquerel designed and performed the research, interpreted results and wrote the manuscript.

Chapter 3

Tissue Specific Diurnal Rhythms of Metabolites and Their Regulation during Herbivore Attack in a Native Tobacco, *Nicotiana attenuata*

Sang-Gyu Kim, Felipe Yon, Emmanuel Gaquerel, **Jyotasana Gulati** and Ian T. Baldwin

Published in Plos One 2011, **6**: e26214

Tissue Specific Diurnal Rhythms of Metabolites and Their Regulation during Herbivore Attack in a Native Tobacco, *Nicotiana attenuata*

Sang-Gyu Kim, Felipe Yon[‡], Emmanuel Gaquerel[‡], Jyotasana Gulati, Ian T. Baldwin*

Department of Molecular Ecology, Max Planck Institute for Chemical Ecology, Jena, Germany

Abstract

Ecological performance is all about timing and the endogenous clock that allows the entrainment of rhythms and anticipation of fitness-determining events is being rapidly characterized. How plants anticipate daily abiotic stresses, such as cold in early mornings and drought at noon, as well as biotic stresses, such as the timing of pathogen infections, is being explored, but little is known about the clock's role in regulating responses to insect herbivores and mutualists, whose behaviors are known to be strongly diurnally regulated and whose attack is known to reconfigure plant metabolomes. We developed a liquid chromatography-mass spectrometry procedure and analyzed its output with model-based peak picking algorithms to identify metabolites with diurnal accumulation patterns in sink/source leaves and roots in an unbiased manner. The response of metabolites with strong diurnal patterns to simulated attack from the specialist herbivore, *Manduca sexta* larvae was analyzed and annotated with in-house and public databases. Roots and leaves had largely different rhythms and only 10 ions of 182 oscillating ions in leaves and 179 oscillating ions in roots were rhythmic in both tissues: root metabolites mainly peaked at dusk or night, while leaf metabolites peaked during the day. Many oscillating metabolites showed tissue-specific regulation by simulated herbivory of which systemic responses in unattacked tissues were particularly pronounced. Diurnal and herbivory-elicited accumulation patterns of disaccharide, phenylalanine, tyrosine, lyciumoside I, coumaroyl tyramine, 12-oxophytodienoic acid and jasmonic acid and those of their related biosynthetic transcripts were examined in detail. We conclude that oscillating metabolites of *N. attenuata* accumulate in a highly tissue-specific manner and the patterns reveal pronounced diurnal rhythms in the generalized and specialized metabolism that mediates the plant's responses to herbivores and mutualists. We propose that diurnal regulation will prove to an important element in orchestrating a plant's responses to herbivore attack.

Citation: Kim S-G, Yon F, Gaquerel E, Gulati J, Baldwin IT (2011) Tissue Specific Diurnal Rhythms of Metabolites and Their Regulation during Herbivore Attack in a Native Tobacco, *Nicotiana attenuata*. PLoS ONE 6(10): e26214. doi:10.1371/journal.pone.0026214

Editor: Daniel J. Kliebenstein, University of California, United States of America

Received: July 2, 2011; **Accepted:** September 22, 2011; **Published:** October 18, 2011

Copyright: © 2011 Kim et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Max Planck Society. 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: baldwin@ice.mpg.de

[‡] These authors contributed equally to this work.

Introduction

Timing is everything for ecological performances. The earth's 24 h rotation on its tilted axis, geographical differences, and interactions with other organisms shape the specific diurnal rhythms of each organism. As sessile organisms, plants can entrain their physiology to abiotic condition of their environment such as day/night cycles and their associated temperature fluctuations. The endogenous plant clock (circadian clock) 'wakes up' the photosynthetic machinery just before sun rise to maximize energy harvesting and regulates guard cells at noon to minimize water loss [1]. It also increases cold tolerance of plants at night and dawn [2]. Plants also synchronize their physiology with tightly associated organisms. For instance, snapdragon flowers emit methyl benzoate during the day to attract day-active pollinating bees [3]. Pathogens that attack at dawn are anticipated by the clock in *Arabidopsis* [4].

Since the pioneering discoveries that the rhythmic behaviors of animals are encoded in their genomes, the molecular components and functions of the plant's endogenous clock have been identified

in the model plant, *Arabidopsis thaliana* [5]. Forward and reverse genetic approaches in *Arabidopsis* have revealed that many diurnal 'behaviors' are controlled by a few clock genes [1,5]. These clock genes regulate 30~40% of total gene expression in *Arabidopsis* [6]. Arrhythmic plants harboring mutations in the clock genes have reduced photosynthetic capacity, growth and competitive ability under normal conditions [7,8]. However, natural mutations in clock genes have been discovered that help entrain a particular accessions' physiology to its local environment which enhances the plant fitness in that area [8].

Plants fix carbon in the shoot using light energy and make numerous metabolites from the products of photosynthesis. Plant metabolites therefore originate from a day/night cycle. Studies of primary metabolites have shown that sugars, starch, amino acids, most of organic acids involved in photosynthesis are circadian-regulated in leaves [9]. Although transcriptomic analyses show that many genes involved in secondary metabolite biosynthesis have diurnal expression patterns [5], the diurnal rhythms of secondary metabolite levels have been less studied in comparison with those of primary metabolites.

Plants are exposed to two completely different environments: the aboveground and the belowground. Aboveground and belowground plant parts thus develop their own endogenous rhythms [10]. Clock components in the shoot and root oscillate in a similar way under light/dark cycles but not under constant light. The roots express approximately four-fold fewer oscillating genes than do shoots in *Arabidopsis* [10]. However, this does not mean that the endogenous clock is less important for root physiology. Important physiological processes such as root bending and lateral root formation occur every 6 h and are controlled by the internal clock [11]. In addition, clock-regulated water contents in roots are reduced during the day and increase during the night [12].

Nicotiana attenuata is a native tobacco plant growing in the Great Basin Desert of southwestern USA. It germinates in post-fire ecosystems and shows diverse intra-specific and inter-specific interactions. We have studied *N. attenuata* growing in its ecological niche in natural habitats for more than two decades and observed several diurnal rhythms in its physiology. *N. attenuata* interacts with different pollinators in a light-dependant manner [13,14]. It produces two kinds of flowers. Night-opening flowers (NoFs) that open their corollas during the night and close during the day. Morning-opening flowers (MoFs) open their corollas during the early morning. NoFs emit benzyl acetone only during the night to attract nocturnal hawkmoth pollinators (*Manduca sexta* and *M. quinquemaculata*). Early morning flower visitors such as humming birds mainly nectar at MoFs and transfer pollen. Plant-herbivore interactions are also regulated in a day/night cycle. The accumulation of two lipoxygenase transcripts (*NaLOX2* and *NaLOX3*) involved in the biosynthesis of green leaf volatiles and jasmonic acid (JA) show diurnal rhythms in the leaf [15]. In addition, the generalist predator, *Geocoris* spp. feeds on the eggs and neonates of the specialist herbivore, *M. sexta*, usually during the day.

Metabolites produced in leaves and roots are essential elements determining the outcome of plants' aboveground and belowground interaction with other organisms. If most of the organisms on earth are governed by their endogenous clocks, the rhythms of the metabolites they produce should help us understand plant-plant, plant-animal interactions. Even though many genes involved in metabolism show diurnal expressions, large-scale screenings of tissue-specific oscillating metabolites and their regulation by herbivory remain largely unknown. Here, we examined tissue-specific diurnal rhythms of metabolites in leaves and roots of *N. attenuata* for two days. To find interconnections among oscillating metabolites and herbivore-induced plant defenses, we treated mechanical wounds in the leaves of *N. attenuata* with oral secretions (OS) from *M. sexta* larvae to mimic herbivore-induced changes [16] and precisely time the onset of elicitation to analyze the changes in oscillating metabolite levels with the oscillations of the transcripts of their associated genes.

Results and Discussion

Experimental Design

N. attenuata plants were grown in 16 h light/8 h dark cycle for 5 weeks, and source/sink leaves and roots were collected every 4 h for two days without any treatment to identify oscillating metabolites (Figure 1A). Diluted OS (1:5 with distilled water) from larvae of the specialist herbivore, *M. sexta*, were applied to puncture wounds in leaves created with a pattern wheel (W+OS) to mimic herbivory at 1 pm on a second day (Figure 1A). We also treated wounds with water (W+W) to distinguish OS-specific responses from wound-induced responses. Two source leaves (at nodes +2, +1) and one transition leaf (at node 0) were treated with

water or OS, and collected to examine local response in oscillating metabolites and transcripts. We also collected two sink leaves (at nodes -1, -2) and roots to examine systemic response in untreated tissues (Figure 1B).

The enormous diversity of metabolites among different plant taxa and their diverse chemical properties means that metabolomic analysis must be optimized for each plant species. In previous studies, we developed an efficient method to extract defense-related metabolites in *N. attenuata* [17,18] and used this 40% methanol-based extraction method (Figure 1C) and separated the metabolites by rapid separation liquid chromatography (RSLC). The separated metabolites were ionized by electrospray in both negative and positive modes and the exact mass to charge ratio of ions was measured with time of flight (TOF)-mass spectrometry. We also used a cDNA library and a microarray system of *N. attenuata* to explore the overall molecular mechanism in plant-herbivore interactions. The three-dimensional data (retention time, mass, intensity) of the mass spectrometry analysis was processed using the peak picking freeware XCMS [19], and diurnally oscillating metabolites were extracted by a model-based peak picking algorithm of the HAYSTACK [20] program. To reduce the information redundancy in the dataset, isotope peaks were clustered and annotated using the pseudo-spectrum deconvolution freeware CAMERA and removed from the analysis. Fragment and adduct ions detected in negative mode were included as there is no certain way to date of selecting only mother ions in large scale experiments, and negative ionization produces fewer daughter ions and adducts compared to positive ionization mode. In the negative ionization analysis, a total of 2209 and 1463 ions were detected from leaves and roots, respectively. With these platforms, we identified diurnally oscillating metabolites and their related transcripts accumulation (Figure 1C).

Leaves and roots have distinctive diurnal patterns

Pattern analysis revealed that 8% of total leaf metabolites and 12% of total root metabolites detected in negative mode had diurnal rhythms (Figure 2, Table S1 and File S1). Oscillating metabolites separated roughly into two groups by hierarchical clustering (Figure 2A and 2C). One group of metabolites was highly induced during the day, and the other group peaked at dusk or night. In leaf extracts, 72% of oscillating metabolites peaked during the day (Figure 2A), whereas 81% of root oscillating metabolites peaked at dusk or night (Figure 2C). The number of ions that show highest accumulations at each time clearly showed distinctive patterns in the two different tissues (Figure 2B and 2D). Moreover, only 10 ions (among 182 in leaves and 179 in roots) had diurnal accumulations in both tissues (Figure 2E) and among them, only one ion had the same diurnal rhythm. After the pattern analysis, we identified the fragment and adduct ions among the oscillating ions by CAMERA and by Pearson correlations dependent upon time and treatments [17] and found that a still smaller number of ions (8 ions among 122 and 132 ions in leaves and roots, respectively) were diurnal-regulated in both tissues (Figure 2F).

The results show that oscillating metabolites in *N. attenuata* accumulated in a tissue-specific manner and a few metabolites were commonly oscillating in both leaves and roots (Figure 2). According to the transcriptome analyses performed in root and shoot of *Arabidopsis*, the root has fewer oscillating genes than does the shoot under constant light condition [10]. However, our analysis of oscillating metabolites, which are the final products of gene regulations, demonstrates that roots also have strong diurnal rhythms which are distinct from those found in the leaves of *N. attenuata*.

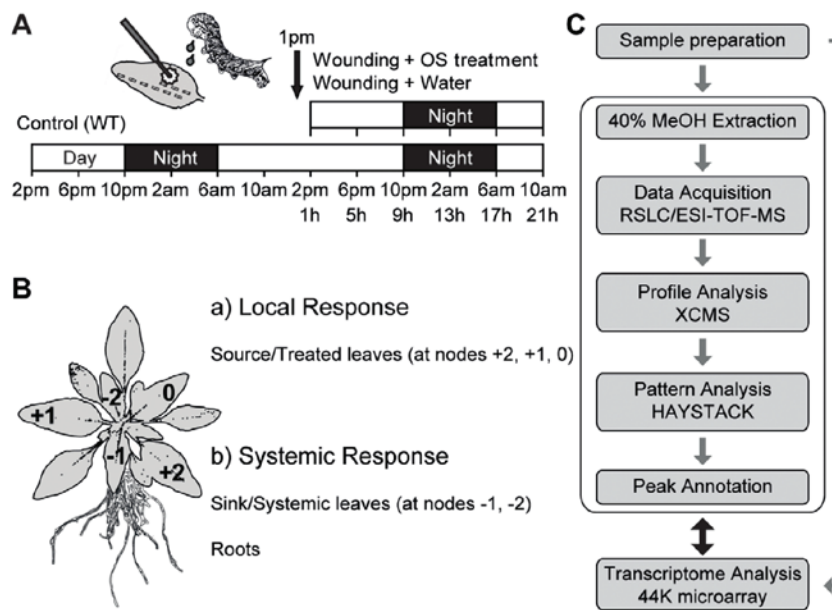


Figure 1. Experimental procedures used to identify oscillating and herbivore-induced metabolites and their associated genes in different tissues of *Nicotiana attenuata*. (A) Wild type (WT) *N. attenuata* plants were harvested every 4 h for two days during the initiation of stem elongation. To mimic herbivory, oral secretions (OS) of the larvae of the specialist herbivore, *M. sexta*, were immediately applied to puncture wounds made in leaves at 1 pm. Water treatment of puncture wounds on separate plants was used to distinguish OS-specific from wound-induced changes in metabolites and transcripts. (B) Metabolites from three different tissues, source leaves, sink leaves, and roots of *N. attenuata* were isolated. The leaf at node 0 had completed the sink to source transition and the leaf at node +1 was older by one leaf position than the leaf at node 0 and so forth. Source leaves (at nodes +2, +1, 0) were wounded with a fabric pattern wheel and treated with 20 μ l of *M. sexta* OS, which was diluted 1:5 with water. Untreated leaves (at nodes -1, -2) and roots were harvested to monitor systemic responses. (C) After sample preparation from six biological replicates, a 40% methanol extraction method optimized for the defense metabolites of *N. attenuata* was used and the metabolites separated with a rapid separation liquid chromatography (RSLC) on a C_{18} column and detected by ESI-TOF-MS (electrospray ionization time-of-flight mass spectrometer) for parents and their daughter ions. Peak picking and alignments were performed with the XCMS package [19]. Diurnal oscillating metabolites were extracted by the pattern matching algorithms of HAYSTACK tool [20]. In-house and public databases were used to identify oscillating metabolites and a 44K Agilent microarray designed for *N. attenuata* was used to examine the expression of metabolite-related genes. doi:10.1371/journal.pone.0026214.g001

Next, we annotated the oscillating ions using in-house and public databases, and analyzed their accumulation after W+W and W+OS treatments. Here, we describe the stories of 7 suites of oscillating metabolites, tales that speak of interesting interactions among the diurnal rhythms of metabolites and plant defense responses against herbivory.

A disaccharide and its related genes

We first focused our attention on two signals corresponding to the m/z 341.11 and its dimer, m/z 683.23, at 90 s both with the same strong diurnal rhythms that peaked at dusk only in roots (Figure 3A and S1). While, the same signals were detected in treated and systemic leaves they did not pass our selection filter in these tissues (Figure 3A). We calculated the elemental formulas (m/z 341.11, $C_{12}H_{21}O_{11}^-$; m/z 683.23, $C_{24}H_{43}O_{22}^-$) using SmartFormula. Molecular mass and standard compound injections verified that m/z 341.11 at 90 s is a disaccharide and m/z 683.23 at 90 s is a dimer of the disaccharide.

Changes in primary metabolism strongly influence plant-herbivore interactions [21]. We therefore examined the regulation of disaccharide levels with their related genes (Figure 3B) after W+W and W+OS treatments. The result showed that only OS-elicitation rapidly reduced disaccharide levels in roots (Figure 3A, green line) within 1 h ($P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). In untreated systemic leaves, disaccharide

levels (Figure 3A, blue line) were significantly reduced by W+OS treatment after 13 h and W+W treatment after 17 h ($P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). However, disaccharide accumulations in treated leaves did not change within 21 h after treatments (Figure 3A, red line), except for a small reduction at 17 h after W+W treatment ($P = 0.0134$, one-way ANOVA followed by Bonferroni *post hoc* test).

Most of the genes involved in sugar metabolism have diurnal rhythms [9,22]. We blasted the Arabidopsis and other plant species homologues of these genes (Figure 3B) against our cDNA database of *N. attenuata* and identified the transcripts with high similarity (Table S2). We observed two distinct diurnal patterns of transcript accumulation for these genes using the HAYSTACK algorithm (Figure 3C and File S1). The first group peaked in the middle of the day and was low during the night (Figure 3C, left). The second group peaked at dawn and remained at low levels for the rest of the day (Figure 3C, right). We analyzed the regulation of these genes after OS-elicitation. Interestingly, the genes involved in sucrose transport (*suc transporter* and sugar exporter, *SWEET* [23]) were significantly up-regulated and several other genes (β -amylase, *suc synthase*, *suc phosphatase* and sugar exporter) were down-regulated by W+OS treatment in treated leaves (Figure 3D; $P < 0.05$, Student's *t*-test), while disaccharide levels in treated leaves (Figure 3A) were not changed. The first group of genes peaked in the day was usually down-regulated in treated

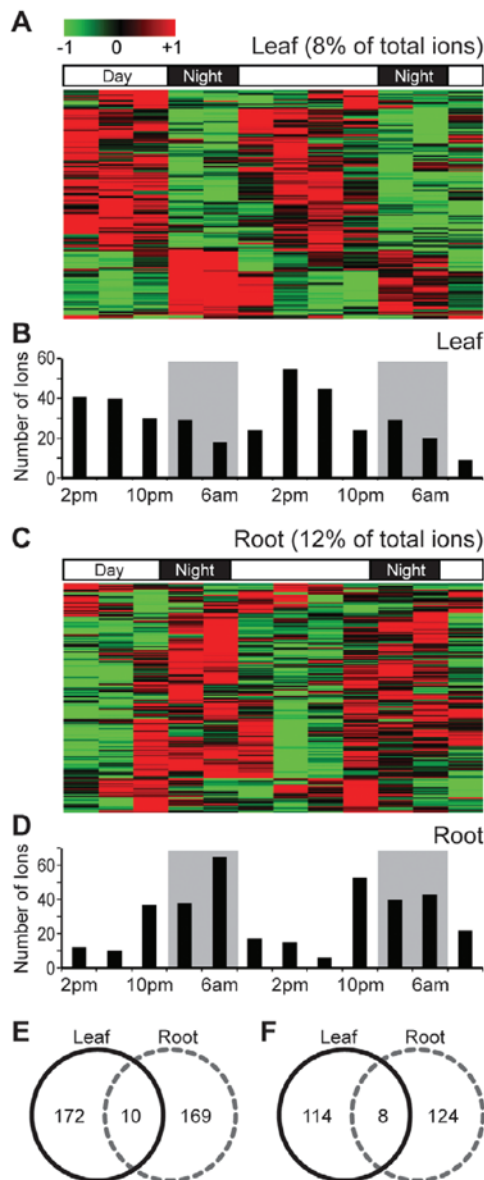


Figure 2. Accumulation of oscillating metabolites in *N. attenuata* show tissue-specific rhythms. Oscillating metabolites in source leaves (A) and roots (C) were roughly divided into two groups, one peaking during the day, and the other peaking at night. The heat map displays all of the Z-transformed oscillating metabolites levels in a false-color scale where green indicates low and red indicates high values. Each metabolite is represented by a single row with the average linkage hierarchical clustering tree obtained using Euclidean distances as metric. We counted the number of ions (y-axis) that peak at a particular harvest times (x-axis) in source leaves (B) and roots (D). Gray boxes depict the dark period. (E) Venn diagram of the oscillating metabolites selected across source leaves (black solid line) and roots (gray dashed line). (F) Venn diagram of the oscillating metabolites after removing adduct and daughter ions. CAMERA package and Pearson correlation [17] were used to select adduct and daughter ions detected from extracts of source leaves (black solid line) and roots (gray dashed line).

doi:10.1371/journal.pone.0026214.g002

leaves by OS-elicitation but the regulation of the second group by OS-elicitation was highly variable (File S1). This pattern of regulation might be affected by the timing of induction (1 pm in this experiment). It will be interesting to evaluate in future studies how elicitation at night influence these patterns. In systemic leaves and roots, these genes were less regulated (Figure 3D), whereas suc levels reduced quickly after treatments (Figure 3A). However, root *SWEET* and *suc invertase* transcript accumulations were rapidly increased within 5 h of OS-treatment (Figure 3D; $P < 0.05$, Student's *t*-test). The increase in *suc invertase* transcript levels is consistent with previous results of increased invertase activity in roots after OS-treatment [24].

Sucrose is synthesized from photosynthesis during the day and loaded into the phloem for transport to sink leaves and roots [25]. Starch is synthesized from sugars during the day and degraded into maltose to provide energy for the metabolic requirements of the dark phase [26]. A plant's clock anticipates the end of the night and regulates the rate of starch degradation to prevent energy limitations at night [27]. The sugar content in plants is known to have a diurnal rhythm [28]. A weak diurnal rhythm of disaccharide in source leaves was observed in early-elongated stage of *N. attenuata* but did not pass our selection filter (Figure 3A). The fact that our analytical procedures do not distinguish sucrose from maltose, which has a rhythm opposite to that of sucrose, may be the explanation, because sucrose accumulations are known to peak at dusk, while maltose peaks at night [27]. Testing this possibility will require calculating the ratio of sucrose and maltose in *N. attenuata* and examining separately their regulation during herbivory. The other reason may be the relatively strong selection filter that we employed. Consistent with this, the kinetic of this disaccharide in source leaves but not in sink leaves was detected as diurnally rhythmic when the stringency was lowered (data not shown).

OS-elicitation significantly altered many sugar-metabolic and transporter genes in treated leaves within 5 h (Figure 3D) but the disaccharide levels in treated leaves was not changed significantly within 21 h (Figure 3A). This indicates that the rapid turnover of disaccharide levels may maintain sugar levels in treated leaves, which is used to produce secondary metabolites for direct defense responses and supply carbon for allocation from shoot to root that would increase a plant's tolerance of herbivory [24,29,30]. In contrast, disaccharide levels in sink tissues, systemic leaves and roots were significantly reduced (Figure 3A). The OS-elicited decrease in disaccharide levels in roots during a day may be mainly due to a reduction of sucrose because maltose levels remain low during the day and increase during the night [26,28]. The reduced disaccharide levels in roots may have increased the gradient force driving sucrose translocation from shoot to root [24]. However, it remains an open question whether the decrease in disaccharide levels in sink leaves is linked to a reduction of sucrose or maltose levels or whether it results from increased sucrose transport from source leaves to sink leaves.

A sugar-containing diterpene glycoside and its related genes

The analysis identified a sugar-containing defense metabolite with strong diurnal patterns of accumulation. We found that the accumulation of lyciumoside I (m/z 629.35 at 339 s, $C_{32}H_{53}O_{12}$), a precursor of various diterpene glycosides (DTGs) in *N. attenuata* [31], peaked at dusk in treated and systemic leaves but this metabolite was not detected in roots (Figure 4A and 4C). Sink leaves contained more lyciumoside I than did source leaves (Figure 4A and 4C). OS-elicitation did not alter the accumulation of lyciumoside I in treated leaves within 21 h, whereas W+W

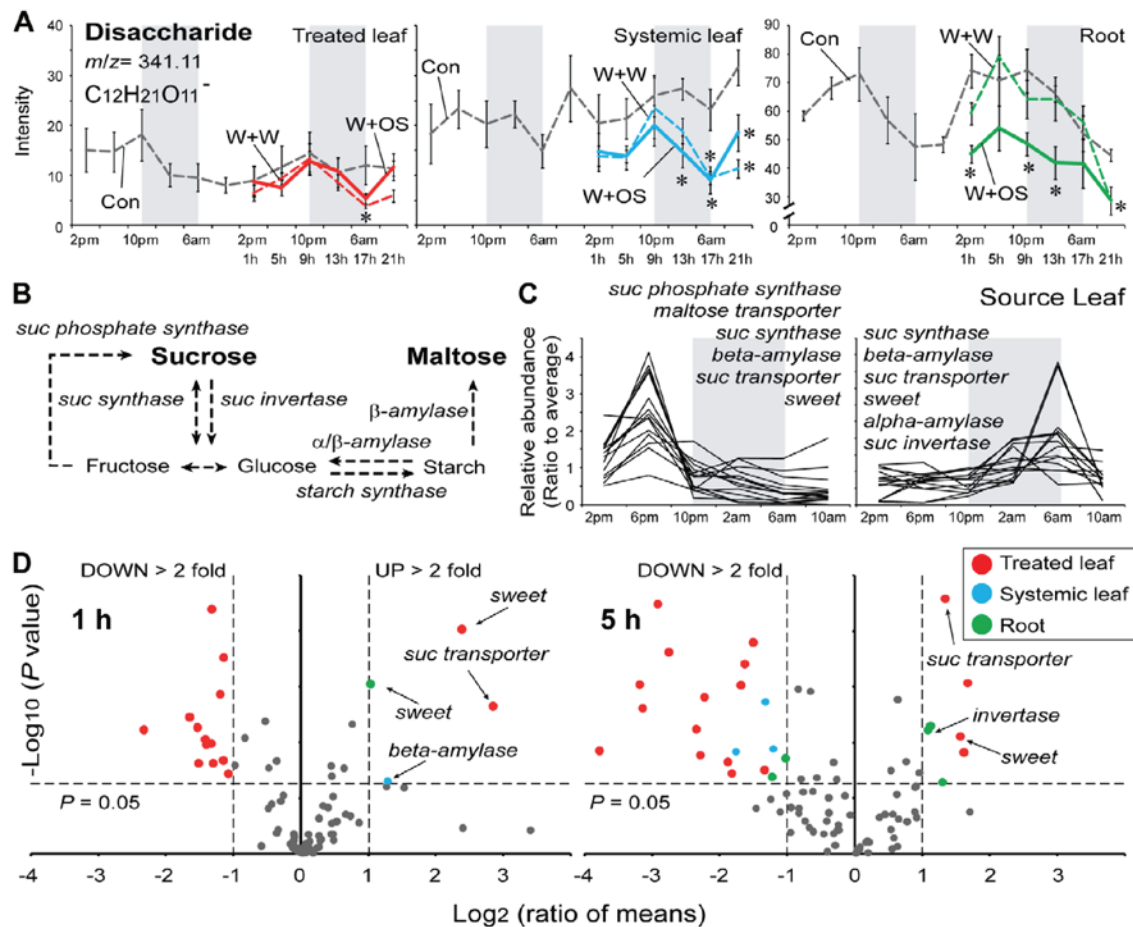


Figure 3. Accumulation of disaccharides and sugar related genes in three different tissues. (A) Mean (\pm SE) levels of normalized intensity of disaccharides (m/z 341.11 at 90 s, $C_{12}H_{21}O_{11}^-$) in source leaves, sink leaves and roots at each harvest time for two days (gray dashed lines) in control (Con) plants. After wounding and treating puncture wounds with either water (W+W, dashed lines with colors) or *M. sexta* OS (W+OS, solid lines with colors), disaccharides levels were examined in treated leaves (red), untreated systemic leaves (blue) and roots (green). Gray boxes depict the dark period. Asterisks indicate significant differences among the treatments at the indicated harvest time (* = $P < 0.05$, one-way ANOVA with Bonferroni *post hoc* test). (B) Schematic overview of sucrose (suc) metabolism. (C) Two diurnal patterns of sugar metabolism-related genes (Table S2) accumulation in source leaves. Gray boxes depict the dark period. Sweet: sugar exporter [23]. Ratio to average: Ratio of transcript abundance at the time point shown, to the mean abundance of the same transcript across all time points. (D) Volcano plot analysis of the transcript levels of sugar-related genes in treated leaves (red dot), systemic untreated leaves (blue dot) and roots (green dot) 1 and 5 h after *M. sexta* OS-elicitation. The \log_2 ratio of mean intensities (OS-elicited/Con, with microarray expression data) plotted against the negative \log_{10} -transformed P value derived from Student's t tests. The horizontal dashed line indicates the threshold for statistically significant expression at $P = 0.05$ and the vertical dashed line, the threshold for two-fold changes in gene expression.
doi:10.1371/journal.pone.0026214.g003

treatment (Figure 4A, red line) increased its levels in treated leaves at 5, 13 and 17 h ($P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). Consistent with these results was the observation that *NaGGPPS* (*N. attenuata geranylgeranyl diphosphate synthase*) transcripts (Figure 4B), which are involved in the biosynthesis of lyciumoside I [31], significantly increased only after W+W treatment ($P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). Systemic signaling also elicited the accumulation of lyciumoside I (Figure 4C) and *NaGGPPS* transcript (Figure 4D) in systemic leaves ($P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). W+W and W+OS treatments resulted in similar increases in lyciumoside I, whereas

OS-elicitation resulted in larger increases in *NaGGPPS* transcripts in systemic leaves after wounding (Figure 4D).

OS-elicitation reduced the wound-induced genes (type I) or amplified the regulation of genes by wounding (type II) [32]. Interestingly, *NaGGPPS* transcript showed a type I expression pattern in treated leaves and a type II expression pattern in systemic leaves. In flowering stage plants, several DTGs accumulate mainly in young leaves and reproductive tissues [31]. Induction of Lyciumoside I by OS-elicitation also followed this pattern which might enhance plant fitness by protecting these young fitness-enhancing tissues first, as predicted by Optimal Defense Theory.

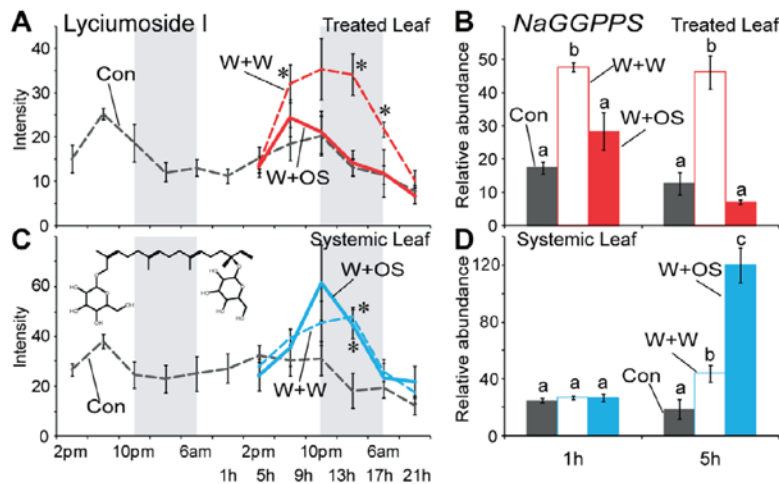


Figure 4. Diurnal rhythms and OS elicitation of glucose-containing secondary metabolites in different tissues. Mean (\pm SE) levels of normalized intensity of lyciumoside I (m/z 629.35 at 339 s, $C_{32}H_{53}O_{12}$) in treated (A), and untreated systemic leaves (C) at each harvest time for two days (gray dotted lines) in control plants. Lyciumoside I was not detected in roots. After W+W (dashed lines with colors) or W+OS (solid lines with colors) treatments, lyciumoside I levels were examined in treated leaves (A) and untreated systemic leaves (C). Gray boxes depict the dark period. Asterisks indicate significant differences among the treatments at the indicated time points (* = $P < 0.05$, one-way ANOVA with Bonferroni *post hoc* test). (B), (D) Effects of W+W and W+OS on relative transcript abundance (\pm SE) of NaGGPPS (*N. attenuata* geranylgeranyl diphosphate synthase), a gene involved in producing the diterpenoid precursor, geranylgeranyl diphosphate [31]. Different letters (a, b and c) reflect significant differences among the treatments at the indicated time points ($P < 0.05$, one-way ANOVA with Bonferroni *post hoc* test). doi:10.1371/journal.pone.0026214.g004

Phenylalanine/tyrosine and their related genes

The aromatic amino acids, phenylalanine (Phe) and tyrosine (Tyr) are well-known primary metabolites with diurnal rhythms [9]. Phe and Tyr, which are synthesized via the shikimate pathway, are used in the production of lignin, anthocyanins, alkaloids, floral scents and defensive metabolites [33]. These compounds were also identified by our method for identifying oscillating metabolites (Figure 5C and 5G). The genes involved in their biosynthesis or catabolism showed similar diurnal patterns as well (Figure 5A, 5B and Table S3). Most of these genes increased during the day and remained at low levels at dusk and night (Figure 5B). Interestingly, Phe accumulation in different tissues peaked at different times: 2 or 6 pm in leaves and 10 pm in roots (Figure 5C and S2). While Phe levels in roots peaked later than in leaves, we did not find a diurnal pattern of the Phe biosynthetic gene transcripts, *arogenate dehydratases* (*NaADTs*) in roots, which are thought to convert arogenate to Phe, (Figure S3A and Table S3). The diurnal rhythm of Phe levels in the roots may be linked to the diurnal accumulation of *NaADT1/2* transcripts in leaves (Figure 5B and 5D) and the translocation of Phe from leaves to roots through the phloem [34–36]. Phe levels in flowers are particularly interesting because it is a precursor of the nocturnally emitted pollinator attractant, benzyl acetone [37,38]. Phe levels in flowers were elevated during the night (Figure S2). Three different tissues (leaf, root and flower) have their own diurnal rhythms of Phe, which would be connected to specific roles of Phe in different tissues. Tyr levels were constitutively high in roots, exceeding the levels found in leaves even during peak accumulations (Figure 5G).

W+W and W+OS treatments induced similar increases of Phe levels in treated leaves (Figure 5C, red line; $P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test), but OS-elicitation resulted in much stronger systemic responses (Figure 5C, blue line; $P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). Phe accumulation was also increased in roots, but the OS-elicited

increase was less than in treated/systemic leaves (Figure 5C, green line). Transcripts of two *NaADTs* were significantly increased within 1 h of elicitation followed by dramatic decreases in treated leaves (Figure 5D, red line; $P < 0.05$, Student's *t*-test), which may explain Phe's peak 5 h after treatment. Transcripts of the same genes (Figure 5D, blue line) were also increased in systemic leaves ($P < 0.05$, Student's *t*-test) but with a 4 h delay. Phe accumulation in systemic leaves peaked twice at 5 and 13 h after OS-treatment but only once at 5 h after W+W treatment. It is possible that a part of the induced Phe accumulation in younger leaves originated from Phe produced in treated leaves.

Phenylalanine ammonia lyase (PAL) is the first enzyme in phenylpropanoid biosynthesis (Figure 5A). *NaPAL1/2* transcript accumulations followed a similar pattern to that of *NaADT1* (Figure 5E) with OS-elicited increases within 1 h in treated leaves and within 5 h in systemic leaves ($P < 0.05$, Student's *t*-test). The other phenylpropanoid pathway gene, *CAH* (*cinnamate-4-hydroxylase*) showed a similar pattern of increase observed in *NaPALs* transcript levels (Figure 5F; $P < 0.05$, Student's *t*-test). The accumulation of *NaADTs*, *NaPALs* and *NaCAHs* transcripts in roots was less affected by W+W and W+OS treatments than in leaves (Figure S3).

W+W and W+OS treatments increased the amount of Tyr in treated and systemic leaves (Figure 5G; $P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test), while OS-treatment increased its level with an 8 h delay in systemic compared to treated leaves (Figure 5G). Only W+W treatment increased Tyr accumulation in roots at 9 and 13 h (Figure 5G; $P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). We identified two *arogenate dehydrogenase* (*TyrA*) genes, which are thought to be involved in Tyr biosynthesis (Figure 5H), with OS-elicited increases in treated and systemic leaves ($P < 0.05$, Student's *t*-test). *NaTyrA1/2* transcript levels increased more in treated leaves compared to systemic leaves (Figure 5H), but induced Tyr levels in both tissues were similar (Figure 5G). The induced Tyr

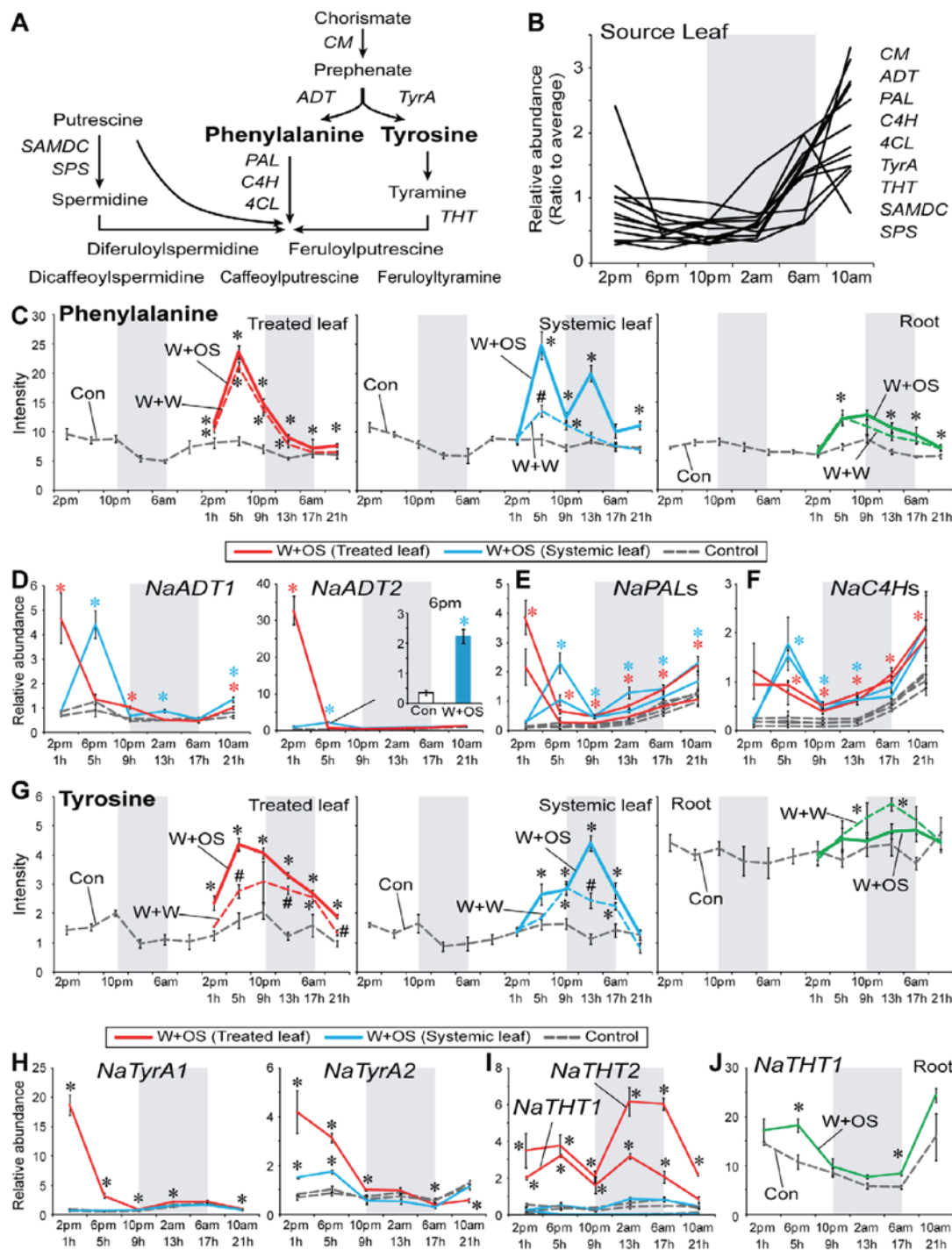


Figure 5. Diurnal rhythms and OS elicitation of Phe and Tyr and their related genes in different tissues. (A) Schematic overview of Phenylalanine (Phe) and Tyrosine (Tyr) metabolism. CM, chorismate mutase; ADT, arogenate dehydratase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate-coa ligase; TyrA, arogenate dehydrogenase; THT, tyramine N-hydroxycinnamoyltransferase; SAMDC, S-adenosylmethionine decarboxylase; SPS, spermidine synthase. (B) Diurnal expression of genes encoding Phe or Tyr metabolism enzymes in source leaves. Gray box depicts the dark period. Ratio to average: Ratio of transcript abundance at the time point shown, to the mean abundance of the same transcript across all time points. Mean (\pm SE) levels of normalized intensity of Phe (C; m/z 164.07 at 192 s, $C_9H_{10}NO_2^-$) and Tyr (G; m/z 180.07 at 144 s, $C_9H_{10}NO_3^-$) in treated leaves, untreated leaves and roots at each harvest time for two days (gray dashed lines) in control plants. After W+W

(dashed lines with colors) or W+OS (solid lines with colors) treatments, Phe (C) and Tyr (G) levels were quantified in treated leaves (red), untreated systemic leaves (blue) and roots (green). Effects of W+W and W+OS treatments on relative transcript abundance (\pm SE) of genes related Phe (D–F) and Tyr (H–J) metabolism. Gray box depicts the dark period. Different symbols (* and #) indicate significant differences among the treatments at the indicated time point ($P < 0.05$, one-way ANOVA with Bonferroni *post hoc* test). doi:10.1371/journal.pone.0026214.g005

accumulation in treated leaves may also be translocated into systemic leaves.

Tyramine N-hydroxycinnamoyltransferase (THT) is an enzyme that conjugates cinnamoyl-CoA, caffeoyl-CoA, or feruloyl-CoA to tyramine [39]. Interestingly, two *NaTHTs* transcript levels (Figure 5I) were increased after OS-elicitation only in treated but not in systemic leaves ($P < 0.05$, Student's *t*-test). *NaTHT1* (Figure 5J) and *NaTHT2* (Figure S4) transcripts accumulated to high levels in roots and only *NaTHT1* transcript displayed diurnal accumulation only in roots. OS-elicitation increased *NaTHT1* (Figure 5J) and decreased *NaTHT2* (Figure S4) transcript levels in roots ($P < 0.05$, Student's *t*-test).

Phenylpropanoid-polyamine conjugates

Coumaroyl tyramine (m/z 284.10 at 165 s, $C_{17}H_{18}NO_3^+$) peaked during the day (Figure 6A) with an OS-elicited increase within 1 h in treated leaves ($P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). The levels of two feruloylamine conjugates [17] were also increased by W+W or W+OS treatments (Figure 6B and 6C). The accumulation of feruloyl putrescine (m/z 265.152 at 212 s, $C_{14}H_{21}N_2O_3^+$) increased in treated and systemic leaves after OS-elicitation but not after W+W treatment (Figure 6B; $P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). N-feruloyl tyramine accumulated (m/z 314.140 at 319 s, $C_{18}H_{20}NO_4^+$) markedly after W+W and W+OS treatments, but only in treated leaves (Figure 6C; $P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). Systemic leaves did not accumulate N-feruloyl tyramine after any treatments was consistent with the low accumulation of *NaTHT1/2* transcripts in systemic leaves (Figure 5I and 6C). However, no change in N-feruloyl tyramine levels was observed in roots (Figure S5) despite the significant increases in *NaTHT1* transcripts in these tissues (Figure 6J).

THT enzyme activity of barley and wheat is known to be stronger in roots than in shoots [40]. N-feruloyl tyramine levels in roots (Figure S5) were higher than in leaves (Figure 6C) before or after W+OS treatment, perhaps a reflection of these *NaTHTs* transcript levels, but did not show diurnal rhythms (Figure S5). Identification of oscillating tyramine conjugates in roots will be helpful to understand plant defenses against root herbivores or pathogens that have their own diurnal activity rhythms [41,42].

OPDA and JA in roots and their related genes

We found two jasmonates with root-specific diurnal patterns of accumulation. OPDA (12-oxophytodienoic acid, m/z 291.20 at 479 s, $C_{18}H_{27}O_3^-$) and JA (m/z 209.12 at 368 s, $C_{12}H_{17}O_3^-$) levels peaked at night only in roots (Figure 7A and 7B). OPDA is a precursor of JA, which is an important phytohormone known to activate defense responses against herbivore attack [43]. JA accumulation is well known to increase after W+W and to be amplified by the OS-elicitation (Figure 7B, red line). Without OS-elicitation, OPDA and JA were detected at higher levels in roots than in leaves (Figure 7A and 7B). W+W and W+OS treatments did not dramatically alter the accumulation of OPDA or JA in roots as they did in treated leaves (Figure 7A and 7B; $P < 0.05$, one-way ANOVA followed by Bonferroni *post hoc* test). Only OS-treatment slightly increased the accumulation of OPDA after 17 h

(Figure 7A; $P = 0.0064$, one-way ANOVA followed by Bonferroni *post hoc* test).

