universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

Molecular Phenotyping of natural accessions of *Arabidopsis thaliana*

verfasst von / submitted by

Mag. Matthias Nagler

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2018 / Vienna, 2018

Studienkennzahl lt. Studienblatt:	A 094 437
Dissertationsgebiet lt. Studienblatt:	Biologie
Betreut von:	Univ.-Prof. Dr. Wolfram Weckwerth

Table of Contents

List of Abbreviations

ANOVA	analysis of variance
GEI	genotype-environment interaction
GIS	geographic information system
GWAS	genome-wide association studies
ODE	ordinary differential equation
PRION	proteinaceous infectious particle
SNP	single nucleotide polymorphism

Abstract

During the last decade, natural variation of the genetic model plant *Arabidopsis thaliana* has been intensively reviewed and discussed in context of many aspects of plant biological research. While numerous studies have analyzed and improved significantly our understanding of genotypic diversity, deriving explicit causal knowledge of molecular mechanisms leading to adaptation and phenotypic plasticity in a natural environment is still challenging. Due to a high degree of variation in plasticity patterns within and among populations as well as the high complexity of plant biochemical systems it is demanding to unambiguously trace back molecular processes resulting in a certain phenotype. Numerous studies have focused on the genetic background of natural variation in *Arabidopsis thaliana* and could successfully identify sets of marker genes and correlate them to traits like flowering time, climatic variables or stress tolerance. Additionally, the fast development of experimental high-throughput techniques being capable of recording thousands of components of transcriptome, proteome and metabolome simultaneously have unraveled an enormous complexity of metabolic regulation and interaction which shape the physiological homeostasis. Yet, because of this intricacy of molecular phenotypes, it is hardly possible to predict metabolism, development or natural variation patterns from current models and data sets. In this context, systems biology research comprising strategies of experimental high-throughput analysis, multivariate statistics and mathematical modeling is a promising approach for comprehensive analysis and interpretation of the systemic behavior of biochemical systems. Here, molecular analysis of metabolic networks was applied to characterize the adaptation of plant metabolism to a changing environment. In a first study, changes in metabolite concentrations, protein and phosphopeptide levels were analyzed to characterize the cold acclimation process in a cold-tolerant and a cold-sensitive natural accession of *Arabidopsis thaliana*. Accession-specific starch dynamics could be explained by a differential reprogramming of the starch degradation pathway in both accessions. To facilitate the interpretation of metabolic time series data with regard to the underlying biochemical network topology, a graph theoretical approach comparing relevant second and first order derivatives with respect to time of interpolated splines on the metabolite levels has been developed. In a third study, we applied metabolomic profiling on primary and secondary metabolites to samples collected in their natural habitat. This and subsequent mathematical modeling of regulatory instances suggests the suitability of molecular phenotyping to classify *in situ* populations of *Arabidopsis thaliana* for generation of hypotheses concerning habitat adaptations, which can be efficiently addressed in targeted follow-up experiments.

Zusammenfassung

Die Grundlagenforschung in biologischen Modellorganismen wie dem ephemeren Kreuzblütler *Arabidopsis thaliana* hat durch die Anwendung moderner molekularbiologischer Techniken und dank der