We identified several oscillating transcripts involved in JA signaling (Figure 7C and Table S4). While the levels of JA in leaves did not show diurnal patterns, the JA biosynthetic genes,

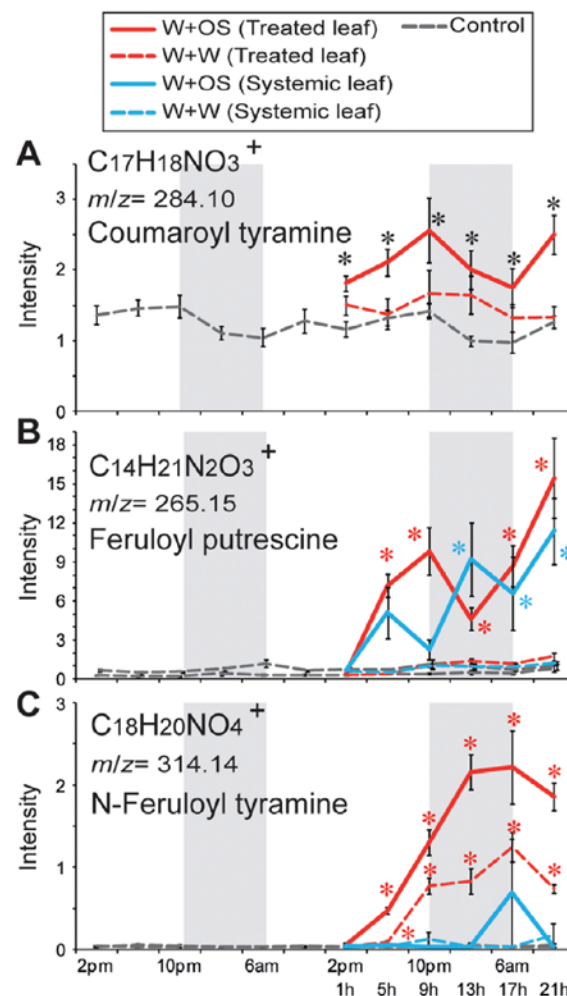


Figure 6. OS-elicitation affects secondary metabolites in the phenylpropanoid pathway. Mean (\pm SE) levels of normalized intensity of coumaroyl tyramine (A; m/z 284.10 at 165 s, $C_{17}H_{18}NO_3^+$), feruloyl putrescine (B; m/z 265.15 at 212 s, $C_{14}H_{21}N_2O_3^+$) and N-feruloyl tyramine (C; m/z 314.14 at 319 s, $C_{18}H_{20}NO_4^+$) in treated and systemic leaves at each harvest time for two days (gray dotted lines) in control plants. Feruloyl putrescine and N-feruloyl tyramine were not detected in roots. After W+W (dashed lines) or W+OS (solid lines) treatments their levels were examined in treated leaves (red) and untreated systemic leaves (blue). Gray boxes depict the dark period. Asterisks indicate significant differences among the treatments at the indicated time point (* = $P < 0.05$, one-way ANOVA with Bonferroni *post hoc* test). doi:10.1371/journal.pone.0026214.g006

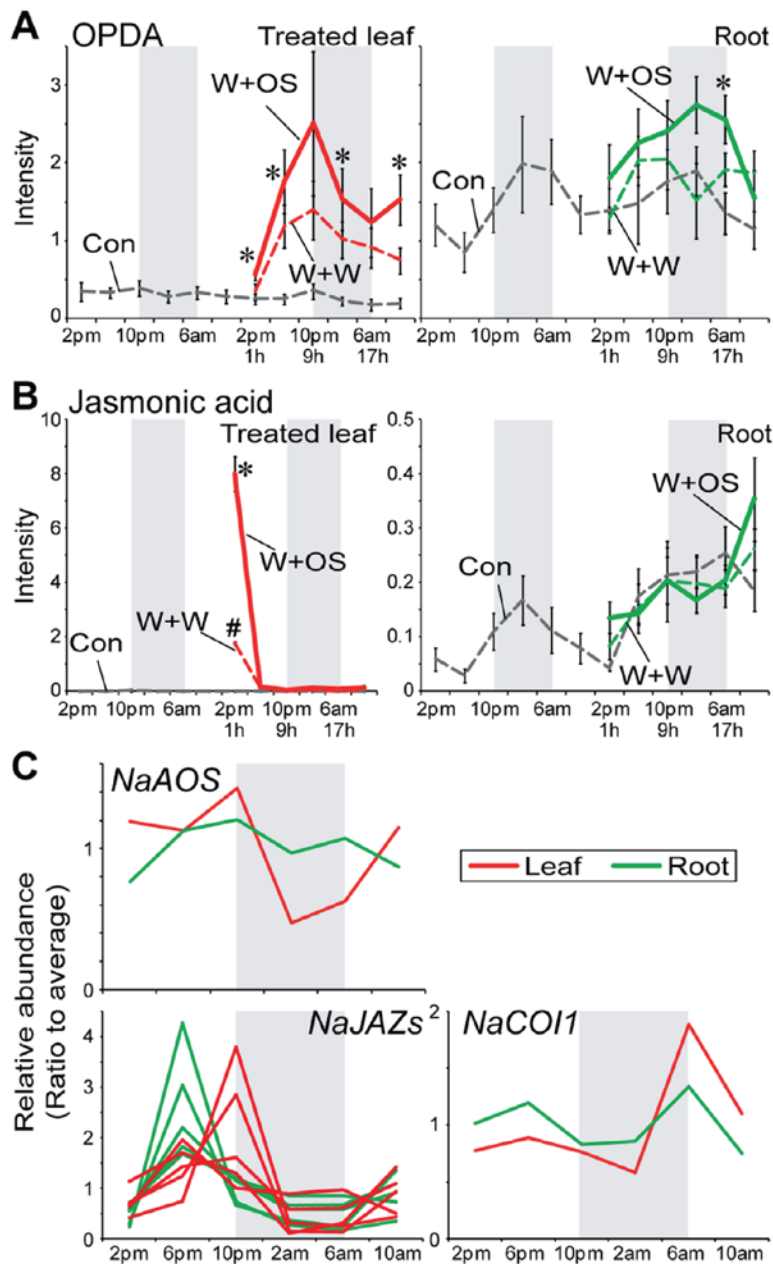


Figure 7. Diurnal rhythms and OS elicitation of OPDA, jasmonic acid and JA-related genes in roots. (A), (B) Mean (\pm SE) levels of normalized intensity of 12-oxophytodienoic acid (OPDA, m/z 291.20 at 479 s, $C_{18}H_{27}O_3^-$) and jasmonic acid (JA, m/z 209.17 at 368 s, $C_{12}H_{17}O_3^-$) in treated leaves and roots at each harvest time for two days (gray dotted lines) in control plants. After W+W (dashed lines with colors) or W+OS (solid lines with colors) treatments, their levels were examined in treated leaves (red) and roots (green). Gray boxes depict the dark period. Different symbols (*) and (#) indicate significant differences among the treatments at the indicated time point ($P < 0.05$, one-way ANOVA with Bonferroni post hoc test). (C) Diurnal rhythms of gene accumulation involved in JA biosynthesis or signaling. AOS, allene oxide synthase; JAZ, jasmonate-ZIM-domain protein; COI1, coronatine insensitive 1. doi:10.1371/journal.pone.0026214.g007

AOS (allene oxide synthase) and two major JA signaling components, COI1 (coronatine insensitive 1) and JAZs (jasmonate-ZIM-domain proteins) transcript accumulations showed diurnal rhythms (Figure 7C, Red line). NaCOI1 transcripts in leaves were induced

at dawn and remained at low levels for the rest of the day (Figure 7C). NaJAZs levels in leaves peaked at 6 or 10 pm in leaves, while the same NaJAZ transcripts in roots peaked at 6 pm (Figure 7C).

Fatty acid-amino acid conjugates (FACs) in *M. sexta* OS induce herbivore-specific defense responses in *N. attenuata* [16,44]. The phytohormone, JA plays an important role in FACs-induced plant defense signaling against herbivore attack [43]. JA biosynthesis occurs in two organelles, the chloroplast and the peroxisome. α -linolenic acid released from membrane is converted into OPDA by LOX, AOS, and AOC (allene oxide cyclase), all enzymes of the chloroplast. The OPDA is then imported into the peroxisome and converted into JA [43]. Even though root cells have no chloroplasts, transcripts encoding LOX, AOS and AOC proteins are expressed in roots of several plant species, likely in the leucoplasts of roots, and jasmonates are detected in their roots [45–48]. To our knowledge, our study represents the first report of a diurnal rhythm of OPDA and JA in roots of plants. Basal levels of JA in leaves did not show a diurnal rhythm, although JA biosynthesis genes, *LOX3* [15] and *AOS* (Figure 7C) transcripts in *N. attenuata*'s leaves did (Figure 7C). It is still unclear whether the oscillation in JA levels in roots results from JA synthesis in leaves and transport into roots or their *de novo* synthesis in the roots. Wang *et al.* treated wounded *N. attenuata*'s leaves with isotope-labeled ($^{13}\text{C}_6$) Ile and detected JA- $^{13}\text{C}_6$ -Ile in treated leaves of *N. attenuata* [49]. However, labeled JA-Ile was not detected in systemic leaves and roots. These data suggest that *N. attenuata*'s roots have the ability to *de novo* synthesize jasmonates [47,49].

Stress-induced biosynthesis of JA initiates direct and indirect defense responses in plants [29]. The diurnal rhythm of JA levels in roots may be linked to the daily occurrence of these stresses in roots. One of the possible stresses is elicited by pathogen attack. The circadian clock component, CCA1 has been recently shown to be critical for plants in the anticipation of leaf pathogen attack and in turn regulate the expression of defense related genes against the pathogen [4]. Arabidopsis *AOS* gene has one CCA1 binding site (AAAAATCT) and one evening element (AAAATATCT) in its promoter region [5,50,51]. *LOX3* in Arabidopsis have also one evening element [50,51]. Homologous genes in *N. attenuata* showed diurnal accumulations (Figure 7C) either in leaves or in roots [15]. It will therefore be extremely interesting to analyze the promoter sequences of JA-related genes in *N. attenuata*. Leaf pathogens usually attack at dawn when temperature and moisture conditions are conducive for pathogen infection. However, roots in the soil are subjected to a completely different microclimate. Diurnal rhythms of JA accumulation in roots therefore may play other roles.

Oscillating JA levels may facilitate root penetration into hard soil. The soil of *N. attenuata*'s native habitat, the Great Basin Desert in Utah is usually dry and hard. We used sand-grown roots, rather than hydroponic culture, to mimic as close as possible these natural conditions. Comparisons of the JA contents of hydroponically-grown and sand-grown roots are not simple, because these different culture conditions produce roots with different patterns of growth and morphology. Nonetheless, when we compared the JA contents in the same mass of roots, we usually observed higher JA levels in sand-grown roots than in hydroponically-grown roots (data not shown). The pattern of JA accumulation that peaked at night may be associated with the growth rhythm of roots and the production of secondary metabolites to protect root cells from infection and attack during growth-associated wounding.

Even though JA levels in leaves did not show a diurnal rhythm, JA signaling components, *COI1* and several *JAZs* transcript levels in *N. attenuata* were diurnal-regulated (Figure 7C). Far-red light treatment increases *AOC* and *JAZ1* transcript accumulation of Arabidopsis and phytochrome A elicits the degradation of *JAZ1* proteins [52], which suppress *COI1*-dependent JA signaling. Extrafloral nectar secretion in lima bean is regulated by JA levels

but also in a light dependent manner [53]. These results suggest that diurnal changes in light quality in nature may influence JA signaling pathway in a time-dependent manner. In other words, diurnal rhythms of JA signaling components in leaves may play a central role in modulating defense responses of plants depending on the time of day when they are attacked.

Future work

The timing of gene regulation has been intensively studied in Arabidopsis. However, a lack of association between diurnal rhythms and one of the major biotic stresses for plants, that of herbivore attack in which secondary metabolites play central roles, led us to examine how oscillating metabolites are influenced by herbivory. Recently, Kerwin *et al.* reported the reverse mechanism, that glucosinolate metabolism affects the output of the circadian clock and clock gene expression [54,55]. Mutations in glucosinolate biosynthetic genes result in the alteration of the circadian clock genes including the *PSUDORESPONSIVE REGULATORs* (*PRRs*). *M. sexta* attack elicits dramatic changes in *N. attenuata* metabolism [17,21,56]. These changes may also feedback on clock gene expression during herbivory. To test this hypothesis, in future work, we will examine the microarray data of *irLOX2* (producing less green leaf volatiles), *irLOX3* (deficient in JA biosynthesis), and *irMTB8* (containing less phenylpropanoid-polyamine conjugates) of *N. attenuata* using single time point analysis described by Kerwin *et al.* [15,54,57].

Materials and Methods

Plant material and treatment

Wild type (WT), *Nicotiana attenuata* plants (30th inbred generation) were grown from seeds originating from a natural population in Utah. The original collection of the seeds was done on private lands and since *N. attenuata* is not an endangered plant species no specific permissions for seed collections were necessary. Moreover, all seeds used in this study were bred in the glasshouse at our institute. Seeds were sterilized and germinated on Gamborg's B5 medium as previously described [58]. Ten-day old seedlings were transferred to small pot (TEKU JP 3050 104 pots, Pöppelmann GmbH & Co. KG, Löhne, Germany) with Klasmann plug soil (Klasmann-Deilmann GmbH, Geesten, Germany) and after 10 days, seedlings were transferred to 1 L pots with sand to facilitate sampling of roots. Plants were watered by flood irrigation system with 200 g $\text{CaNO}_3\cdot 4\text{H}_2\text{O}$, 200 g Flory B1 in 400 L water and grown in the glasshouse at 26–28°C under 16 h supplemental light from Master Sun-T PIA Agro 400 or Master Sun-T PIA Plus 600 W Na lights (Philips, Turnhout, Belgium).

We used early elongated stage of WT *N. attenuata* plants for metabolomics and transcriptomic analyses. Six biological replicates (plants) were harvested every 4 h for 2 d from each treatment group. Before freezing the samples in liquid nitrogen, roots were washed in a water tank for a few seconds to remove sand. *M. sexta* oral secretions (OS) collected from fourth- or fifth-instar larvae were diluted 1:5 with deionized water.

Metabolite analysis

We used a 40% methanol extraction procedure optimized for the extraction of a wide range of our interesting metabolites of in *N. attenuata* [17]. 4 μL of the resulting leaf extracts and 6 μL of the resulting root extracts were injected onto a C18 column (Acclaim, 2.2 μm particle size, 150 mm \times 2.1 mm inner diameter, Dionex Corporation, Sunnyvale, USA) and separated using a RSLC system (Dionex). Solvent A consisted of deionized water containing

0.1% (v/v) acetonitrile (Baker, HPLC grade) and 0.05% (v/v) formic acid. Solvent B consisted of acetonitrile and 0.05% (v/v) formic acid. The following gradient conditions were used for the chromatography: 0–0.5 min 10% B, 0.5–6.5 min linear gradient 80% B, 6.5–10 min 80% B, and re-equilibration at 10% B for 3 min. The flow rate was 300 μ L/min.

An ESI-TOF mass spectrometer (Bruker Daltonic, Bremen, Germany) was used to determine the molecular mass of ionized molecular fragments and the amounts of the eluted analytes. The capillary voltage was 4500 V, and dry gas (200°C) flow rate was 8 L/min. Detected ion range was from m/z 200 to 1400 at a repetition rate of 1 Hz. The mass calibration was achieved using a sodium formate solution (10 mM sodium hydroxide and 0.2% formic acid in isopropanol/water 1:1, v/v).

Raw data processing

Raw data files from Bruker software (Data Analysis v4.0) were exported as netCDF format, and processed using the XCMS R package [19] (http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Peak_Alignment/xcms/). Peak detection was performed using the centWave algorithm with the following parameter settings: ppm = 20, snthresh = 10, peakwidth = c(5,18). Retention time correction was accomplished using the XCMS retcor function with the following parameter settings: mzwid = 0.01, minfrac = 0.5, bw = 3. Missing peak data were filled using the fillPeaks function. The CAMERA package was used to annotate isotope and adduct ions (<http://bioconductor.org/packages/devel/bioc/html/CAMERA.html>). After 75 percentile normalization and log2 transformation, XCMS output files were processed using Microsoft Excel and the Statview software for statistical test, and TIGR's Multiexperiment Viewer software for visualization and clustering.

Diurnal pattern analysis and metabolites annotation

The model-based HAYSTACK [20] algorithm was used to identify diurnal-regulated metabolites (<http://haystack.cgrb.oregonstate.edu/>). Isotope ions detected using CAMERA were removed before processing using pattern matching algorithm of HAYSTACK. With the models from HAYSTACK we selected oscillating metabolites, typically with the following values in the selection filters: correlation cutoff 0.8, fold cutoff 1.5, P -value cutoff 0.05. Molecular formulas were generated using the SmartFormular algorithm in Data Analysis v4.0 software (Bruker). The following maximum elemental composition $C_aH_bN_cO_dNa_eK_f$ and restrictions were used: $1 \leq b/a \leq 3$; $e = 0$ or 1; $f = 0$ or 1; a, b, c and d not limited. Rings plus double bonds values from -0.5 to 40, the nitrogen rule and ions of even electron configuration were considered. The structural annotation based on tandem MS measurements of the precursor ions of selected diterpene glucosides and feruloyl putrescine has been published by our group in Gaquerel *et al.* [17]. Tandem MS measurements for these metabolites are available in the supplemental online material associated with this article. Lyciumoside I has initially been characterized in *N. attenuata* by means of NMR in Heiling *et al.* [31]. Phenylalanine, tyrosine, 12-oxo-phytodienoic acid and jasmonic acid were identified after comparison with authentic standard material. Manual annotation using metabolite information from the literature was performed in the case of tyramine conjugates. Feruloyl tyramine and coumaroyl tyramine are well described in tobacco cells' metabolites whose biosynthesis highly increases by wounding [59,60]. Besides elemental formula calculation, the annotation of these metabolites was facilitated by typical ion signatures corresponding to coumaroyl- (m/z 147.04) and feruloyl- (m/z 177.04) residues released after the break of the

ester bound linking them to a core tyramine molecule during in-source fragmentation.

Microarray data and analysis

Three biological replicates among the six replicates harvested at each harvest time used for metabolites analysis were also used for RNA isolation (six harvest time for control and W+OS treatments, three harvest time for W+W treatments). Total RNA was isolated with TRIZOL reagent and labeled cRNA with the Quick Amp labeling kit (Agilent). Each sample was hybridized on Agilent single color technology arrays (4×44K 60-mer oligonucleotide microarray designed for *N. attenuata* transcriptome analysis, <http://www.agilent.com>, GEO accession number GPL13527). Agilent microarray scanner (G2565BA) and Scan Control software were used to obtain intensity of the spots. All microarray data with each probe name were deposited in the NCBI GEO database (accession number GSE30287). We confirm that all details are MIAME compliant. The resulting gene expression profiles were analyzed using GeneSpring GX software (Silicon Genetics, Redwood City, CA). Raw intensities were normalized using the 75th percentile value and log2 and baseline transformed prior statistical analysis. Probes were filtered based on their Quality control Metrics.

The HAYSTACK algorithm was also used to examine the diurnal rhythm of genes after we selected genes of interest involved in biosynthesis or signaling of oscillating metabolites. With the models from HAYSTACK we examine the diurnal rhythm of gene of interest, typically with the following values in the selection filters: correlation cutoff 0.8, fold cutoff 2, P -value cutoff 0.05.

Supporting Information

Figure S1 Accumulation of $m/z = 683.23$ at retention time, 90 s, in roots. Mean (\pm SE) levels of normalized intensity of $m/z = 683.23$ at 90 s in roots. Calculated molecular formula ($C_{24}H_{43}O_{22}^-$) and retention time (90 s) indicated that it is a disaccharide dimer. After W+W (dashed lines with colors) or W+OS (solid lines with colors) treatments, compound levels were examined in roots (green). Gray boxes depict the dark period. Asterisks indicate significant differences among the treatments at the indicated time point (* = $P < 0.05$, one-way ANOVA with Bonferroni *post hoc* test). (TIF)

Figure S2 Accumulation of Phe in leaves, roots and open flowers. Mean (\pm SE) levels of normalized intensity of Phe in different tissues. Phe accumulation was quantified in leaves (red), roots (green), and flowers (black). To calculate relative accumulation, 75th percentile normalized intensity at each harvest time was divided by average value over all time points. We collected open flowers from 7 week-old plants and extracted metabolites with a 40% methanol extraction method. Gray box depicts the dark period. (TIF)

Figure S3 Transcript abundance of *NaADT1/2*, *NaPAL1/2*, and *NaC4H1/2* in roots. Mean (\pm SE) levels of normalized intensity of *NaADTs*, *NaPALs* and *NaC4Hs* in roots. ADT, arogenate dehydratase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase. Gray boxes depict the dark period. Asterisks indicate significant differences between control plants and *M. sexta* oral secretions-treated plants (W+OS) at indicated time points (* = $P < 0.05$, as determined by Student's *t*-test). (TIF)

Figure S4 Transcript abundance of *NaTHT2* in roots. Mean (\pm SE) levels of normalized intensity of *NaTHT2* in roots. THT, tyramine N-hydroxycinnamoyltransferase. Gray box depicts the dark period. Asterisks indicate significant differences between control plants and *M. sexta* oral secretions-treated plants (W+OS) at indicated time points (* = $P < 0.05$, as determined by Student's *t*-test). (TIF)

Figure S5 Accumulation of N-feruloyl tyramine in roots. Mean (\pm SE) levels of normalized intensity of N-feruloyl tyramine in roots. Gray boxes depict the dark period. Asterisks indicate significant differences between control plants and *M. sexta* oral secretions-treated plants (W+OS) at indicated time points (* = $P < 0.05$, as determined by Student's *t*-test). (TIF)

Table S1 Proposed molecular formulas for oscillating compounds detected in leaves and roots. (TIF)

Table S2 Oscillating transcripts involved in sugar metabolism of *N. attenuata*. (TIF)

Table S3 Oscillating transcripts involved in Phe/Tyr metabolism of *N. attenuata*. (TIF)

Table S4 Oscillating transcripts involved in jasmonic acid metabolism and signaling of *N. attenuata*. (TIF)

File S1 Complete list and expression value of oscillating metabolites and its related transcripts. (XLSX)

Acknowledgments

We thank Matthias Schöttner for technical assistance, and Rothe Eva for sample preparation.

Author Contributions

Conceived and designed the experiments: SK ITB. Performed the experiments: SK FY EG. Analyzed the data: SK EG JG. Contributed reagents/materials/analysis tools: FY. Wrote the paper: SK ITB EG.

References

- Pruneda-Paz JL, Kay SA (2010) An expanding universe of circadian networks in higher plants. *Trends in Plant Science* 15: 259–265.
- Espinoza C, Degenkolbe T, Caldana C, Zuther E, Leisse A, et al. (2010) Interaction with diurnal and circadian regulation results in dynamic metabolic and transcriptional changes during cold acclimation in *Arabidopsis*. *PLoS One* 5: e14101.
- Kolosova N, Gorenstein N, Kish CM, Dudareva N (2001) Regulation of circadian methyl benzoate emission in diurnally and nocturnally emitting plants. *The Plant Cell* 13: 2333–2347.
- Wang W, Barnaby JV, Tada Y, Li H, Tor M, et al. (2011) Timing of plant immune responses by a central circadian regulator. *Nature* 470: 110–114.
- Doherty C, Kay S (2010) Circadian control of global gene expression patterns. *Annual Review of Genetics* 44: 419–444.
- Covington ME, Maloof JN, Straume M, Kay SA, Harmer SL (2008) Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biology* 9: R130.
- Dodd AN, Salathia N, Hall A, Kévei E, Tóth R, et al. (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* 309: 630–633.
- Michael TP, Salomé PA, Yu HJ, Spencer TR, Sharp EL, et al. (2003) Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* 302: 1049–1053.
- Gibon Y, Usadel B, Blasing OE, Kamlage B, Hoehne M, et al. (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in *Arabidopsis* rosettes. *Genome Biology* 7: R76.
- James AB, Monreal JA, Nimmo GA, Kelly CL, Herzyk P, et al. (2008) The circadian clock in *Arabidopsis* roots is a simplified slave version of the clock in shoots. *Science* 322: 1832–1835.
- Moreno-Risueno MA, Van Norman JM, Moreno A, Zhang J, Ahnert SE, et al. (2010) Oscillating gene expression determines competence for periodic *Arabidopsis* root branching. *Science* 329: 1306–1311.
- Takase T, Ishikawa H, Murakami H, Kikuchi J, Sato-Nara K, et al. (2011) The circadian clock modulates water dynamics and aquaporin expression in *Arabidopsis* roots. *Plant and Cell Physiology* 52: 373–383.
- Kessler D, Baldwin IT (2007) Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of *Nicotiana attenuata*. *The Plant Journal* 49: 840–854.
- Kessler D, Diezel C, Baldwin IT (2010) Changing pollinators as a means of escaping herbivores. *Current Biology* 20: 237–242.
- Allmann S, Halitschke R, Schuurink RC, Baldwin IT (2010) Oxylinin channelling in *Nicotiana attenuata*: lipoxygenase 2 supplies substrates for green leaf volatile production. *Plant Cell and Environment* 33: 2028–2040.
- Halitschke R, Schittko U, Pohnert G, Boland W, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. III. Fatty acid-amino acid conjugates in herbivore oral secretions are necessary and sufficient for herbivore-specific plant responses. *Plant Physiology* 125: 711–717.
- Gaquerel E, Heiling S, Schoettner M, Zurek G, Baldwin IT (2010) Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. *Journal of Agricultural and Food Chemistry* 58: 9418–9427.
- Keinänen M, Oldham NJ, Baldwin IT (2001) Rapid HPLC screening of jasmonate-induced increases in tobacco alkaloids, phenolics, and diterpene glycosides in *Nicotiana attenuata*. *Journal of Agricultural and Food Chemistry* 49: 3553–3558.
- Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* 78: 779–787.
- Michael TP, Mockler TC, Breton G, McEntee C, Byer A, et al. (2008) Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genetics* 4: e14.
- Schwachtje J, Baldwin IT (2008) Why does herbivore attack reconfigure primary metabolism? *Plant Physiology* 146: 845–851.
- Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, et al. (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* 290: 2110–2113.
- Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, et al. (2010) Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* 468: 527–532.
- Schwachtje J, Minchin PE, Jahnke S, van Dongen JT, Schittko U, et al. (2006) SNF1-related kinases allow plants to tolerate herbivory by allocating carbon to roots. *Proceedings of the National Academy of Sciences of the United States of America* 103: 12935–12940.
- Stitt M, Lunn J, Usadel B (2010) *Arabidopsis* and primary photosynthetic metabolism - more than the icing on the cake. *The Plant Journal* 61: 1067–1091.
- Lu Y, Gehan JP, Sharkey TD (2005) Daylength and circadian effects on starch degradation and maltose metabolism. *Plant Physiology* 138: 2280–2291.
- Graf A, Schlereth A, Stitt M, Smith AM (2010) Circadian control of carbohydrate availability for growth in *Arabidopsis* plants at night. *Proceedings of the National Academy of Sciences of the United States of America* 107: 9458–9463.
- Blasing OE, Gibon Y, Günther M, Höhne M, Morcuende R, et al. (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *The Plant Cell* 17: 3257–3281.
- Heil M, Baldwin IT (2002) Fitness costs of induced resistance: emerging experimental support for a slippery concept. *Trends in Plant Science* 7: 61–67.
- Kaplan I, Halitschke R, Kessler A, Rehill BJ, Sardanelli S, et al. (2008) Physiological integration of roots and shoots in plant defense strategies links above- and belowground herbivory. *Ecology Letters* 11: 841–851.
- Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, et al. (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranylinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *The Plant Cell* 22: 273–292.
- Schittko U, Hermsmeider D, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. II. Accumulation of plant mRNAs in response to insect-derived cues. *Plant Physiology* 125: 701–710.
- Vogt T (2010) Phenylpropanoid biosynthesis. *Molecular Plant* 3: 2–20.
- Gattolin S, Newbury HJ, Bale JS, Tseng HM, Barrett DA, et al. (2008) A diurnal component to the variation in sieve tube amino acid content in wheat. *Plant Physiology* 147: 912–921.
- Hunt E, Gattolin S, Newbury HJ, Bale JS, Tseng HM, et al. (2010) A mutation in amino acid permease AAP6 reduces the amino acid content of the *Arabidopsis*

Diurnal Rhythms in Plant Defense Metabolites

- sieve elements but leaves aphid herbivores unaffected. *Journal of Experimental Botany* 61: 55–64.
36. Turgeon R, Wolf S (2009) Phloem transport: cellular pathways and molecular trafficking. *Annual Review of Plant Biology* 60: 207–221.
 37. Abe I, Takahashi Y, Morita H, Noguchi H (2001) Benzalacetone synthase. *European Journal of Biochemistry* 268: 3354–3359.
 38. Euler M, Baldwin IT (1996) The chemistry of defense and apparency in the corollas of *Nicotiana attenuata*. *Oecologia* 107: 102–112.
 39. Hagel JM, Facchini PJ (2005) Elevated tyrosine decarboxylase and tyramine hydroxycinnamoyltransferase levels increase wound-induced tyramine-derived hydroxycinnamic acid amide accumulation in transgenic tobacco leaves. *Planta* 221: 904–914.
 40. Louis V, Negrel J (1991) Tyramine hydroxycinnamoyl transferase in the roots of wheat and barley seedlings. *Phytochemistry* 30: 2519–2522.
 41. Muroi A, Ishihara A, Tanaka C, Ishizuka A, Takabayashi J, et al. (2009) Accumulation of hydroxycinnamic acid amides induced by pathogen infection and identification of agmatine coumaroyltransferase in *Arabidopsis thaliana*. *Planta* 230: 517–527.
 42. Newman MA, von Roepenack-Lahaye E, Parr A, Daniels MJ, Dow JM (2001) Induction of hydroxycinnamoyl-tyramine conjugates in pepper by *Xanthomonas campestris*, a plant defense response activated by *hrp* gene-dependent and *hrp* gene-independent mechanisms. *Molecular Plant-Microbe Interactions* 14: 785–792.
 43. Wu J, Baldwin IT (2010) New insights into plant responses to the attack from insect herbivores. *Annual Review of Genetics* 44: 1–24.
 44. Bonaventure G, Vandoorn A, Baldwin IT (2011) Herbivore-associated elicitors: FAC signaling and metabolism. *Trends in Plant Science* 16: 294–299.
 45. Abdala G, Miersch O, Kramell R, Vigliocco A, Agostini E, et al. (2003) Jasmonate and octadecanoid occurrence in tomato hairy roots. Endogenous level changes in response to NaCl. *Plant Growth Regulation* 40: 21–27.
 46. Baldwin IT, Schmelz EA, Ohnmeiss TE (1994) Wound-induced changes in root and shoot jasmonic acid pools correlate with induced nicotine synthesis in *Nicotiana sylvestris* spigazzini and comes. *Journal of Chemical Ecology* 20: 2139–2157.
 47. Bonaventure G, Schuck S, Baldwin IT (2011) Revealing complexity and specificity in the activation of lipase-mediated oxylipin biosynthesis: A specific role of the *Nicotiana attenuata* GLA1 lipase in the activation of JA biosynthesis in leaves and roots. *Plant Cell & Environment* 34: 1507–1520.
 48. Gao X, Starr J, Göbel C, Engelberth J, Feussner I, et al. (2008) Maize 9-lipoxygenase ZmLOX3 controls development, root-specific expression of defense genes, and resistance to root-knot nematodes. *Molecular Plant-Microbe Interactions* 21: 98–109.
 49. Wang L, Allmann S, Wu J, Baldwin IT (2008) Comparisons of LIPOXYGENASE3- and JASMONATE-RESISTANT4/6-silenced plants reveal that jasmonic acid and jasmonic acid-amino acid conjugates play different roles in herbivore resistance of *Nicotiana attenuata*. *Plant Physiology* 146: 904–915.
 50. O'Connor TR, Dyrneson C, Wyrick JJ (2005) Athena: a resource for rapid visualization and systematic analysis of *Arabidopsis* promoter sequences. *Bioinformatics* 21: 4411–4413.
 51. Pan Y, Michael TP, Hudson ME, Kay SA, Chory J, et al. (2009) Cytochrome P450 Monooxygenases as Reporters for Circadian-Regulated Pathways. *Plant Physiology* 150: 858–878.
 52. Robson F, Okamoto H, Patrick E, Harris SR, Wasternack C, et al. (2010) Jasmonate and phytochrome A signaling in *Arabidopsis* wound and shade responses are integrated through JAZ1 stability. *The Plant Cell* 22: 1143–1160.
 53. Radhika V, Kost C, Mithofer A, Boland W (2010) Regulation of extrafloral nectar secretion by jasmonates in lima bean is light dependent. *Proceedings of the National Academy of Sciences of the United States of America* 107: 17228–17233.
 54. Kerwin RE, Jimenez-Gomez JM, Fulop D, Harner SL, Maloof JN, et al. (2011) Network quantitative trait loci mapping of circadian clock outputs identifies metabolic pathway-to-clock linkages in *Arabidopsis*. *The Plant Cell* 23: 471–485.
 55. Ueda HR, Chen W, Minami Y, Honma S, Honma K, et al. (2004) Molecular-timetable methods for detection of body time and rhythm disorders from single-time-point genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 101: 11227–11232.
 56. Gaquerel E, Weinhold A, Baldwin IT (2009) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. VIII. An unbiased GCxGC-ToFMS analysis of the plant's elicited volatile emissions. *Plant Physiology* 149: 1408–1423.
 57. Kaur H, Heinzl N, Schöttner M, Baldwin IT, Gális I (2010) R2R3-NaMYB8 regulates the accumulation of phenylpropanoid-polyamine conjugates, which are essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*. *Plant Physiology* 152: 1731–1747.
 58. Krügel T, Lim M, Gase K, Halitschke R, Baldwin IT (2002) *Agrobacterium*-mediated transformation of *Nicotiana attenuata*, a model ecological expression system. *Chemoecology* 12: 177–183.
 59. Guillet G, De Luca V (2005) Wound-inducible biosynthesis of phytoalexin hydroxycinnamic acid amides of tyramine in tryptophan and tyrosine decarboxylase transgenic tobacco lines. *Plant Physiology* 137: 692–699.
 60. Pearce G, Marchand PA, Griswold J, Lewis NG, Ryan CA (1998) Accumulation of feruloyltyramine and p-coumaroyltyramine in tomato leaves in response to wounding. *Phytochemistry* 47: 659–664.

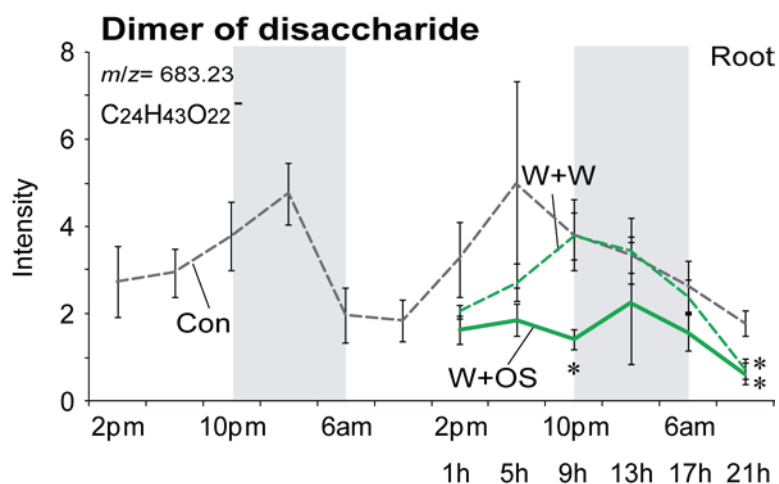


Figure S1.

Accumulation of $m/z = 683.23$ at retention time, 90 s, in roots. Mean (\pm SE) levels of normalized intensity of $m/z = 683.23$ at 90 s in roots. Calculated molecular formula ($C_{24}H_{43}O_{22}^-$) and retention time (90 s) indicated that it is a disaccharide dimer. After W+W (dashed lines with colors) or W+OS (solid lines with colors) treatments, compound levels were examined in roots (green). Gray boxes depict the dark period. Asterisks indicate significant differences among the treatments at the indicated time point (* = $P < 0.05$, one-way ANOVA with Bonferroni *post hoc* test).

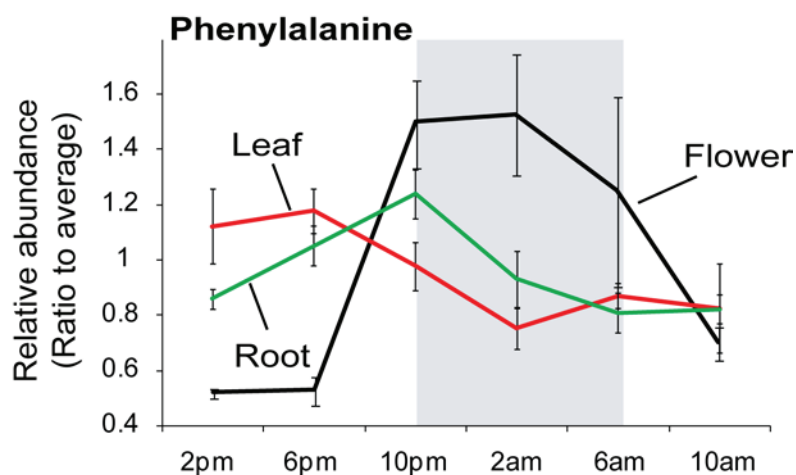


Figure S2.

Accumulation of Phe in leaves, roots and open flowers. Mean (\pm SE) levels of normalized intensity of Phe in different tissues. Phe accumulation was quantified in leaves (red), roots (green), and flowers (black). To calculate relative accumulation, 75th percentile normalized intensity at each harvest time was divided by average value over all time points. We collected open flowers from 7 week-old plants and extracted metabolites with a 40% methanol extraction method. Gray box depicts the dark period.

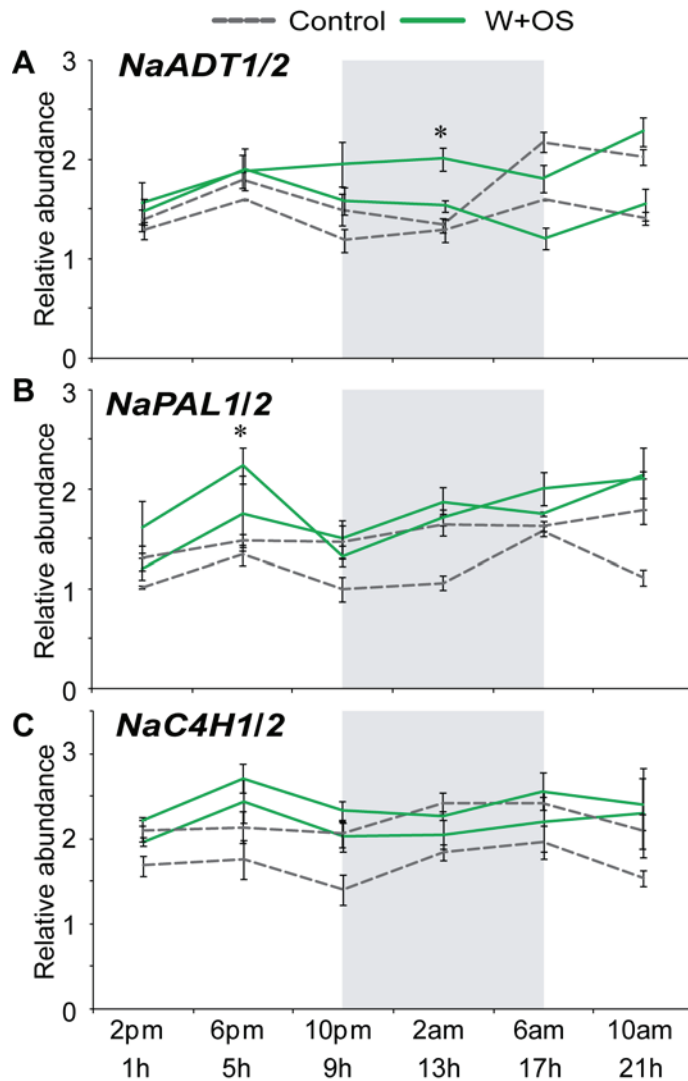


Figure S3.

Transcript abundance of *NaADT1/2*, *NaPAL1/2*, and *NaC4H1/2* in roots. Mean (\pm SE) levels of normalized intensity of *NaADTs*, *NaPALs* and *NaC4Hs* in roots. ADT, arogenate dehydratase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase. Gray boxes depict the dark period. Asterisks indicate significant differences between control plants and *M. sexta* oral secretions-treated plants (W+OS) at indicated time points (* = $P < 0.05$, as determined by Student's *t*-test).

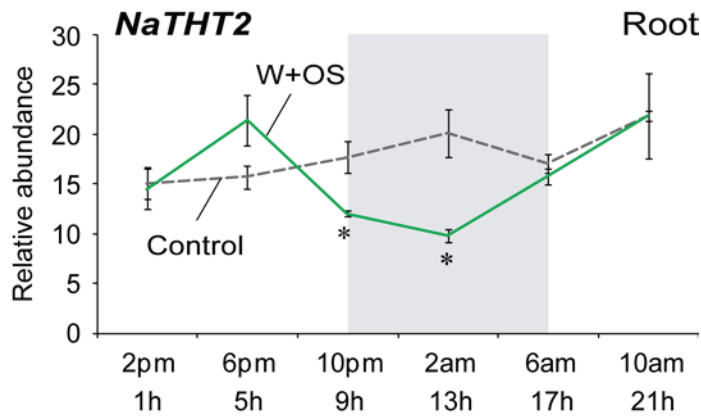


Figure S4.

Transcript abundance of *NaTHT2* in roots. Mean (\pm SE) levels of normalized intensity of *NaTHT2* in roots. THT, tyramine N-hydroxycinnamoyltransferase. Gray box depicts the dark period. Asterisks indicate significant differences between control plants and *M. sexta* oral secretions-treated plants (W+OS) at indicated time points (* = $P < 0.05$, as determined by Student's *t*-test).

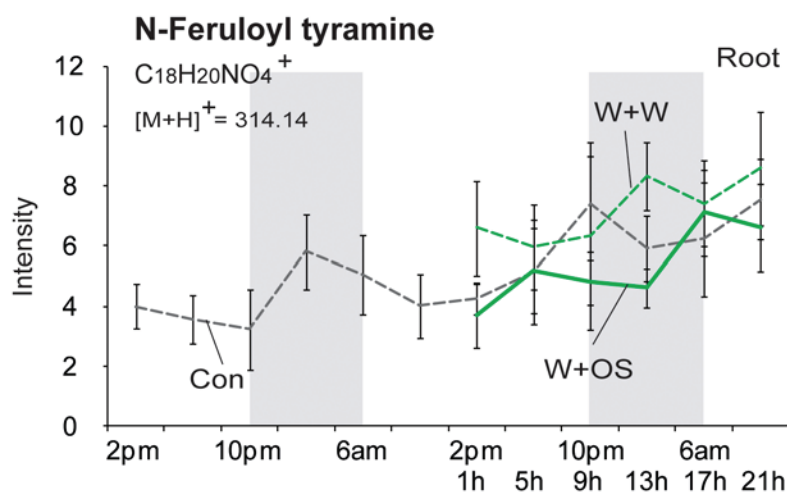


Figure S5.

Accumulation of N-feruloyl tyramine in roots. Mean (\pm SE) levels of normalized intensity of N-feruloyl tyramine in roots. Gray boxes depict the dark period. Asterisks indicate significant differences between control plants and *M. sexta* oral secretions-treated plants (W+OS) at indicated time points (* = $P < 0.05$, as determined by Student's *t*-test).

Table S1. Proposed molecular formulas of oscillating compounds detected in leaves and roots

Peak No.	R.T.	m/z	M.F.	Error (mDa)	SigmaFit (mσ)	Name
L1	129 s	191.02	C ₆ H ₇ O ₇ ⁻	0.76	4.9	Citric acid
L2/R3	137 s	243.06	C ₉ H ₁₁ N ₂ O ₆ ⁻	0.28	19.7	
L3	143 s	180.07	C ₉ H ₁₀ NO ₃ ⁻	0.85	19.1	Tyrosine
L4	167 s	282.08	C ₉ H ₁₆ NO ₉ ⁻	-0.56	24.1	
L5/R5	192 s	164.07	C ₉ H ₁₀ NO ₂ ⁻	-1.20	10.7	Phenylalanine
L6	261 s	463.22	C ₂₁ H ₃₅ O ₁₁ ⁻	1.99	15.3	
L7	306 s	691.34	C ₃₂ H ₅₁ O ₁₆ ⁻	-0.28	26.7	
L8	317 s	227.13	C ₁₂ H ₁₉ O ₄ ⁻	1.44	10.7	
L9	325 s	673.34	C ₃₃ H ₅₃ O ₁₄ ⁻	1.61	60.5	
L10	340 s	629.35	C ₃₂ H ₅₃ O ₁₂ ⁻	-0.79	17.2	Lyciumoside I
L11	352 s	715.35	C ₃₅ H ₅₅ O ₁₅ ⁻	-3.08	16.8	
L12	361 s	689.34	C ₃₃ H ₅₃ O ₁₅ ⁻	2.63	12.2	
R1	88s	341.10	C ₁₂ H ₂₁ O ₁₁ ⁻	1.27	3.3	Dissacharide
R2	99 s	259.02	C ₈ H ₇ N ₂ O ₈ ⁻	0.07	13.8	
R4	138 s	243.06	C ₉ H ₁₁ N ₂ O ₆ ⁻	0.71	6.7	
R6	197 s	315.07	C ₁₃ H ₁₅ O ₉ ⁻	0.71	3.7	
R7	354 s	329.23	C ₁₈ H ₃₃ O ₅ ⁻	0.07	1.3	
R8	367 s	209.12	C ₁₂ H ₁₇ O ₃ ⁻	0.81	6.6	Jasmonic acid
R9	396 s	327.22	C ₁₈ H ₃₁ O ₅ ⁻	0.32	2.4	
R10	433 s	309.21	C ₁₈ H ₂₉ O ₄ ⁻	0.35	83.8	
R11	437 s	341.23	C ₁₉ H ₃₃ O ₅ ⁻	1.08	11.8	
R12	446 s	311.22	C ₁₈ H ₃₁ O ₄ ⁻	1.01	21.8	
R13	458 s	311.22	C ₁₈ H ₃₁ O ₄ ⁻	1.41	4.8	
R14	478 s	291.19	C ₁₈ H ₂₇ O ₃ ⁻	1.04	56.6	OPDA

Molecular formulas (M.F.) were calculated using Bruker SmartFormula software. Blanks in name column indicate unknown compounds. L, leaf; R, root; R.T., retention time.

Table S2. Oscillating transcripts involved in sugar metabolism of *N. attenuata*.

Probe Name	Peak	Gene Annotation
Na_454_03214	Night	alpha-amylase
Na_454_06724	Night	beta-amylase
Na_454_08612	Night	beta-amylase
Na_454_10779	Night	beta-amylase
Na_454_35633	Night	beta-amylase
Na_454_12641	Night	sucrose invertase
Na_454_41936	Night	sucrose invertase
Na_454_01065	Night	sucrose synthase
Na_454_05247	Night	sucrose transporter
Na_454_02704	Night	sugar exporter (SWEET)
Na_454_05391	Night	sugar exporter (SWEET)
Na_454_16634	Night	sugar exporter (SWEET)
Na_454_28050	Night	sugar exporter (SWEET)
Na_454_14081	Day	beta-amylase
Na_454_18724	Day	beta-amylase
Na_454_20797	Day	beta-amylase
Na_454_22181	Day	beta-amylase
Na_454_34221	Day	beta-amylase
Na_454_39081	Day	beta-amylase
Na_454_00622	Day	maltose transporter
Na_454_11568	Day	sucrose phosphate synthase
Na_454_39731	Day	sucrose phosphate synthase
Na_454_40855	Day	sucrose phosphate synthase
Na_454_16096	Day	sucrose synthase
Na_454_27568	Day	sucrose synthase
Na_454_04103	Day	sugar exporter (SWEET)
Na_454_05017	Day	sugar exporter (SWEET)
Na_454_06723	Day	sugar exporter (SWEET)

Table S3. Oscillating transcripts involved in Phe/Tyr metabolism of *N. attenuata*.

Probe Name	Gene Annotation
Na_454_02400	arogenate dehydratase (ADT)
Na_454_09000	arogenate dehydrogenase (TyrA)
Na_454_00178	Phe ammonia lyase 1 (PAL1)
Na_454_00083	Phe ammonia lyase 2 (PAL2)
Na_454_00021	4-coumarate:CoA ligase (4CL)
Na_454_00101	cinnamic acid-4-hydroxylase (C4H)
Na_454_00269	cinnamic acid-4-hydroxylase (C4H)
Na_454_00315	S-adenosylmethionine decarboxylase (SAMDC)
Na_454_00797	S-adenosylmethionine decarboxylase (SAMDC)
Na_454_36134	spermidine synthase (SPS)
Na_454_07429	tyramine N-hydroxycinnamoyltransferase (THT)

Table S4. Oscillating transcripts involved in jasmonic acid metabolism and signaling of *N. attenuata*.

Probe Name	Gene Annotation
Na_454_00565	allen oxide synthase (AOS)
Na_454_04958	coronatine insensitive 1 (COI1)
Na_454_02978	jasmonate-ZIM-domain (JAZ)
Na_454_13225	jasmonate-ZIM-domain (JAZ)
Na_454_14633	jasmonate-ZIM-domain (JAZ)
Na_454_19886	jasmonate-ZIM-domain (JAZ)
Na_454_41896	jasmonate-ZIM-domain (JAZ)

Chapter 4

Deciphering Herbivory-Induced Gene-to-Metabolite Dynamics in *Nicotiana attenuata* Tissues Using a Multifactorial Approach

Jyotasana Gulati, Sang-Gyu Kim, Ian T. Baldwin and Emmanuel Gaquerel

Published in *Plant Physiology* 2013, **162**: 1042-1059

Deciphering Herbivory-Induced Gene-to-Metabolite Dynamics in *Nicotiana attenuata* Tissues Using a Multifactorial Approach^{1[C][W][OA]}

Jyotasana Gulati, Sang-Gyu Kim, Ian T. Baldwin, and Emmanuel Gaquerel*

Department of Molecular Ecology, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany

In response to biotic stresses, such as herbivore attack, plants reorganize their transcriptomes and reconfigure their physiologies not only in attacked tissues but throughout the plant. These whole-organismic reconfigurations are coordinated by a poorly understood network of signal transduction cascades. To explore tissue-based interdependencies in the resistance of *Nicotiana attenuata* to insect attack, we conducted time-series transcriptome and metabolome profiling of herbivory-elicited source leaves and unelicited sink leaves and roots. To probe the multidimensionality of these molecular responses, we designed a novel approach of combining an extended self-organizing maps-based dimensionality reduction method with bootstrap-based nonparametric analysis of variance models to identify the onset and context of signaling and metabolic pathway activations. We illustrate the value of this analysis by revisiting dynamic changes in the expression of regulatory and structural genes of the oxylipin pathway and by studying nonlinearities in gene-metabolite associations involved in the acyclic diterpene glucoside pathway after selectively extracting modules based on their dynamic response patterns. This novel dimensionality reduction approach is broadly applicable to capture the dynamic rewiring of gene and metabolite networks in experimental design with multiple factors.

Plants adapt to environmental stresses through large-scale transcriptional reprogramming, which involves intricate signaling pathways (Hahlbrock et al., 2003; Nakashima et al., 2009; Zeller et al., 2009; Walley and Dehesh, 2010). These transcriptional adjustments can be captured by studying changes in the expression of genes in different tissues in order to elucidate the influence of particular pathways as well as the relative contribution of a given tissue to the whole-organism response. Although poorly understood, signaling networks controlling these transcriptional responses have been shown to be highly stress condition specific, as clearly illustrated by the large number of studies that have demonstrated differences in plant responses to mechanical wounding and herbivory (Baldwin, 1990; Alborn et al., 1997; Halitschke et al., 2003; Reymond et al., 2004; Wu et al., 2007). Experiments designed to study such intricate networks often have a complex factorial structure, obtained from different conditions/

treatments, tissue types, or genetic contexts. In addition, time-series experiments are often employed to capture dynamic expression profiles that distinguish primary from secondary responses to stress in gene regulatory networks (GRN).

Basic statistical tests and clustering algorithms based on Pearson correlation to analyze multifactorial experiments are often plagued by the problem of gene prioritization and large numbers of false positives (Bittner et al., 1999; Getz et al., 2000). First, clustering algorithms classify genes on the basis of their expression under all experimental conditions, whereas signaling pathways underlying these gene expression responses are generally affected only by a subset of the experimental conditions (Swindell, 2006). Additionally, connections in GRN computed by considering all samples for different tissues, treatments, and time points together in a single analysis fail to recognize the transient gene associations found in early stress-responsive pathways that only appear in a subset of treatment types. Thus, the synchronous coregulation of genes representing intermediate biological states cannot be captured using collective information studies, and this represents a major challenge for the identification of the exact mechanisms of stress adaption in many organisms. Bioinformatic approaches such as mutual information (Priness et al., 2007) and biclustering (Dharan and Nair, 2009) have been developed to address this limitation. Moreover, interaction patterns deduced from these coexpression studies represent the static wiring of the network, whereas networks will assemble dynamically as the organism adapts to external stimuli.

¹ This work was supported by the Max Planck Society and by the European Research Council (grant no. 293926 to I.T.B.).

* Corresponding author; e-mail egaquerel@ice.mpg.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Emmanuel Gaquerel (egaquerel@ice.mpg.de).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.113.217588

To best capture the temporal dimension as a variable affecting the structure of GRN, several efforts have been published that identify patterns in time-series data. Park et al. (2003) used a modified ANOVA approach taking time as a factor along with other conditions. Wang and Kim (2003) used mixed-effect ANOVA to identify genes with different temporal profiles for different stress conditions in *Caenorhabditis elegans*. Tai and Speed (2006) used an empirical Bayes approach to introduce moderation, defined as the effect of moving gene-specific variances toward a common value estimated from a whole gene set, to reduce the number of false positives and false negatives. Storey et al. (2005) developed a statistical method that identifies genes showing differential temporal expression profiles by assigning a statistic calculated using spline-based methods. Zhou et al. (2010) developed a method to simultaneously analyze experiments involving more than one factor measured across time series by finding the significant direction in the time course across different conditions.

Nicotiana attenuata, a wild tobacco species native to the Great Basin Desert in the United States, is among the few model plants for which different omics technologies have been applied to understand its complex ecophysiological responses (Halitschke et al., 2003; Giri et al., 2006; Gaquerel et al., 2010). *N. attenuata* germinates in the postfire environment from long-lived seed banks to form monocultures in nitrogen-rich soils (Baldwin and Morse, 1994; Baldwin et al., 1994). As a consequence of its peculiar germination system, this plant is an ideal model for understanding the traits that native plants have evolved to cope with stresses characteristic of the agricultural niche: intense intra-specific competition and highly variable biotic and abiotic stress regimes (Baldwin, 2001). By germinating into the postfire environment, the plant becomes the focus of herbivores that colonize open habitats. This large and unpredictable herbivore community has provided a major evolutionary selective pressure that has likely sculpted many aspects of the plant's GRN. Some of the essential nodes in the plant's transcriptome and metabolome responses to attack from larvae of the specialist lepidopteran herbivore, *Manduca sexta*, have been functionally characterized (for review, see Wu and Baldwin, 2010). Feeding by this herbivore or its simulation by the application of its oral secretions (OS) into puncture wounds results in profound reorganizations of the plant's metabolic and growth processes, in the de novo production of defense compounds, such as the accumulation and mobilization of nicotine (Steppuhn et al., 2004), phenolic derivatives (Kaur et al., 2010), and acyclic diterpene glycosides (Heiling et al., 2010), and in the activation of tolerance mechanisms essential for survival (carbon and nitrogen bunkering in roots; Schwachtje et al., 2006). The jasmonic acid (JA) signaling cascade represents the core pathway controlling these responses (Halitschke and Baldwin, 2003). Its activity mediates, via a set of largely unknown pathway-specific

transcription factors, profound changes in the expression of regulatory and structural genes (Halitschke and Baldwin, 2003; Wang et al., 2008).

To investigate the dynamics of activation in time and space of herbivory-induced changes in gene-to-metabolite networks, we conducted replicated global expression profiling using identically treated wild-type *N. attenuata* plants for three tissues and two stress conditions (mechanical wounding and simulated herbivory) with a regular time series of six time points for both metabolomics and transcriptomics data. We then employed a bootstrap-based nonparametric ANOVA model to find the projected significant direction in the time vector for two factors (stress and tissue type) for both genes and metabolites (Zhou et al., 2010; Zhou and Wong, 2011). To further characterize the coexpression modules of different tissues, we imposed the structure using batch-learning (BL) self-organizing maps (SOM). From this analysis, we framed the concept of interactive motifs, which are defined as patterns of interconnections between genes and metabolites that are differentially perturbed in local and systemic tissues in response to stress, with additional information of their time of action obtained from projected data on time series termed "ANOVA directions." We extracted and studied motifs defining gene-gene and gene-metabolite interactions involved in early and late responses in systemic leaf tissues. The premise of this approach is that genes or metabolites with similar ANOVA directions and, therefore, similar dynamic responses during herbivory are likely to be involved in similar biochemical pathways and/or are under the control of a common transcriptional regulatory mechanism. This broadly applicable approach allowed us to identify nonlinear relationships in gene-metabolite interactions with a high level of accuracy and robustness.

RESULTS

Work Flow for Analyzing the *N. attenuata* Transcriptome and Metabolome

The objective of this study was to identify tissue-specific gene-gene and gene-metabolite associations recruited in response to chewing insect herbivory in *N. attenuata* plants. Therefore, we conducted time-course transcriptome analysis and broadly targeted ultra-high performance liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer (UHPLC-qTOFMS) metabolome measurements of source/sink transition leaves and roots (Fig. 1). Each sample was harvested from an independent control or treated plant every 4 h during a 21-h period to capture early and late activity phases of OS-elicited responses. In treated plants, *M. sexta* feeding was simulated by wounding a leaf with a fabric pattern wheel on both sides of the midrib of a source leaf and immediately applying *M. sexta* OS to the fresh puncture wounds (W+OS). This procedure, hereafter referred to as OS elicitation, recapitulates most changes in the *N.*

Gulati et al.

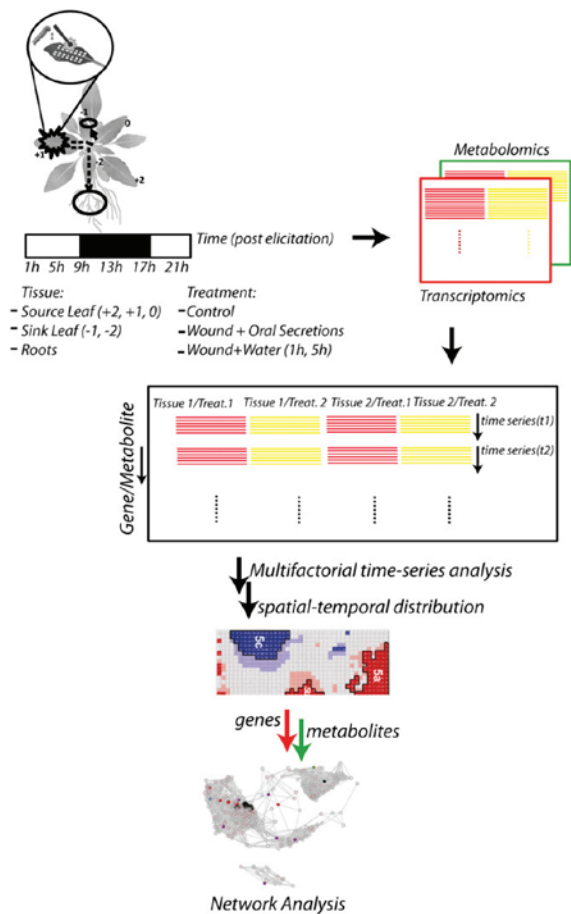


Figure 1. Work flow for analyzing changes in the transcriptome/metabolome landscape of *N. attenuata* elicited by insect herbivory. Leaf and root tissues of wild-type *N. attenuata* plants were harvested 1, 5, 9, 13, 17, and 21 h after leaves were wounded with a fabric pattern wheel and immediately treated with *M. sexta* (W+OS) to study herbivory-induced responses or with water (W+W) to study OS-specific induction. Replicated transcriptomic and metabolomic data were analyzed using multifactorial analysis, with both factors (tissue and treatment) taken together across the time series to identify modules showing differential OS elicitation. These modules were spatiotemporally resolved using BL-SOM. Herbivory-elicited interactive effects among transcript and metabolite levels were analyzed, with a special emphasis on identifying short- and medium-term changes in metabolism. To this end, interactive motifs from resolved maps were extracted and analyzed using network properties. [See online article for color version of this figure.]

attenuata transcriptome and metabolome, which are repeatedly activated during continuous insect feeding, and provides a convenient means of accurately standardizing herbivore elicitation and dependent responses in *N. attenuata* plants (McCloud and Baldwin, 1997). This standardization of herbivore elicitations is critical to conduct replicated microarray and metabolomics analyses. Additionally, this procedure allows

inferring OS-specific responses. For this, we also sampled plants with leaves similarly wounded but in which the puncture wounds were treated with water, referred to as the W+W treatment. Microarray and metabolomics data sets were then preprocessed and normalized using pipelines described in “Materials and Methods.” A pivotal step in this work flow consisted of analyzing processed transcript and metabolite data for two factors using nonparametric ANOVA models in order to generate distinct clusters, separated by their combined treatment effects and time behaviors. Instead of modeling effects of the three factors (time, tissue type, and treatment applied) together, we used time-vector space to find the most significant effects from two-way ANOVA and used this information to conduct explorative studies on tissue- and treatment-specific responses.

Multifactorial Time Response Features of OS-Elicited Gene Expression

The nature and amplitude of herbivore responses in untreated leaves and roots are controlled by short- and long-distance systemic signaling networks (Wu et al., 2007). As response pattern and timing of activation of individual genes likely underlie the differences observed between systemic and localized herbivore responses, we simultaneously analyzed time series with two binary factors (treatment and tissue type) using bootstrap-based nonparametric ANOVA models, according to the methods implemented in the R package TANNOVA. This method has been designed specifically to handle multifactorial microarray experiments with the aim of extracting gene-specific responses across time series based on their dependency on experimental factors used for comparison (Zhou et al., 2010). We conducted such dynamic response analyses for two tissue comparisons: treated (source) leaf versus untreated systemic (sink) leaf (TvS comparison), to explore differential gene expression patterns activated during shoot systemic signaling; and treated leaf versus untreated root (TvR comparison), to obtain novel insights into root-specific responses.

Using a series of statistical tests on factor effects (false discovery rate [FDR] = 0.05, bootstraps = 200), we obtained five mutually exclusive groups of genes showing their best ANOVA structure along optimal directions in the time series (Fig. 2A). The four resulting structures represent interactive (tissues behaving differently in response to OS elicitation across the time series), additive (herbivore responses independent of tissue type), or corresponding main effects (major treatment effects in both treated and untreated tissue or significant differences in tissue type with no response to treatment; Supplemental Fig. S1). Briefly, ANOVA structures with main effects correspond to the sets of genes that are influenced by only one of the factors. The best-fit model for two-way ANOVA with interaction is observed when the effect of level change

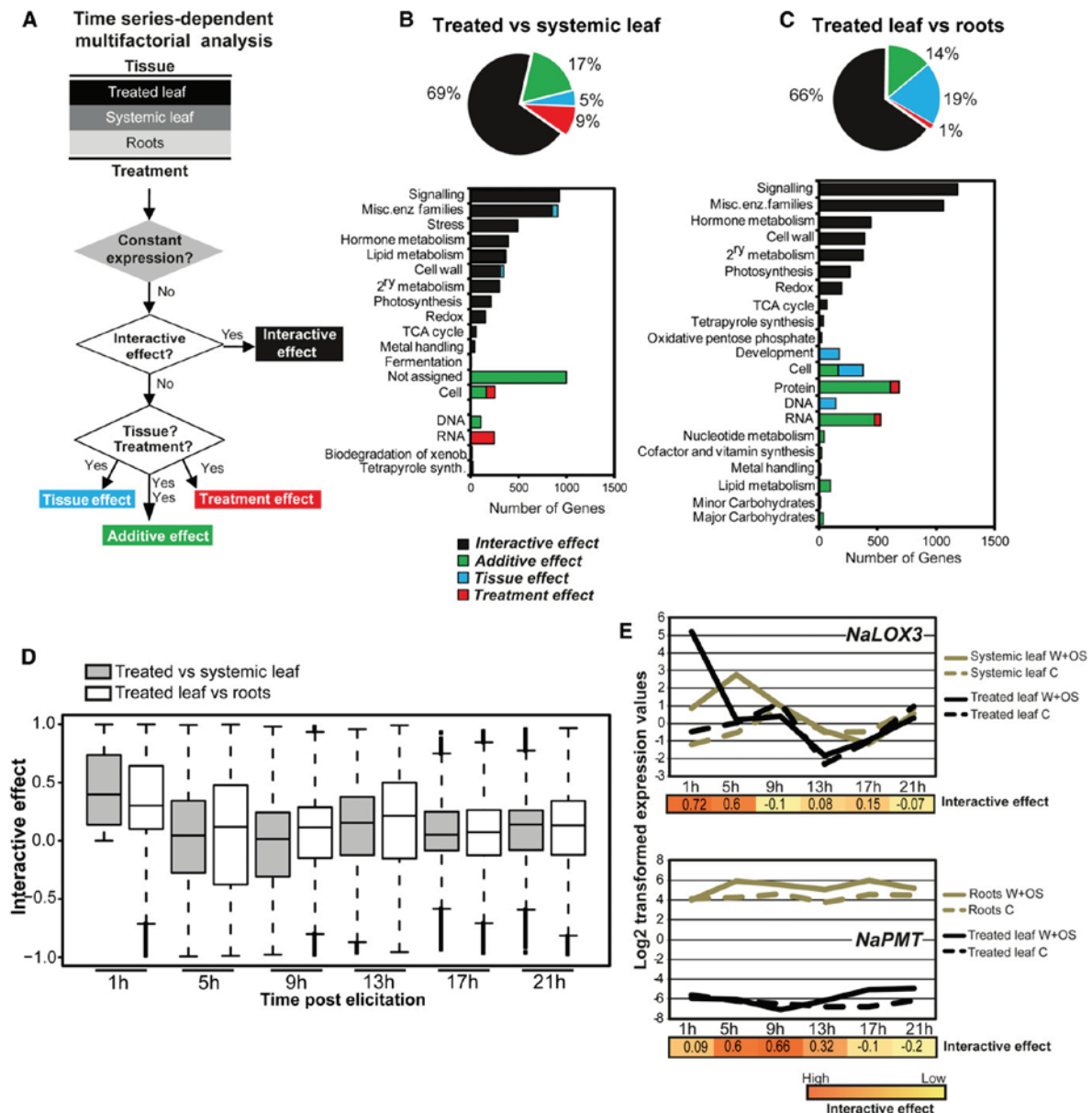


Figure 2. Time series-dependent multifactorial classification of transcripts and encoded molecular processes. A, We employed a bootstrap-based nonparametric ANOVA model to classify transcript levels based on combined significant responses in different tissues and for different treatments across the time series. Transcripts with no significant changes in expression across different conditions were excluded from further analysis. B and C, Functional categorization of transcripts. GO classification was performed for processes with a bin-specific enrichment statistic of $F < 0.05$. TCA, Tricarboxylic acid. D, Box plots of ANOVA directions, described as a time-response metric representing the significant interactive effect along the time series while comparing systemic tissues (unelicited leaves and roots) with the treated leaves. E, Temporal profiles of two representative genes (*NaLOX3* and *NaPMT*) obtained from clusters exhibiting interactive effects between treated and systemic tissues using ANOVA model analysis on transcript time-response behaviors.

for one factor depends on the level of the other factor; therefore, structures resulting from interactive models correspond to the sets of genes with significant

responses to OS elicitation that are different between the two tissues compared. By contrast, the additive model assumes that the effect of level change for one

factor does not depend on the level of other factors and, therefore, includes genes showing both main effects (treatment and tissue differences). The estimated optimal direction was biased toward the time points with strong effects estimated by fitting different models. Figure 2E shows the temporal profiles of representative genes from the sets of genes showing interactive response patterns for the two different comparisons between elicited and unelicited tissues. *NaLOX3*, whose gene product catalyzes the first committed step in JA biosynthesis (Halitschke and Baldwin, 2003), shows a strong interactive effect for the TvS comparison for the first two time points (1 and 5 h after elicitation), while *NaPMT1*, consistent with its requirement for the formation of the pyrrolidine ring of nicotine in roots (Steppuhn et al., 2004), shows a strong interactive effect for the TvR comparison at 5 and 9 h after elicitation. The optimal direction for interactive effects was estimated by finding the first eigenvector of the residuals, obtained by eliminating the main effects, using the following ANOVA model:

$$Y_{ij} = \mu_{ij} + \alpha_{ij} + \beta_{ij} + \gamma_{ij} + \epsilon_{ij}$$

where Y_{ij} is gene expression vector, μ_{ij} is mean expression vector, α_{ij} and β_{ij} are main tissue and treatment effects, and ϵ_{ij} and γ_{ij} are residual and interactive effects between tissue and treatment type.

Genes displaying an interactive response pattern were our major interest and were used for further analysis. The set of genes showing interactive effects between treated and untreated leaf tissues constituted 69% of the total probe set, while those showing interactive effects between treated leaf and untreated root tissue constituted 66% (Supplemental File S1). We analyzed the distribution of interactive effects detected in leaf and root tissues using box plots and observed larger effects at 1 and 5 h for the TvS comparison and at 1, 5, and 13 h after elicitation for the TvR comparison (Fig. 2D). To assess the functional significance of these gene clusters, we computed the enrichment of Gene Ontology (GO) terms for genes within each group defined by the multifactorial analysis using hypergeometric tests ($F < 10e-10$; Fig. 2, B and C). Consistent with previous transcriptomic studies (Hui et al., 2003; Schittko and Baldwin, 2003; Schmidt et al., 2005), the group of genes showing interactive effects in shoots (TvS comparison) was highly enriched for processes associated with signaling, stress responses, hormone metabolism, and secondary metabolism. GO term representations for groups of genes showing additive and treatment effects clearly contrasted with sets of genes with interactive effects. GO terms for these gene classes likely reflect major reconfigurations, coordinated between treated and systemic shoot tissues, of transcriptional and cell cycle machineries. For the TvR comparison, GO terms significantly enriched for interactive effects were from signaling and secondary and hormone metabolism; as expected from the root's highly specialized physiology, we also observed

significant differences between tissue type for a large set of genes with functional terms enriched for development, transcription, and cell cycle machineries.

Biotic and abiotic stresses activate profound reorganizations of the general phenylpropanoid metabolism. Genes in this pathway were among those exhibiting the largest interactive effects in shoots (Supplemental Fig. S2). We evaluated the potential of the time-response metric to improve the detection of coexpression patterns among genes of this pathway compared with the use of average expression data. Supplemental Figure S2 presents Pearson correlation coefficients between *NaPAL1*, characterized for the first committed step in the phenylpropanoid pathway, and downstream functionally characterized genes. More significant coexpression values between functionally related genes were obtained when using the response metric from the factorial analysis, which was specifically biased at time points showing the best ANOVA structure.

Interactive Time Response Analysis Highlights Branch-Specific Functional Organization and Transition Points in the Oxylin Gene Network

Genes involved in a common biological process tend to be coregulated and thus under the control of a shared regulatory system (Saito et al., 2008). Since interactive effect vectors capture as a whole the dynamic response of a gene in more than one tissue, we hypothesized that this information could be used as a metric to assess the differential regulation of members of a gene family based on their functional associations with other known genes. To illustrate the value of this approach in delineating pivotal signaling pathways for herbivory responses, we extracted gene interactive responses within the oxylin pathway and analyzed the activation transition points between the different branches of this pathway (Fig. 3).

Oxylin are bioactive lipids rapidly produced from enzymatic and nonenzymatic fatty oxidation cascades upon various stresses. Lipoxygenases (LOXs) are iron-containing enzymes that catalyze the dioxygenation of fatty acids, the first committed step in oxidation cascades connecting most oxylin biosynthetic pathways. In most plant systems, different LOX isoforms with different tissue expression and cellular compartmentalization control specific branches of the oxylin metabolic network. Three LOX genes responsive to OS elicitation have been functionally characterized in *N. attenuata* (Halitschke and Baldwin, 2003). Activation of *NaLOX1* leads in *N. attenuata* to the production of fatty acid 9-hydroperoxides, which serve as substrates for the formation of fatty acid divinyl ethers by a divinyl ether synthase (Bonaventure et al., 2011). *NaLOX2* and *NaLOX3* control the biogenesis of two distinct chloroplastic linolenic acid 13-hydroperoxide (13-HPOT) pools (Allmann et al., 2010). *NaLOX2*-based 13-HPOTs are cleaved by

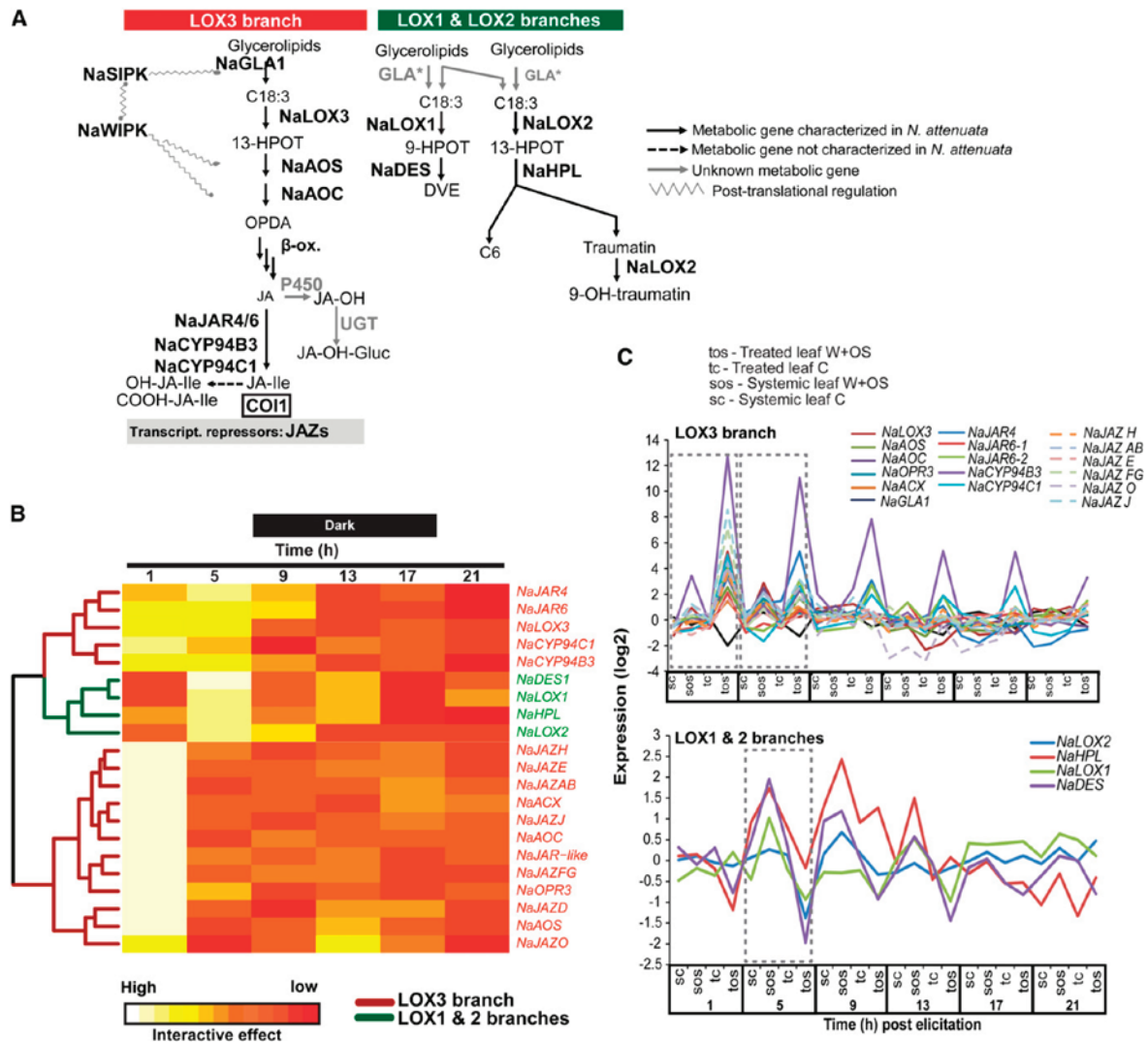


Figure 3. Separation of LOX-dependent branches of the oxylipin pathway based on their ANOVA directions in the time dimension. **A**, Schematic representation of the three main LOX-based branches of the oxylipin pathway. Free linolenic acid released by the induced activity of a glycerolipase (NaGLA1) is oxygenated by NaLOX3 enzymes (LOX) and converted by a multienzymatic cascade into JA. JA-Ile formed after enzyme-dependent conjugation of JA to Ile and is the bioactive jasmonate interacting with the F-box receptor protein NaCOI1 and NaJAZ transcriptional repressor proteins. C12 and C6 aliphatic chains are produced by the cleavage of NaLOX2-dependent hydroperoxides. Divinyl ether (DVE) oxylipins are produced by enzymatic rearrangement of NaLOX1-dependent hydroperoxides. **B**, HCA of ANOVA directions for interactive effects for genes involved in the two different branches of the LOX pathway. Elements of the NaLOX3 branch (red) exhibit differential regulation in leaf tissues 1 h after OS elicitation. NaLOX2 and NaLOX1 branches (green) exhibit high interactive effects 5 h after elicitation. **C**, Temporal profiles for genes from the NaLOX3 pathway showing a high degree of coordination at 1 h and from the NaLOX2 and NaLOX1 pathways showing a high degree of coordination at 5 h.

hydroperoxide lyase enzymes into C12 and C6 aliphatic chains, the latter ones being further metabolized into a large array of green leaf volatiles. NaLOX3-based 13-HPOTs are cyclized by the combined action of allene oxide synthase (NaAOS) and allene oxide cyclase (NaAOC) enzymes to serve as precursors for

the production of JA and its Ile conjugate (JA-Ile; Fig. 3A), the signaling molecules mediating most of the changes in gene expression occurring during OS elicitation (Wang et al., 2008).

Consistent with the functional diversification of these three LOXs, the temporal profiles of NaLOX1,

NaLOX2, and *NaLOX3* and functionally associated genes showed distinct responses to OS elicitation in the two leaf tissues that were compared (Fig. 3B): *NaLOX3* exhibits interactive effects largely at 1 h after elicitation, while *NaLOX1* and *NaLOX2* show significant interactive effects 5 h after elicitation. To assess the organization of LOX-dependent oxylipin branches, we also extracted the time-response patterns for genes that had been functionally associated with each LOX branch and assembled them using hierarchical clustering analysis (HCA) based on Euclidean distance measure (Fig. 3C). HCA revealed clearly distinguishable clusters, each with a high degree of coordination among genes involved in the same LOX branch. Interactive effects for all genes of the LOX3 branch were detected 1 h after elicitation, while downstream elements of the LOX1 and LOX2 branches showed significant interactive effects at 5 h. Interestingly, within the LOX3 branch, interactive responses between leaf tissues were highly transient for most structural genes as well as for *NaJAZ* transcriptional repressors, but *NaLOX3* and genes involved in the formation and catabolism of JA-Ile (*NaJAR4*, *NaJAR6*, *NaCYP94C1*, and *NaCYP94B3*) showed more prolonged interactive effects. A striking observation was the coordination of *NaGLA1* (for glycerolipase1), the lipase controlling wound- and herbivory-induced jasmonate levels in *N. attenuata* leaves (Kallenbach et al., 2010) but not LOX1- and LOX2-dependent oxylipin formation (Bonaventure et al., 2011), with *NaLOX3* and associated genes. This pattern, which has not previously been detected in correlational studies of gene expression data, is consistent with the rapid conversion of fatty acid released during herbivory into LOX3-dependent hydroperoxide derivatives (Kallenbach et al., 2010).

SOM-Based Spatiotemporal Resolution of Interactive Responses Identifies the Sequential Activation of Functional Gene Associations

As seen above, clustering genes using their interactive responses as an associative metric clearly delineates the multiple branches of a biological process. To further resolve the temporal distribution of OS-elicited processes, we supplemented the time-response metrics with additional information from the differences in fold change between the two tissues that were compared in factorial analysis as described in “Materials and Methods.” We then used BL-SOM (Hirai et al., 2004) to impose structure on the scaled data. With grids of size 40×16 and 40×18 for comparisons between treated leaf and systemic tissues (leaves and roots), nodes were mapped into a six-dimensional space, initially based on principal component analysis. On subsequent iterations (200), a data point was selected and its nearest nodes were adjusted by moving toward the selected point, finally converging to distinctive and tight clusters (Fig. 4A). Maps produced for each comparison were colored so that red clusters

represent major up-regulated elements in treated leaves and minor down-regulated elements in untreated systemic tissue. Similarly, blue clusters represent major up-regulation in untreated systemic tissue with minor down-regulated elements in treated leaves. Each cluster, termed an interactive motif, contains a set of nodes with genes showing interactive responses of similar amplitude (Fig. 4B). Since regulation patterns could differ slightly for genes involved in one particular process, functionally associated genes tend to localize not necessarily within the same node but within a common interactive motif. The nodes in these spatiotemporally resolved motifs with similar ANOVA directions along time series may be biologically related, as their dynamic responses are highly coordinated.

Examination of major interactive motifs extracted from the SOM grid underscored fundamental trends in the spatiotemporal activation of OS-elicited transcriptional rewiring. Consistent with the interactive effect concept, gene assemblies according to interactive effect metrics appeared highly tissue specific and greatly plastic over time (Fig. 4A). Thus, we observed few topological overlaps between interactive motifs formed at each time point and in different tissues. For the TvS comparison, we observed a clear reduction in the size of the treated leaf interactive motifs (red clusters) along the time series, suggesting a rapid dampening of local herbivory responses concomitant with an increase in the size of the systemic leaf interactive motifs (blue clusters) starting after 1 h and attaining maximum values 9 h after OS elicitation. A similar trend of dampening of responses in treated leaves was observed for the TvR comparison, except for 13 h after OS elicitation in the middle of the dark phase. Mapping genes known to be regulated by fatty acid-amino acid conjugates (FACs; Gilardoni et al., 2010), well-established elicitors contained in *M. sexta* OS (Halitschke et al., 2001), revealed significant enrichment of these FAC-regulated genes in motifs labeled as 1a, 5a, and 5c for the TvS comparison and r1a, r5a, r1b, and r5b for the TvR comparison, indicating that FAC signaling had a strong contribution to the activation of these interactive motifs. GO enrichment analyses for these motifs (1a, 1b, 5a, 5b, r5a, and r5b) are well represented by processes involved in signaling cascades and secondary metabolism (Supplemental Fig. S4).

Plant tissue-specific responses to insect herbivory are thought to be controlled by tissue-specific activation and interaction between hormone gene networks (Meldau et al., 2012). Consistent with this view, our factorial analysis revealed significant differences in the amplitude and timing of the activation of hormone gene networks between damaged and systemic leaf tissues. Thus, the intensity of the ANOVA signals for genes encoding for JA (maximum at 1 h; Fig. 3), ethylene (1, 5, and 9 h), auxin (1, 5, and 9 h), cytokinin (1 h), and brassinosteroid (1 h) synthesis and connected signaling components was higher in treated leaf tissues (Supplemental Figs. S3 and S4). Additionally, we observed specific sets of genes encoding notably auxin-

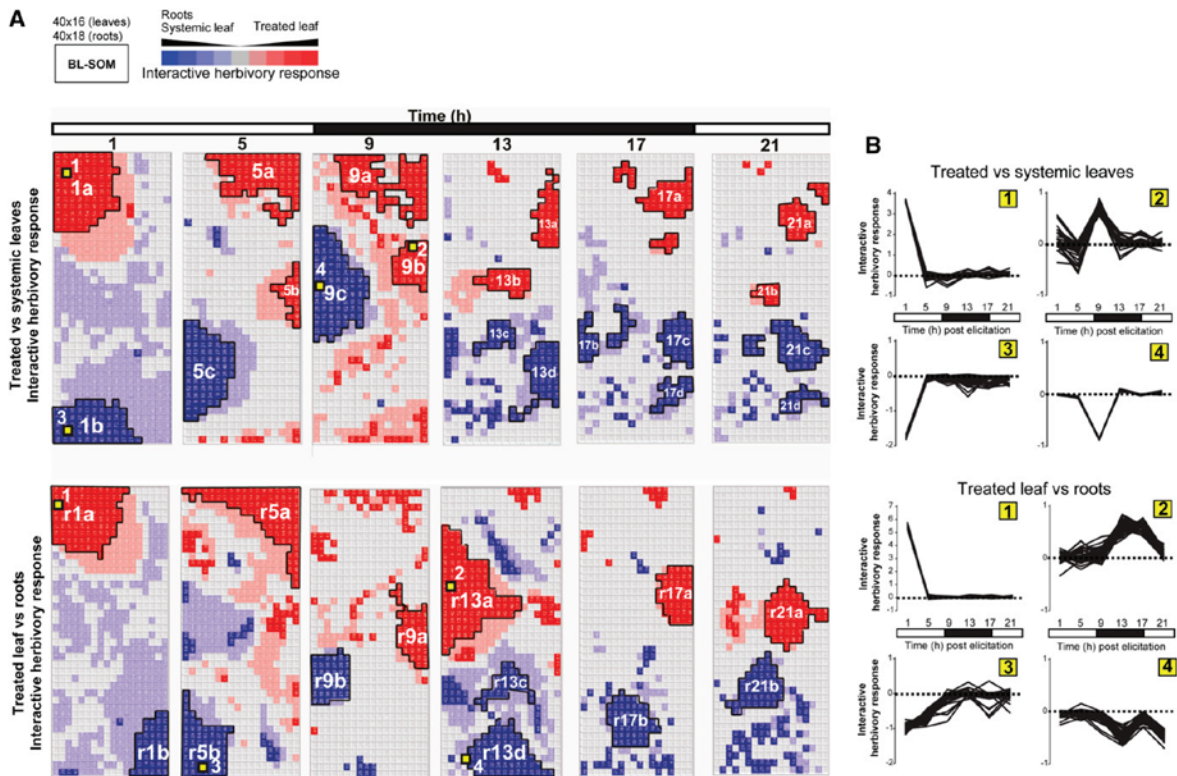


Figure 4. Spatiotemporal resolution of herbivory-regulated interactive motifs in gene expression. A, Time-response metrics, derived from bootstrap-based nonparametric ANOVA models for genes with tissue-specific responses to OS elicitation, were scaled and classified by BL-SOM. Red and pink clusters represent sets of genes showing up-regulation in treated leaves or down-regulation in unelicited tissues (roots or systemic leaves). Blue and pale blue clusters represent sets of genes showing up-regulation in unelicited systemic tissues (roots or systemic leaves) and down-regulation in elicited leaves. B, Each node in the grid represents a cluster of genes with similar amplitude of responses to simulated herbivory, defined as the product of the response metric with fold change differences between tissues that were compared. GO term enrichment for these motifs is presented in Supplemental Figure S4.

and ethylene-related signaling processes that displayed high responses to OS elicitation in untreated systemic leaf tissues (Supplemental Figs. S3 and S4).

Studying OS Responses in Systemic Leaf Tissues Using Interactive Motifs: the Acyclic Diterpene Glycoside Pathway as a Case Study

Coordinated pattern analysis based on dynamic response information from factorial analysis suggested better clustering with fewer false positives and separated components of biological machineries deployed at different times after elicitation. Next, we analyzed gene-gene networks to understand how different biological functions are orchestrated. To understand the performance of entire parts of the OS-elicited transcriptome, we isolated interactive motifs from the SOM grids and analyzed their dynamic behaviors using network analysis.

We used this network-based approach to analyze the regulation of OS-elicited changes in the acyclic

diterpene glycoside pathway. 17-Hydroxygeranylinalool diterpene glycosides (17-HGL-DTGs) are a group of sugar-containing defense metabolites active against a wide spectrum of herbivores (Heiling et al., 2010). The high abundance of these metabolites and their rapid de novo production in systemic tissues during herbivory are controlled by important transcriptional changes that are not completely understood (Heiling et al., 2010). Three experimentally validated genes, *NaDXS*, *NaDXR*, and *NaGGPPS2*, of the nonmevalonate pathway (the 2-C-methyl-D-erythritol 4-P/1-deoxy-D-xylulose 5-P pathway) involved in the biosynthesis of geranylgeranyl pyrophosphate (GGPP), forming the backbone of 17-HGL-DTGs (Jassbi et al., 2008), as well as three other genes (*NaHDR*, *NaHDS*, and *NaISPD*) predicted based on their homology to *Arabidopsis thaliana* isoprenoid genes, were mapped to the SOM grid and found to be localized in the same interactive motif (5c), the largest one showing a huge response to OS elicitation in systemic leaves. GO terms enriched ($F < 10e-10$)

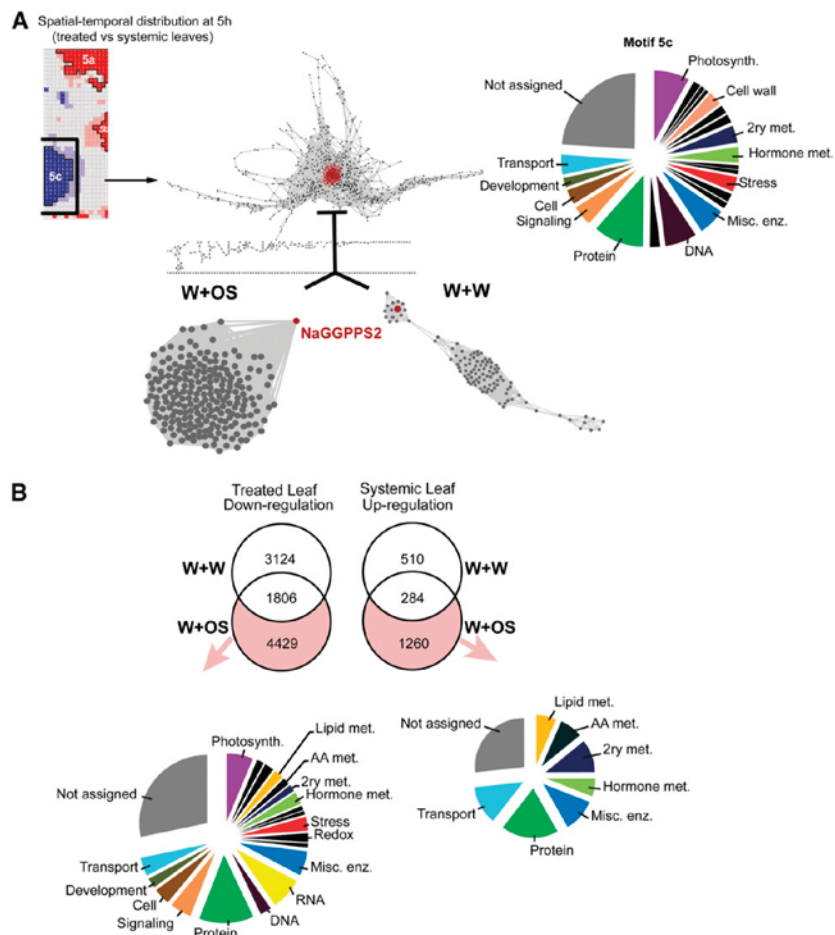
for this motif are overrepresented by those involved in photosystem, hormone metabolism, secondary metabolism, and stress, with a major section of them having unknown functions.

Using a module-centric approach, we extracted this interactive motif to construct a gene-gene network using a statistically sound two-stage coexpression detection algorithm with FDR = 0.05 and a minimal acceptable strength of 0.7 using the GeneNT package in R (Fig. 5A). We found that this network shares properties of other biological networks. The scale-free topology is of 0.96, which suggests that the network is composed of many genes with few connections but a few genes with many connections. Additionally, the clustering coefficient, which provides a measure of cliquishness (0.516), and the measure of network heterogeneity, which reflects the tendency of a network to contain hub nodes (1.40), are within ranges expected for biological networks (Supplemental Fig. S5). Interestingly, we observed the highest degree of connectivity for the six genes of the nonmevalonate pathway (Supplemental Fig. S6), suggesting tight coregulation

between this process and a large array of gene networks activated in untreated systemic leaves or suppressed in treated leaves 5 h after OS elicitation (interactive motif 5c). Network analysis also confirmed differential functional activation of the three *NaGGPPS* genes present in *N. attenuata*, with *NaGGPPS1* showing no significant interactive effect in the TvS comparison while the other two showed large differences in their degree of connectivity in the motif 5c network [$\text{deg}(\text{NaGGPPS2}) = 224$, $\text{deg}(\text{NaGGPPS3}) = 90$].

To infer OS elicitation-specific gene associations, we constructed a partial coexpression network using the first neighbors of *NaGGPPS2* for the W+W treatment from the set of genes analyzed with the same parameters of the factorial analysis and compared it with the subnetwork from the W+OS treatment. A large number of genes uniquely expressed in the W+OS condition failed to show interactive behavior in the W+W treatment, which reduced the number of connections in the W+W subnetwork (Fig. 5A). GO terms overrepresented in the OS elicitation-specific subnetwork include those from the photosynthetic pathway and

Figure 5. OS elicitation selectively activates interactive motif formation in unelicited systemic leaves. A, Distribution of enriched GO for a gene set extracted from motif 5c, which displayed major transcriptional changes in unelicited systemic leaves 5 h after elicitation. Extracted genes were used to construct a gene-gene network representing connectivities screened with FDR = 0.05 and a minimum Pearson correlation of 0.98. A known gene with the highest degree of nodal connectivity (*NaGGPPS2*) is used to illustrate temporal overlap between the two different treatments (W+OS/C and W+W/C). B, Analysis of differentially expressed genes between the control and W+OS for a single time point (5 h after elicitation) for treated and systemic leaves independently. The pie chart represents the distribution of enriched GO terms for both down-regulated pathways in elicited tissues and up-regulated pathways in unelicited systemic tissues. 2ry, Secondary; AA, amino acid; Misc. enz., miscellaneous enzymes; met., metabolism. [See online article for color version of this figure.]



from secondary metabolite and phytohormone metabolism, reinforcing the conclusion that OS-based signaling controls the recruitment and coherent activation of molecular processes otherwise loosely connected.

To further support this conclusion, we conducted a single-time-point differentially expressed genes analysis for 5-h elicitation samples of both tissues independently and detected a large number of genes uniquely up-regulated in systemic leaves and down-regulated in treated leaves (Fig. 5B). As expected, major processes highlighted by this analysis included secondary metabolism and hormone metabolism as up-regulated pathways in untreated systemic leaf tissue and genes from photosynthetic pathways as down-regulated processes in treated leaf tissue. The biosynthesis of plastid isoprenoids by the nonmevalonate is essential for photosynthesis and chloroplast function (Vranová et al., 2012). Dense connectivities were shared between the nonmevalonate/17-HGL-DTG pathways and photosynthetic genes, suggesting yet-unknown coordination mechanisms between these two processes and consistent with the previously reported transcriptional down-regulation of photosynthesis (Halitschke et al., 2001; Hui et al., 2003; Mitra and Baldwin, 2008) as an integral part of the mechanisms facilitating defense compound production.

Factorial Analysis Reveals OS-Elicited Metabolic Changes with Diverse Response Patterns

To facilitate the interpretation of the large transcriptomic differences observed between elicited and unelicited tissues, we profiled downstream metabolite responses using a broadly targeted UHPLC-qTOFMS metabolomic approach (Gaquerel et al., 2010). Often in such large-scale profiling, several compounds are not completely chromatographically resolved or are observed with shifts in their retention times, so mathematical procedures involving deconvolution and retention time corrections need to be applied to extract accurate mass spectra with resolved chromatographic peaks for further comparisons. These procedures require samples with somewhat similar metabolomes; therefore, we only processed treated leaf and untreated systemic leaf tissues for all time points, and the two conditions together for both positive and negative modes, using the XCMS package in R.

The first stage in the identification of differences between local and systemic responses is the projection of the ANOVA structure obtained by fitting the same ANOVA models used in the transcriptomic analysis. The experimental factors used were treatment (control and W+OS) and tissue type (treated and untreated leaves). As with the transcriptomic analysis and using the same parameters (FDR = 0.05, bootstraps = 200), the sampling time was not taken as another factor but was used to find the best ANOVA structure along the optimal direction resulting in response metrics. In addition to the induced changes in responses to OS (treatment effect), a set of metabolites of interest also

showed small differences between the two leaf tissues (tissue effect); hence, they were classified in an additive bin. Therefore, in order to increase metabolite coverage, we combined both bins of ions showing interactive and additive effects for further analysis. A total of 19.2% (1,610 of 8,357) of mass-to-charge ratio (m/z) signals from the positive mode and 26.9% (720 of 2,676) of m/z signals from the negative mode showed major differences in induced responses between leaf tissues for at least one time point in the series (Supplemental Files S2 and S3). This is severalfold fewer than the transcriptomic differences observed between the responses of the two leaf tissues. One reason for this lower coverage could be the use of a targeted approach for profiling major classes of secondary metabolites; moreover, some metabolites showed large constitutive biological variation, which thwarted the detection of significant differences between control and W+OS conditions with FDR = 0.05. For further analysis, we also combined positive and negative mode m/z lists. We extracted and plotted response metrics obtained from the factorial analysis for both bins and observed larger differences between elicited and unelicited leaves at 9 and 13 h after elicitation (Fig. 6A, box plots). Finally, we scaled the interactive response metrics with additional data of the fold change differences between elicited and unelicited leaf tissue, as described in “Materials and Methods,” and visualized these differences using HCA. We observed a large number of small clusters showing very diverse patterns across combinations of time points after elicitation. Larger accumulation of metabolites in systemic tissues was observed at 9, 13, and 17 h after elicitation, reflected by the red part of the color scheme used in the HCA. This difference in the timing of activation of transcriptomic changes in systemic tissues (1 and 5 h after elicitation; Fig. 4, blue motifs 1b and 5b) and metabolic changes (Fig. 6A) was likely due to a time lag in the biosynthesis of metabolites compared with underlying gene expression or to the transport of metabolites from treated to untreated leaf tissues.

To identify major classes of compounds showing larger differences between the two leaf tissues, we included the additional information of retention time and mapped interactive effects onto the chromatographic scale. Figure 6B highlights three main regions of the chromatogram showing differential OS responses in treated and systemic leaves at 9 h after elicitation for both positive and negative modes, corresponding to phenolic derivatives, 17-HGL-DTGs, and *O*-acyl sugars, well-studied classes of defense metabolites. Next, we ordered ions by their retention time and visualized the scaled data obtained from factorial analysis using heat maps. Since we reduced the number of ions based on the ANOVA model for their differential behavior between the two tissues, we created bins of retention times for these selected ions to match the chromatogram for visualization. To further illustrate the application of this approach, we focused on responses detected at 9 h after OS elicitation. This

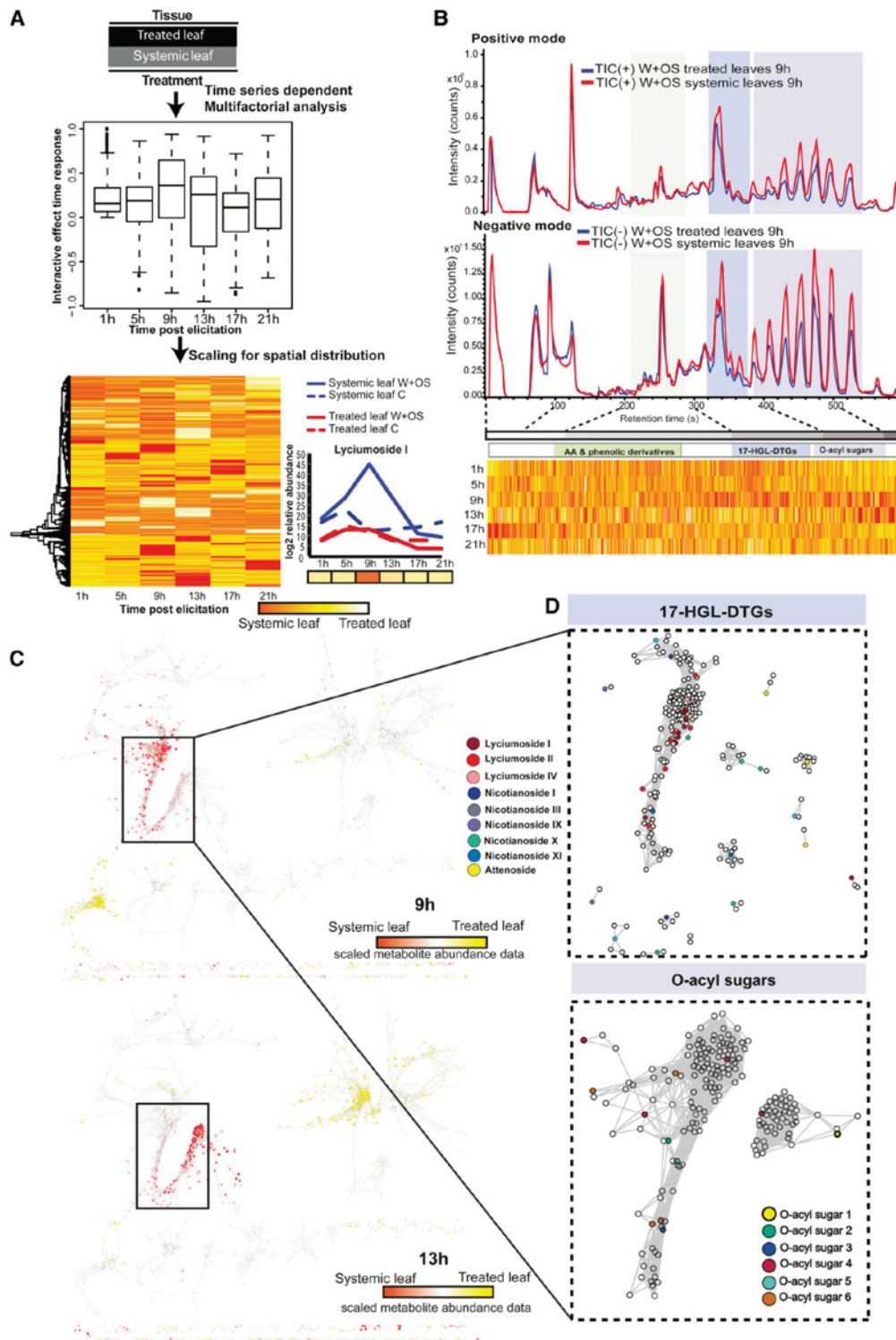


Figure 6. Factorial analysis revealed OS-elicited large-scale metabolic changes with diverse response patterns across the time series. **A**, Elicited and unelicited leaf tissues were compared for control and W+OS treatments using multifactorial analysis. Box plots represent the distribution of interactive response metrics across the time series. Ions with interactive effects were scaled and used for HCA. **B**, Chromatograms for treated leaf and untreated systemic leaf tissues obtained 9 h after W+OS elicitation

exploratory analysis covers the spanned red regions across the classes of 17-HGL-DTGs and *O*-acyl sugars for 9 h after elicitation, which suggests tandem induction and/or metabolic cross talk among several of the metabolic pathways affected by OS elicitation in systemic tissues. To understand such concerted increases in metabolites, we constructed metabolite networks based on associations among metabolites (Pearson correlation ≥ 0.98 ; Fig. 6C). We mapped scaled data on the network to visualize distinctly occupied branches of the network in which the red regions spanned, containing different sets of ions with high OS responses in systemic leaves at 9 and 13 h. We isolated a branch enriched with *m/z* signals showing induction in systemic leaves at 9 h and detected the presence of different known 17-HGL-DTGs; interestingly, in the same network branch, we also found several *m/z* signals corresponding to *O*-acyl sugars (Supplemental Fig. S7). Consistent with the power of this approach to cluster biochemically connected metabolites, we additionally observed highly coordinated dynamic responses between shikimate pathway-derived amino acids and downstream metabolites produced within the phenylpropanoid pathway (Supplemental Fig. S2).

Time-Lag-Corrected Correlation on the Interactive Response Metric Supports the Reconstruction of Gene-to-Metabolite Networks

The nature of the coregulation of functionally related genes and metabolites could vary depending on the biological activity of the studied metabolic class and the experimental conditions applied. The patterns of correlation between metabolite and transcript data have been successfully analyzed in a few studies using high-resolution time-course analyses (Hirai et al., 2004), but most studies in this field, including our work (Fig. 7), have reported significant differences in the temporal dynamics of transcriptomes and metabolomes (Walther et al., 2010; Takahashi et al., 2011). Considering the complex interdependencies between metabolites and transcripts, we sought to detect and interpret gene-metabolite interactions at the level of isolated pathways for TvS comparisons using the interactive responses of genes and metabolites as associative metrics and used this approach to study previously reviewed oxylipin and 17-HGL-DTG pathways.

Although a tight correlation pattern is usually not expected for metabolites that are rapidly consumed by subsequent reactions, as is the case for jasmonate production, we observed that genes of the LOX3 pathway correlated well with JA levels as well as associated metabolic intermediates, based on the interactive

response patterns observed while comparing two leaf tissues together (Fig. 2; Supplemental Fig. S8). We observed a fundamentally different picture for the 17-HGL-DTG pathway. Figure 7A shows expression profiles with their time-response metric for interactive effects in a color-coded scheme of known genes and metabolites of this pathway. A comparison of averaged interactive responses suggests a shift in the observed interactive effect behavior of 17-HGL-DTG metabolites compared with that of their underlying genes. Therefore, to address this time lag, we applied lagged Pearson correlation (PC), estimated as follows:

$$PC(G, M) = PC\{(g_1, \dots, g_{T-1}), (m_2, \dots, m_T)\}$$

Where (g_1, \dots, g_{T-1}) indicates gene expression from 1 h onward and (m_2, \dots, m_T) indicates metabolite level from 5 h onward.

We constructed a gene-to-metabolite network using lagged time-specific data with Pearson correlation ≥ 0.98 . Figure 7B presents a magnification of the network region containing the first neighbors of the six known genes of the nonmevalonate/17-HGL-DTG pathways in strong associations with *m/z* signals derived from 17-HGL-DTGs. We observed strong coordination between 17-HGL-DTGs, *O*-acyl sugars, genes from the nonmevalonate/17-HGL-DTG pathways, and photosynthetic genes. Unknown *m/z* signals reported in this network represent potential molecular fragments of the aforementioned compounds that will need to be confirmed by additional mass spectrometry-based fragmentation analysis.

DISCUSSION

The comprehensive classification of leaf and root herbivory-activated gene expression levels presented here significantly expands our knowledge of the multidimensionality of herbivory-activated transcriptional and metabolic reprogramming. The use of SOM on interactive effect-response metrics derived from factorial analysis, instead of raw expression signatures, enabled a reconstruction of gene-gene, metabolite-metabolite, and gene-metabolite associations with a high degree of predictive power. Spatiotemporal maps produced in this study not only underscore the high plasticity of OS elicitation responses but also provide a powerful data platform for the functional genomics of novel regulatory and structural genes involved in antiherbivory processes. The two metabolic routes investigated in greater detail in this study, oxylipin signaling and 17-HGL-DTG pathways, illustrate fundamental working aspects of gene-metabolite networks in terms of activation and

Figure 6. (Continued.)

and aligned with heat maps representing scaled interactive effect data ordered by retention time. C, Visualization of systemic responses occurring 9 and 13 h after elicitation localized on different branches of the metabolite-metabolite network. D, Partial network representations of the first neighbors of known 17-HGL-DTG and *O*-acyl sugar *m/z* signals extracted from the same branch of the network. AA, amino acid; TIC, total ion current.

Gulati et al.

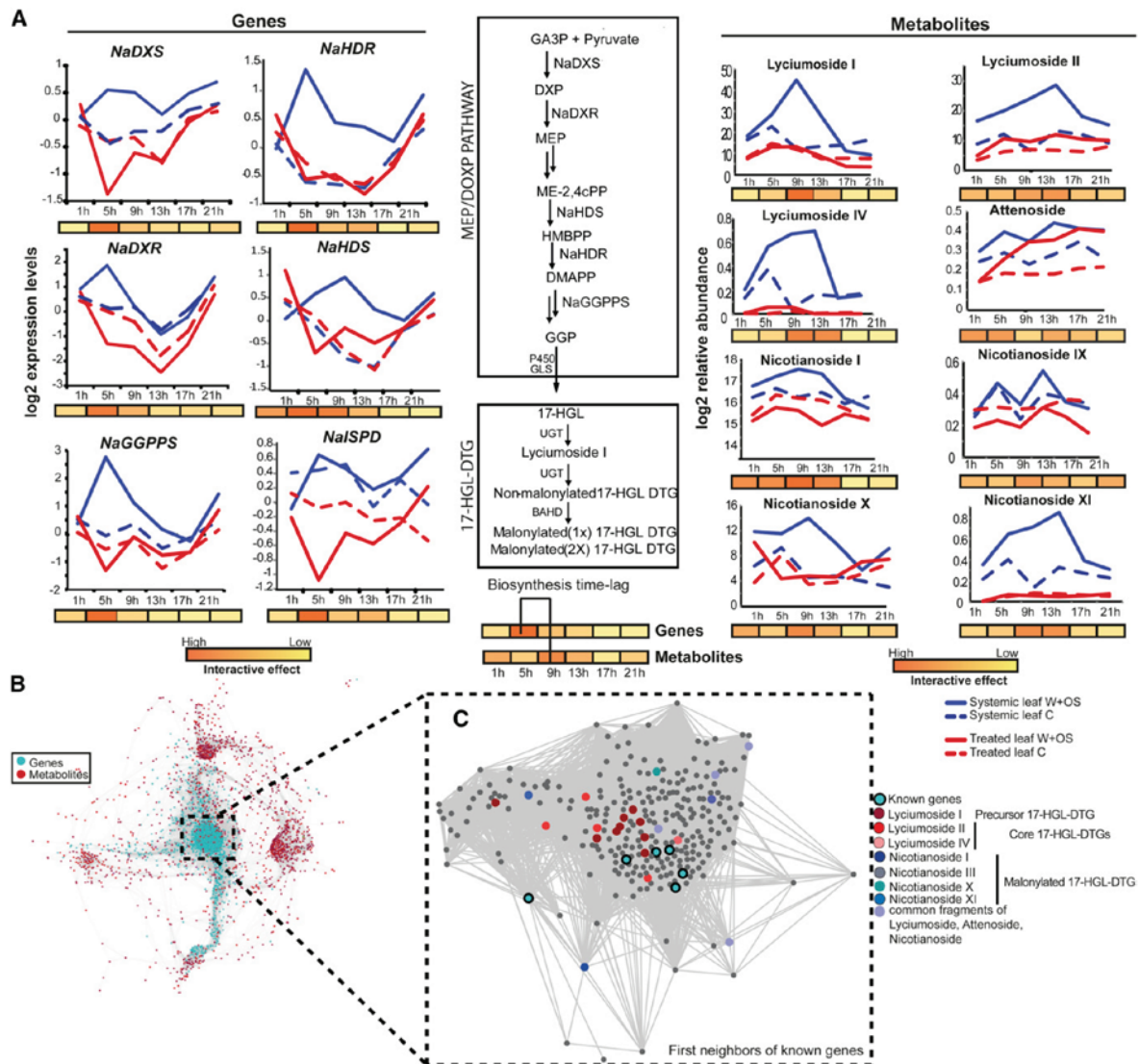


Figure 7. Gene-metabolite associations underlying the formation of 17-HGL-DTG in *N. attenuata*. **A**, Temporal profiles for identified acyclic 17-HGL-DTG and known hub genes in the biosynthetic network of these metabolites, with the schematic of the diterpene glycoside synthetic pathway. **B**, Network constructed after time-lagged correlation on the interactive patterns of the module representing major transcriptomic changes in systemic leaves 5 h after elicitation and the entire set of metabolites classified for differential changes in treated and systemic leaves in response to simulated herbivory. **C**, Partial coexpression network representing the first neighbors of identified 17-HGL-DTG in *N. attenuata*. The cluster represents the strong connectivity between 17-HGL-DTG metabolites and known biosynthetic genes in *N. attenuata*. Other previously characterized *m/z* features present in this cluster are enriched for *O*-acyl sugar-associated signals. BAHD, acyltransferase; GLS, geranylinalool synthase; MEP/DOXP, non-mevalonate pathway or 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway; UGT, UDP-glucosyltransferase.

transition points during herbivory and shared patterns of regulation with other physiological processes.

Elucidating groups of genes involved in the sequential reorganization of biological networks is extremely challenging using the available bioinformatics

approaches, but we successfully studied such a reorganization by systematically combining information about when to respond (time points) with information about how to respond (fold changes). This method could be used to query different gene families of

interest as input or to discover regulatory motifs for each separated module from the SOM grid. The applicability of this method can be further extended to study diurnal rhythms perturbed in response to herbivore attack (Kim et al., 2011) by considering the ordering of time points before deriving the optimal direction for the best ANOVA structure. Here, we demonstrate that SOM-based herbivory-elicited interactive motifs are highly tissue specific, highly versatile, and decrease in size for later time points, reflecting specific sequential responses deployed in different tissues. We speculate that this increased modularity in time reflects the functional specialization of the sequential responses deployed in the different tissues. These observations are in line with smaller scale transcriptional screens published by several groups indicating that insect herbivory results in major shifts in almost every aspect of a plant's physiology (Hermsmeider et al., 2001; Schmidt et al., 2001). The fact that large-sized interactive motifs are more pronounced in systemic tissues at 5, 9, and 13 h after elicitation indicates the involvement of organismic-level responses specifically triggered by OS-based signaling, with a large proportion of these also attributable to unexplored root-specific processes. Previous work has shown that the elicitation in *N. attenuata* shoots of defense-related transcriptional and metabolic programs relies on the rapid translocation, specifically triggered by OS perception, of signaling molecules from the sites damaged by insects to systemic undamaged tissues (Wu et al., 2007). Jasmonate signaling through the plant vasculature has major functions in this process and is controlled by rapid transcriptional and posttranscriptional changes (Wu and Baldwin, 2010).

Analysis of interactive effects and response timings for jasmonates and oxylipin genes illustrates that the extraction of multifactorial statistical information in terms of response patterns not only facilitates the clustering of genes involved in similar biochemical pathways, as also shown for the phenylpropanoid (Supplemental Fig. S2) and 17-HGL-DTG pathways (Figs. 6 and 7), but also the identification of transition points in the activation of certain pathways (Fig. 3). Most structural genes involved in the biogenesis of jasmonates (*NaAOS*, *NaAOC*, and β -oxidation components), associated upstream signaling components such as mitogen-activated protein kinase (*NaSIPK* and *NaWIPK*), posttranslationally activating linolenic acid release and hydroperoxidation by *NaLOX3* (Kallenbach et al., 2010), and downstream signaling components such as transcriptional repressors (*NaJAZs*) and *NaMYC2*, show transient interactive responses. In clear contrast, metabolic systems involved in the initiation of JA synthesis (*NaLOX3*), its conjugation to form bioactive JA-Ile (*NaJAR4/NaJAR6*), and the catabolism of this signaling molecule (*NaCYP94C1* and *NaCYP94B3*; Koo et al., 2011; Heitz et al., 2012) exhibited a prolonged interactive effect. Experimental work supported by mathematical modeling (Banerjee and Bose, 2011) has demonstrated that the transient nature of the jasmonate

pulse and most of the underlying gene expression is an outcome of two opposing influences. Jasmonate accumulation activates a positive feedback loop in which the expression of jasmonate biosynthesis genes is activated after degradation of a specific set of JAZ proteins inhibiting MYC2 transcriptional activity (Chung et al., 2008). This positive loop is counterbalanced by the MYC2-dependent increased expression of JAZs repressing jasmonate signaling transduction (Chini et al., 2007). Prolonged interactive responses for certain genes that contrast with the pulse effects, however, are consistent with the existence, as suggested by different experimental efforts (Howe et al., 2000; Strassner et al., 2002), of additional regulatory loops modulating transition points in pathway activation.

To analyze the metabolic output of these large transcriptional changes, we also profiled metabolites for the same experimental conditions and tissue types (Fig. 6). Even though instrumental in highlighting the breadth of the complete metabolic profile being affected by the treatment, the mining of these data are extremely challenging, notably due to compound-specific in-source fragmentation resulting in redundant signal detection (Dettmer et al., 2007) and also because of the scarcity of methods available for analyzing such complex factorial time-series data. Here, we demonstrate the advantage of constructing a dynamic correlation network based on response features captured by the factorial analysis. This conclusion is based on the inference of compound-family-wise grouping of *m/z* signals exhibiting similar interactive effects (e.g. for 17-HGL-DTG and O-acyl sugar fragmentation patterns) and on the detection of a significant coexpression, after time-lag correction of interactive effect metrics, of structural genes and metabolites of the 17-HGL-DTG pathway (Fig. 7). To our knowledge, this analysis represents the first successful example of spatiotemporal categorization on a pathway scale of induced changes in plant transcriptomics and metabolomics data.

The analysis of OS-specific interactive motifs activated in systemic tissues as well as network reconstructions placed known genetic elements of the nonmevalonate pathway at the center of the largest gene network formed in systemic leaf tissues following OS elicitation (Fig. 5). This is consistent with the role of the nonmevalonate pathway as a central metabolic provider for the production of structurally diverse plastidic isoprenoids such as carotenoids, phytol (a side chain of chlorophylls), isoprene, and monoterpenes and diterpenes (Vranová et al., 2012). In *N. attenuata*, this plastidial metabolic route also supplies the production of 17-HGL-DTG, an important defensive molecule increasing in concentration during insect herbivory (Heiling et al., 2010). The organization of the molecular processes within this OS-specific gene network supports known metabolic interactions between the nonmevalonate pathway, photosynthesis, and chloroplast functions (Vranová et al., 2012) and, more importantly, highlights that the induction of 17-HGL-DTG structural genes represents one of the major

reorientations of the gene network dynamics triggered by OS elicitation. The de novo production of 17-HGL-DTGs during OS elicitation documented in this study relies on the biogenesis of the GGPP backbone by a specific GGPPS, *NaGGPPS2* (Jassbi et al., 2008). Here, we show that expression of the gene coding for this protein is part of the hub region of a gene network, while the two other *NaGGPPS* copies (*NaGGPPS1* and *NaGGPPS3*) exhibit reduced connectivity with this hub region. We predicted the subcellular location of these three proteins based on the presence of chloroplast transit peptide, mitochondrial targeting peptide, or secretory pathway signal peptide (Emanuelsson et al., 2007). Therefore, *NaGGPPS2* and *NaGGPPS3* with high chloroplast transit peptide scores (0.369 and 0.457) are most likely localized in the chloroplast, while *NaGGPPS1* showed insignificant scores for chloroplast transit peptide, mitochondrial targeting peptide, and secretory pathway signal peptide. These predictions further support our findings on the differential activation of these three GGPPS genes in response to OS elicitation.

The ability to reconstruct tissue-specific gene-to-metabolite dynamics with high confidence underscores the importance of extracted time-response patterns for features of interest, which single-point analysis or pooled data using time series as a third factor would fail to identify (Supplemental Fig. S6). This novel method of combining multifactorial analysis with the information extracted from time-series data is broadly applicable to investigate signaling/metabolic pathways in other biological systems with time-series data to deduce the activation times of particular response elements required for understanding complex physiological processes. This strategy can further be used for improving clustering analyses by applying a dimension reductionist approach, for different cell/tissue types, developmental stages, or genetic backgrounds, based on the time-response metric.

MATERIALS AND METHODS

Plant Material

Wild-type *Nicotiana attenuata* from an inbred line in its 30th generation was used for all gene expression and metabolite profiling experiments. All seeds were first sterilized and incubated with diluted smoke and 0.1 M GA₃, as described (Krugel et al., 2002), and then germinated on plates containing Gamborg B5 medium. Ten-day-old seedlings were transferred to small pots (TEKU JP 3050 104 pots; Pöppelmann) with Klasmann plug soil (Klasmann-Deilmann), and after 10 d, seedlings were transferred to 1-L pots with sand to facilitate the sampling of roots. Plants were grown in the glasshouse with a day/night cycle of 16 h (26°C–28°C)/8 h (22°C–24°C) under supplemental light from Master Sun-T PIA Agro 400 or Master Sun-TPIA Plus 600-W sodium lights (Philips Sun-T Agro).

Plant Treatment Experimental Design

A scheme for plant treatments and the work flow for data collection/analysis are shown in Figure 1. Briefly, plant treatments (no treatment, mechanical wounding, and simulation of *Manduca sexta* feeding) were randomly applied to plants in the early elongating stage of growth. To simulate *M. sexta* feeding, the laminae of three leaves per plant (two source leaves at nodes +2 and +1 and one source-sink transition leaf at node 0) were mechanically wounded with a fabric pattern wheel on both sides of the midrib, and immediately, 20 μ L of *M. sexta* OS

(diluted 1:10 in water) was applied to the fresh wounds (W+OS). This procedure, which is referred to as OS elicitation, has been shown to recapitulate most of the reconfigurations occurring in the leaf transcriptome, proteome, and metabolome during *M. sexta* feeding (for review, see Wu et al., 2010). *M. sexta* OS were collected from third to fourth instar larvae reared on *N. attenuata* wild-type leaves as described (Roda et al., 2004), and eggs of this insect species were obtained from North Carolina State University. Responses inherent to the mechanical damage were monitored by applying 20 μ L of deionized water onto the wounds (W+W). For each time point (1, 5, 9, 13, 17, and 21 h after treatment), treated leaves or control ones at the same nodal positions, systemic leaves (two sink leaves at nodes –1 and –2), and the complete root system were collected from six plants and immediately flash frozen in liquid nitrogen. Roots were washed in a water tank for a few seconds to remove sand.

Analysis of Leaf Metabolites by UHPLC-ESI/qTOFMS

Metabolites were extracted from all local and systemic leaf samples. We used a 40% methanol extraction procedure optimized for the recovery of a wide range of metabolites of interest in *N. attenuata* (Gaquerel et al., 2010). Approximately 100-mg aliquots of liquid nitrogen-ground leaf powder were extracted by adding 1 mL of acidified 40% methanol prepared with 0.5% acetic acid water to each sample in 2-mL microcentrifuge tubes with metal balls. The samples were homogenized in a ball mill (Genogrinder 2000; SPEX CertiPrep) for 45 s at 250 strokes min^{–1}. Homogenized samples were then centrifuged at 16,000g and 4°C for 30 min, and supernatants were transferred into 1.5-mL microcentrifuge tubes. Two microliters of the extracts prepared as above was separated using a Dionex rapid separation liquid chromatography system equipped with a Dionex Acclaim 2.2- μ m, 120-Å, 2.1- \times 150-mm column, applying the following binary gradient: 0 to 6 min, isocratic 70% A (deionized water, 0.1% [v/v] acetonitrile [Baker; HPLC grade], and 0.05% formic acid), 30% B (acetonitrile and 0.05% formic acid); 6 to 13 min, isocratic 20% A, 80% B and 70% A, 30% B; 13 to 18 min, isocratic 70% A, 30% B. Flow rate was 200 μ L min^{–1}. Eluted compounds were detected by a MicroToF quadrupole time-of-flight mass spectrometer (qTOFMS; Bruker Daltonics) equipped with an electrospray ionization source. To maximize metabolome coverage, mass spectral detection was performed in both positive and negative ionization modes. Typical instrument settings were as follows: capillary voltage, 4,500 V; capillary exit, 130 V; dry gas temperature, 200°C; and dry gas flow, 8 L min^{–1}. *m/z* values were detected within a range from *m/z* 200 to 1,400 at a repetition rate of 1 Hz. Mass calibration was performed using sodium formate clusters (10 mM solution of NaOH in 50:50 [v/v] isopropanol:water containing 0.2% formic acid).

Microarray Data Analysis

Three biological replicates from six harvested replicates were used for RNA isolation. Total RNA was isolated with TRIZOL reagent and labeled copy RNA with the Quick Amp labeling kit (Agilent). Each sample was hybridized on Agilent single-color technology arrays (4 \times 44K 60-mer oligonucleotide microarray designed for *N. attenuata* transcriptome analysis; <http://www.agilent.com>; accession no. GPL13527). All microarray data with each probe name were deposited in the National Center for Biotechnology Information Gene Expression Omnibus database (accession no. GSE30287). Raw intensities were log₂ and baseline transformed and normalized to their 75th percentile using the R software package, prior to statistical analysis.

Processing of Metabolomics Data

Raw data files were converted to netCDF format using the export function of the Data Analysis version 4.0 software (Bruker Daltonics) and processed using the XCMS package in R (Smith et al., 2006). Peak detection was performed using the centwave algorithm with the following parameter settings: ppm = 20, snthresh = 10, peakwidth = c(5,18). Retention time correction was accomplished using the XCMS retcor function with the following parameter settings: mzwid = 0.01, minfrac = 0.5, bw = 3. Areas of missing features were estimated using the fill-Peaks method. Annotation of compound spectra derived from in-source fragmentation during ionization and corresponding ion species was performed with the BioConductor package CAMERA (version 1.9.8; Kuhl et al., 2012). Compound spectra were built with CAMERA according to the retention time similarity, the presence of detected isotopic patterns, and the strength of the correlation values among extracted ion currents of coeluting *m/z* features. CAMERA grouping and correlation methods were used with default parameters. Clustered features were annotated based on the match (\pm 5 ppm) of

calculated m/z differences versus an ion species and neutral loss transitions rule set (Supplemental File S2). Consistent mass features that were at least present in four out of the six biological replicates with a retention time greater than 1 min were considered for further analysis. Zero values, which remained after application of the “filling in” function in XCMS, were replaced by one-half of the minimum positive value of the ion across all time points and conditions in the original data. Raw intensity values were 75th percentile normalized before statistical analysis. Metabolite fragmentation patterns were annotated as described in Supplemental Materials and Methods S1.

Statistical Analysis and Data Visualization

Multifactorial analysis was carried out using the methods implemented in the R package TANOVA (Zhou and Wong, 2011). Genes and ions were filtered by fitting the following models in sequential order for identifying nonconstantly expressed elements, elements showing interactive effects, and elements showing major effects of tissue or treatment type:

$$Y_{ij} = \mu_{ij} + \epsilon_{ij} \quad (1)$$

$$Y_{ij} = \mu_{ij} + \alpha_{ij} + \beta_{ij} + \gamma_{ij} + \epsilon_{ij} \quad (2)$$

$$Y_{ij} = \mu_{ij} + \alpha_{ij} + \epsilon_{ij} \quad (3)$$

$$Y_{ij} = \mu_{ij} + \beta_{ij} + \epsilon_{ij} \quad (4)$$

where Y_{ij} is gene expression vector, μ_{ij} is mean expression vector, α_{ij} and β_{ij} are main tissue and treatment effects, and ϵ_{ij} and γ_{ij} are residual and interactive effects between tissue and treatment type. To mine the major biological processes perturbed in response to OS elicitation, we functionally annotated probe sets using the best BLASTX hit of The Arabidopsis Information Resource 6 proteome with an e-value cutoff of $1e-15$. Next, using MAPMAN classification of biological processes for Arabidopsis (*Arabidopsis thaliana*), we assigned classes to each probe identifier of our microarray data set. Enrichment analysis of GO biological processes based on a hypergeometric test was performed using R.

Significant enrichments were those with $F < 10e-10$. For spatial categorization, we applied the following scaling method for time-response metrics obtained from factorial analysis for genes and m/z signals, filtered for their interactive effect:

$$E_i = [F_1 \dots F_6]_i * [R_1 \dots R_6]_i^2$$

where E_i is scaled expression, F_i is difference in fold changes (OS elicitation/control) between treated and untreated tissues, and R_i is response timing.

BL-SOM for transcriptomic data were constructed using BL-SOM software (http://prime.psc.riken.jp/?action=blsom_index) with X coordinates sized 40. Networks representing associations between genes and metabolites were visualized with Cytoscape (<http://www.cytoscape.org/>) using organic layouts. Topological properties of the networks were analyzed using the NetworkAnalyzer plugin in the Cytoscape software. HCA for all heat maps is based on Euclidean distance measures and average linkage aggregation methods. All heat maps and box plots were created using R.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Examples of the structures obtained (interactive, additive, and major effects) by fitting ANOVA model.

Supplemental Figure S2. Interactive effect responses of phenylpropanoid pathway genes are highly coordinated.

Supplemental Figure S3. SOM-based classification of FAC-responsive and hormone and secondary metabolism genes in leaves.

Supplemental Figure S4. Enriched GO terms for a few important motifs extracted from SOM analysis unravels large OS-specific gene expression responses in treated leaves and untreated systemic tissues (leaves and roots).

Supplemental Figure S5. Power law distribution plot of the network representation obtained for genes extracted from motif 5c, showing

many genes with few connections but a small set of genes with many connections.

Supplemental Figure S6. Gene-gene Pearson correlation for six genes from the nonmevalonate pathway using five different metrics.

Supplemental Figure S7. Temporal profiles for m/z signals corresponding to the class O-acyl sugars exhibit high correlation and colocalize with those of 17-HGL-DTG in the network.

Supplemental Figure S8. Temporal profiles of *NaLOX3*, functionally characterized for the production of JA and m/z signal of JA.

Supplemental File S1. Multifactorial analysis and BL-SOM results for gene expression data.

Supplemental File S2. UHPLC-TOFMS metabolomic positive and negative mode data.

Supplemental File S3. Multifactorial analysis results for metabolomic data.

Supplemental Materials and Methods S1. Identification of metabolites from UHPLC-TOFMS data.

ACKNOWLEDGMENTS

We thank Eva Roth and Felipe Yon for their help with sample preparation as well as Dr. Matthias Schöttner, Dr. Klaus Gase, and Wibke Kroker for technical assistance with metabolomics and microarray analyses.

Received March 12, 2013; accepted April 25, 2013; published May 8, 2013.

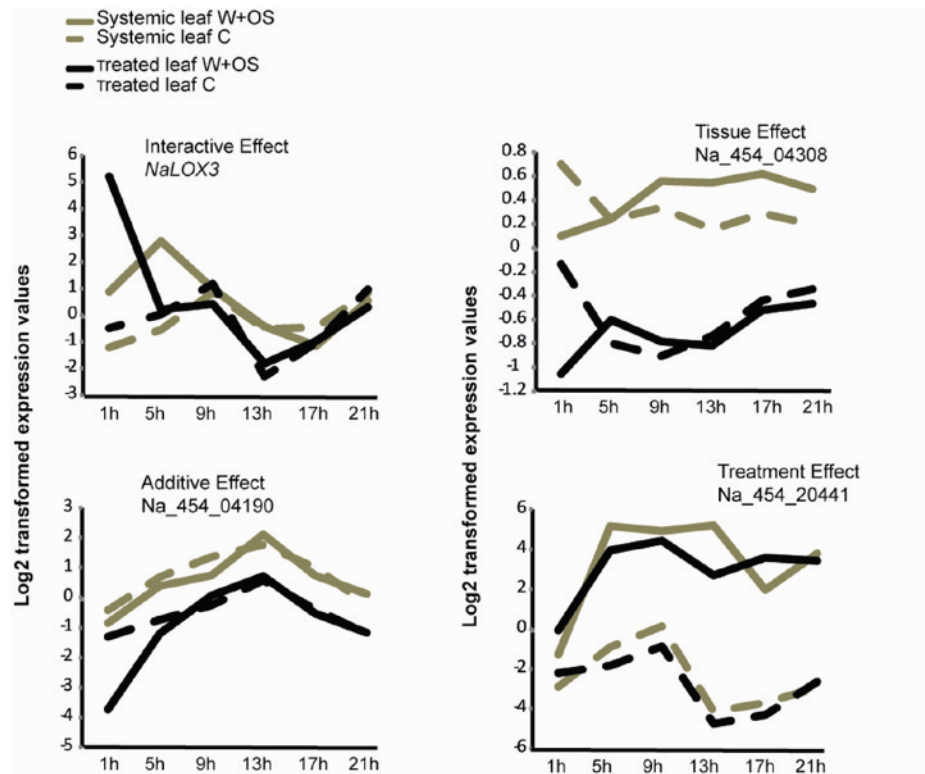
LITERATURE CITED

- Alborn HT, Turlings TCJ, Jones TH, Stenhagen G, Loughrin JH, Tumlinson JH (1997) An elicitor of plant volatiles from beetle armyworm oral secretion. *Science* **276**: 945–949
- Allmann S, Halitschke R, Schuurink RC, Baldwin IT (2010) Oxylinipin channelling in *Nicotiana attenuata*: lipoxygenase 2 supplies substrates for green leaf volatile production. *Plant Cell Environ* **33**: 2028–2040
- Baldwin IT (1990) Herbivory simulations in ecological research. *Trends Ecol Evol* **5**: 91–93
- Baldwin IT (2001) An ecologically motivated analysis of plant-herbivore interactions in native tobacco. *Plant Physiol* **127**: 1449–1458
- Baldwin IT, Morse L (1994) Up in smoke. 2. Germination of *Nicotiana attenuata* in response to smoke-derived cues and nutrients in burned and unburned soils. *J Chem Ecol* **20**: 2373–2391
- Baldwin IT, Stasakozinski L, Davidson R (1994) Up in smoke. 1. Smoke-derived germination cues for postfire annual, *Nicotiana attenuata* Torr ex Watson. *J Chem Ecol* **20**: 2345–2371
- Banerjee S, Bose I (2011) Transient pulse formation in jasmonate signaling pathway. *J Theor Biol* **273**: 188–196
- Bittner M, Meltzer P, Trent J (1999) Data analysis and integration: of steps and arrows. *Nat Genet* **22**: 213–215
- Bonaventure G, Schuck S, Baldwin IT (2011) Revealing complexity and specificity in the activation of lipase-mediated oxylinipin biosynthesis: a specific role of the *Nicotiana attenuata* GLA1 lipase in the activation of jasmonic acid biosynthesis in leaves and roots. *Plant Cell Environ* **34**: 1507–1520
- Chini A, Fonseca S, Fernández G, Adie B, Chico JM, Lorenzo O, García-Casado G, López-Vidriero I, Lozano FM, Ponce MR, et al (2007) The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448**: 666–671
- Chung HS, Koo AJK, Gao XL, Jayanty S, Thines B, Jones AD, Howe GA (2008) Regulation and function of Arabidopsis JASMONATE ZIM-domain genes in response to wounding and herbivory. *Plant Physiol* **146**: 952–964
- Dettmer K, Aronov PA, Hammock BD (2007) Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26**: 51–78
- Dharan S, Nair AS (2009) Biclustering of gene expression data using reactive greedy randomized adaptive search procedure. *BMC Bioinformatics (Suppl 1)* **10**: S27
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**: 953–971

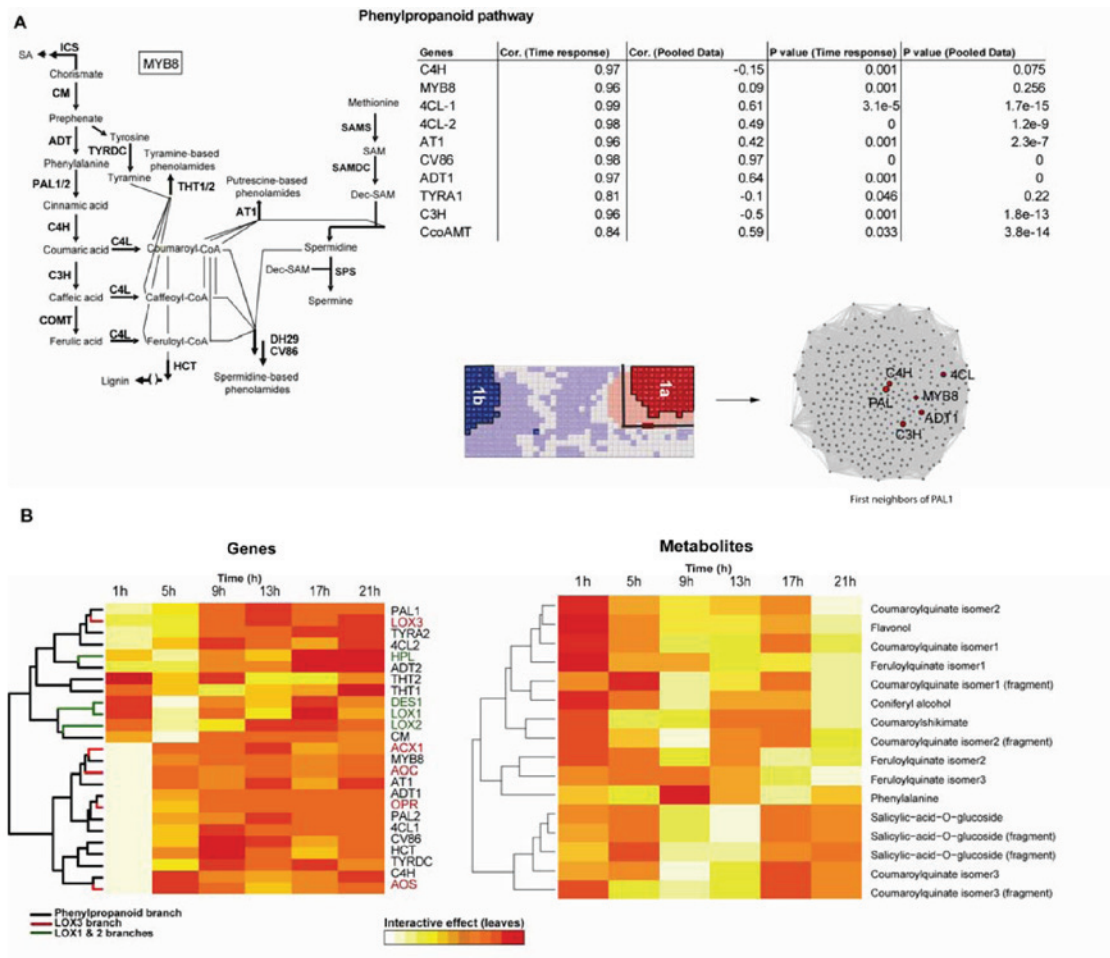
- Gaquerel E, Heiling S, Schoettner M, Zurek G, Baldwin IT (2010) Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. *J Agric Food Chem* 58: 9418–9427
- Getz G, Levine E, Domany E (2000) Coupled two-way clustering analysis of gene microarray data. *Proc Natl Acad Sci USA* 97: 12079–12084
- Gilardoni PA, Schuck S, Jungling R, Rotter B, Baldwin IT, Bonaventure G (2010) SuperSAGE analysis of the *Nicotiana attenuata* transcriptome after fatty acid-amino acid elicitation (FAC): identification of early mediators of insect responses. *BMC Plant Biol* 10: 66
- Giri AP, Wünsche H, Mitra S, Zavala JA, Muck A, Svatos A, Baldwin IT (2006) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. VII. Changes in the plant's proteome. *Plant Physiol* 142: 1621–1641
- Hahlbrock K, Bednarek P, Ciolkowski I, Hamberger B, Heise A, Liedgens H, Logemann E, Nürnberger T, Schmelzer E, Somssich IE, et al (2003) Non-self recognition, transcriptional reprogramming, and secondary metabolite accumulation during plant/pathogen interactions. *Proc Natl Acad Sci USA* (Suppl 2) 100: 14569–14576
- Halitschke R, Baldwin IT (2003) Antisense LOX expression increases herbivore performance by decreasing defense responses and inhibiting growth-related transcriptional reorganization in *Nicotiana attenuata*. *Plant J* 36: 794–807
- Halitschke R, Gase K, Hui DQ, Schmidt DD, Baldwin IT (2003) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. VI. Microarray analysis reveals that most herbivore-specific transcriptional changes are mediated by fatty acid-amino acid conjugates. *Plant Physiol* 131: 1894–1902
- Halitschke R, Schittko U, Pohnert G, Boland W, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. III. Fatty acid-amino acid conjugates in herbivore oral secretions are necessary and sufficient for herbivore-specific plant responses. *Plant Physiol* 125: 711–717
- Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranylinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *Plant Cell* 22: 273–292
- Heitz T, Widemann E, Lugan R, Miesch L, Ullmann P, Désaubry L, Holder E, Grausem B, Kandel S, Miesch M, et al (2012) Cytochromes P450 CYP94C1 and CYP94B3 catalyze two successive oxidation steps of plant hormone jasmonoyl-isoleucine for catabolic turnover. *J Biol Chem* 287: 6296–6306
- Hermesmeier D, Schittko U, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. I. Large-scale changes in the accumulation of growth- and defense-related plant mRNAs. *Plant Physiol* 125: 683–700
- Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 101: 10205–10210
- Howe GA, Lee GI, Itoh A, Li L, DeRocher AE (2000) Cytochrome P450-dependent metabolism of oxylipins in tomato: cloning and expression of allene oxide synthase and fatty acid hydroperoxide lyase. *Plant Physiol* 123: 711–724
- Hui DQ, Iqbal J, Lehmann K, Gase K, Saluz HP, Baldwin IT (2003) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. V. Microarray analysis and further characterization of large-scale changes in herbivore-induced mRNAs. *Plant Physiol* 131: 1877–1893
- Jassbi AR, Gase K, Hettenhausen C, Schmidt A, Baldwin IT (2008) Silencing geranylgeranyl diphosphate synthase in *Nicotiana attenuata* dramatically impairs resistance to tobacco hornworm. *Plant Physiol* 146: 974–986
- Kallenbach M, Alagna F, Baldwin IT, Bonaventure G (2010) *Nicotiana attenuata* SIPK, WIPK, NPR1, and fatty acid-amino acid conjugates participate in the induction of jasmonic acid biosynthesis by affecting early enzymatic steps in the pathway. *Plant Physiol* 152: 96–106
- Kaur H, Heinzel N, Schöttner M, Baldwin IT, Gális I (2010) R2R3-NaMYB8 regulates the accumulation of phenylpropanoid-polyamine conjugates, which are essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*. *Plant Physiol* 152: 1731–1747
- Kim SG, Yon F, Gaquerel E, Gulati J, Baldwin IT (2011) Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native tobacco, *Nicotiana attenuata*. *PLoS ONE* 6: e26214
- Koo AJK, Cooke TF, Howe GA (2011) Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. *Proc Natl Acad Sci USA* 108: 9298–9303
- Krugel T, Lim M, Gase K, Halitschke R, Baldwin IT (2002) Agrobacterium-mediated transformation of *Nicotiana attenuata*, a model ecological expression system. *Chemoecology* 12: 177–183
- Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* 84: 283–289
- McCloud ES, Baldwin IT (1997) Herbivory and caterpillar regurgitants amplify the wound-induced increases in jasmonic acid but not nicotine in *Nicotiana sylvestris*. *Planta* 203: 430–435
- Meldau S, Erb M, Baldwin IT (2012) Defence on demand: mechanisms behind optimal defence patterns. *Ann Bot (Lond)* 110: 1503–1514
- Mitra S, Baldwin IT (2008) Independently silencing two photosynthetic proteins in *Nicotiana attenuata* has different effects on herbivore resistance. *Plant Physiol* 148: 1128–1138
- Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. *Plant Physiol* 149: 88–95
- Park T, Yi SG, Lee S, Lee SY, Yoo DH, Ahn JI, Lee YS (2003) Statistical tests for identifying differentially expressed genes in time-course microarray experiments. *Bioinformatics* 19: 694–703
- Prinss I, Maimon O, Ben-Gal I (2007) Evaluation of gene-expression clustering via mutual information distance measure. *BMC Bioinformatics* 8: 111
- Reymond P, Bodenhausen N, Van Poecke RMP, Krishnamurthy V, Dicke M, Farmer EE (2004) A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell* 16: 3132–3147
- Roda A, Halitschke R, Steppuhn A, Baldwin IT (2004) Individual variability in herbivore-specific elicitors from the plant's perspective. *Mol Ecol* 13: 2421–2433
- Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics: 'majority report by precogs.' *Trends Plant Sci* 13: 36–43
- Schittko U, Baldwin IT (2003) Constraints to herbivore-induced systemic responses: bidirectional signaling along orthostichies in *Nicotiana attenuata*. *J Chem Ecol* 29: 763–770
- Schmidt DD, Voelckel C, Hartl M, Schmidt S, Baldwin IT (2005) Specificity in ecological interactions: attack from the same lepidopteran herbivore results in species-specific transcriptional responses in two solanaceous host plants. *Plant Physiol* 138: 1763–1773
- Schwachtje J, Minchin PEH, Jahnke S, van Dongen JT, Schittko U, Baldwin IT (2006) SNF1-related kinases allow plants to tolerate herbivory by allocating carbon to roots. *Proc Natl Acad Sci USA* 103: 12935–12940
- Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78: 779–787
- Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004) Nicotine's defensive function in nature. *PLoS Biol* 2: E217
- Storey JD, Xiao WZ, Leek JT, Tompkins RG, Davis RW (2005) Significance analysis of time course microarray experiments. *Proc Natl Acad Sci USA* 102: 12837–12842
- Strassner J, Schaller F, Frick UB, Howe GA, Weiler EW, Amrhein N, Macheroux P, Schaller A (2002) Characterization and cDNA-microarray expression analysis of 12-oxophytodienoate reductases reveals differential roles for octadecanoid biosynthesis in the local versus the systemic wound response. *Plant J* 32: 585–601
- Swindell WR (2006) The association among gene expression responses to nine abiotic stress treatments in *Arabidopsis thaliana*. *Genetics* 174: 1811–1824
- Tai YC, Speed TP (2006) A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann Statist* 34: 2387–2412
- Takahashi H, Morioka R, Ito R, Oshima T, Altaf-Ul-Amin M, Ogasawara N, Kanaya S (2011) Dynamics of time-lagged gene-to-metabolite networks of *Escherichia coli* elucidated by integrative omics approach. *OMICS* 15: 15–23
- Vranová E, Coman D, Gruissem W (2012) Structure and dynamics of the isoprenoid pathway network. *Mol Plant* 5: 318–333

Herbivory-Activated Gene-Metabolite Networks

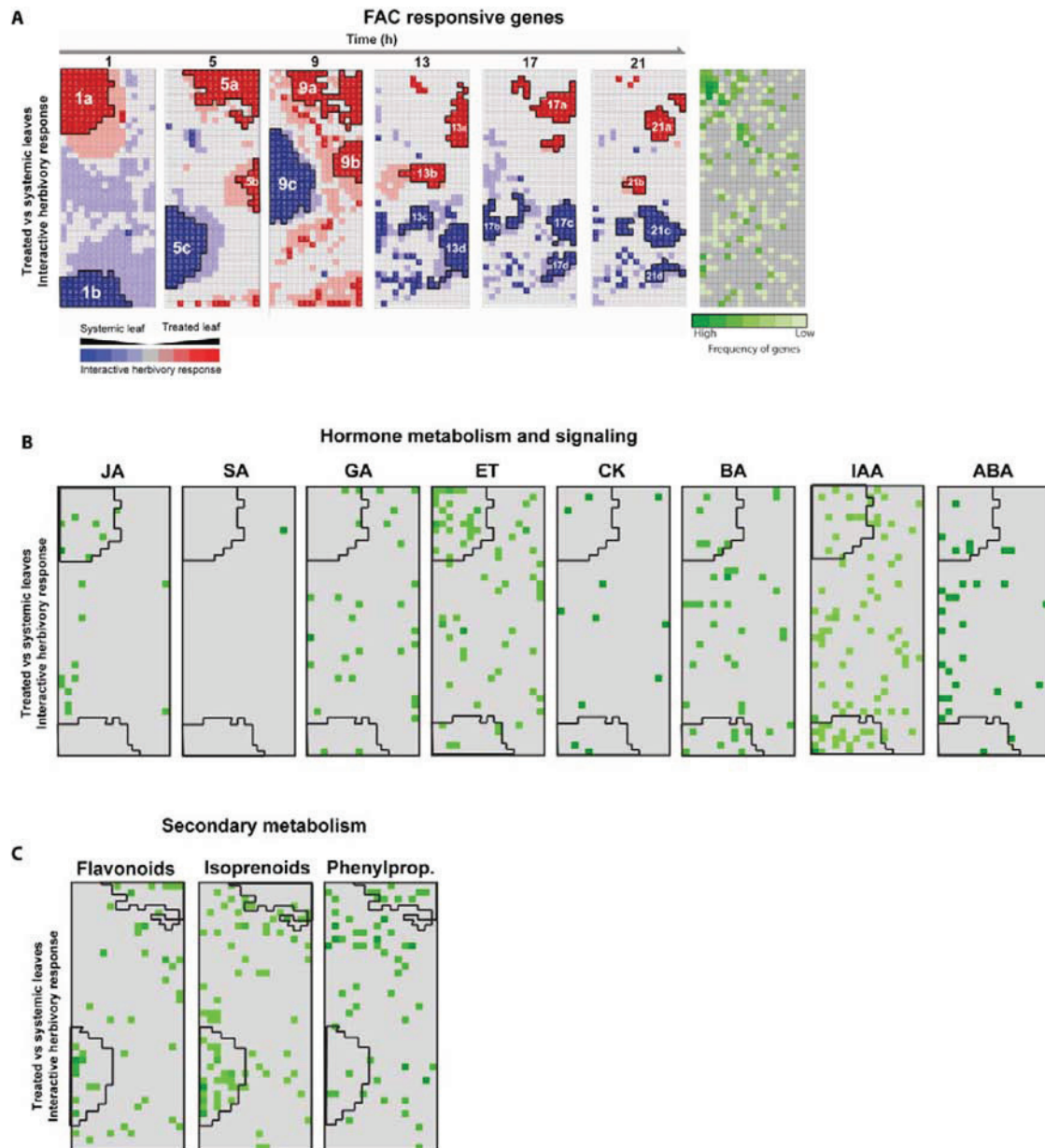
- Walley JW, Dehesh K (2010) Molecular mechanisms regulating rapid stress signaling networks in *Arabidopsis*. *J Integr Plant Biol* **52**: 354–359
- Walther D, Strassburg K, Durek P, Kopka J (2010) Metabolic pathway relationships revealed by an integrative analysis of the transcriptional and metabolic temperature stress-response dynamics in yeast. *OMICS* **14**: 261–274
- Wang L, Allmann S, Wu JS, Baldwin IT (2008) Comparisons of LIPOXYGENASE3- and JASMONATE-RESISTANT4/6-silenced plants reveal that jasmonic acid and jasmonic acid-amino acid conjugates play different roles in herbivore resistance of *Nicotiana attenuata*. *Plant Physiol* **146**: 904–915
- Wang J, Kim SK (2003) Global analysis of dauer gene expression in *Caenorhabditis elegans*. *Development* **130**: 1621–1634
- Wu JQ, Baldwin IT (2010) New insights into plant responses to the attack from insect herbivores. *Annu Rev Genet* **44**: 1–24
- Wu JQ, Hettenhausen C, Meldau S, Baldwin IT (2007) Herbivory rapidly activates MAPK signaling in attacked and unattacked leaf regions but not between leaves of *Nicotiana attenuata*. *Plant Cell* **19**: 1096–1122
- Zeller G, Henz SR, Widmer CK, Sachsenberg T, Rätsch G, Weigel D, Laubinger S (2009) Stress-induced changes in the *Arabidopsis thaliana* transcriptome analyzed using whole-genome tiling arrays. *Plant J* **58**: 1068–1082
- Zhou BY, Wong WH (2011) A bootstrap-based non-parametric ANOVA method with applications to factorial microarray data. *Statist Sinica* **21**: 495–514
- Zhou BY, Xu WH, Herndon D, Tompkins R, Davis R, Xiao WZ, Wong WH, Toner M, Warren HS, Schoenfeld DA, et al (2010) Analysis of factorial time-course microarrays with application to a clinical study of burn injury. *Proc Natl Acad Sci USA* **107**: 9923–9928



Supplementary Figure 1: Examples for the structures obtained (interactive, additive and major effects) by fitting ANOVA model.

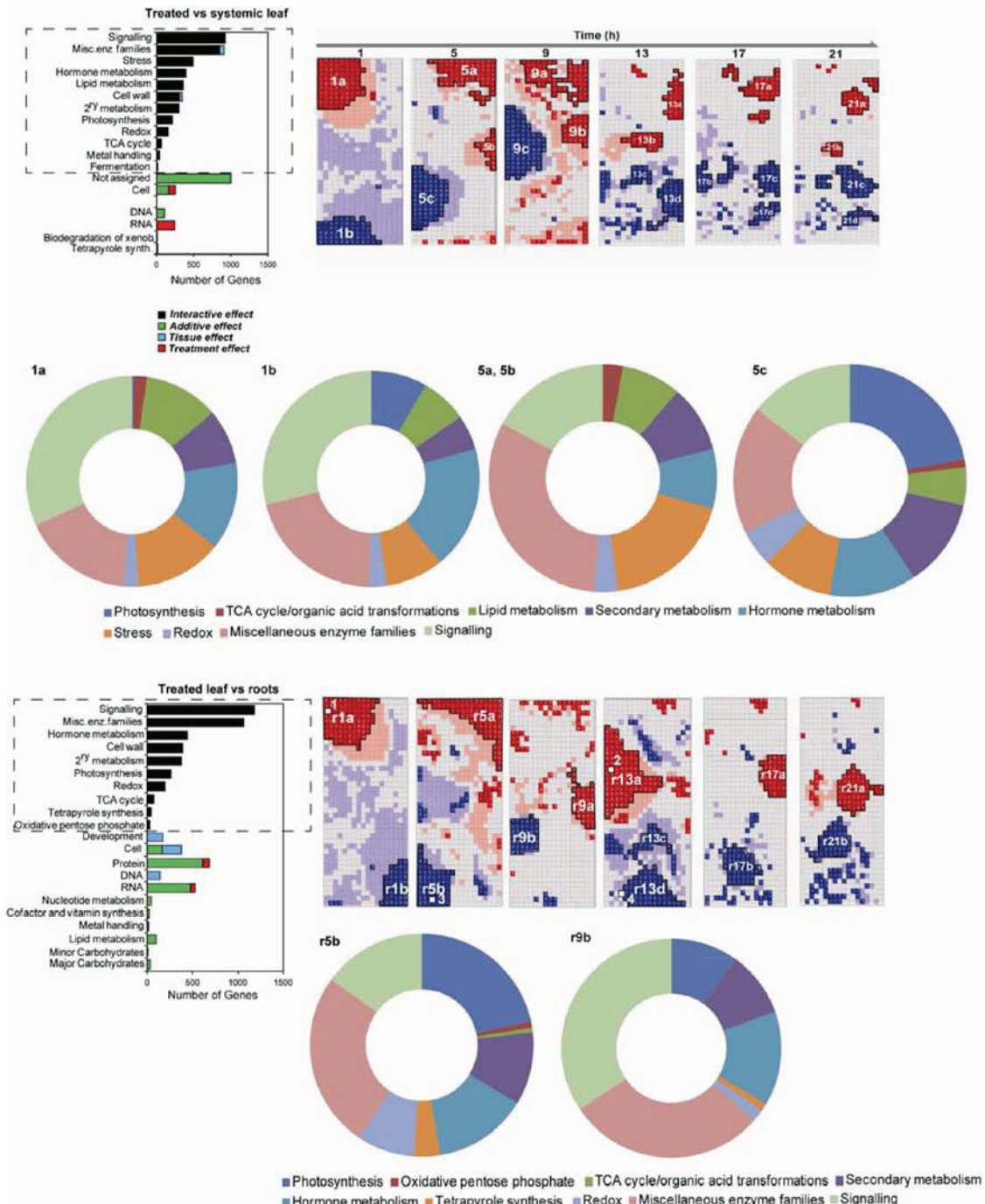


Supplementary Figure 2: Interactive effect responses of phenylpropanoid pathway genes are highly coordinated. (A) Schematic representation of the phenylpropanoid pathway and Pearson correlation coefficients (PC), calculated using two different associations metrics, between NaPAL, controlling cinnamate biosynthesis, and downstream genes of the phenylpropanoid pathway. Higher PC values obtained when using the interactive effect response metric allows the identification of biologically-relevant gene interactions which are weak or insignificant when drawn using pooled expression values across time series and all experimental conditions. **(B)** HCA (Hierarchical Clustering Analysis) of genes involved in the oxylipin and phenylpropanoid pathways exhibit high interactive effects 1h after OS-elicitation. But in contrast with similar analysis for metabolites, we observed interactive effect for later time points (9h and 13h after elicitation) for known phenylpropanoids.

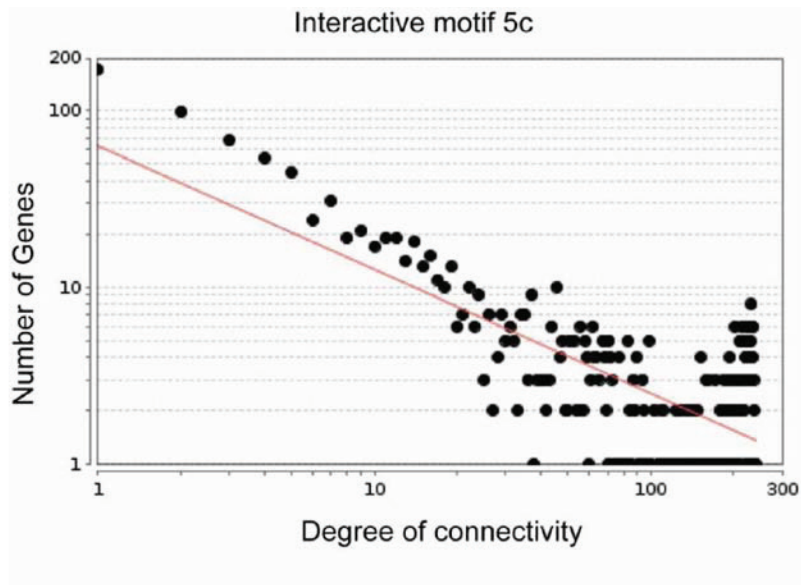


Supplementary Figure 3: Self-organizing maps based classification of fatty acid amino acid conjugate-responsive and hormone and secondary metabolism genes in leaves. (A) Probes representing unique sequences, taken from published SuperSAGE library, known to be induced by fatty acid amino acid conjugates (FAC) elicitors in *N. attenuata* are mapped onto SOM grid. They are abundantly located in motif labeled as 1a, showing early response in treated leaf. **(B)** Genes with GO terms for different hormone and signaling pathways were extracted and mapped

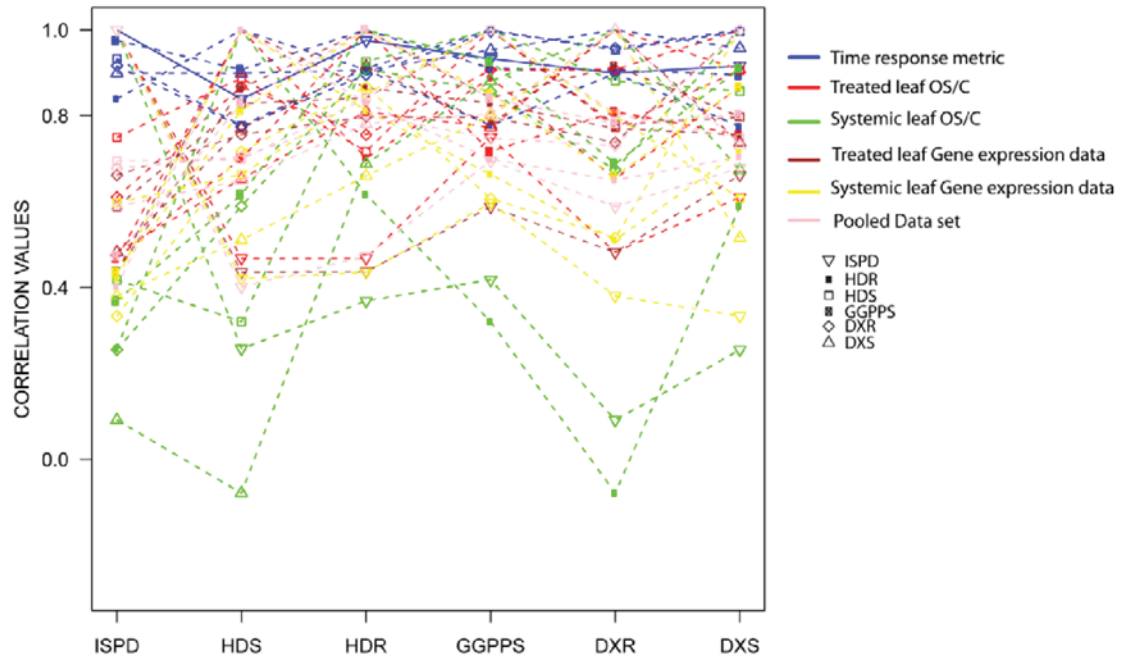
onto SOM grid. While genes of the JA (jasmonic acid) and ET (ethylene) signaling pathways were found overrepresented in the motif labeled as 1a and showing early responses in the treated leaf, those of the auxin pathway were more abundantly located in the motif labeled as 1b with major activation in the untreated systemic leaf. **(C)** Similarly, genes with GO terms involved in the flavonoid, isoprenoid, and phenylpropanoid pathways were mapped onto the SOM grid and found overrepresented in motifs labeled as 5a and 5c, indicating the importance of OS-specific metabolic responses activated 5h after elicitation.



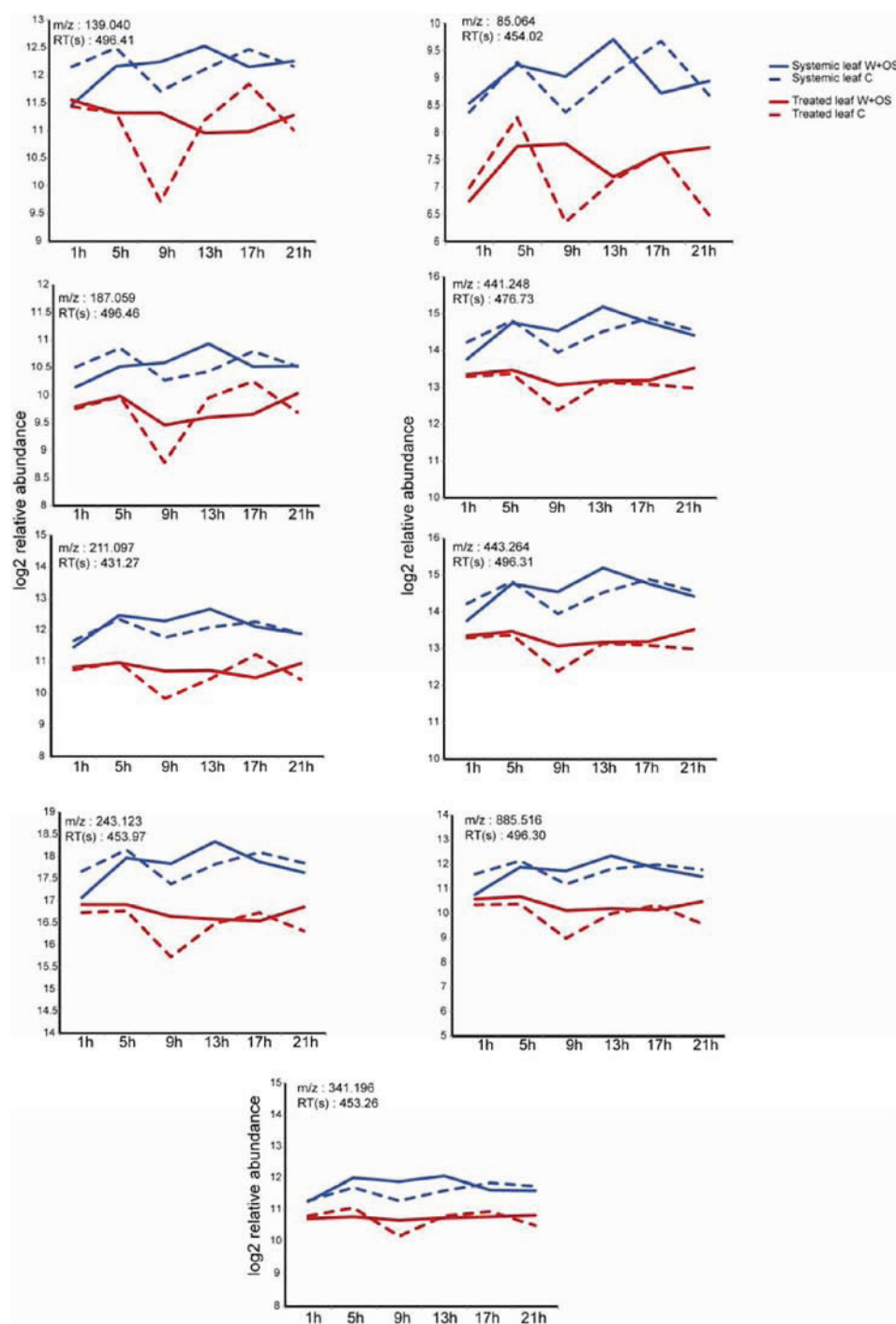
Supplementary Figure 4: Enriched GO terms for few important motifs extracted from SOM analysis unravels large OS-specific gene expression responses in treated leaves and systemic tissues (leaves and roots).



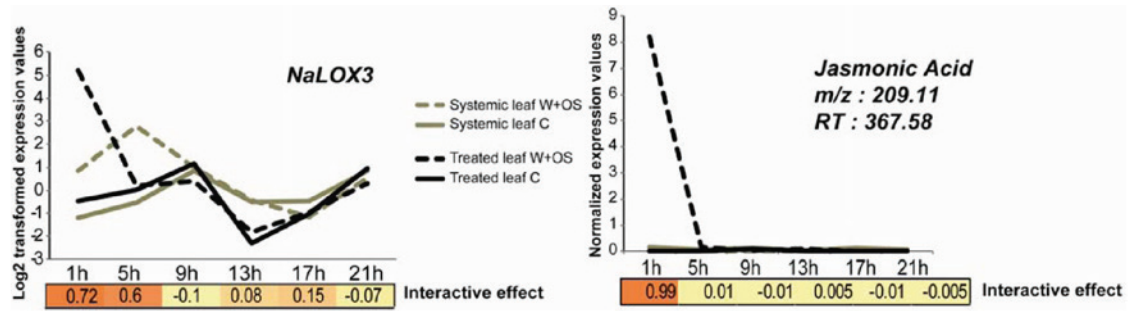
Supplementary Figure 5: Power law distribution plot of the network representation obtained for genes extracted from motif 5c showing many genes with few connections, but a small set of genes with many connections.



Supplementary Figure 6: Gene-gene Pearson Correlation (PC) for 6 genes from the non-mevalonate pathway using 5 different metrics: The time response metric-based PC calculations for all gene pairs are recovered with a threshold of 0.75 but not in the case of pooled data or with expression data obtained from just one tissue type.



Supplementary Figure 7: Temporal profiles for m/z signals corresponding to the class O-acyl sugars exhibit high correlation and co-localize with those of 17-HGL-DTG in the network.



Supplementary Figure 8: Temporal profiles of *NaLOX3*, functionally characterized for the production of jasmonic acid and *m/z* signal of jasmonic acid. High values at 1h in time response metric for interactive effect for treated leaf vs untreated systemic leaf comparison were observed for both gene (*NaLOX3*) and metabolite (jasmonic acid).

Supplementary Method

Metabolite identification

Targeted structure elucidation analyses on metabolites presented in this study have previously been published (Gaquerel et al., 2010; Heiling et al., 2010; Weinhold et al., 2011). Briefly, each 17-HGL-DTG was purified, analyzed by MS2 high resolution HPLC-ESI/TOF-MS and its structure identified by NMR (Heiling et al., 2010). Fragmentation patterns of O-acyl sugars characterized in *N. attenuata* exhibit typical losses of acetylated and non-acetylated fructose, previously observed for O-acyl sugars described in other Solanaceae. Short fatty acid chains involved in *N. attenuata* O-acyl sugars were identified, after trans-methylation, by GC-MS and comparison with authentic compounds (Weinhold et al., 2011). However, the position of these acyl moieties could not be inferred from the MS2 analyses conducted on O-acyl sugars (Weinhold et al., 2011). Herbivory-inducible phenolic derivatives produced by *N. attenuata* have been identified by tandem MS analysis in a previous study (Okonkesung et al., 2012).

During the analysis of xcms-processed metabolomic matrices, m/z signals corresponding to molecular fragments of 17-HGL-DTG, O-acyl sugars and phenolic derivatives were annotated based on elemental formula predictions and analysis of tandem MS high resolution fragmentation patterns obtained by Gaquerel et al (2010). MS detection was carried out with a maXis ESIqTOF mass spectrometer operated in electrospray positive mode. Typical instrument settings were as follows: capillary voltage 4500 V, dry gas temperature 200 °C, dry gas flow of 8 L/min, capillary exit 117V and funnel RF 300Vpp.

Below are examples of tandem MS records obtained for one 17-HGL-DTG and one O-acyl sugars reported in this study, others are available in the Supplemental material of Gaquerel et al. (2010).

Nicotianoside I, $[M+Na]^+$ (+MS2 m/z 885.4090, $C_{41}H_{66}O_{19}Na^+$, 18.4 min, 49 eV): 885.4098 (35.3 %, $C_{41}H_{66}O_{19}Na^+$, calc. 885.4090), 841.4192 (13.7 %, $C_{40}H_{66}O_{17}Na^+$, calc. 841.4192), 739.3487 (8.6 %, $C_{35}H_{56}O_{15}Na^+$, calc. 739.3511), 559.2874 (6.3 %, $C_{29}H_{44}O_9Na^+$, calc. 559.2877), 491.2973 (22.3 %, $C_{26}H_{44}O_7Na^+$, calc. 491.2979), 475.3023 (21.9 %, $C_{26}H_{44}O_6Na^+$, calc. 475.3031), 473.2868 (23.5 %, $C_{26}H_{42}O_6Na^+$, calc. 473.2873), 435.1103 (100 %, $C_{15}H_{24}O_{13}Na^+$, calc. 435.1109), 433.0949 (37.6 %, $C_{15}H_{22}O_{13}Na^+$, calc. 433.0952), 289.0530 (8.5 %, $C_9H_{14}O_9Na^+$, calc. 289.0530), 271.2424 (4.1 %, $C_{20}H_{31}^+$, calc. 271.2420).

Massbank best hit: -

Authentic standard: purified from *N. attenuata* leaves and analyzed by NMR.

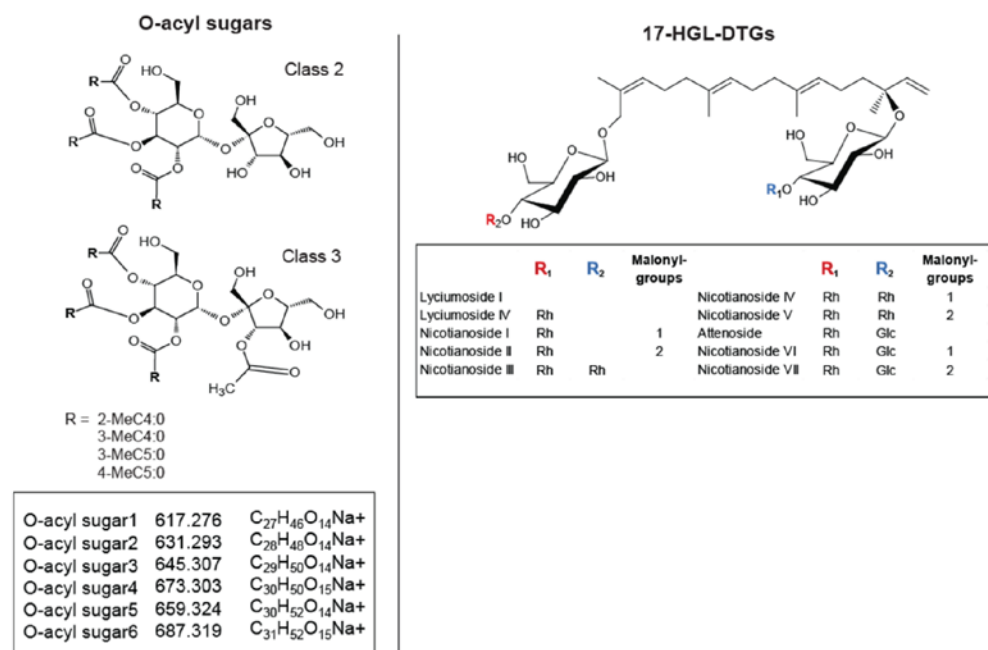
O-acyl sugar 1, $[M+Na]^+$ (+MS2 m/z 603.2608, $C_{26}H_{44}O_{14}Na^+$, 19.08 min, 46 eV): 603.2608 ($C_{26}H_{44}O_{14}Na^+$, calc. 603.2623), 442.2119 (23.1 %, $^{13}C_1C_{19}H_{34}O_9Na^+$, calc. 442.2128), 441.2083

(100 %, $C_{20}H_{34}O_9Na^+$, calc. 441.2095), 353.1558 (10.5 %, $C_{16}H_{26}O_7Na^+$, calc. 353.1571), 339.1399 (10.1 %, $C_{15}H_{24}O_7Na^+$, calc. 339.1414), 325.1252 (12.1 %, $C_{14}H_{22}O_7Na^+$, calc. 325.1257), 185.0419 (10.2 %, $C_6H_{10}O_5Na^+$, calc. 185.0420), 127.0389 (9.6 %, $C_6H_7O_3^+$, calc. 127.0389).

Massbank best hit: -

Authentic standard: -

Chemical structure of O-acyl sugars and 17-HGL-DTGs



O-acyl sugars are esterified by short chain fatty acids ranging from three to six carbons in chain length. Type III sucrose esters are acetylated at their fructose moiety. 17-HGL-DTGs differ in their sugar (Rh: rhamnose, Glc: glucose) and malonyl groups.

References:

Gaquerel E, Heiling S, Schoettner M, Zurek G, Baldwin IT (2010) Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. *J Agr Food Chem* **58**: 9418-9427

Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *Plant Cell* **22**: 273-292

Onkokesung N, Gaquerel E, Kotkar H, Kaur H, Baldwin IT, Galis I (2012) MYB8 Controls Inducible Phenolamide Levels by Activating Three Novel Hydroxycinnamoyl-Coenzyme A:Polyamine Transferases in *Nicotiana attenuata*. *Plant Physiol* **158**: 389-407

Weinhold A, Baldwin IT (2011) Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *P Natl Acad Sci USA* **108**: 7855-7859

Chapter 5

An integrative statistical method to explore herbivory-specific responses in plants

Jyotasana Gulati, Ian T. Baldwin and Emmanuel Gaquerel

Accepted for Publication in Plant Signaling & Behavior 2013, 8:e25638; PMID: 23857359

An integrative statistical method to explore herbivory-specific responses in plants

Jyotasana Gulati, Ian T. Baldwin, and Emmanuel Gaquerel*

Department of Molecular Ecology; Max Planck Institute for Chemical Ecology; Jena, Germany

Keywords: metabolomics, network analysis, plant stress responses, systems biology, transcriptomics

Abbreviations: OS, oral-secretion; SOM, self-organizing maps; ANOVA, analysis of variance

Spatial-temporal coordination between multiple processes/pathways is a key determinant of whole-organism transcriptome and metabolome reconfigurations in plant's response to biotic stresses. To explore tissue-based interdependencies in *Nicotiana attenuata*'s resistance to insect attack, we performed time course analyses of the plant's transcriptome and metabolome in herbivory-elicited source leaves and unelicited sink leaves and roots. To dissect the multidimensionality of these responses, we have recently designed a novel approach of constructing interactive motifs by combining an extended self-organizing maps (SOM) based dimensionality reduction method with bootstrap-based non-parametric ANOVA models. In this previous study, we used this method to study nonlinearities in gene-metabolite associations involved in the acyclic diterpene glucoside pathway. Here, we extend the application of this method to the extraction of genes showing herbivory-elicitation specifically in systemic (distal from the treatment sites) tissues using motif analysis for different combinations of treatment applied to *Nicotiana attenuata*.

Plants have evolved efficient defense strategies which involve rapid changes in intricately connected signaling and metabolic networks.¹⁻⁴ Experiments designed to study such intricate networks often have a complex factorial structure, obtained by assessing plant responses in different conditions/treatments, tissue types or genetic contexts. Among the major problems associated with the statistical analysis of multifactorial experimental designs are those of gene prioritization and of large numbers of false positives.^{5,6} Since signaling pathways underlying major stress responses are generally affected only by a subset of the experimental conditions⁷ and are based on transient gene associations, therefore these cannot be captured using collective information studies. Although bioinformatic approaches such as mutual information⁸ and biclustering⁹ have been developed to address this limitation, gene networks assemble dynamically as the organism adapts to external stimuli and therefore their analysis necessitates the mining of multifactorial time series experiments. Several efforts including those by Park et al.,¹⁰ Wang and Kim¹¹ and Tai and Speed¹² have been published to assess single factorial effect on gene expression in a time course experiment. Zhou et al.¹³ have developed a method to simultaneously analyze experiments involving more than one factor measured across time series by finding the significant direction in the time course across different conditions.

Nicotiana attenuata is an annual fire-chasing plant native to the Great Basin Desert of the southwestern United States of America which has evolved a large number of specific induced responses against generalist and specialist herbivores.^{14,15} Some of the essential nodes in the plant's transcriptome and metabolome responses to attack from larvae of the specialist lepidopteran herbivore, *Manduca sexta*, have been functionally characterized.¹⁶ Feeding by this specialist herbivore or its simulation by the application of its oral secretions (OS) into puncture wounds produced by mechanical wounding activate rapid changes in the plant's metabolic and growth processes in order to facilitate de novo production of defense compounds.

In a recent study,¹⁷ we profiled the transcriptome and metabolome of identically treated wild type *Nicotiana attenuata* plants for 3 tissues and 2 stress conditions (mechanical wounding and simulated herbivory) with a regular time series of 6 time points. To investigate the dynamics of activation in time and space of herbivory-induced changes in gene-to-metabolite networks, we employed a bootstrap-based non-parametric ANOVA (NANOVA) model designed to find gene/metabolite-specific responses across the time series based on their dependency on experimental factors used for comparison.¹³ We conducted dynamic response analyses taking control (Ctrl) and OS treated

*Correspondence to: Emmanuel Gaquerel; Email: egaquerel@ice.mpg.de

Submitted: 06/25/2013; Accepted: 07/04/2013

Citation: Gulati J, Baldwin IT, Gaquerel E. An integrative statistical method to explore herbivory-specific responses in plants. *Plant Signal Behav* 2013; 8: e25638; <http://dx.doi.org/10.4161/psb.25638>

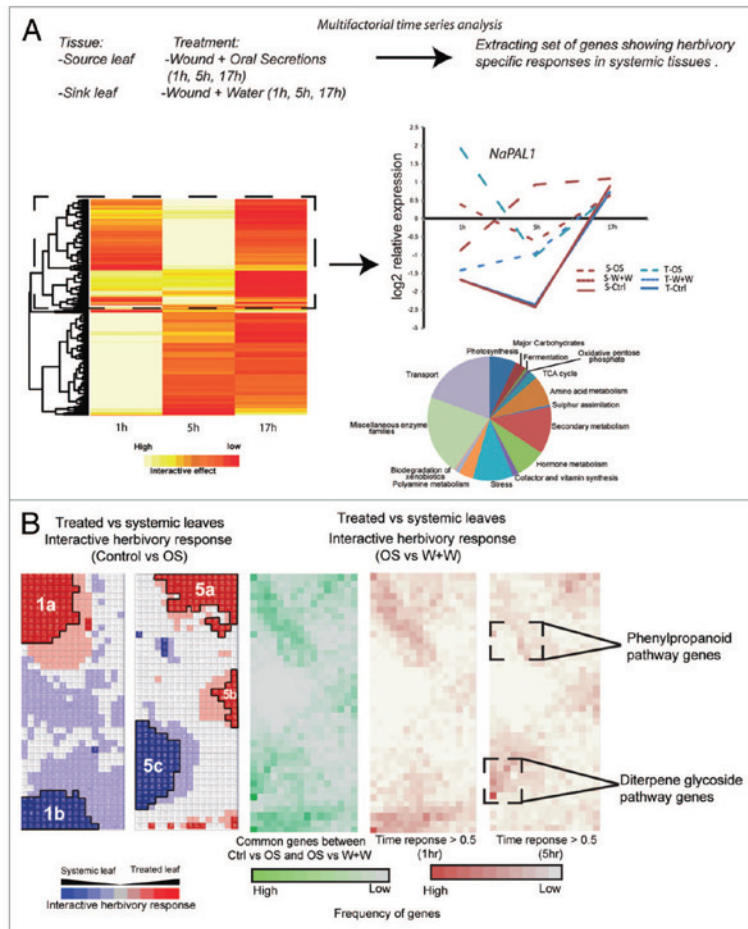


Figure 1. (A) Implementation of a statistical method designed to study OS-elicitation specific responses in untreated tissues in *Nicotiana attenuata*: Replicated transcriptomic and metabolomic data were analyzed using multi-factorial analysis with both factors (tissue and treatment) taken together across the time series to identify modules showing differential OS-elicitation. Heatmap represents hierarchical cluster analysis of genes showing interactive effect for W+OS and W+W condition when compared for treated and untreated leaf tissue. Temporal profile of NaPAL1 represents the specific pattern, OS specific response at 5 h after elicitation, extracted using this method. Pie chart represents the distribution of enriched GO terms for genes showing interactive effect. (B) Localization of genes showing OS-specific responses dependent on tissue type in different interactive motifs: Probes from interactive bin for the analysis comparing 2 factors (treatment, W+OS vs. W+W; tissue, Treated leaf vs. systemic leaf) were mapped onto SOM grids and found overrepresented in motifs 1a, 1b, 5a and 5c. This set is further divided into 2 sub-groups based on their time response. (A) greater than 0.5 at 1h after treatment; (B) greater than 0.5 at 5h after treatment. Set (B) is represented by 2 well-studied secondary metabolic pathways—the acyclic diterpene glycoside and phenylpropanoid pathways.

comparisons: treated (source) leaf vs. untreated systemic (sink) leaf (TvS comparison) to explore differential gene expression patterns activated during shoot systemic signaling, and treated leaf vs. untreated roots (TvR comparison) to obtain novel insights into root specific responses. Using a series of statistical tests on factor effects,¹³ we divided the transcriptome/metabolome into 4 mutually exclusive groups showing their best ANOVA structure

along the estimated optimal direction in the time series. The 4 resulting structures represent interactive (tissues behaving differently in response to OS-elicitation across the time series), additive (herbivore responses independent of tissue type), or corresponding main effects on gene expression (major treatment effects in both treated and untreated tissue or significant differences in tissue type with no response to treatment). Set of genes and metabolites displaying interactive response patterns were further studied. With this approach, we captured the dynamic response of a gene in more than one tissue in terms of a single metric which was then used to delineate elements of signaling pathway and to analyze activation transition points between different sub-branches of a single pathway. Next, we imposed structure on the data using batch learning self-organizing maps (BL-SOM)¹⁸ to obtain interactive motifs which are defined as patterns of interconnections between genes and metabolites that are differentially perturbed in local and systemic tissues in response to stress, additional information of their time of action having been obtained from projected data on time series termed as ANOVA directions. Since dynamic responses for signaling/metabolic pathways are considered highly coordinated, we hypothesized that the nodes in these spatio-temporally resolved motifs with similar ANOVA directions along the time series may reflect biological organization. We isolated interactive motifs from the SOM grids for genes involved in the biosynthesis of defense metabolites of the 17-hydroxygeranyllinalool diterpene glycoside (17-HGL-DTG) class^{19,20} and analyzed their dynamic behaviors using network analysis. Metabolic analyses conducted in parallel further supported the advantage of constructing dynamic correlation network based on response features captured by the factorial analysis. Specifically, we identified multiple clusters of biochemically-connected metabolites that shared similar time response metric and grouped according to their inferred compound-family-wise grouping of *m/z* ions for 17-HGL-DTG, O-acyl sugars, shikimate

pathway-derived amino acids and downstream metabolites produced within the phenylpropanoid pathway.

Here, we illustrate the extension of this statistical method to study herbivory-specific plant responses in treated and untreated leaves. Additionally to the time course transcriptome data collected from the W+OS treatment type, we considered transcriptome data for leaf tissues (treated and untreated leaves) that had

collected at 3 time points (2 early time points, 1 h and 5 h after treatment; one late time point, 17 h after treatment; Fig. 1A). By applying this novel method we differentiated OS-specific systemic response (W+OS) from those inherent to the mechanical wounding (W+W) by extracting the set of genes showing an interactive effect to the treatment type along the time series for the comparison between treated and untreated leaf tissues. OS-specific responses dependent on tissue type are more clearly visible for early time points (1 h and 5 h after treatment) in a hierarchical cluster analysis of genes showing interactive effect (Fig. 1A). To assess the functional significance of these genes, we computed the enrichment of GO terms using hyper-geometric tests ($f < 0.05$) (Fig. 1A). As expected from previously characterized mechanisms of biotic stress adaptation in this plant species, the group of genes showing OS-specific systemic responses were highly enriched for processes associated with stress responses, hormone metabolism, secondary metabolism, photosynthetic pathway and amino-acid metabolism. We next mapped the set of genes showing an interactive effect to the W+OS treatment type when compared with the W+W condition onto the SOM map which was generated earlier for the analysis of the TvS (treated vs systemic leaf response) comparison between Control and W+OS treatment type¹⁷ and detected these genes of interest being localized into the nodes from interactive motifs 1a, 1b, 5a, and 5c which have been shown earlier to be enriched for stress, signaling and secondary metabolic pathways. Since we found OS-specific systemic responses being more pronounced at 1 h and 5 h, we resolved the map by filtering genes based on their time response metrics ($> = 0.5$) for these 2 time points. Genes showing large effects at 1 h are in motifs 1a and 1b with many of them being up-regulated and few down-regulated in treated leaf tissues while those showing large effects at 5 h are in motifs 5c and only few from motif 1a. The pathway for diterpene glycoside, present in motif 5c, is well studied and has been presented in an earlier study.¹⁷ Figure 2 presents the comparative time response behavior of genes involved in the phenylpropanoid pathway for 2 comparisons: (A) Control vs. W+OS (combined herbivory and mechanical wounding responses), (B) W+OS vs. W+W (OS-specific responses). Activation of gene expression in response to OS in treated leaves 1 h post elicitation is supported by the high value of the response metric at 1 h while the differential activation of these genes specifically to OS at 5 h in untreated leaf tissue in response to application of OS is reflected by the high value of response metric obtained from W+OS vs. W+W comparison at 5 h.

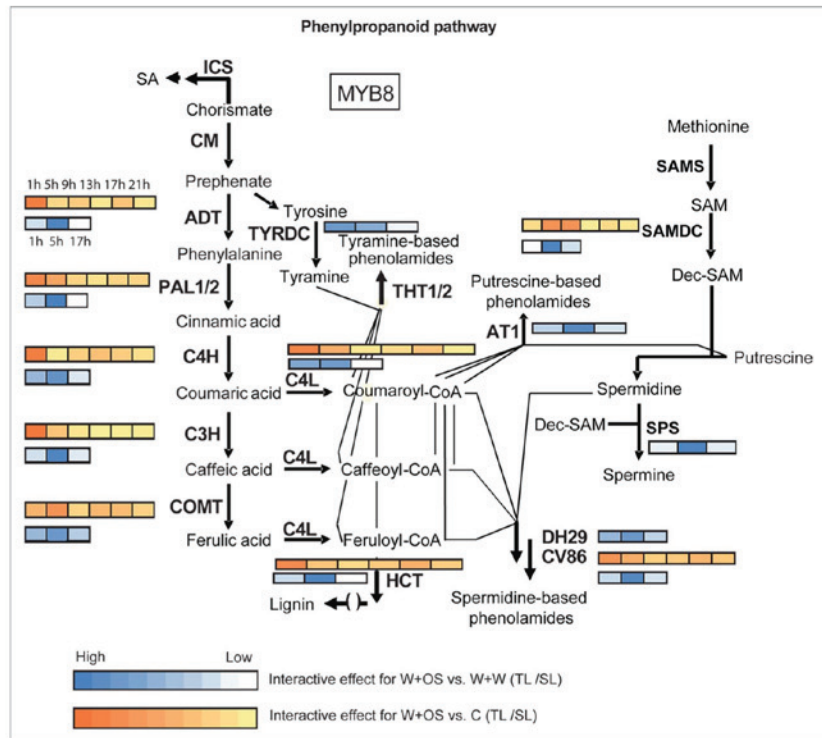


Figure 2. OS elicitation selectively activates genes in the phenylpropanoid pathway at 5 h after elicitation. Schematic representation of the phenylpropanoid pathway with the color coded representation of interactive effect response metric for two different comparisons: (A) Control vs. W+OS, (B) W+OS vs. W+W. Activation of gene expression in response to OS in treated leaves 1 h post elicitation is supported by the high value of the response metric at 1 h while the differential activation of these genes specifically to OS at 5 h in untreated leaf tissue in response to application of OS is reflected by the high value of response metric obtained from W+OS vs. W+W comparison at 5 h.

by the high value of the response metric at 1 h while the differential activation of these genes specifically to OS at 5 h in untreated leaf tissue in response to application of OS is reflected by the high value of response metric obtained from W+OS vs. W+W comparison at 5 h.

This broadly applicable approach allows identifying complex interdependencies between metabolites and transcripts with a high level of accuracy and robustness. The application presented here for the identification of herbivory-specific responses in leaves distal to the plant treatment sites provides an additional support for the importance of finely-tuned changes in metabolism throughout the complete plant during stress adaptation.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Eva Roth and Felipe Yon for their help with sample preparation, Dr Matthias Schöttner, Dr Klaus Gase and Wibke Krober for technical assistance with metabolomics and microarray analyses and the Max Planck Society for funding.

References

- Hahlbrock K, Bednarek P, Ciolkowski I, Hamberger B, Heise A, Liedgens H, et al. Non-self recognition, transcriptional reprogramming, and secondary metabolite accumulation during plant/pathogen interactions. *Proc Natl Acad Sci USA* 2003; 100:14569-76; PMID:12704242; <http://dx.doi.org/10.1073/pnas.0831246100>
- Zeller G, Henz SR, Widmer CK, Sachsberg T, Ratsch G, Weigel D, et al. Stress-induced changes in the *Arabidopsis thaliana* transcriptome analyzed using whole-genome tiling arrays. *Plant J* 2009; 58:1068-82; PMID:19222804; <http://dx.doi.org/10.1111/j.1365-3113.2009.03835.x>
- Walley JW, Dehesh K. Molecular mechanisms regulating rapid stress signaling networks in *Arabidopsis*. *J Integr Plant Biol* 2010; 52:354-9; PMID:20377697; <http://dx.doi.org/10.1111/j.1744-7909.2010.00940.x>
- Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. *Plant Physiol* 2009; 149:88-95; PMID:19126699; <http://dx.doi.org/10.1104/pp.108.129791>
- Bittner M, Meltzer P, Trent J. Data analysis and integration: of steps and arrows. *Nat Genet* 1999; 22:213-5; PMID:10391202; <http://dx.doi.org/10.1038/10265>
- Getz G, Levine E, Domany E. Coupled two-way clustering analysis of gene microarray data. *Proc Natl Acad Sci USA* 2000; 97:12079-84; PMID:11035779; <http://dx.doi.org/10.1073/pnas.210134797>
- Swindell WR. The association among gene expression responses to nine abiotic stress treatments in *Arabidopsis thaliana*. *Genetics* 2006; 174:1811-24; PMID:17028338; <http://dx.doi.org/10.1534/genetics.106.061374>
- Priness I, Maimon O, Ben-Gal I. Evaluation of gene-expression clustering via mutual information distance measure. *BMC Bioinformatics* 2007; 8; PMID:17212835
- Dharan S, Nair AS. Biclustering of gene expression data using reactive greedy randomized adaptive search procedure. *BMC Bioinformatics* 2009; 10; PMID:19133123
- Park T, Yi SG, Lee S, Lee SY, Yoo DH, Ahn JI, et al. Statistical tests for identifying differentially expressed genes in time-course microarray experiments. *Bioinformatics* 2003; 19:694-703; PMID:12691981; <http://dx.doi.org/10.1093/bioinformatics/btg068>
- Wang J, Kim SK. Global analysis of dauer gene expression in *Caenorhabditis elegans*. *Development* 2003; 130:1621-34; PMID:12620986; <http://dx.doi.org/10.1242/dev.00363>
- Tai YC, Speed TP. A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann Stat* 2006; 34:2387-412; <http://dx.doi.org/10.1214/009053606000000759>
- Zhou BY, Xu WH, Herndon D, Tompkins R, Davis R, Xiao WZ, et al. Inflammation and Host Response to Injury Program. Analysis of factorial time-course microarrays with application to a clinical study of burn injury. *Proc Natl Acad Sci USA* 2010; 107:9923-8; PMID:20479259; <http://dx.doi.org/10.1073/pnas.1002757107>
- Hermesmeier D, Schirto U, Baldwin IT. Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. I. Large-scale changes in the accumulation of growth- and defense-related plant mRNAs. *Plant Physiol* 2001; 125:683-700; PMID:11161026; <http://dx.doi.org/10.1104/pp.125.2.683>
- Giri AP, Wünsche H, Mitra S, Zavala JA, Muck A, Svatos A, et al. Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. VII. Changes in the plant's proteome. *Plant Physiol* 2006; 142:1621-41; PMID:17028148; <http://dx.doi.org/10.1104/pp.106.088781>
- Wu JQ, Baldwin IT. New insights into plant responses to the attack from insect herbivores. *Annu Rev Genet* 2010; 44:1-24; PMID:20649414; <http://dx.doi.org/10.1146/annurev-genet-102209-163500>
- Gulati J, Kim SG, Baldwin IT, Gaquerel E. Deciphering herbivory-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. *Plant Physiol* 2013; 162:1042-59; PMID:23656894; <http://dx.doi.org/10.1104/pp.113.217588>
- Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, et al. Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 2004; 101:10205-10; PMID:15199185; <http://dx.doi.org/10.1073/pnas.0403218101>
- Jassbi AR, Gase K, Hettnerhausen C, Schmidt A, Baldwin IT. Silencing geranylgeranyl diphosphate synthase in *Nicotiana attenuata* dramatically impairs resistance to tobacco hornworm. *Plant Physiol* 2008; 146:974-86; PMID:17965175; <http://dx.doi.org/10.1104/pp.107.108811>
- Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, et al. Jasmonate and ppH-systemin regulate key Malonylation steps in the biosynthesis of 17-Hydroxygeranylinalool Diterpene Glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. *Plant Cell* 2010; 22:273-92; PMID:20081114; <http://dx.doi.org/10.1105/tpc.109.071449>

Chapter 6

The roots of plant defenses: Integrative multivariate analyses uncover dynamic behaviors of roots' gene and metabolic networks elicited by leaf herbivory

Jyotasana Gulati, Ian T. Baldwin and Emmanuel Gaquerel

In Review: The Plant Journal 2013

Abstract

Background

High-throughput analyses have frequently been used to characterize herbivory-induced reconfigurations in plant primary and secondary metabolism in above and below-ground tissues but the conclusions drawn from these analyses are often limited by the univariate methods used to analyze the data.

Results

Here we use a multivariate time series data analysis to evaluate the simulated leaf herbivory-elicited transcriptional and metabolic dynamics in the roots of *Nicotiana attenuata*. We observed large, but transient, systemic responses in the roots that contrasted with the pattern of co-linearity observed in the up- and down-regulation of genes and metabolites across the entire time series in treated and systemic leaves. Using a newly developed approach for the analysis of whole-plant molecular responses in a time course multivariate data-set, we simultaneously analyzed stress responses in leaves and roots in response to the elicitation of a leaf. We found that transient systemic responses in roots resolved into two principal trends characterized by: (a) an inversion of root-specific semidiurnal (12h) gene oscillations and (b) transcriptional changes with major amplitude effects that translated into a distinct suite of root-specific secondary metabolites (e.g. alkaloids synthesized in the roots of *N. attenuata*).

Conclusions

These findings underscore the importance of understanding tissue-specific stress responses in the correct day-night phase context and provide a holistic framework for the important role played by roots in aboveground stress responses.

Background

Advances in high-throughput “-omics” technologies have enabled several efforts to characterize plant physiological responses by analyzing complex networks of interactions at different levels and scales. Most of these analyses aim at deciphering how environmental perturbations sensed by one part of the network impact the entire network, a dynamic process which is often captured by static snapshots framed in a time course experiment. To explain the underlying biochemical pathways of these complex networks, simplified representations are generated using a top-down approach based on the principle of “guilt-by-association” [1]. The observation that genes involved in a common biological process tend to be co-regulated has enabled the inference of functional associations among genes and different biological pathways in many studies of plant responses [1-4]. This conceptual approach is also the basis of the widely known databases offering genome-wide representations of networks of co-function in *Arabidopsis* (AraNet; [5]) and *Rice* (RiceNet; [6]). Although these representations are useful, a major drawback of integrating multiple data-sets to predict co-functional relationships is the lack of context specificity in which functional information depend only on a subset of key interactions rather than on the entire data set [7]. To overcome this limitation, condition-dependent approaches which limit co-functional studies of genes or metabolites to a single tissue are applied to enhance the statistical confidence for inferring tissue-specific regulatory networks. A database named SeedNet [3], which uses gene co-expression data exclusively from *Arabidopsis* seeds, is one such example. Such approaches, however, are not appropriate for studies that seek to understand the networks at the scale of the entire organism.

Plants have evolved sophisticated mechanisms to withstand herbivore attack which can be classified as defense responses that limit the extent of damage and a suite of more poorly understood tolerance mechanisms that mitigate the negative fitness effects of herbivore attack [8, 9]. Dramatic reconfigurations of these metabolic pathways spread rapidly throughout a plant during herbivore attack which suggests the existence of complex inter-dependencies in the way that each plant tissue adjusts its intrinsic physiology as part of the whole-plant response to the attack. In the past, to unravel these defense-tolerance trade-off strategies, emphasis was placed only on the response that leaves deploy, but more recently roots have been recognized as playing an integral role in a plant’s aboveground defense mechanisms as well as serving as a dynamic storage organ following herbivory-induced resource (re)-allocation [10].

Several studies have highlighted the importance of roots in synthesizing nitrogen-rich secondary metabolites involved in leaf defenses, such as nicotine in the roots of tobacco plants [11], tropane alkaloids in various *Solanaceae* species [12] and pyrrolizidine alkaloids in *Astraceae* species [13]. These systemically-activated changes in root metabolism can have a profound impact on leaf attackers [14]. The production of proteins with shoot defensive functions has also been reported in the roots of herbivore-attacked plants such as- mir1-CP in *Zea mays* [15]. The role of roots as a sink tissue that sequesters partitioned assimilates to facilitate regrowth after herbivore attack has also been uncovered as a key tolerance strategy to herbivory in many plants and involves specific herbivory signaling networks [10]. Compared to leaf-mediated aboveground defenses, the regulatory mechanisms by which roots contribute to these defenses have however remains underexplored and “-omic” approaches have recently been applied in a few cases [16, 17].

Another important area of research for plant's responses to environmental stress involves the analysis of the physiological importance of oscillations in genes and metabolites controlled by the circadian clock. The diurnal regulation of primary metabolic pathways, such as those of starch and sugar metabolism in leaves and of nitrogen metabolism in roots, involves multiple clock components as demonstrated by previous studies using plants with disrupted clock functions [18-21]. Interestingly, genes and metabolites involved in secondary metabolic pathways have also been shown to oscillate [22, 23] but the implication of these rhythms in the roots for rapid defense induction is unknown. A recent systems biology based study on root transcriptional activity identified root-specific short-rhythms of gene expression that determine the periodicity in lateral root development [24]. This study clearly highlights the power of holistic approaches in identifying unknown key elements in roots in a time course experiment. Our understanding of the biochemical pathways underlying root-based mixed tolerance-resistance strategies would certainly be enriched by applying the above mentioned top-down approaches in co-functional studies.

In order to gain a comprehensive picture of the role of roots in defense-tolerance trade-offs during shoot herbivory, we studied the dynamic behavior of genes and metabolites and their interactions in the roots of *Nicotiana attenuata* plants to which herbivore attack had been simulated to a leaf. This plant has been well-studied for its *de novo* production of defense compounds, such as phenolic derivatives [25] and acyclic diterpene glycosides in leaves following insect attack [26], responses which are controlled by signaling pathways that have been

partly elucidated. Roots of this plant synthesize alkaloids for aboveground defense [14] and control tolerance mechanisms essential for survival, such as the bunkering of C and N in roots [10].

To overcome problems associated with gene module detection using integrated data sets or the loss of information that typically accompanies the analysis of context-specific inferences, we investigated leaf herbivory responses in both treated leaves and roots in parallel using a coordinated multivariate time series analysis, as reported in [27]. From this work, we established the importance of deriving temporal information from multiple factors for the analysis of systemic responses in roots that are elicited by simulated herbivory to leaves. Distinct functional modules identified by this analysis are used to illustrate the changes in root-specific entrained gene circadian rhythms elicited by simulated leaf-herbivory and the switching of amplitude effects between leaves and roots. These findings highlight the power of this approach in depicting the multiple roles of roots in aboveground defense responses.

Results and discussion

A mosaic of co-linear up- and down-regulations in transcripts and metabolites spreads throughout the plant after OS-elicitation

To study tissue-based responses to leaf herbivory, we used a data set consisting of 134 published (Kim et al., 2011) microarray profiles of source/sink leaves and roots collected every 4 h from 3 biological replicates of treated and control *Nicotiana attenuata* plants. In treated plants, diluted oral secretion (W+OS) from larvae of the specialist herbivore *Manduca sexta* was applied into mechanically produced puncture wounds in leaves to mimic herbivory by this insect [22]. This procedure used to simulate leaf herbivory is hereafter referred to as “OS-elicitation”. Principal component analysis (PCA) clearly separated control and treated leaves at 1 and 5h and roots of control and treated plants at 9 and 13h after OS-elicitation (Figure S1). We investigated OS-elicited transcriptomic and metabolic changes in elicited (treated leaves) and un-elicited tissues (untreated systemic leaves and roots) of the same plant in a time course experiment using differentially expressed genes (DEG) analysis at each time point (FDR=0.05, $-1 \Rightarrow \text{fold change} \geq 1$). After plotting the results, we observed a strong coordination between the number of up- and down-regulated transcripts and metabolites in leaf tissues (treated and untreated) across the entire time series (Figure1). While the number of up and down-regulated transcripts decreased

after 1h of OS elicitation in treated leaves, there was a clear increase in the number of both up and down-regulated transcripts in systemic (untreated) leaves which further decreased in both the leaf tissues 9h post elicitation, the harvest time-point which marks the beginning of the dark phase. The maximum number of up and down-regulated m/z features derived from metabolite mass spectrometry (MS)-based analysis peaked in the dark phase in treated and untreated leaves. However, the number of significantly regulated metabolic features was similar in both treated and systemic leaves, although larger differences were observed in the number of regulated genes in these two tissues. We observed a common pattern of delayed responses that materialized in a subset of transcripts and m/z features, peaking 13h after OS-elicitation in both elicited and un-elicited leaves.

In clear contrast with leaf tissues, these patterns of co-linearity were not observed for the time series of both the metabolomes and transcriptomes of root tissues. Interestingly, systemic signaling elicited larger transcriptomic changes in roots compared to the systemic (untreated) leaves with direct vascular connection with treated leaves. Despite these larger transcriptomic responses, the metabolic responses of roots reached almost the same magnitude as those of systemic leaves, but the differentially regulated metabolites were more frequently found in the positive than in the negative ionization mode of the MS analysis (Figure 1, pattern b). We infer that this pattern translates from root-specific changes in nitrogen-containing metabolites which ionize poorly in the negative mode of the MS.

Surprisingly, we observed a pattern unique to root tissues that manifest itself in a similar number of induced and suppressed genes separated by a short time lag (Figure 1, pattern a). This temporal uncoupling between the up- and down-regulation of a root's gene expression machinery clearly contrasted with the highly collinear responses described above for leaves. To identify the degree of overlap between the transcriptional and metabolic responses in treated and systemic tissues, we calculated the percentage of genes showing significant differential expression in the two tissues and found higher overlap between leaf tissues (elicited and systemic) than between elicited leaves and roots, suggesting a role for distinct molecular players and pathways in roots that respond to simulated leaf herbivory (Figure S2). To assess the representation of molecular processes at each time point, we computed the enrichment of gene ontology (GO) terms for gene sets obtained by DEG analysis at each time point using the MapMan classification of biological processes for Arabidopsis (TAIRv6) [28]. Enriched GOs for up-regulated genes in elicited and systemic leaves were related to “stress”, “oxidative pentose pathway” (OPP), “lipid metabolism”,

and “cell wall” (Figure 2). The GO class that showed the largest difference between treated and untreated leaves corresponded to “secondary metabolism” and contributed to a significant number of the up-regulated transcripts at all the time points in the series. It is conceivable from the above trends that the underlying regulatory networks might have evolved to provide tight temporal control over resource partitioning in leaves. The enriched GOs for up-regulated root transcriptome signatures occurring 5h and 13h after elicitation that showed a significant match to Arabidopsis proteome included “OPP”, “stress”, “lipid metabolism”, “developmental programs” and “cellular functions” (cell wall, cell cycle/organization, DNA) as main processes. Comparatively, less information was inferred based on homology searches with Arabidopsis regarding enriched gene processes for down-regulated transcripts in the roots. To robustly disentangle the kinetics of these mechanisms, we employed informatics strategies involving coordinated multivariate analysis *without merging the time variable as a single functional entity*.

Simulated leaf herbivory triggers largely unexplored metabolic changes in roots

To interpret the downstream effects of the large transcriptional changes observed in roots upon leaf herbivory, we analyzed the time series root metabolomics data set. Figure 3A presents the overlaid chromatograms for roots of treated and untreated plants, obtained 21h after OS-elicitation of leaves. We visualized the large metabolic responses by extracting regions of differential accumulation using the DISSECT algorithm of the Data Analysis software which allows for the deconvolution of mass feature-specific traces into compound specific spectra. This step in the MS data analysis allows for the mapping of different statistically significant patterns of regulation in ion intensities of compound-specific precursor ions and fragments onto predicted deconvoluted mass spectra. For ease of interpretation, we visualized metabolites exhibiting pronounced changes in relative levels (Figure 3A), including many unknown ion peaks, on the chromatogram. The calculation of predicted molecular formulae for many of these unknown ions revealed that the simulated leaf herbivory-regulated root metabolome was replete with many nitrogen containing compounds, which may provide an explanation for the much larger number of induced ions detected in the positive which were not detected in the negative mode. This is consistent with the well-established phenomena [29] that many of nitrogen-containing molecules are only efficiently ionized in the positive ionization mode.

To facilitate the annotation of unknown molecular ions to particular metabolic pathways, we classified patterns for the set of 1728 differentially regulated m/z features (FDR=0.05,

0.5=>FC>=2) using the time series clustering software SplineCluster [30]. For simplicity, we presented 10 of the 15 clusters retrieved from this analysis (Figure 3B). The profiles of these clusters indicated that changes in metabolite accumulation in roots upon leaf OS-elicitation are extremely dynamic and detected at each time point; we therefore labeled the resulting clusters based on the time point at which highest fold-change effects were detected. The set of clusters labeled (a,b,c), (e,f,g) and h and j contained ions which accumulated respectively at 1, 5, 9 and 13h after OS-elicitation. A few of these clusters showed induced accumulations at more than one time point, especially cluster “i” that contained a large number of ions with statistically significant differential accumulation at 9, 17 and 21h after OS elicitation. Pronounced accumulation patterns were detected for free tyramine (Figure 3C) and its conjugates to phenolic derivatives, phenolic conjugates of putrescine with known defensive functions in leaves [25] and free amino acids (tryptophan and phenylalanine) (Figure S3). We also detected clear elevations in the levels of glucoside conjugates of 12-hydroxy-jasmonic acid and salicylic acid. It is unclear whether these glucoside conjugates of active defense hormones were produced in attacked leaves and transported to roots. Figure S4 represents the temporal profiles of a few unknown m/z features from cluster “i” with their predicted molecular formulae, which show an increasingly induced accumulation in the time series.

Multivariate time series analysis captures sequential transcriptomic changes in roots

To dissect the pattern of up- and down-regulated genes shifted by a harvest time in roots, we compared the transcriptional responses at 5h and 9h after elicitation and found a large overlap (68%) in the gene identities of those 2 groups, but the genes were largely of unknown function with only 40% having close homologs in Arabidopsis. Those that did have close homologs were enriched in lipid metabolism, OPP, cell wall and cell cycle associated pathways (Figure 4A). Considering this transient up-regulation of a set of genes in the root transcriptome, we postulated that the simultaneous analysis of all 6 harvest times would help to deduce the significance of this pattern and to explore new ones from the series. Studies have shown that responses to shoot herbivory in roots are controlled by signaling pathways from aboveground tissues [10]; therefore understanding shoot-root systemic defense signaling necessitated the coordinated analysis of profiled transcriptomic responses in both treated leaves and roots. In a previous study of systemic signaling in aboveground tissues [27], we developed a method to mine dynamic changes in the expression of genes by simultaneously analyzing time series with 2 binary factors (tissue type

and treatment). We used this method to characterize co-expression motifs in elicited leaves and roots. By applying the first step of this method (Figure 4B, step 1), we obtained 4 clusters of genes derived from bootstrap-based non-parametric ANOVA models [31] representing statistical structures referred to as “interactive” (treated leaves and roots behaving differently in response to OS-elicitation), “additive” (treatment responses independent of tissue type), “tissue effects” (significant difference in leaves and roots with no response to OS-elicitation), or “treatment effects” (equivalent response to OS-elicitation in both treated leaves and untreated roots).

To understand the differences in the transcriptional responses of elicited leaves and roots, we focused on the set of genes displaying an “interactive” response pattern. Figure 4B (step 2) provides a heatmap representation of the amplitudes of the time response metrics obtained from the above analyses. The time response metric represents the projection of strong effects (interactive) along the time series and was estimated by fitting different ANOVA models. Consistent with results from the above single time point analysis, we observed large differences in the responses of treated leaves and roots at 5 and 13h after elicitation. The end result of the application of our novel method was the generation of spatio-temporally-resolved OS-elicited gene clusters. These clusters were obtained by superimposing Self Organizing Maps (BL-SOM) on the scaled data which include information about differences in fold changes of OS responses in both treated leaves and roots and about the amplitude of the differences of responses between the two tissues at each time point in the series (see Methods section). We designated each cluster on the obtained maps (40x18 cells) as “interactive” motifs. As shown in our previous work using a similar comparison of treated leaves and systemic leaves [27], these interactive motifs readily identify major trends in the spatio-temporal activation of OS-elicited responses in systemic tissues.

We mined these interactive motifs to understand the role of the root transcriptome in the responses of aboveground tissues to herbivory. To this end, we mapped the genes showing the specific pattern described above of significant induction at 5h and immediate suppression at next time point (9h) onto the spatio-temporal maps and found these genes to localize entirely in the motifs labeled “R5a”. Next, we used these “interactive” motifs to refine our interpretation of the transient transcriptional changes that were unique to roots and also to identify new patterns that were not captured by the univariate approach of single time point analyses but appeared in motifs labeled R5b, R9 and R13.

Simulated leaf herbivory elicits an inversion in root-specific transcriptome rhythms

The endogenous circadian clock regulating biological rhythms allows plants to anticipate fluctuations in environmental conditions as well as certain biotic stresses [32, 33] and to regulate its physiology accordingly. Importantly, above and belowground tissues of a plant possess autonomous circadian clocks that adapt the rhythmic expression of genes and metabolites [24, 34]. Here we report and discuss the identification of novel circadian transcripts that displayed a treatment and root-specific inversion of their rhythm in response to leaf OS-elicitation.

To facilitate the interpretation of the very large transcriptomic responses observed in roots 5h after leaf elicitation, we extracted “interactive” motifs assembled at this time-point from the SOM grids and analyzed their dynamic behaviors. To this end, we scaled expression data so that both the information representing the extent of differential statistical response to the OS treatment in treated leaves and roots throughout the time series as well as the differences in fold change responses were combined into a single metric. Using these data, we achieved spatio-temporal resolution of gene expression using the SOM grids and detected two dominating motifs at 5h that reflected two different modes of gene regulation (Figure 5A). The motif labeled “R5a” represents the set of genes with large differences in fold changes (OS/C) between elicited leaves and roots but weak time response metrics for the “interactive” effect. As expected, this motif contained all the genes that showed significant differential expression in the single time point analysis ($FDR=0.05$, $F>2$). To assess the functional significance of this first group of genes, we computed the enrichment of GO terms using MapMan classification of biological processes in Arabidopsis (hyper-geometric test, $F<0.05$). With only 48% of genes showing a significant match to Arabidopsis genes, this motif appeared enriched with genes implicated in only two main processes that corresponded to transport and cell wall metabolism.

The detection of rapid modulations in the expression of transport-related genes is not surprising considering the fundamental roles that roots play in vascular transport and translocation functions, particularly during stress adaptations. It is clearly established that leaf OS-elicitation triggers the partitioning of recently fixed photoassimilates from the damaged sites to sink tissues, including roots. Passive unloading of sucrose (symplasmically or apoplasmically) diverted actively from attacked leaves has been proposed earlier, but the fact that roots of herbivory-elicited plants recruit sugar much more efficiently than do roots of control plants suggests the importance of transporter activity [10]. We screened sugar-related genes based on literature search and found few genes known to regulate sugar translocation exhibiting rapid

modulations characteristic of motif R5a (Figure S6). Reconfigurations of transporter-mediated root functions also contribute to the defense of shoots, especially in species which have evolved the capacity to defend leaves with root-produced secondary metabolites. This is especially well described in *N. attenuata*, in which nicotine synthesized in roots are rapidly transported to aboveground tissues during herbivore attack [35, 36]. JAT1, a gene from the multidrug and toxic compound extrusion transporter family, has been proposed to act as a secondary transporter responsible for unloading of nicotine in the aerial parts of the xylem and loading it into the vacuoles [37]. The expression of this transporter did not follow the mode of regulation found in motif R5a.

The overrepresentation of processes associated with cell wall metabolism in motif “R5a” suggests that leaf OS-elicitation systemically activates physical changes in roots’ cell wall. The plant cell wall is a dynamic structure that plays important roles in growth and in the interactions of plants with their environment. Simulated leaf herbivory has been shown to negatively impact short-term dynamics of root growth [38, 39]. Transcriptomic changes controlling this phenomenon have not been explored yet but may involve high amplitude effects in cell wall related genes such as those observed in motif “R5a”. Large changes in the expression of genes encoding structural components of the cell wall have been reported as part of the developmental pathway response elicited in roots during abiotic stress [40]. The consecutive changes in the cell wall structure can impact root exudation [41], have signaling functions and may also reflect changes in root “foraging behavior” that are elicited by leaf herbivory.

To better visualize process-specific gene dynamics, we further classified motif R5a into 3 clusters using k-means clustering with averaged gene expression in roots and treated leaves for all time points. Interestingly, cluster labeled “R5a-1” which covers almost 85% of the motif “R5a” was dominated by genes showing rhythmic expression. To objectively determine which genes exhibited robust circadian expression, we tested for statistically significant ($P < 0.05$) correlation ($PC > 0.75$) between the temporal expression profiles of each gene and defined model types using HAYSTACK [42]. This analysis considered parameters for minimal changes in amplitude and the signal strength as -3 (log transformed) and 0.01 respectively. According to these criteria, 3204 transcripts (89% of cluster “R5a-1”) were classified as cycling with a best fit to the “box” model type (Figure S4) in roots of control plants. In a previous study, we reported perturbations in the oscillation of genes involved in secondary metabolic pathways in response to simulated leaf herbivory in *N. attenuata* [22]. In contrast to the secondary metabolic genes that

had conventional circadian rhythms of 24h period lengths, these root transcripts oscillated with a 12h period length and all clustered into a single group with identical phases. Similarly, 2214 (60%) of these genes also showed a circadian rhythm with a 12h period length in leaves of control plants. Root-specific, short-term fluctuations of gene expression with a 6h period have previously been reported to coordinate lateral root formation in *Arabidopsis* [24] and semidiurnal (12h) oscillations controlling starch-related gene expression have been described in Cassava storage roots [43]. One very striking observation from our analysis was that 41% of these rhythmic transcripts following the “box” model type showed a phase inversion of their expression to the “spike” model (Figure S4) in the roots of OS elicited plants. In other words, while still maintaining a 12h period rhythm, these transcripts peaked now both at dusk and at dawn. These oscillating transcripts showed a fold change (OS/C) greater than 2 at 5h which decreased along the time series but the rhythm inversion was observed for all time points. This inversion of the 12h rhythms was not found in treated leaves suggesting that this rhythm change in response to leaf OS-elicitation might be driven by root-specific circadian clock components and/or effects.

To identify whether this phenomenon involved the specific recognition of insect’s OS, we compared the root transcriptional responses to OS-elicitation with those obtained after mechanical wounding alone (W+W treatment) and found no overlap, which reinforced the conclusion that these responses are highly selective to OS-elicitation. Next, we checked for the presence of a similar pattern of inversion in the transcriptional responses in systemic leaves to OS-elicitation but no transcripts showed statistically significant rhythmic patterns in systemic leaves after OS elicitation. This demonstrates that the responsiveness is selective to the systemic cues directed towards roots. Taken together, these data can be summarized by a model in which long distance signals triggered by herbivore feeding onto leaves profoundly change the expression pattern of a core set of genes in the roots by shifting their rhythmic behavior. Further experiments involving a longer time series experiment with wild-type and transgenic lines impaired in known upstream nodes of herbivory-signaling networks will be required to examine the time required for the rhythms to revert back to their original “box” model type and to understand which OS-elicited components regulate the 12h gene rhythms of roots.

An interactive motif with opposite amplitude changes in treated leaves and roots

The second component of the large transcriptional responses in roots visualized by the spatial-temporal SOM grids 5h after elicitation corresponded to motif – “R5b” (Figure5B). This

motif is characterized by a stronger “interactive” behavior compared to fold change differences between elicited leaves and roots, with up-regulation in roots and down-regulation of the same genes in treated leaves. We further partitioned motif “R5b” using Euclidean distance-based k-means clustering and analyzed the two main patterns. A cluster named – “R5b-1” represents the set of genes with a similar inversion of a 12h periodic rhythm in roots in response to OS-elicitation but which exhibits significant down-regulation in gene expression in treated leaves 5h after elicitation (heatmap of Figure 5B). Compared to other clusters, a cluster named – “R5b-2” consists of few genes each of which showed a higher constitutive level in roots compared to leaves and is well represented by genes implicated in the nicotine biosynthetic pathway. Temporal profiles of these nicotine biosynthetic genes (*NaPMT1*, *NaPMT2*, *NaA622*, *NaQPT*, *NaDAO*) correspond to a rhythmic pattern with peaks at 9h and 17h. As previously observed, OS-elicitation in leaves induced an increase in the expression of these genes which lasted for all time points but we also noticed that peaks shifted from 9 to 5h compared to control plants. Although, we did not observe particular rhythms in treated leaves, expression of these genes showed a significant and consistent down-regulation in this tissue. Nicotine synthesis takes place in the roots and a role for the residual expression of nicotine structural genes in other plant parts has not been assigned yet. Consistent with our observations, previous work has reported the down-regulation of *PUTRESCINE N-METHYL TRANSFERASE (PMT)* transcript levels in treated leaves in *Nicotiana tabacum* [44]. Investigating other genes in the cluster “R5b-2” with similar leaf-to-root expression pattern as that of nicotine biosynthetic genes may help identifying unknown elements of this pathway or make connections with other biochemical pathways.

A chronological perspective on the arrangement of root gene expression motifs using SOM

The main benefit of retaining the temporal information of a time series experiment rather than merging it as a single variable within the multivariate analysis is the ability to directly visualize the chronology of activation of known metabolic pathways central to signaling, tolerance and defense in *N. attenuata*. Figure 6A reveals the location of a few examples of known elements in these pathways onto “interactive” motifs of the SOM grid. For instance, *NaGAL83*, the β -subunit of the SNF1-related kinase which mediates herbivory-induced allocation of sugars to roots of *N. attenuata* [10], is detected in the motif named “R1”. We found sugar metabolic enzymes and transporters located in motifs “R5b” and “R5a” respectively, reinforcing the role of these motifs in adapting the root physiology to the tolerance response. Known and putative

metabolism-based defense mechanisms are more represented in later established gene interactive motifs: *NaPMT1* from the nicotine biosynthetic pathway is located in motifs “R5b” and “R9” and *NaLOX1* which initiates the production of fatty acid 9-hydroperoxides and hydroxides, which play central roles in root signaling [45] and in the production of defensive oxylipins -- is found in motif “R13”.

Conclusions and perspectives

Approaches used to prioritize genes for functional characterization aim at identifying the most promising genes among a larger pool of candidates through integrative computational analyses. The rationale behind these methods is first to partition genes into modules or clusters, so that these clusters can be readily queried for criteria such as their involvement into a given cellular process. But often, the vector time corresponding to the genes’ temporal dynamics is considered as another variable.

In this article, we implemented a similar approach to mine biologically important transcriptomic patterns in roots by implementing a two step partition. In the first step, we simultaneously handled the time course and the two binary factors (tissue type: treated leaves and roots, treatment: Control and W+OS) and obtained four exclusive clusters. This approach facilitated the retention of information about the timings of gene activation. A second step of classification was applied to the cluster of genes showing an “interactive” effect and we visualized the dynamic behavior of genes separated across the times series and the tissue type (treated leaves and untreated roots) by employing SOM. From this analysis, distinct motifs reflecting various broad functions were identified. As a proof of principle, we studied motifs – “R5a” and “R5b” and discovered two major trends – OS-specific inversions in 12h root-specific rhythms and amplitude effects in nicotine biosynthetic and other metabolic genes.

We also recovered 2 additional clusters which should be studied intensively in the future to understand the role of roots in orchestrating leaf responses to attack from herbivores (Figure 6B): a cluster with an “additive” effect (genes’ responses to OS elicitation that are independent of the tissue type) and one with a “treatment” effect (genes showing responsiveness to OS elicitation in both treated leaves and untreated roots). *NaLOX6* shows a significant up-regulation in treated leaves at 1h after OS elicitation and in roots at later time-points. Consistent with a major function of this gene in roots, a close homolog in Arabidopsis has recently been characterized for stress-induced jasmonate accumulation in roots [46]. Combining together these sets of interactive,

additive and treatment effects allowed us to decipher the complex dynamics that occur in the transcriptomes and metabolomes of roots in response to leaf OS-elicitation with an unprecedented level of resolution.

Materials and methods

Microarray Data Sets and processing

We analyzed 134 published microarray expression data of *Nicotiana attenuata* from the Gene Expression Omnibus database (accession no. GSE30287), reflecting responses in treated leaves and untreated leaves and roots to simulated *M. sexta* feeding (by applying diluted insect's oral secretions into freshly created puncture wounds to a specific leaf) and a W+W treatment with the identical leaf damage (in which water was applied to the puncture wounds) for 6 whole-plant harvest time points (1, 5, 9, 13, 17, 21h after treatment). Raw intensities were log₂ and baseline transformed and normalized to their 75th percentile using the R software package, prior to statistical analysis.

Processing of Metabolomics Data

Metabolites were extracted from root samples using an extraction procedure optimized for the recovery of a wide range of metabolites of interest in *N. attenuata* [29]. Raw data files were converted to netCDF format using the export function of the Data Analysis version 4.0 software (Bruker Daltonics, Bremen, Germany) and processed using the XCMS package in R [47]. Peak detection was performed using the centwave algorithm with the following parameter settings: ppm=20, snthresh=10, peakwidth=c(5,18). Retention time correction was accomplished using the XCMS retcor function with the following parameter settings: mzwid=0.01, minfrac=0.5, bw=3. Areas of missing features were estimated using the fillPeaks method. Annotation of compound spectra derived from in-source fragmentation during ionization and corresponding ion species was performed with the BioConductor package CAMERA (v1.9.8) [48]. Compound spectra were built with CAMERA according to the retention time similarity, the presence of detected isotopic patterns and the strength of the correlation values among extracted ion currents (EICs) of co-eluting *m/z* features. CAMERA grouping and correlation methods were used with default parameters. Raw intensity values were 75th percentile normalized before statistical analysis. The

DISSECT algorithm implemented in DataAnalysis was used to map deconvoluted compound spectra onto the chromatographic scale.

Statistical Analysis and Data Visualization

All statistical tests for DEG analyses were carried out using R software. To mine the major biological processes perturbed in response to OS elicitation, we functionally annotated probe sets using best BlastX hit of Arabidopsis TAIR6 Proteome with an e value cut off of 1e-15. Next, using MapMan classification of biological processes for Arabidopsis, we assigned classes to each probe id of our microarray data set. Enrichment analysis of gene ontology biological processes based on hyper-geometric test was performed using R. Significant enrichments were those with $F < 10e-10$. Multifactorial data analysis was carried out using the methods implemented in the R package TANOVAs [31]. The cluster showing interactive effect in gene expression between treated leaves and roots was scaled as

$$E_i = [F_1 \dots F_6]_i * [R_1 \dots R_6]_i^2$$

Where E_i : scaled expression, F_i : difference in fold changes (OS-elicitation/control) between treated leaves and untreated roots, R_i : response timing. Results of the multifactorial analysis, including gene annotation and scaled data used for the Self Organizing Maps are available as a spreadsheet (217588Supplemental_File1.xls) in the TvR sheet at: <http://www.plantphysiol.org/content/early/2013/05/09/pp.113.217588/suppl/DC1>. The Self Organizing Maps were constructed using BL-SOM software (http://prime.psc.riken.jp/?action=blsom_index). Hierarchical Clustering Analyses for all heatmaps are based on Euclidean distance measures and average linkage aggregation methods. All heatmaps and box-plots were created using R. The model based HAYSTACK algorithm [42] was used to identify periodicity in root and leaf transcriptomes. Euclidean distance-based k-means clustering was obtained using R.

Abbreviations

ANOVA: Analysis of variance, SOM: Self-organizing map, OS: Oral-secretion, DEG: Differentially expressed genes

Acknowledgments

We thank Dr. Matthias Schöttner, Dr. Klaus Gase and Wibke Krober for technical assistance with metabolomics and microarray analyses and Dr. Sang-Gyu Kim for scientific discussion. This work is supported by the Max Planck Society, the European Research Council advanced grant ClockworkGreen (No. 293926) to ITB and the Global Research Lab program (2012055546) of the National Research Foundation of Korea.

Competing interests

The authors declare that they have no competing interests.

Authors' contribution

JG, ITB and EG: designed research and data analysis, performed research and interpreted results, wrote the paper.

References

1. Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ: **Co-expression tools for plant biology: opportunities for hypothesis generation and caveats.** *Plant Cell Environ* 2009, **32**(12):1633-1651.
2. Persson S, Wei H, Milne J, Page GP, Somerville CR: **Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets.** *Proc Natl Acad Sci U S A* 2005, **102**(24):8633-8638.
3. Bassel GW, Lan H, Glaab E, Gibbs DJ, Gerjets T, Krasnogor N, Bonner AJ, Holdsworth MJ, Provart NJ: **Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions.** *Proc Natl Acad Sci U S A* 2011, **108**(23):9709-9714.
4. Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W: **A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis.** *Plant Cell* 2012, **24**(9):3506-3529.

5. Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY: **Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*.** *Nat Biotechnol* 2010, **28**(2):149-156.
6. Lee I, Seo YS, Coltrane D, Hwang S, Oh T, Marcotte EM, Ronald PC: **Genetic dissection of the biotic stress response using a genome-scale gene network for rice.** *Proc Natl Acad Sci U S A* 2011, **108**(45):18548-18553.
7. Gillis J, Pavlidis P: **"Guilt by association" is the exception rather than the rule in gene networks.** *PLoS Comput Biol* 2012, **8**(3):e1002444.
8. Schwachtje J, Baldwin IT: **Why does herbivore attack reconfigure primary metabolism?** *Plant Physiol* 2008, **146**(3):845-851.
9. Erb M, Lenk C, Degenhardt J, Turlings TC: **The underestimated role of roots in defense against leaf attackers.** *Trends Plant Sci* 2009, **14**(12):653-659.
10. Schwachtje J, Minchin PE, Jahnke S, van Dongen JT, Schittko U, Baldwin IT: **SNF1-related kinases allow plants to tolerate herbivory by allocating carbon to roots.** *Proc Natl Acad Sci U S A* 2006, **103**(34):12935-12940.
11. Dawson RF, Solt ML: **Estimated Contributions of Root and Shoot to the Nicotine Content of the Tobacco Plant.** *Plant Physiol* 1959, **34**(6):656-661.
12. Ziegler J, Facchini PJ: **Alkaloid biosynthesis: metabolism and trafficking.** *Annu Rev Plant Biol* 2008, **59**:735-769.
13. Ober D, Kaltenegger E: **Pyrrolizidine alkaloid biosynthesis, evolution of a pathway in plant secondary metabolism.** *Phytochemistry* 2009, **70**(15-16):1687-1695.
14. Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT: **Nicotine's defensive function in nature.** *PLoS Biol* 2004, **2**(8):E217.
15. Mohan S, Ma PW, Pechan T, Bassford ER, Williams WP, Luthe DS: **Degradation of the *S. frugiperda* peritrophic matrix by an inducible maize cysteine protease.** *J Insect Physiol* 2006, **52**(1):21-28.
16. Marti G, Erb M, Boccard J, Glauser G, Doyen GR, Villard N, Robert CA, Turlings TC, Rudaz S, Wolfender JL: **Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots.** *Plant Cell Environ* 2013, **36**(3):621-639.
17. Kaplan I, Halitschke R, Kessler A, Sardanelli S, Denno RF: **Constitutive and induced defenses to herbivory in above- and belowground plant tissues.** *Ecology* 2008, **89**(2):392-406.
18. Lu Y, Gehan JP, Sharkey TD: **Daylength and circadian effects on starch degradation and maltose metabolism.** *Plant Physiol* 2005, **138**(4):2280-2291.

19. Gibon Y, Usadel B, Blaesing OE, Kamlage B, Hoehne M, Trethewey R, Stitt M: **Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes.** *Genome Biol* 2006, **7**(8):R76.
20. Gutierrez RA, Stokes TL, Thum K, Xu X, Obertello M, Katari MS, Tanurdzic M, Dean A, Nero DC, McClung CR *et al*: **Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1.** *Proc Natl Acad Sci U S A* 2008, **105**(12):4939-4944.
21. Espinoza C, Degenkolbe T, Caldana C, Zuther E, Leisse A, Willmitzer L, Hinch DK, Hannah MA: **Interaction with diurnal and circadian regulation results in dynamic metabolic and transcriptional changes during cold acclimation in Arabidopsis.** *PLoS One* 2010, **5**(11):e14101.
22. Kim SG, Yon F, Gaquerel E, Gulati J, Baldwin IT: **Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native tobacco, *Nicotiana attenuata*.** *PLoS One* 2011, **6**(10):e26214.
23. Doherty CJ, Kay SA: **Circadian control of global gene expression patterns.** *Annu Rev Genet* 2010, **44**:419-444.
24. Moreno-Risueno MA, Van Norman JM, Moreno A, Zhang J, Ahnert SE, Benfey PN: **Oscillating gene expression determines competence for periodic Arabidopsis root branching.** *Science* 2010, **329**(5997):1306-1311.
25. Kaur H, Heinzel N, Schottner M, Baldwin IT, Galis I: **R2R3-*NaMYB8* regulates the accumulation of phenylpropanoid-polyamine conjugates, which are essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*.** *Plant Physiol* 2010, **152**(3):1731-1747.
26. Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT: **Jasmonate and ppHsystemin regulate key Malonylation steps in the biosynthesis of 17-Hydroxygeranyllinalool Diterpene Glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*.** *Plant Cell* 2010, **22**(1):273-292.
27. Gulati J, Kim SG, Baldwin IT, Gaquerel E: **Deciphering Herbivory-Induced Gene-to-Metabolite Dynamics in *Nicotiana attenuata* Tissues Using a Multifactorial Approach.** *Plant Physiol* 2013, **162**(2):1042-1059.
28. Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M: **MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** *Plant J* 2004, **37**(6):914-939.
29. Gaquerel E, Heiling S, Schoettner M, Zurek G, Baldwin IT: **Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced**

- changes in *Nicotiana attenuata* leaves during simulated herbivory. *J Agric Food Chem* 2010, **58**(17):9418-9427.
30. Heard NA, Holmes CC, Stephens DA: **A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: An application of Bayesian hierarchical clustering of curves.** *J Am Stat Assoc* 2006, **101**(473):18-29.
 31. Zhou BY, Wong WH: **A Bootstrap-Based Non-Parametric Anova Method with Applications to Factorial Microarray Data.** *Stat Sinica* 2011, **21**(2):495-514.
 32. Izawa T: **Physiological significance of the plant circadian clock in natural field conditions.** *Plant Cell Environ* 2012, **35**(10):1729-1741.
 33. Wang W, Barnaby JY, Tada Y, Li H, Tor M, Caldelari D, Lee DU, Fu XD, Dong X: **Timing of plant immune responses by a central circadian regulator.** *Nature* 2011, **470**(7332):110-114.
 34. James AB, Monreal JA, Nimmo GA, Kelly CL, Herzyk P, Jenkins GI, Nimmo HG: **The circadian clock in Arabidopsis roots is a simplified slave version of the clock in shoots.** *Science* 2008, **322**(5909):1832-1835.
 35. Baldwin IT: **Inducible nicotine production in native Nicotiana as an example of adaptive phenotypic plasticity.** *J Chem Ecol* 1999, **25**(1):3-30.
 36. Morita M, Shitan N, Sawada K, Van Montagu MCE, Inze D, Rischer H, Goossens A, Oksman-Caldentey KM, Moriyama Y, Yazaki K: **Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in Nicotiana tabacum.** *P Natl Acad Sci USA* 2009, **106**(7):2447-2452.
 37. Morita M, Shitan N, Sawada K, Van Montagu MC, Inze D, Rischer H, Goossens A, Oksman-Caldentey KM, Moriyama Y, Yazaki K: **Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in Nicotiana tabacum.** *Proc Natl Acad Sci U S A* 2009, **106**(7):2447-2452.
 38. Schmidt L, Hummel GM, Schottner M, Schurr U, Walter A: **Jasmonic acid does not mediate root growth responses to wounding in Arabidopsis thaliana.** *Plant Cell and Environment* 2010, **33**(1):104-116.
 39. Hummel GM, Naumann M, Schurr U, Walter A: **Root growth dynamics of Nicotiana attenuata seedlings are affected by simulated herbivore attack.** *Plant Cell Environ* 2007, **30**(10):1326-1336.
 40. Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN: **Cell identity mediates the response of Arabidopsis roots to abiotic stress.** *Science* 2008, **320**(5878):942-945.

41. Badri DV, Vivanco JM: **Regulation and function of root exudates.** *Plant Cell Environ* 2009, **32**(6):666-681.
42. Michael TP, Mockler TC, Breton G, McEntee C, Byer A, Trout JD, Hazen SP, Shen R, Priest HD, Sullivan CM *et al*: **Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules.** *PLoS Genet* 2008, **4**(2):e14.
43. Baguma Y, Sun C, Boren M, Olsson H, Rosenqvist S, Mutisya J, Rubaihayo PR, Jansson C: **Sugar-mediated semidian oscillation of gene expression in the cassava storage root regulates starch synthesis.** *Plant Signal Behav* 2008, **3**(7):439-445.
44. Sachan N, Falcone DL: **Wound-induced gene expression of putrescine N-methyltransferase in leaves of *Nicotiana tabacum*.** *Phytochemistry* 2002, **61**(7):797-805.
45. Vellosillo T, Martinez M, Lopez MA, Vicente J, Cascon T, Dolan L, Hamberg M, Castresana C: **Oxylipins produced by the 9-lipoxygenase pathway in Arabidopsis regulate lateral root development and defense responses through a specific signaling cascade.** *Plant Cell* 2007, **19**(3):831-846.
46. Grebner W, Stingl NE, Oenel A, Mueller MJ, Berger S: **Lipoxygenase6-dependent oxylipin synthesis in roots is required for abiotic and biotic stress resistance of Arabidopsis.** *Plant Physiol* 2013, **161**(4):2159-2170.
47. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G: **XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification.** *Anal Chem* 2006, **78**(3):779-787.
48. Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S: **CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets.** *Anal Chem* 2012, **84**(1):283-289.

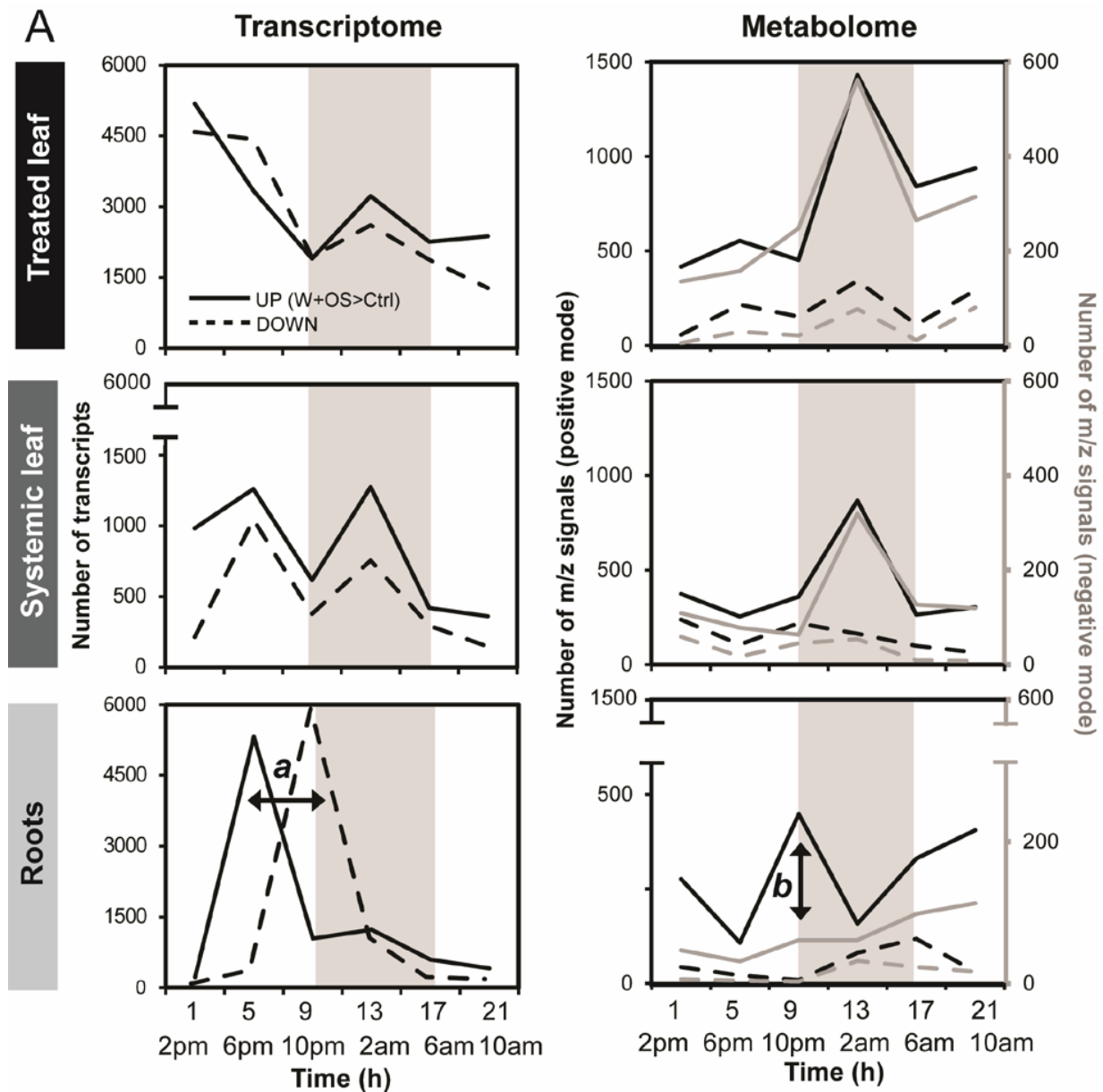


Figure 1: Tissue-specific temporal patterns of up- and down-regulated transcriptional and metabolic changes after simulated leaf herbivory in *Nicotiana attenuata*. Leaf herbivory-elicited differentially regulated transcripts and *m/z* features derived from metabolite MS analysis (Wound + OS/Control, $P \leq 0.05$ and $-1 \geq \text{fold change} \geq 1$) were identified for treated (leaves) and untreated systemic (leaves and roots) tissues using normalized and log2 transformed expression values at each harvest time. Briefly, a trend of co-linearity between up- and down-regulated transcripts and *m/z* features was observed for both treated and systemic leaf tissues across the entire time series. Interestingly, root transcriptional responses were pronounced at 5, 9

and 13h after elicitation and were as large as or larger than those in systemic leaves. The amplitude of the metabolic responses in roots was comparable to that of systemic leaves for all the time points. Two properties were particularly striking: **(a)** a pronounced temporal shift in the peaks of up- and down-regulation in gene expression that were unique to roots and **(b)** a predominance in the regulation of the positively-charged metabolome.

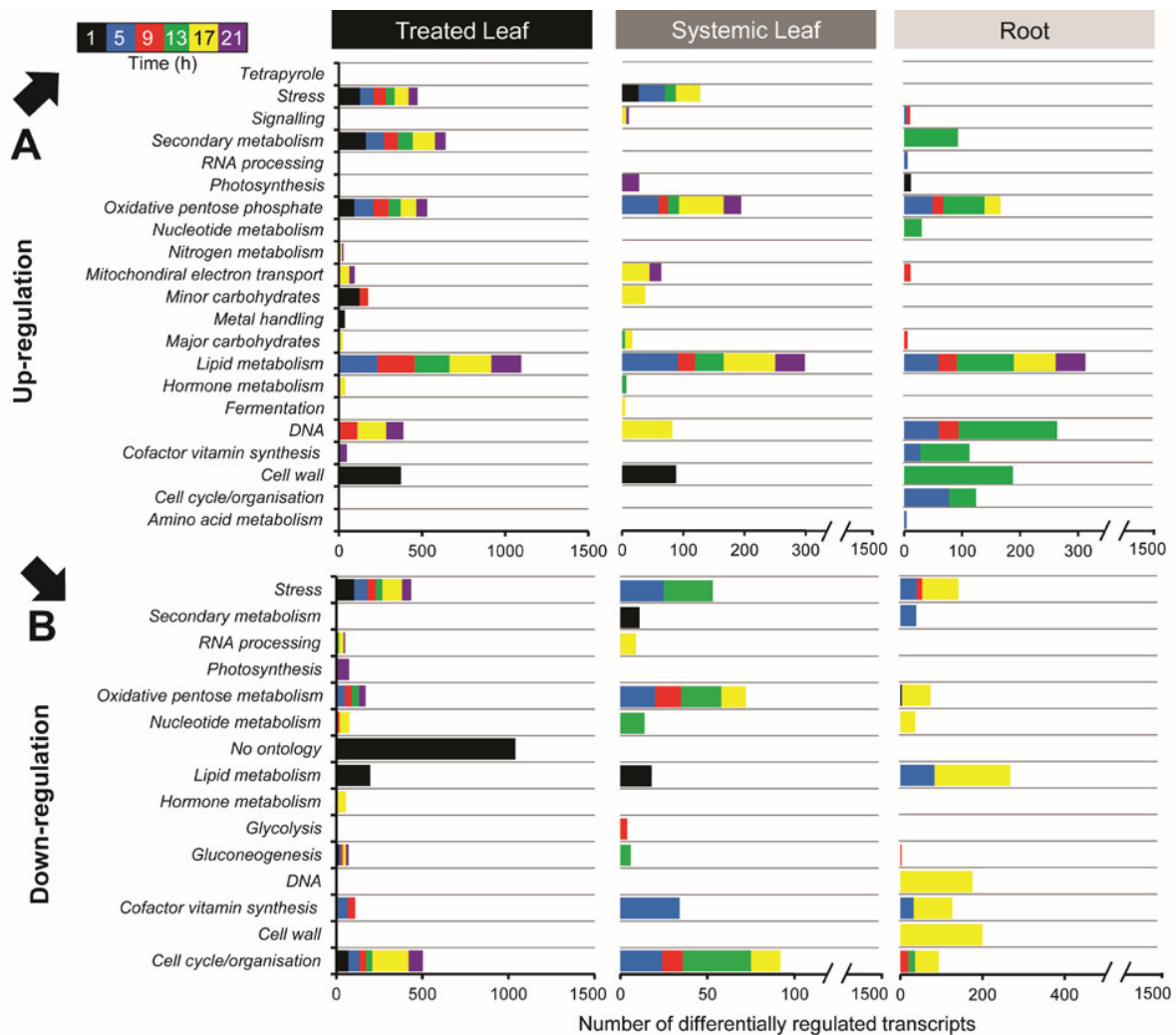
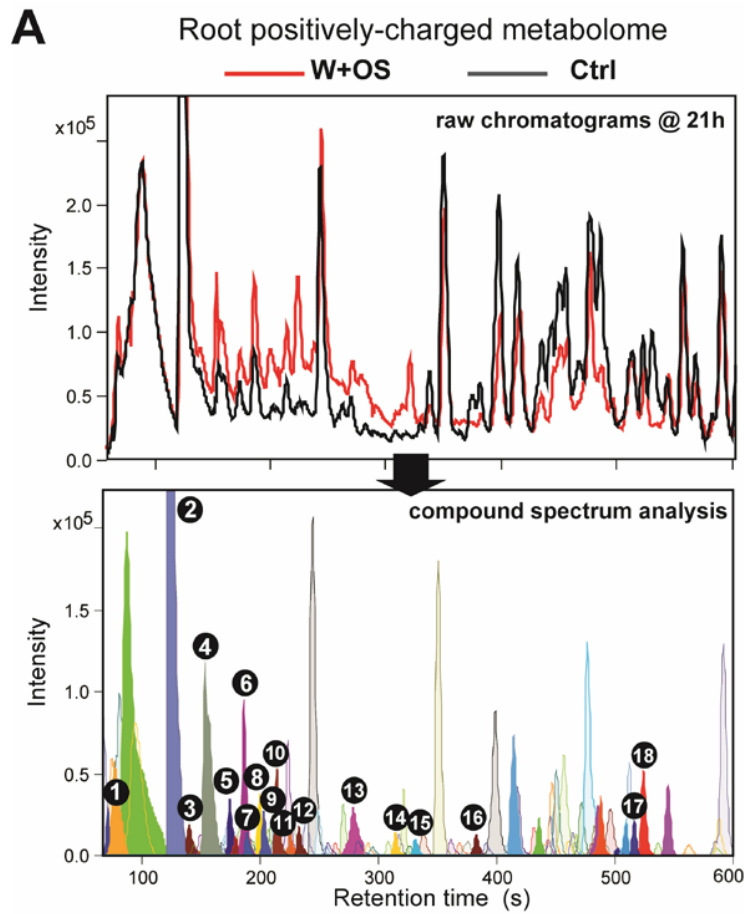


Figure 2: Functional categorization of OS-elicited whole organism transcriptional responses: A gene ontology (GO) enrichment (Hyper-geometric test, $F < 10e-10$) was performed to identify significantly up- (A) and down-regulated (B) transcripts in treated leaves and untreated systemic tissues (leaves and roots) using MapMan classification of biological processes for Arabidopsis (TAIRv6). Most perturbed processes in treated and systemic leaves correspond to GO terms involved in stress, secondary metabolism and lipid metabolism. GO classification for roots using MapMan revealed that only a small percentage of the genes exhibiting large transcriptional responses (Figure 1) have functionally characterized homologs in Arabidopsis. Overrepresented processes that were reconfigured in roots notably involved lipid metabolism, transcription and cell wall pathways.



- 1 [M+H]⁺, 341.17, C₁₆H₂₅N₂O₆
- 2 [M+H]⁺, 163.12, Nicotine
- 3 [M+H]⁺, 165.05, *p*-Coumaric acid
- 4 [M+H]⁺, 121.06, Tyramine
- 5 [M+H]⁺, 220.06, C₅H₁₀N₅O₅
- 6 [M+H]⁺, 166.08, Phenylalanine
- 7 [M+H]⁺, 251.13, *N*-caffeoylputrescine
- 8 [M+H]⁺, 122.10, C₈H₁₁N
- 9 [M+H]⁺, 265.14, *N*-Feruloylputrescine
- 10 [M+H]⁺, 205.09, Tryptophan
- 11 [M+Na]⁺, 323.07, Salicylic acid glucoside
- 12 [M+Na]⁺, 411.17, Hydroxyjasmonic acid glucoside
- 13 [M+H]⁺, 193.05, C₁₀H₉O₄
- 14 [M+H]⁺, 284.12, C₁₇H₁₈NO₃
- 15 [M+Na]⁺, 395.20, C₂₆H₂₈NaO₂
- 16 [M+Na]⁺, 397.25, C₂₀H₃₈NaO₆
- 17 [M+Na]⁺, 349.23, C₁₉H₃₄NaO₄
- 18 [M+H]⁺, 375.22, C₂₂H₃₁O₅

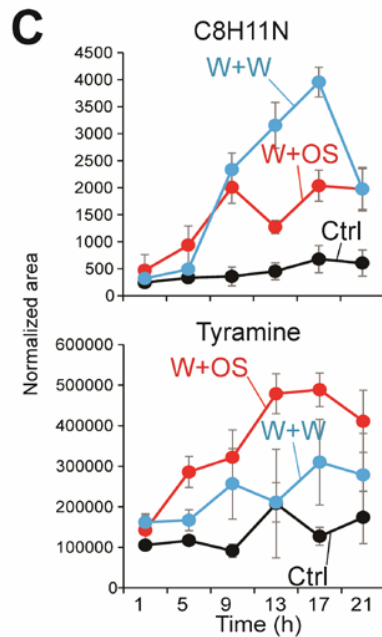
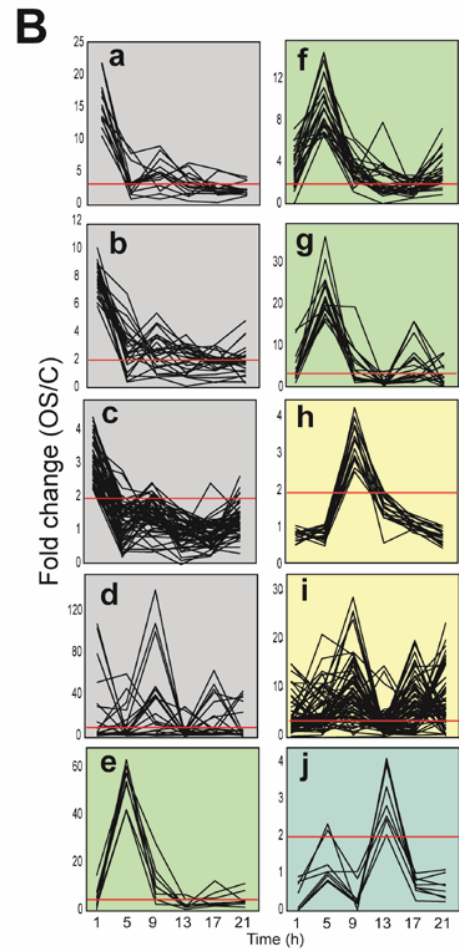


Figure 3: Leaf OS-elicitation triggered metabolic reconfigurations in roots. (A) Representative metabolic profiles obtained from the analysis by ultrahigh performance liquid chromatography quadrupole-time-of-flight mass spectrometry (UHPLC-qTOFMS) of root samples collected after 21 h from elicited and control plants. As most pronounced metabolic reconfigurations were detected in the positive ionization mode, only results obtained from this analysis are presented. In-source fragmentation of ionized molecules is common for many metabolites and the resulting fragmentation patterns are a source of redundant information for the statistical analysis. To assemble metabolite-derived fragmentation patterns and thereby facilitate metabolite annotation, we used two deconvolution tools, the R package CAMERA and the DISSECT algorithm implemented in Data Analysis (Bruker). Colored peaks correspond to compound spectra detected by DISSECT that were differentially regulated (W+OS/C, Figure 1). Annotations based on previously reported chemical analyses and elemental formula prediction are provided for 20 metabolites. **(B)** We used SplineCluster to classify the dynamics shared by different groups of metabolites. Ten of the 15 clusters are presented. Clusters are colored according to the time of maximum differential regulation and red lines denote the threshold used to identify up-regulated metabolites. **(C)** Temporal profiles of three representative m/z features from clusters labeled – “l” and “k” whose accumulation was differentially amplified in roots by the OS-elicitation in leaves and mechanical wounding treatments compared to controls.

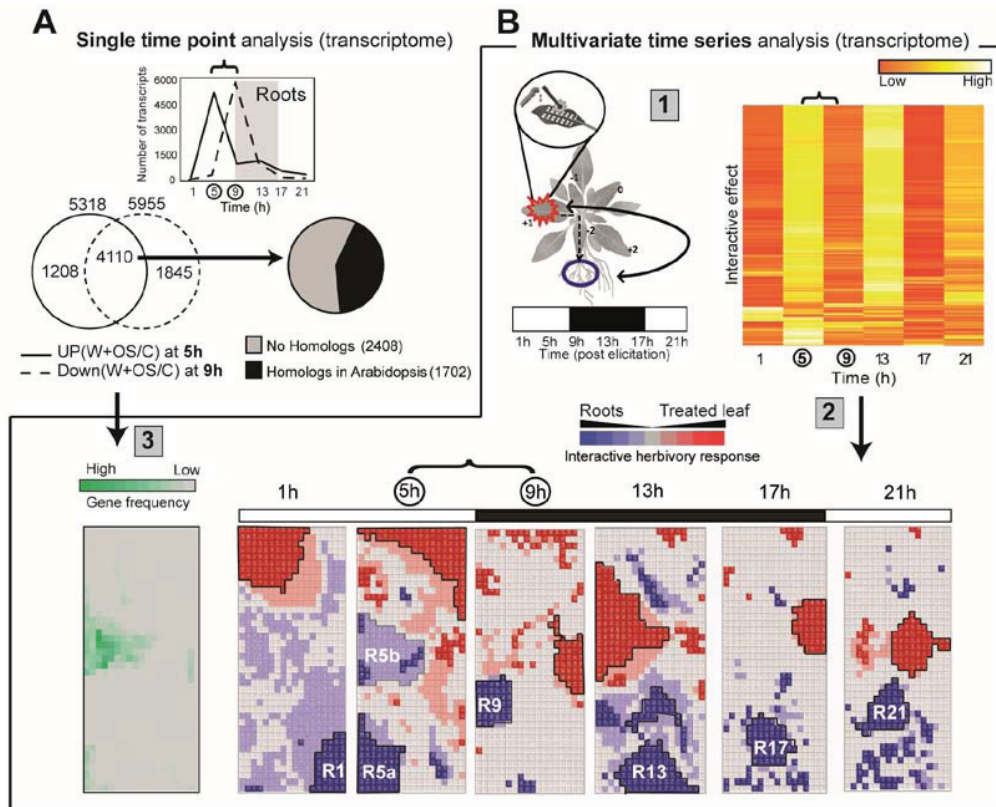


Figure 4: Multivariate analysis deciphers the complex root-specific transcriptional responses activated 5, 9 and 13h after OS-elicitation. (A) Temporally-shifted peaks of up- and down-regulation in gene expression unique to roots (Figure 1, pattern a) were observed via a single time point analysis. A large overlap (68%) is observed between induced transcripts at 5h and those suppressed at 9h. Remarkably, only 40% of these common transcripts have close homologs in *Arabidopsis*. (B) Using multivariate time series analysis, a coordinated comparison of transcriptional responses in treated leaves and untreated roots in the time series was implemented following the novel method described in (Gulati et al., 2013). Step 1 summarizes the experimental design of the microarray analysis. The heatmap depicts the extracted time response metrics which represent the significant interactive effect -- genes with statistically different transcriptional responses in treated leaves and untreated roots -- along the time series. In step 2, the genes showing interactive effect were spatio-temporally-resolved after appropriate scaling and classified by Self Organizing Maps (SOM). In step 3, transcripts common between roots' transcriptional responses at 5 and 9h that were extracted using single time point analysis were found localized in motifs labeled as "R5b" in the SOM produced by multivariate analysis.

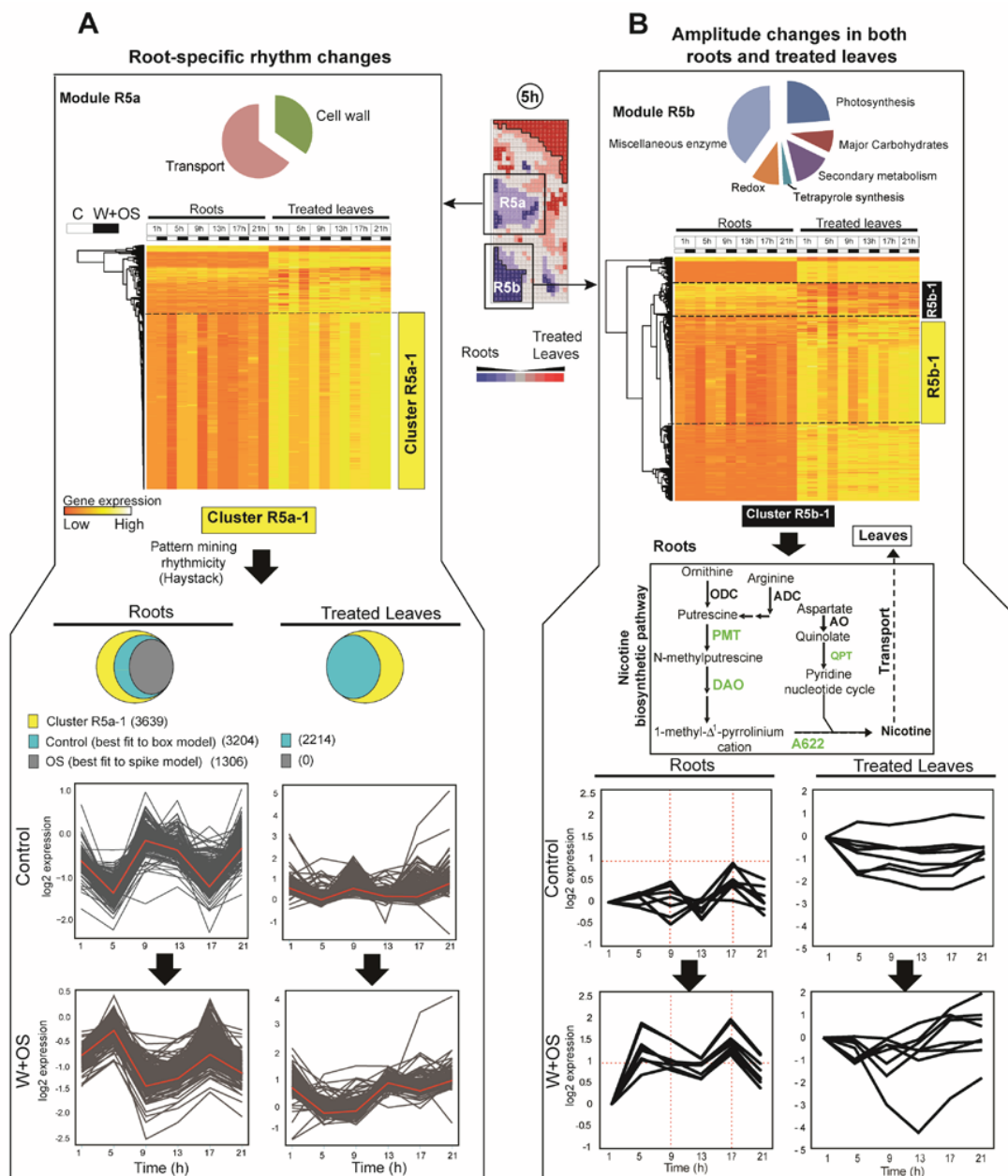


Figure 5: Induced root responses involve changes of transcriptional rhythms and amplitude effects inferred from a coordinated multivariate time series analysis. (A) Simulated leaf-herbivory elicits an inversion of the rhythm in the roots transcriptome. Interactive motif “R5a” from the leaf-to-root SOM is represented by only 48% genes showing homologs in *Arabidopsis* and is enriched with genes implicated in “transport” and “cell wall” processes. A major section of this motif, represented by sub-cluster “R5a-1” showed strong induction in roots 5h after elicitation. 89% of the genes in sub-cluster “R5a-1” oscillated in control plants and their

pattern of expression best fit the “box” model type (Haystack, $PC > 0.75$, $pvalue < 0.05$). Strikingly, the expression of 41% of these genes showed an inverted pattern of peak expression timing to dusk and dawn in response to simulated leaf herbivory. This trend is completely absent in treated leaves. **(B) A transcriptomic response with major amplitude changes in both treated leaves and roots.** Interactive motif “R5b” includes genes with strong significant interactive behavior between treated leaves and roots (gene expression being induced in roots while suppressed in treated leaves) compared to fold change (W+OS/C) and hence were not identified by the single time point analysis. Enriched GO terms are represented by 63% of the genes having close homologs in *Arabidopsis*. Sub-cluster “R5b-1” is exclusively represented by genes of the nicotine biosynthetic pathway. These genes show a root-specific shift of their peaking time only at 5h after elicitation but a major amplitude effect for all the time points in the series.

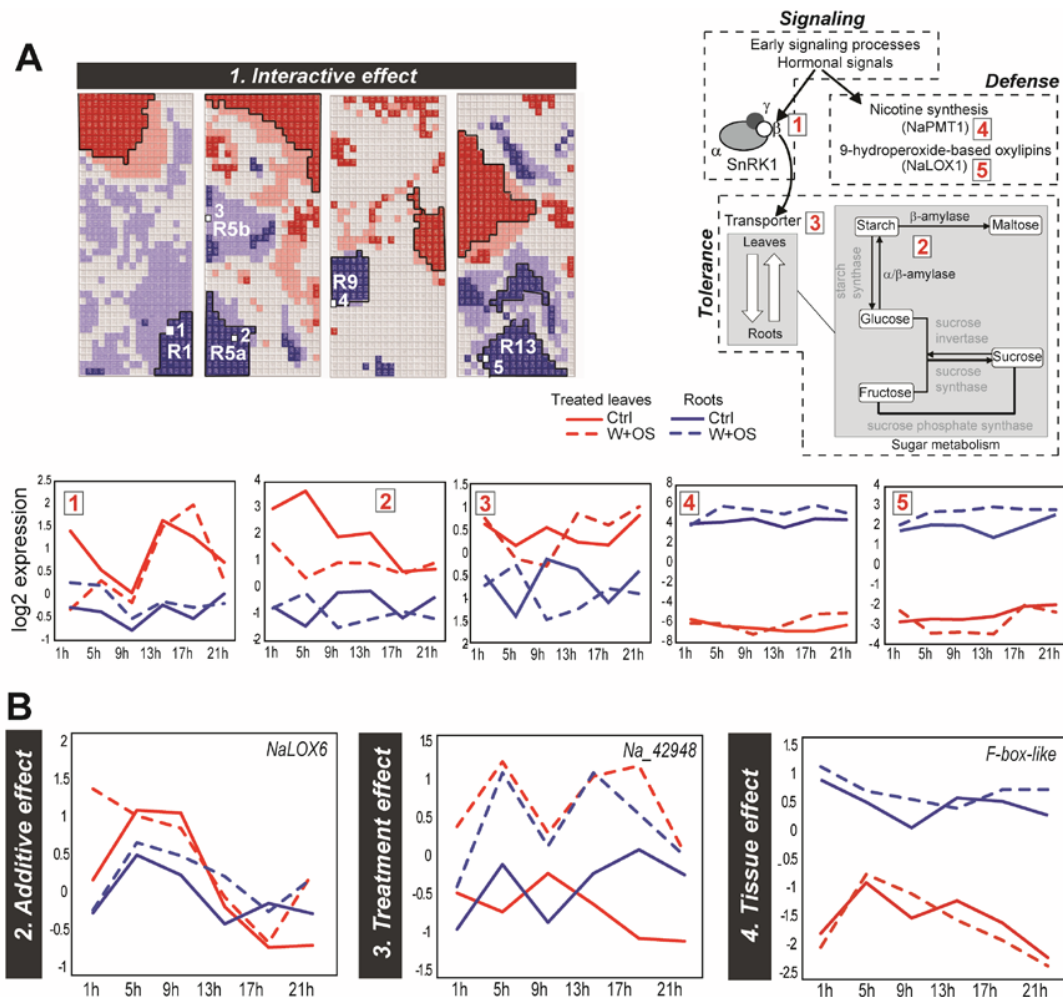
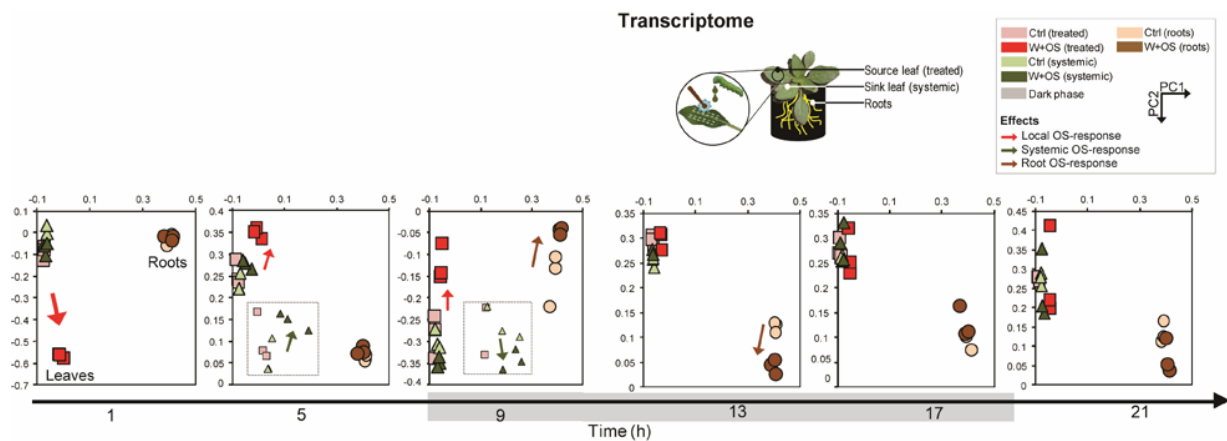
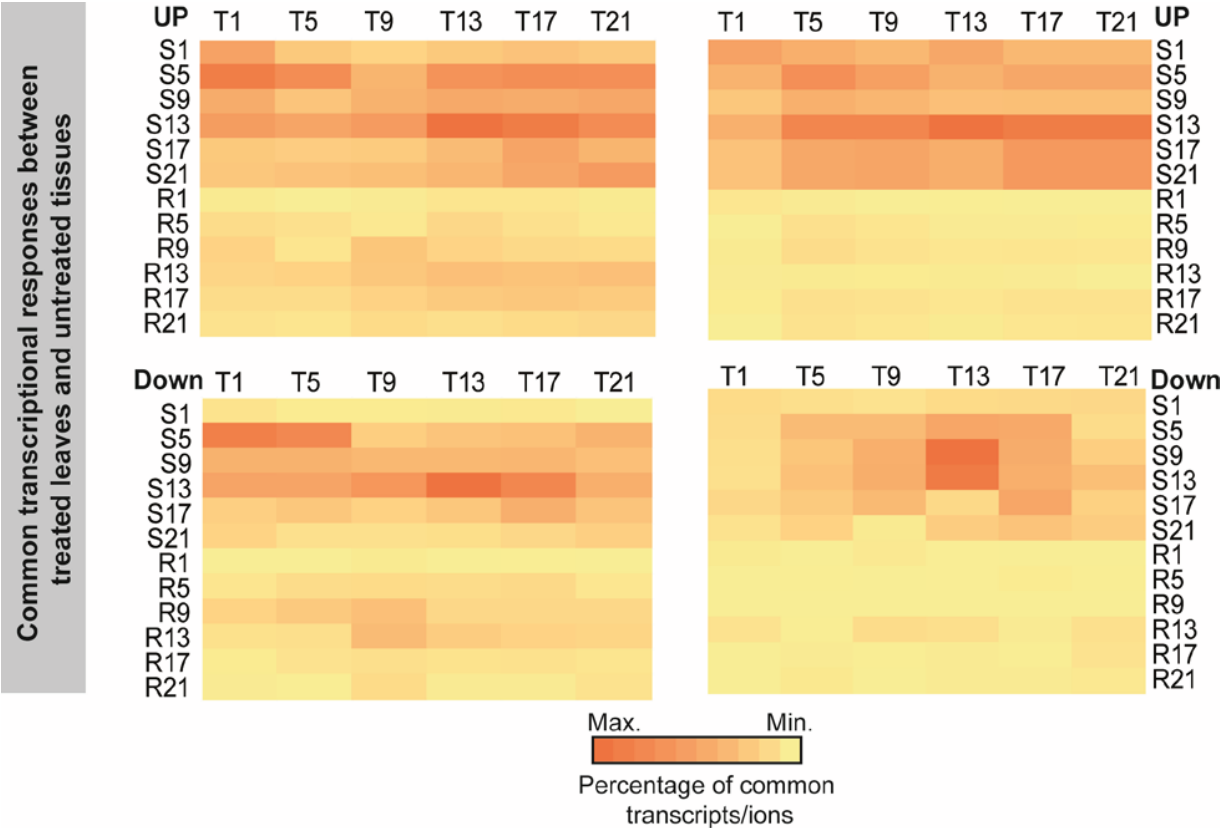


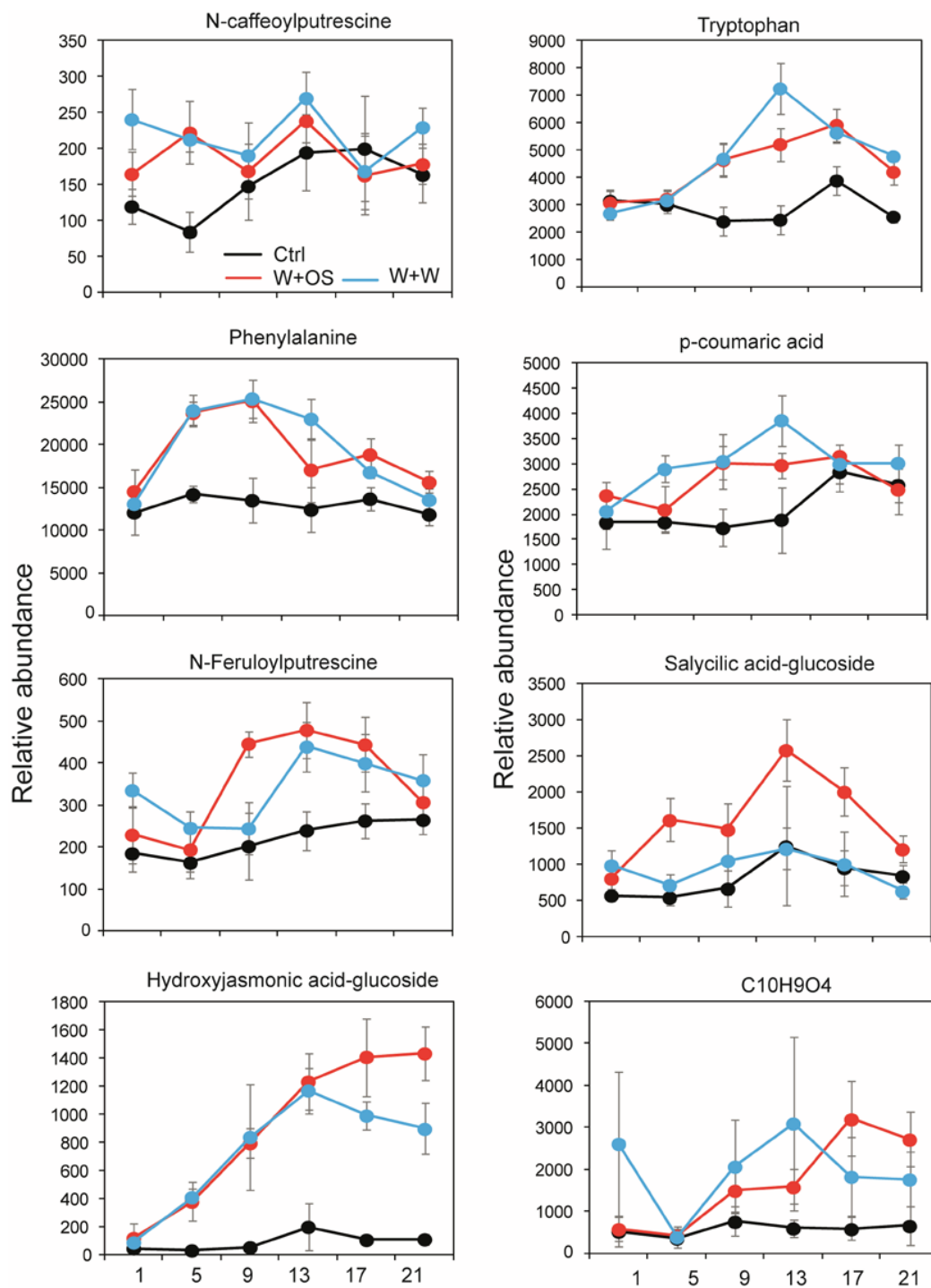
Figure 6: Sequential arrangement of major transcriptional responses identified using a multivariate time series analysis: (A) Interactive effect: Three significant components of plant responses to biotic stress – signaling, tolerance and defense-related processes –were represented by distinct interactive motifs in spatio-temporally resolved SOM. **(B) Temporal profiles of *NaLOX6*, *Na_F-Box-like* and *Na_42948*** represent three additional clusters with genes showing – Additive, Tissue and Treatment effects obtained from multivariate time series analysis.



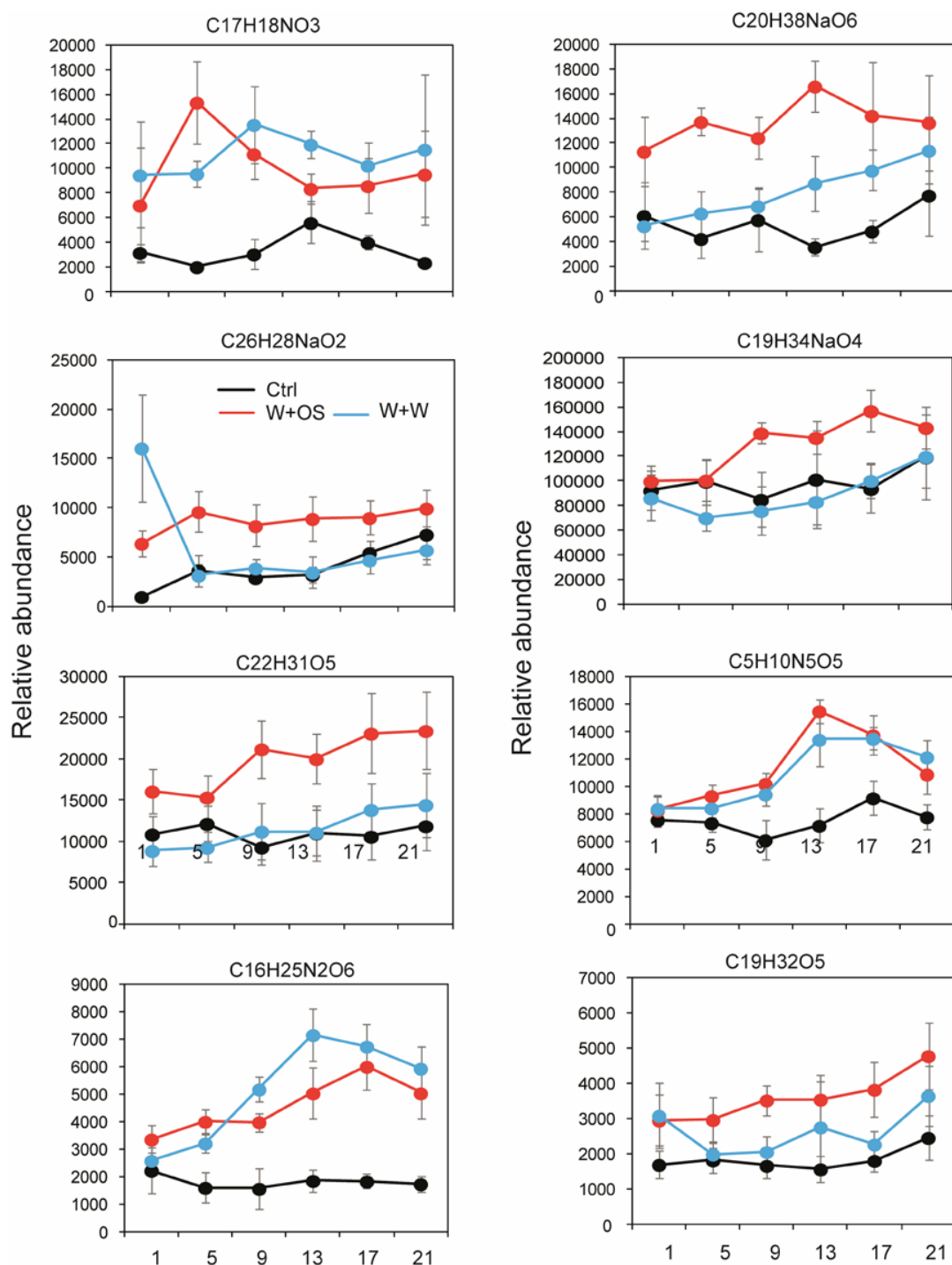
Supplementary_Figure_1. Principal component analysis of constitutive and OS-elicited modulations of leaf and root transcriptomes in *Nicotiana attenuata*. Normalized microarray data were Log₂-transformed and analyzed using factor analysis. Score plots generated for the two first principal components were used for plotting. Dashed boxes are magnifications of selected regions of the score plot. The size of the arrows is proportional to the strength of OS-elicited responses specific to a given tissue.



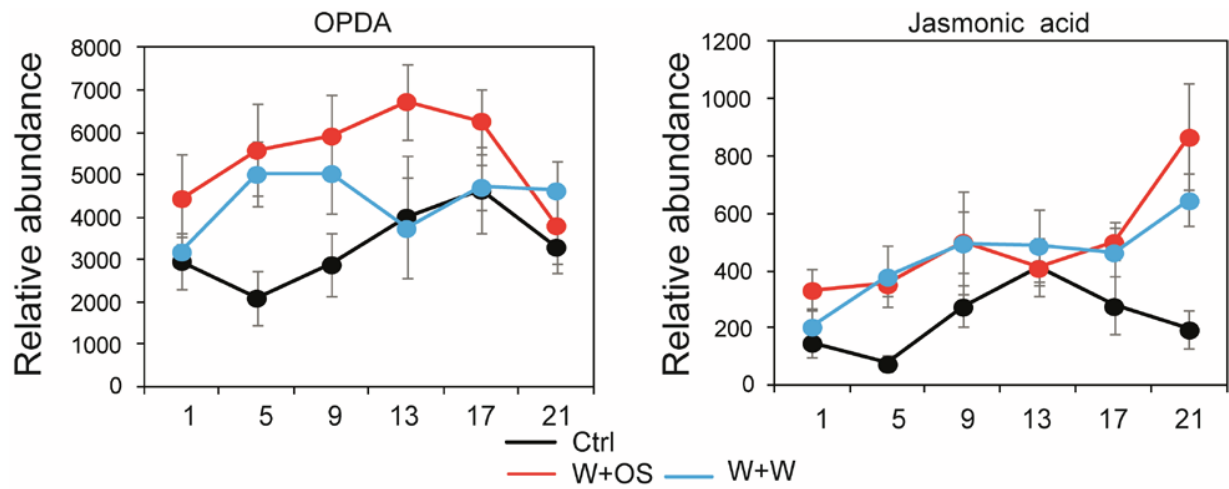
Supplementary_Figure_2. Overlap in transcriptional and metabolic responses to OS-elicitation in treated leaves, and systemic leaves and roots. Heatmap represents the percentage of the intersection set of differentially expressed genes and *m/z* features between treated and systemic tissues.



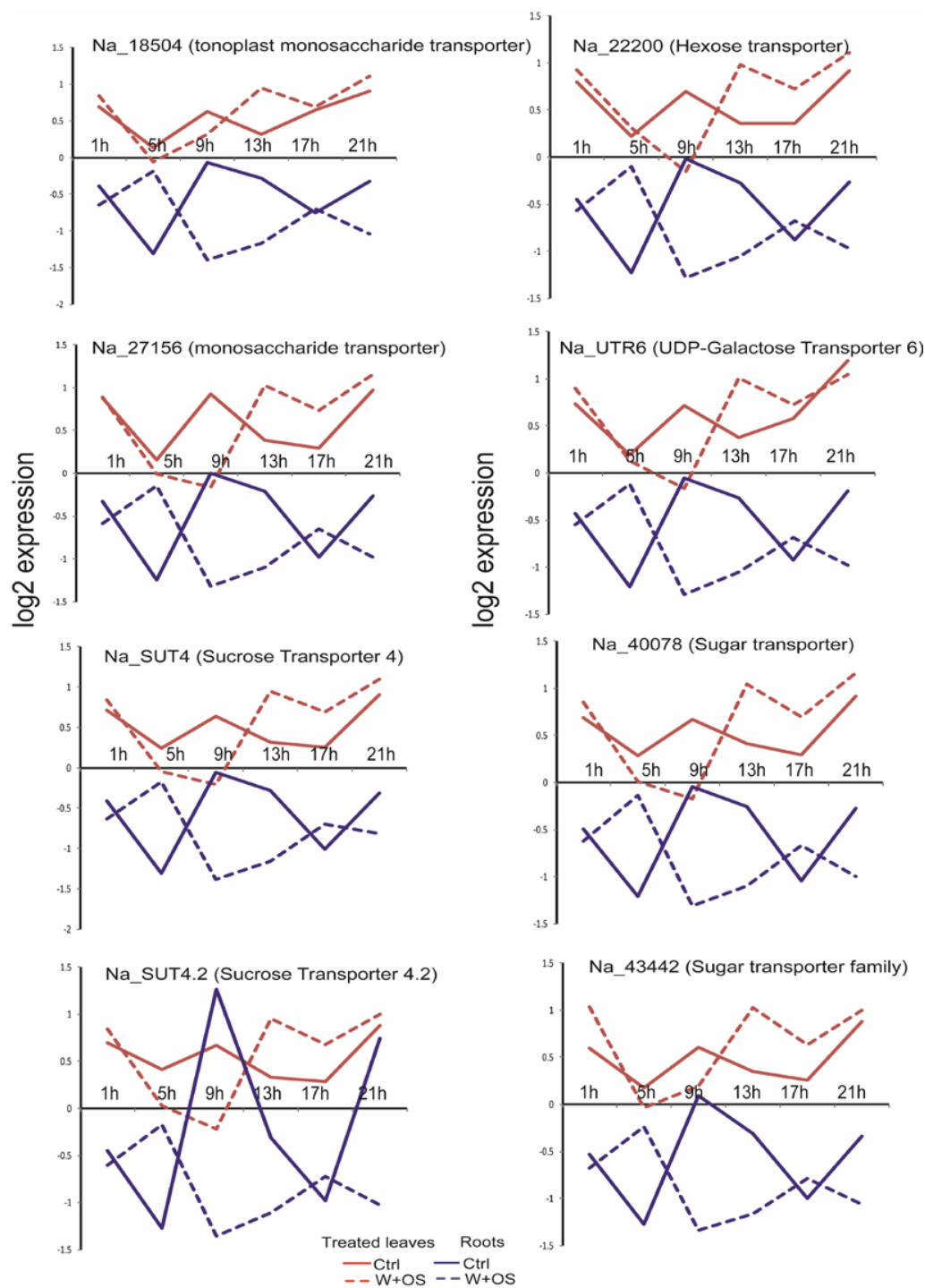
Supplementary_Figure_3. Temporal profiles of induced m/z signals in roots corresponding to free amino acids (tryptophan and phenylalanine), phenolic conjugates of putrescine, glucoside conjugates of 12-hydroxy-jasmonic acid and salicylic acid.



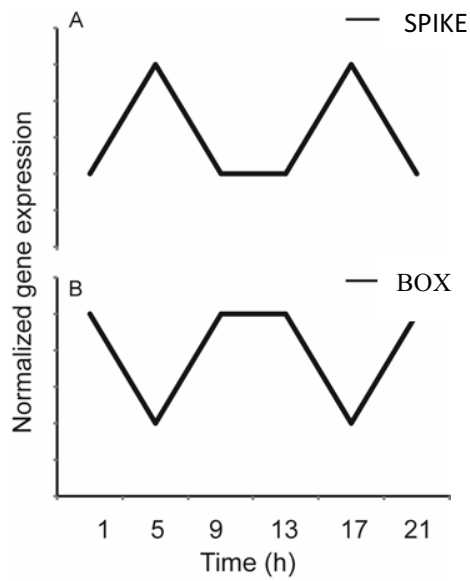
Supplementary_Figure_4. Temporal profiles of unknown m/z features showing an increasingly induced accumulation across the time course in roots in response to OS-elicitation.



Supplementary_Figure_5. Temporal profiles of oxo-phydienic acid (OPDA), the jasmonic acid metabolic precursor, and jasmonic acid, showing induction in roots in response to OS-elicitation.



Supplementary_Figure_6. Temporal profiles of genes known to be involved in sugar transport.



Supplementary_Figure_7. Models describing the oscillating root transcriptome. (A) Spike model type showing a significant match to the responses of the root transcriptome of treated plants. (B) Box model type showing a significant match to the responses of the root transcriptome of untreated plants.

Chapter 7

Discussion

Phenotypic plasticity, the ability of an organism to alter its morphology, biochemical or physiological properties in response to environmental changes, allows plants to maximize their potential fitness in variable environments (Sultan, 2000). The costs and benefits of these alternative phenotypes determine the adaptive value of phenotypic variation to environmental heterogeneity (van Kleunen and Fischer, 2005). Studies on the trade-offs associated with induced plant responses following insect herbivory suggests that induced plant defenses are examples of adaptive behavior (Mole, 1994), where defenses enhance plant fitness only under insect attack but reduce its fitness in the absence of herbivory (Agrawal, 1999). The inducible production of nicotine in *Nicotiana attenuata* during herbivory is an example illustrating adaptive phenotypic plasticity in plant defenses (Baldwin, 1999).

Plants have evolved sophisticated crisis management techniques, which include tight regulatory control over type, amount and duration of defense responses, communicating their stress status to un-attacked organs, and re-allocation of resources to undamaged sites (Wu and Baldwin, 2010). The biosynthesis of defensive compounds, as part of inducible defense systems (with direct and indirect defensive functions), is expensive and hence plants have developed strategies to discern herbivory from casual mechanical wounding which includes the perception of elicitors in insect's oral secretions. A large number of studies in *Nicotiana attenuata* have demonstrated differences in plant responses to mechanical wounding and herbivory (Halitschke et al., 2001; Wu et al., 2007; Gaquerel et al., 2009; Gilardoni et al., 2010). It has also been shown that herbivory induces defense responses not only in the damaged leaf but also in the distal systemic leaves in *N. attenuata*. It is also well established that roots of *N. attenuata* control tolerance mechanism essential for survival, such as herbivory-induced allocation of sugars to roots (Baldwin, 1990; Halitschke et al., 2001; Schwachtje et al., 2006; Wu et al., 2007; Wu and Baldwin, 2010). Additionally, periodicity in plant biology is believed to confer advantages to plants in anticipating fitness-determining environmental changes (Doherty and Kay, 2010; Pruneda-Paz and Kay, 2010). Underlying all of these complex strategies are the multiple interacting genes and gene products and their dynamics in time and space, many of which can be detected using systems biology framework.

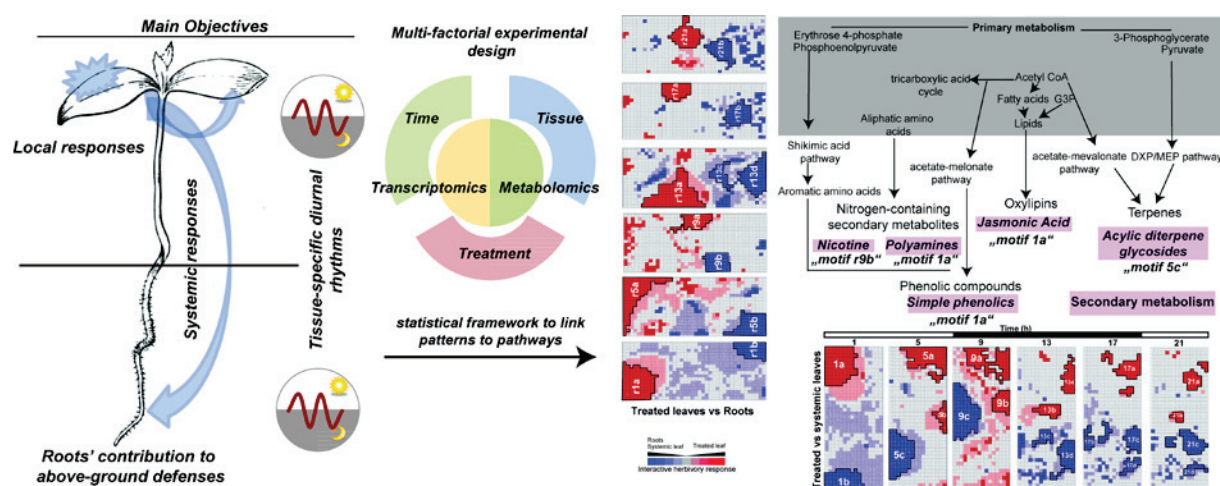


Figure 1: The main objectives of this thesis were to identify tissue-specific diurnal rhythms and their reconfigurations during leaf herbivory, extract molecular features recruited for local and systemic responses to insect attack and the contribution of roots to above-ground defenses. I designed a multi-factorial experiment consisting of transcriptomics and metabolomics profiling for three parameters: time, tissue and treatment type. I designed and implemented a statistical framework to capture metabolic reconfigurations in different tissues to elucidate their contribution to the whole-organism response and to identify the influence of particular biochemical pathways in plant-insect interactions by linking the patterns to the networks of known pathways. Spatio-temporal maps produced in this study have provided clusters of connected genes, which are separated based on their timings of activation, for major secondary metabolic pathways in *N. attenuata*

Defining structure in large scale molecular data and detecting the emergence of complexity in plant defenses are the key objectives in the study of plant-insect interactions. The objective of this study was to identify tissue-specific gene-gene and gene-metabolite associations underlying some of the above mentioned strategies which are recruited in response to insect attack in *N. attenuata*. The manuscripts presented in the thesis are based on findings of the analyses conducted on a time-course factorial designed experiment consisting of transcriptome and metabolome profiling of source/sink leaves and roots of plants with leaves treated by the application of oral secretions (OS) of *M. sexta* (Figure 1). I designed a statistical framework to identify significant and biologically meaningful patterns that characterized the transcriptional and metabolic adjustments in different tissues. This facilitated linking relevant patterns to biochemical pathways which are perturbed or activated

in plant-herbivore interactions. The statistical formulation described in this thesis allowed us to explore the three main avenues of the application of systems biology to the study of plant-insect interactions which are described in rest of the discussion section: (a) Dimensionality reduction approach, (b) “Omics” integration, (c) Emergent network properties.

1. Dimension-reduction approach to identify tissue-specific stress responsive modules

The past few years in the plant biological sciences have been characterized by the development of high-throughput technologies which provide quantitative measurements of thousands of genes and metabolites for biological samples. The increased availability of these high-dimensional observations has necessitated the development of important data-mining techniques to identify, extract and interpret biological insights from these large “omics” data sets. Techniques involving data transformation, statistical modeling and cluster analysis and aims at reducing the dimension of the data, based on the fact that low-dimensional representations often enhance the power of explanation and the prediction of biological mechanisms that govern the observed phenotype.

In chapter 4, I introduced a novel dimensionality reduction method to probe multi-dimensional molecular responses that spreads in the different tissue compartments of *N. attenuata* during herbivory. Exploratory analysis on transcriptomic and metabolic changes in treated and untreated leaves following a two group comparisons (OS vs. control) for each time point showed a pattern of strong coordination between the number of up- and down-regulated genes and metabolites across the entire time series. The above trends suggest the existence of regulatory networks that provide a tight temporal control over defense induction and resource partitioning in leaves during herbivory. This motivated the design of a statistical framework to capture the sequential arrangement of transcriptional and metabolic networks in order to identify activated/perturbed biochemical pathways and their timing of action. Since the nature and amplitude of herbivore responses in distal untreated tissues are controlled by systemic signaling spreading from treated tissues, I simultaneously analyzed dynamic responses of elicited and un-elicited tissues from same plants. The developed statistical framework is a two step procedure and involves a combination of an extended self-organizing maps-based dimensionality reduction method with bootstrap-based nonparametric analysis of variance models. The factorial structure of our experiment consisted of two binary factors – treatment type (control and W+OS treated plants) and tissue type (treated leaves, and untreated leaves

and roots). In the first step, I simultaneously analyzed time series and different factors using ANOVA structures that considered above mentioned factors and their interaction terms (Zhou et al., 2010). The most critical aspect of the method is the estimation of the optimal projected direction in the time vector space that detects the strongest significant influence of the structure. This procedure was iterated to detect the best model type for each gene and resulted in the classification of the entire gene set into four exclusive groups representing structures named as interactive, additive, major treatment and tissue effects. Genes from the interactive effect set that displayed differential responses to OS treatment in treated and untreated tissues from same plants were further evaluated and their corresponding projected vectors referred to as “time response metrics” were used in replacement of the original gene expression trajectories. The second step of our framework consisted in further resolving the temporal distribution of OS-elicited processes by superimposing self organizing maps (BL-SOM) on the scaled data. The scaling combined the information about “when to respond” (obtained from projected vector) with the information about “how to respond” (obtained from fold changes). The obtained spatio-temporal maps represent the classification of herbivory-regulated genes with patterns of differential activation in treated and untreated tissues in six dimensional time space termed as “interactive motifs”. This approach of generating sequential arrangement of functionally related gene clusters facilitated the analysis of highly dynamic responses that are characteristics of plant’s stress responses.

This method facilitated the study of dynamic behavior in more than one tissue at a time, the extraction of small sets of meaningful patterns of interactive behavior between elicited and un-elicited tissues from large data set, and the quantification of the differences in response to treatment in terms of a projected vector. The projected vector is tilted towards the time points with strongest effect and hence suggested the time of activation for each gene. The information from projected vectors when used as associative metric, improved the co-expression patterns of genes involved in similar biochemical pathways in comparison to correlations based on temporal profiles. The method helped in identifying differential regulation of members of a gene family based on their functional associations with other known genes. It also assisted the detection of coordination between members of different but connected biochemical pathways which could not be detected based entirely on their expression data. I reported such coordination for *NaGLA1* with *NaLOX3* and associated genes while analyzing oxylipin signaling pathway. Further analyses using combination of a comparative and functional genomics approach using phylogeny and time response metric could assist the identification of oxylipin branch-specific glycerolipases. Thus, the spatio-

temporal maps provide a platform for identifying novel regulatory and structural genes that mediate plant defense responses.

The use of the framework was exemplified with the analysis of systemic signaling induced pathways in untreated leaves. I isolated and analyzed a motif from SOM maps, characterized for huge OS responses in systemic leaves, using network inference methods. The motif was enriched with genes implicated in the synthesis of acyclic diterpene glycosides (17-HGL-DTGs) which are *de novo* synthesized in systemic tissues during herbivory (Heiling et al., 2010). These interactive motifs represent a group of network components that collectively contribute to the observed OS responses in the leaves at each time point; therefore they might overlap with “topological” or “functional modules” which are the set of genes involved in the same biochemical pathway. Therefore, the gene-gene network constructed for this motif showed dense connectivities between 17-HGL-DTG pathway and photosynthetic genes, supporting the previously reported but unknown coordination mechanisms between the two processes (Halitschke et al., 2001; Hui et al., 2003; Mitra and Baldwin, 2008). This example highlights the power of this approach in detecting interactions between different metabolic pathways activated at the same time in response to a perturbation.

In chapter 5, I extended the application of this method to study herbivory-specific responses in treated and un-treated leaves by differentiating induced responses to oral secretions (from the larvae of *M. sexta*) from those elicited by mechanical wounding. Tissue type dependent OS-specific responses showed large interactive effect at early time points (1h and 5h after treatment) and were enriched for two well studied secondary metabolic pathways for the biosynthesis of 17-HGL-DTGs and phenylpropanoids (Heiling et al., 2010; Kaur et al., 2010; Onkokesung et al., 2012). The functionally characterized elements of these pathways show stronger W+OS elicitation in systemic leaves compared to local leaves at 5h after elicitation. Together these results illustrate the applicability of dimension reduction techniques in identifying relevant modules.

2. “Omics” integration at spatio-temporal resolution

In chapter 4, I demonstrated that our method can also be applied on metabolomics data sets for which very few multivariate approaches have been developed. Since pre-processing strategies involving deconvolution and retention time corrections require somewhat similar metabolomes, therefore I restricted the comparison between treated and untreated leaves and identified metabolites showing major differences in induced responses between two leaves in

both positive and negative mode of ionization while using the same parameters that were used for transcriptomic analysis. The identified interactive responses for metabolic data sets were largely observed at 9 and 13h after elicitation which is later than the time points observed for transcriptional changes in systemic leaves (1 and 5h after elicitation). I supplemented the information of retention time to the interactive effects and recovered three major classes of compounds – phenolic derivatives, 17-HGL-DTG, O-acyl sugars. As with the transcriptomic analysis for a selected motif, the method facilitated the detection of tandem induction or metabolic cross talk among several metabolic pathways activated by OS elicitation in systemic tissues. Also, consistent with the power of this approach to cluster biochemically connected genes, I observed highly coordinated dynamic responses between different types of DTGs, shikimate pathway-derived amino acids and downstream metabolites produced within the phenylpropanoid pathway.

The challenge of integrating “omics” data sets is the identification of unknown or complex relationships between different molecular data types. The complexity arises because the nature of the coregulation among functionally related genes, metabolites or proteins could vary depending on the biological activity of the studied metabolic classes and on the applied experimental conditions. I showed in chapter 4 that the technique of reducing the number of features in both transcriptome and metabolome space with reference to a common six dimensional time vector allows the detection and interpretation of gene-metabolite interactions at the level of isolated motifs which are in turn used to mine activated metabolic pathways. I used the interactive responses of genes and metabolites as associative metrics to study 17-HGL-DTG pathway which I previously detected as the central metabolic pathway activated in the interactive motif exhibiting the most pronounced systemic responses. A comparison of averaged interactive responses for genes and metabolites of this pathway suggested a shift in their interactive effect. Therefore, I applied a time lag Pearson Correlation and constructed a gene-to-metabolite network which highlighted a strong coordination between 17-HGL-DTGs and their biosynthetic genes. I also reported unknown m/z signals which are likely to be involved in this pathway. These potential candidates need to be confirmed by additional mass spectrometry-based methods.

Early studies of integration of genes and metabolites to identify global dynamic responses during sulphur and nitrogen depletion in *Arabidopsis* are some of the successful examples of identifying genes' function based on their underlying gene-to-metabolite networks (Hirai et al., 2004; Hirai et al., 2005). However, the study used a hard threshold on

log ratio values for reducing the number of features before studying their dynamic responses. After filtering, both genes and metabolites were classified on the same temporal maps assuming no time lag between the activation of genes and metabolites of the same pathway. Additionally, they studied stress responses in roots and leaves separately because of which features showing coordinated dynamic responses in both tissues at the same or different time point could not be extracted from the same temporal maps. However, our multifactorial approach reduced the number of features based on their ANOVA structure (interactive effect) and the timing of activation of a “motif”, therefore it efficiently extracts the biologically meaningful information based on influence of the factors. Additionally, it allows the identification of time lag for genes and metabolites based on the few known elements of the same pathway and it collectively studies dynamic responses in treated leaves and untreated tissues which helped in the identification of emergent properties in roots.

3. Emergent network properties in plant responses to herbivory

When two or more independent analyses are combined together, the resultant properties are not always additive but infact new properties emerge that cannot be observed by single analysis. Such properties are called as emergent properties. The observation that interaction between entities in lower space constrain their behavior suggests that it is highly expected to detect emergent features in a lower dimensional space compared to uncoordinated dynamics in a bigger space (Boschetti et al., 2005). Methods to detect emergence include techniques based on inductive inference that identify low dimensional space in higher dimensional data sets. These techniques are commonly used in data compression strategies in engineering sciences but now have also been extensively used in data mining in biological sciences with potential application in the study of emergence (Long et al., 2008; Moreno-Risueno et al., 2010). One of the goals of systems biology is the identification of emergent properties via the integration of different data types within a system. In a comprehensive study to understand the important roles played by roots in above-ground herbivory using our dimensionality reduction method, I discovered an emergent physiological property of roots in the way they adapt the periodicity of its transcriptional activity to leaf OS-elicitation (Chapter 6).

From single time point analyses, I observed large transcriptional and metabolic changes in roots which surprisingly were either comparable or even more than those observed in systemic leaves. In particular, I observed a pattern unique to root tissues that consisted of the same set of induced and suppressed genes separated by a short time lag. Since a recent

study in *N. attenuata* has shown that responses to shoot herbivory in roots are controlled by hormone signaling pathways from aboveground tissues (Machado et al., 2013), I utilized our previously designed statistical framework to characterize co-expression “interactive” motifs derived on SOM maps by simultaneously analyzing elicited leaves and un-elicited roots of the same plant. I discovered two principal trends that were characterized by an inversion of root-specific semidiurnal (12h) gene oscillations and transcriptional changes with major amplitude effects while analyzing dynamic behaviors of “interactive” motifs assembled on SOM maps at 5h after leaf elicitation. The first set of semidiurnal rhythms consists of processes associated with cell wall metabolism, sugar metabolism and transport which relates to the phenomenon of root growth reduction by simulated herbivory and herbivory-induced allocation of sugars to roots observed in *N. attenuata* (Schwachtje et al., 2006; Hummel et al., 2007). The influence of signaling components on OS-responsive semidiurnal rhythms and their regulation by circadian clock components can be studied using transgenic lines impaired in known upstream nodes of herbivory-signaling networks and clock oscillators in *N. attenuata*. The second set is well represented by nicotine biosynthetic genes (*NaPMT1*, *NaPMT2*, *NaA622*, *NaQPT*, *NaDAO*) which show a rhythmic pattern in roots and induction in response to OS-elicitation.

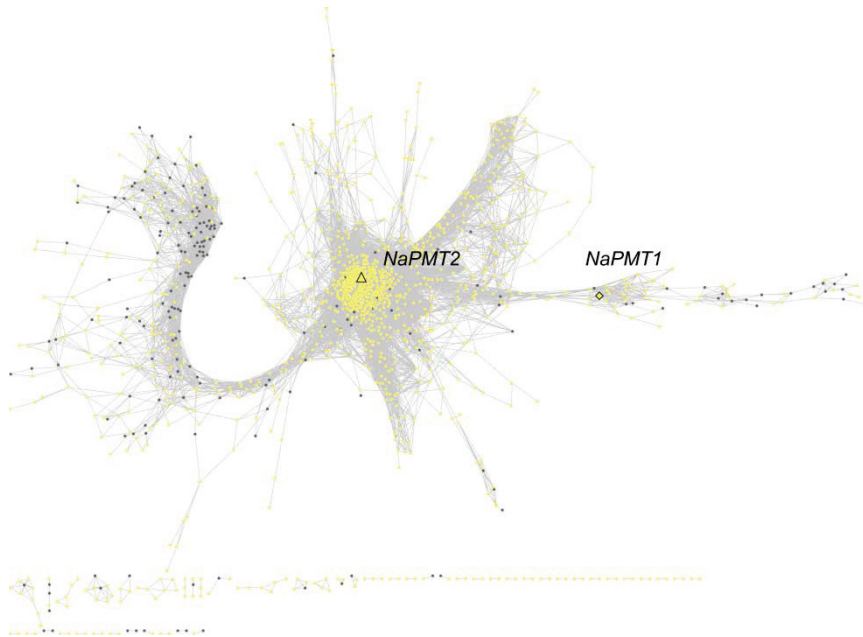


Figure 2: Gene-gene network representation of motif “R9” (connectivity screened with FDR = 0.05 and a minimum Pearson Correlation of 0.98) highlighted the differences in the degree of connectivity between *NaPMT1* and *NaPMT2* within the root gene network.

I also extracted “interactive motifs” marked on SOM maps as “R9” and analyzed using network inference algorithms which based on the differences in the network-based positions

suggested functional diversification of two *NaPMT* genes (Figure 2) that are known to be involved in nicotine biosynthetic pathway (Steppuhn et al., 2004). Similarly, I analyzed other “interactive motifs” obtained on SOM maps and visualized the chronology of activation of known pathways central to signaling, tolerance and defense in *N. attenuata* which would assist identifying other key elements from the similar or related activated pathways.

I obtained similar prospective for root tissues while analyzing the tissue-specific oscillation in primary and secondary metabolism and how these rhythms are reconfigured during insect attack in *Nicotiana attenuata* using an informatics theoretic approach of pattern recognition which finds a best match to periodic models with different amplitude and phase. I reported in chapter 3 the identification of strong diurnal rhythms in roots which were distinct from those found in leaves. While leaf metabolites mainly peaked during the day, root metabolites peaked at dusk or night. In contrast to previously reported secondary metabolic genes (Doherty and Kay, 2010) along with the metabolites reported in chapter 3 with conventional circadian rhythms of 24h period length, the set of root transcripts identified showing emergent property oscillated with a 12h period length. Other studies have also reported short periodic rhythms in roots such as root-specific, short-term fluctuations of gene expression with a 6h period length have been reported to coordinate lateral root formation in *Arabidopsis* (Moreno-Risueno et al., 2010).

Collectively, these results underscore the importance of understanding the complex physiology of roots for interpreting whole organism defense responses to leaf herbivory. These findings provide an exemplary study for identifying emergent properties that govern plant stress responses by analyzing large scale data using correctly framed biological questions in a statistical model.

Summary

The challenge to analyze and interpret large dimensional “omics” output in a variety of experimental settings using systems biology framework calls for an unprecedented integration of biological understanding and statistical methods. This work focuses on the development of methods to select features (genes and metabolites) exhibiting induced local and systemic defense responses to insect attack in *Nicotiana attenuata* along with the extraction of additional information regarding their timing of action. Using our methods, we captured the sequential transcriptional and metabolic adjustments in different tissues which elucidated the relative contribution of a given tissue to the whole-organism response.

Using pattern recognition methods on a time course experiment, we first studied tissue-specific oscillation in metabolite levels and how these rhythms are reconfigured during insect attack in *Nicotiana attenuata*. The rhythmic metabolites largely showed systemic responses in un-attacked leaves and roots. We also observed strong diurnal rhythms in roots which were distinct from those found in leaves. To further characterize the dynamics of activation in time and space of herbivory-induced responses, we designed a statistical framework by combining methods previously developed for feature selection and extraction to identify activated network motifs. These motifs are the set of features that are differentially perturbed in local and systemic tissues in response to herbivory. The extraction of multifactorial statistical information in terms of time response variable simultaneously captured the dynamic response of a gene/metabolite in more than one tissue and therefore helped in identifying tissue-specific activation of biochemical pathways during herbivory, their transition points and shared patterns of regulation with other physiological processes. Specifically, we studied gene-gene and gene-metabolite interactions in selected motifs to identify activated metabolic pathways in untreated leaf tissues that control the deployment of defense metabolism in these systemic tissues. This allowed us to identify non-linear relationships in gene-metabolite interactions of a systemically activated diterpene glycoside biosynthetic pathway. We extended the application of this method to study herbivory-specific responses in treated and un-treated leaves by differentiating induced responses to oral secretions (from the larvae of *Manduca sexta*) from those elicited by mechanical wounding. Following the observation of strong diurnal rhythms in the roots of *N. attenuata*, we utilized our framework to evaluate the transcriptional and metabolic dynamics in the roots to investigate their role in aboveground stress responses. We discovered an emergent property of

an inversion in root-specific semidiurnal (12h) rhythms in response to simulated leaf herbivory. In addition, we illustrated the benefits of our statistical framework, used for generating spatio-temporally resolved transcriptional/metabolic maps, by visualizing the chronology of the activation of pathways central to signaling, tolerance and defense in *N. attenuata*. The research described in this thesis, in addition to being valuable in deciphering dynamic responses to insect attack in a whole plant context, lays the foundation for future analyses in which statistical modeling of these networks assisted with experimental data could predict the logical rules governing these dynamic interactions.

Zusammenfassung

Die Analyse und Interpretation von multidimensionalen „omics“-Daten in einer Vielzahl experimenteller Ansätze unter Nutzung systembiologischer Grundstrukturen erfordert eine beispiellose Integration von biologischem Verständnis und statistischen Methoden. Die vorliegende Arbeit befasst sich mit der Entwicklung von Methoden zur Auswahl von Komponenten (Genen und Metaboliten), die eine induzierte und systemische Verteidigungsantwort auf Insektenfraß in *Nicotiana attenuata* hervorrufen sowie mit der Gewinnung zusätzlicher Informationen über den zeitlichen Ablauf der Reaktion. Mit unseren Methoden haben wir aufeinanderfolgende transkriptionale und metabolische Regulationen in verschiedenen Geweben erfasst, um den relativen Beitrag eines bestimmten Gewebes zur Antwort des Gesamtorganismus aufzuklären.

Mit Hilfe der Mustererkennung untersuchten wir zuerst über einen bestimmten Zeitverlauf die gewebespezifische Oszillation metabolischer Konzentrationen und wie dieser Rhythmus während des Insektenfraßes in *Nicotiana attenuata* rekonfiguriert wurde. Die rhythmischen Metabolite zeigten weitgehend systemische Antworten in nicht befallenen Blättern und Wurzeln. Zur weiteren Charakterisierung der zeitlichen und räumlichen Aktivierungsdynamik der herbivorie-induzierten Antworten entwarfen wir eine statistische Grundstruktur, indem wir die zuvor entwickelten Methoden zur Auswahl und Extraktion von Komponenten konfigurierten, um aktivierte Netzwerk motive zu identifizieren. Diese Motive entsprechen dem Satz von Eigenschaften, die differentiell in lokalen und systemischen Geweben als Reaktion auf Herbivorie gestört werden. Die Extraktion der multifaktoriellen statistischen Information hinsichtlich der variablen Antwort in Abhängigkeit von der Zeit erfasste gleichzeitig die dynamische Antwort eines Gens/Metabolits in mehr als einem Gewebe und trug daher zur Identifikation von gewebspezifischer Aktivierung biochemischer Stoffwechselwege während der Herbivorie, ihrer Umschlagpunkte und gemeinsamen Regulationsmuster mit anderen physiologischen Prozessen bei. Insbesondere untersuchten wir Gen-Gen- und Gen-Metabolit-Interaktionen innerhalb ausgewählter Motive, um aktivierte metabolische Stoffwechselwege in systemischen Blattgeweben zu untersuchen, die die Verwendung von Metaboliten zur Verteidigung in diesen Geweben steuern. Damit konnten wir nicht-lineare Zusammenhänge in

Gen-Metabolit-Interaktionen eines systemisch aktivierten Diterpen-Glycosid-Biosynthesewegs identifizieren. Wir haben diese Methode erweitert, um bei herbivorie-spezifischen Antworten in behandelten und unbehandelten Blättern zwischen induzierten Antworten auf orale Sekrete (von den Larven von *Manduca sexta*) und Antworten, die durch mechanisches Verwunden induziert werden, zu unterscheiden. Nach der Beobachtung starker Tagesrhythmen in den Wurzeln von *N. attenuata* nutzten wir unsere Grundstruktur zur Evaluierung der transkriptionalen und metabolischen Dynamik in den Wurzeln und zur Untersuchung ihrer Rolle in oberirdischen Stressantworten. Wir entdeckten das Auftreten einer Inversion des wurzelspezifischen halbtäglichen (12 h) Rhythmus als Antwort auf simulierte Blattherbivorie. Außerdem verdeutlichten wir den Nutzen unserer statistischen Grundstruktur für die Erzeugung räumlich-zeitlich aufgelöster transcriptionaler/metabolischer Karten durch Visualisierung der Chronologie der Aktivierung von zentralen Stoffwechselwegen für Signalübertragung, Toleranz und Verteidigung in *N. attenuata*. Die Forschung, die in dieser Dissertation beschrieben wird, ist nicht nur wertvoll für das Entschlüsseln dynamischer Antworten auf Insektenfraß im Kontext der gesamten Pflanze, sondern schafft auch die Grundlage für zukünftige Analysen, in denen die statistische Modellierung dieser Grundstruktur, unterstützt von experimentellen Daten, die logischen Regeln vorhersagen könnte, die für diese Interaktionen maßgeblich sind.

References

(Introduction and discussion)

- Alm E, Arkin AP** (2003) Biological networks. *Current Opinion in Structural Biology* **13**: 193-202
- Auffray C, Imbeaud S, Roux-Rouquie M, Hood L** (2003) Self-organized living systems: conjunction of a stable organization with chaotic fluctuations in biological space-time. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* **361**: 1125-1139
- Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S** (2008) Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* **320**: 938-941
- Baldwin IT** (2001) An ecologically motivated analysis of plant-herbivore interactions in native tobacco. *Plant Physiology* **127**: 1449-1458
- Baldwin IT, Morse L** (1994) Up in Smoke .2. Germination of *Nicotiana-Attenuata* in Response to Smoke-Derived Cues and Nutrients in Burned and Unburned Soils. *Journal of Chemical Ecology* **20**: 2373-2391
- Baldwin IT, Staszakozinski L, Davidson R** (1994) Up in Smoke .1. Smoke-Derived Germination Cues for Postfire Annual, *Nicotiana-Attenuata* Torr Ex Watson. *Journal of Chemical Ecology* **20**: 2345-2371
- Bar-Joseph Z, Gerber G, Simon I, Gifford DK, Jaakkola TS** (2003) Comparing the continuous representation of time-series expression profiles to identify differentially expressed genes. *Proc Natl Acad Sci U S A* **100**: 10146-10151
- Bennett RN, Wallsgrove RM** (1994) Secondary Metabolites in Plant Defense-Mechanisms. *New Phytologist* **127**: 617-633
- Breeze E, Harrison E, McHattie S, Hughes L, Hickman R, Hill C, Kiddle S, Kim YS, Penfold CA, Jenkins D, Zhang C, Morris K, Jenner C, Jackson S, Thomas B, Tabrett A, Legaie R, Moore JD, Wild DL, Ott S, Rand D, Beynon J, Denby K, Mead A, Buchanan-Wollaston V** (2011) High-resolution temporal profiling of transcripts during *Arabidopsis* leaf senescence reveals a distinct chronology of processes and regulation. *Plant Cell* **23**: 873-894

- Csete ME, Doyle JC** (2002) Reverse engineering of biological complexity. *Science* **295**: 1664-1669
- Giri AP, Wunsche H, Mitra S, Zavala JA, Muck A, Svatos A, Baldwin IT** (2006) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. VII. Changes in the plant's proteome. *Plant Physiology* **142**: 1621-1641
- Gulati J, Kim SG, Baldwin IT, Gaquerel E** (2013) Deciphering herbivory-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. *Plant Physiology* **162**: 1042-1059
- Halitschke R, Gase K, Hui D, Schmidt DD, Baldwin IT** (2003) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, sphingidae) and its natural host *Nicotiana attenuata*. VI. Microarray analysis reveals that most herbivore-specific transcriptional changes are mediated by fatty acid-amino acid conjugates. *Plant Physiology* **131**: 1894-1902
- Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT** (2010) Jasmonate and ppHsystemin Regulate Key Malonylation Steps in the Biosynthesis of 17-Hydroxygeranyllinalool Diterpene Glycosides, an Abundant and Effective Direct Defense against Herbivores in *Nicotiana attenuata*. *Plant Cell* **22**: 273-292
- Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K** (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. *Journal of Biological Chemistry* **280**: 25590-25595
- Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K** (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 6478-6483
- Ideker T** (2004) Systems biology 101 - what you need to know. *Nature Biotechnology* **22**: 473-475
- Ideker T, Galitski T, Hood L** (2001) A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics* **2**: 343-372

- Johnstone IM, Titterton DM** (2009) Statistical challenges of high-dimensional data INTRODUCTION. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences **367**: 4237-4253
- Joyce AR, Palsson BO** (2006) The model organism as a system: integrating 'omics' data sets. Nature Reviews Molecular Cell Biology **7**: 198-210
- Kang JH, Wang L, Giri A, Baldwin IT** (2006) Silencing threonine deaminase and JAR4 in *Nicotiana attenuata* impairs jasmonic acid-isoleucine-mediated defenses against *Manduca sexta*. Plant Cell **18**: 3303-3320
- Kaur H, Heinzl N, Schottner M, Baldwin IT, Galis I** (2010) R2R3-Myb8 Regulates the Accumulation of Phenylpropanoid-Polyamine Conjugates, Which Are Essential for Local and Systemic Defense against Insect Herbivores in *Nicotiana attenuata*. Plant Physiology **152**: 1731-1747
- Kessler A** (2004) Silencing the jasmonate cascade: Induced plant defenses and insect populations (vol 305, pg 665, 2004). Science **306**: 2042-2042
- Kessler A, Baldwin IT** (2002) Plant responses to insect herbivory: the emerging molecular analysis. Annu Rev Plant Biol **53**: 299-328
- Kitano H** (2002) Systems biology: a brief overview. Science **295**: 1662-1664
- Liberman LM, Sozzani R, Benfey PN** (2012) Integrative systems biology: an attempt to describe a simple weed. Current Opinion in Plant Biology **15**: 162-167
- Long TA, Brady SM, Benfey PN** (2008) Systems approaches to identifying gene regulatory networks in plants. Annu Rev Cell Dev Biol **24**: 81-103
- Mithofer A, Boland W** (2012) Plant Defense Against Herbivores: Chemical Aspects. Annual Review of Plant Biology, Vol 63 **63**: 431-450
- Moreno-Risueno MA, Busch W, Benfey PN** (2010) Omics meet networks - using systems approaches to infer regulatory networks in plants. Current Opinion in Plant Biology **13**: 126-131
- Moreno-Risueno MA, Van Norman JM, Moreno A, Zhang JY, Ahnert SE, Benfey PN** (2010) Oscillating Gene Expression Determines Competence for Periodic Arabidopsis Root Branching. Science **329**: 1306-1311
- Onkokesung N, Gaquerel E, Kotkar H, Kaur H, Baldwin IT, Galis I** (2012) MYB8 Controls Inducible Phenolamide Levels by Activating Three Novel Hydroxycinnamoyl-Coenzyme A:Polyamine Transferases in *Nicotiana attenuata*. Plant Physiology **158**: 389-407

- Park T, Yi SG, Lee S, Lee SY, Yoo DH, Ahn JI, Lee YS** (2003) Statistical tests for identifying differentially expressed genes in time-course microarray experiments. *Bioinformatics* **19**: 694-703
- Romagnolo B, Jiang M, Kiraly M, Breton C, Begley R, Wang J, Lund J, Kim SK** (2002) Downstream targets of let-60 Ras in *Caenorhabditis elegans*. *Developmental Biology* **247**: 127-136
- Sawada Y, Kuwahara A, Nagano M, Narisawa T, Sakata A, Saito K, Hirai MY** (2009) Omics-Based Approaches to Methionine Side Chain Elongation in Arabidopsis: Characterization of the Genes Encoding Methylthioalkylmalate Isomerase and Methylthioalkylmalate Dehydrogenase. *Plant and Cell Physiology* **50**: 1181-1190
- Schuman MC** (2012) Herbivory-induced volatiles function as defenses increasing fitness of the native plant *Nicotiana attenuata* in nature. *eLife*, 1: e00007. doi:10.7554/eLife.00007.
- Schwachtje J, Minchin PEH, Jahnke S, van Dongen JT, Schittko U, Baldwin IT** (2006) SNF1-related kinases allow plants to tolerate herbivory by allocating carbon to roots. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 12935-12940
- Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT** (2004) Nicotine's defensive function in nature. *PLoS Biol* **2**: E217
- Tai YC, Speed TP** (2006) A multivariate empirical Bayes statistic for replicated microarray time course data. *Annals of Statistics* **34**: 2387-2412
- Tusher VG, Tibshirani R, Chu G** (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**: 5116-5121
- Van Norman JM, Benfey PN** (2009) Arabidopsis thaliana as a model organism in systems biology. *Wiley Interdisciplinary Reviews-Systems Biology and Medicine* **1**: 372-379
- Van Regenmortel MHV** (2004) Biological complexity emerges from the ashes of genetic reductionism. *Journal of Molecular Recognition* **17**: 145-148
- Van Regenmortel MHV** (2004) Reductionism and complexity in molecular biology. *Embo Reports* **5**: 1016-1020
- Wang J, Kim SK** (2003) Global analysis of dauer gene expression in *Caenorhabditis elegans*. *Development* **130**: 1621-1634
- Wu JQ, Baldwin IT** (2010) New Insights into Plant Responses to the Attack from Insect Herbivores. *Annual Review of Genetics*, Vol 44 **44**: 1-24

- Zhou BY, Xu WH, Herndon D, Tompkins R, Davis R, Xiao WZ, Wong WH, Injury IHR**
(2010) Analysis of factorial time-course microarrays with application to a clinical study of burn injury. Proceedings of the National Academy of Sciences of the United States of America **107**: 9923-9928

Eigenständigkeitserklärung

Entsprechend der geltenden, mir bekannten Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena erkläre ich, daß ich die vorliegende Dissertation eigenständig angefertigt und alle von mir benutzten Hilfsmittel und Quellen angegeben habe. Personen, die mich bei der Auswahl und Auswertung des Materials sowie bei der Fertigstellung der Manuskripte unterstützt haben, sind am Beginn eines jeden Kapitels genannt. Es wurde weder die Hilfe eines Promotionsberaters in Anspruch genommen, noch haben Dritte für Arbeiten, welche im Zusammenhang mit dem Inhalt der vorliegenden Dissertation stehen, geldwerte Leistungen erhalten. Die vorgelegte Dissertation wurde außerdem weder als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung noch als Dissertation an einer anderen Hochschule eingereicht.

Jyotasana Gulati

Jena, August 22, 2013

JYOTASANA GULATI

Born: 16th March 1985, India

Nationality: Indian

Email: jyotasana@gmail.com

Education

2010–2013: Ph. D. (Plant systems biology)

International Max Planck Research School, Jena, Germany

Dissertation: “Integrative statistical methods for decoding molecular responses to insect herbivory in *Nicotiana attenuata*”

2006–2008: M. Sc (Bioinformatics), Gold Medal

Jamia Milia Islamia Central University, Delhi, India

Dissertation: “Open Source Development of a Distribution Using LFS (Linux from Scratch) specialized for molecular modeling and drug designing”

2003–2006: B. Sc (Life Sciences), Distinction

Maharshi Dayanand Univeristy, Haryana, India

Research Experience

- Junior Research Fellow, 2008–2009, Institute of Genomics and Integrative Biology, Delhi, India

1. Integrated insilico analysis of surface proteins from *Mycobacterium tuberculosis*.
2. A Comprehensive Immunoinformatics Analysis of *M.Tuberculosis* H37Rv and H37Ra Proteins.

- Project Intern, 2008, M. Sc, Jamia Milia Islamia Central University, Delhi, India

1. Analysis of the insertion sites of class I Non - LTR (Long Terminal Repeat) retrotransposons in Human Genome using in silico approaches.

Publications:

- **Gulati J**, Baldwin IT, Gaquerel E (2013) The roots of plant defenses: Integrative multivariate analyses uncover dynamic behaviors of roots' gene and metabolic networks elicited by leaf herbivory. (in review, The Plant Journal)
- Farhan A, **Gulati J**, Grosse-Wilde E, Vogel H, Hansson B, Knaden M. (2013). The CCHamide 1 receptor modulates sensory perception and olfactory behavior in starved Drosophila. Scientific Reports 3(2765)
- **Gulati J**, Baldwin IT, Gaquerel E (2013) An integrative statistical method to explore herbivory-specific responses in plants. Plant Signal Behav 8
- **Gulati J**, Kim SG, Baldwin IT, Gaquerel E (2013) Deciphering Herbivory-Induced Gene-to-Metabolite Dynamics in *Nicotiana attenuata* Tissues Using a Multifactorial Approach. Plant Physiol **162**: 1042–1059

- Kim SG, Yon F, Gaquerel E, **Gulati J**, Baldwin IT (2011) Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native tobacco, *Nicotiana attenuata*. PLoS One 6: e26214

Poster presentations:

- **Gulati J**, Kim S.G., Baldwin. I.T, Gaquerel E. Leaf herbivory elicits distinct asynchronous shifts in the root transcriptome of *Nicotiana attenuata*, 12th IMPRS Symposium, MPI for Chemical Ecology-2013, Jena, Germany.
- Gaquerel E., Brütting C., Schäfer M., Stanton M., Ullmann-Zeunert L., **Gulati J**, Erb M., Schöttner M., Baldwin I.T. New tools for the *Nicotiana attenuata* system: 'Real Time' genetic manipulation in nature, transcriptome-metabolome networks, fluxomics, and new imaging procedures. SAB Meeting 2012, MPI for Chemical Ecology, Jena, Germany.
- **Gulati J**, Gaquerel E., Kim S.G., Baldwin. I.T. Finding transcriptomic modules mediating plant defense responses to herbivory using systems biology. Intelligent systems for molecular biology 2012, CA, USA.
- **Gulati J**, Gaquerel E., Kim S.G., Baldwin. I.T. Studying chemical ecology of *Nicotiana attenuata* using system level factorial time course OMICS data, International Conference on Systems Biology-2011, Mannheim, Germany.
- **Gulati J**, Gaquerel E., Kim S.G., Baldwin. I.T. Inferring Stress-Responsive biochemical networks in *Nicotiana attenuata* using Integrative systems biology approaches, International Conference on Systems Biology-2010, Edinburgh center for bioinformatics.
- **Gulati J**, Gaquerel E., Kim S.G., Baldwin. I.T. Exploring molecular signatures underlying stress-responsive pathways in *Nicotiana attenuata* using Systems biology approach, 10th IMPRS Symposium, MPI for Chemical Ecology-2010, Dornburg, Germany.
- **Gulati J**, Ramachandran S., Integrated in silico identification and analysis of ligands binding to Mtb protein Antigen MPT64, Institute of Microbial Technology 2009, Chandigarh, India.

Oral presentations:

- **Gulati J**, Tissue and treatment specific transcriptomic modules directing chemical ecology in *Nicotiana attenuata*. 11th IMPRS Symposium, MPI for Chemical Ecology-2012, Dornburg, Germany.
- **Gulati J**, Computational Genomics: An Edge to Functional Genomics in Plant Systems Biology. Kick off meeting, Max-Planck Partner group-Kolkata, IISER-Kolkata, 2011.

Language Proficiency

Hindi: mother tongue

English: fluent

Acknowledgments

I would like to express my deep gratitude to my supervisors Professor Ian T. Baldwin and Dr. Emmanuel Gaquerel for their excellent guidance, never ending enthusiasm, constructive suggestions and full-time support. Thank you Ian! Thank you for trusting me and giving me the opportunity to work in such an interesting and “data-rich” research area. Your passion for Science has always motivated me. I consider myself lucky for being in your interdisciplinary lab. Thank you Emmanuel for your unflagging patience to listen to my results, bringing them in finest shape; and I believe that I will bring that fineness to the science I will do in future. You were a great resource for answering my scientific questions and I have been fortunate to enjoy your invaluable suggestions mixed with your good sense of humor.

I would also like to thank Dr. Sang-Gyu Kim, Dr. Klaus Gase, Wibke Krober, Thomas Hahn, Dr. Matthias Schottner and Felipe Yon. Thesis would not have been the same without their precious support in generating such a wonderful molecular data. I would also like to thank Dr. Karin Groten and Evelyn Classen for their ever ready support and guidance that I received since the day I arrived in Jena.

Thanks to Variluska for making my change of research from tuberculosis to tobacco plant so easy. Thanks Vari! Thanks to Michael for always coming to help whenever I needed. Thanks to Sagar and Pavan for introducing me to the world of small RNAs.

I would like to thank all the members of the department of molecular ecology for wonderful scientific discussions. I feel privileged having had the opportunity to work with them. I deeply appreciate the “instant” help provided by the wonderful people in IT department, library team, administrative departments and the IMPRS office.

Finally, I would like to thank the Max Planck Society for giving me such a big opportunity and for the financial support.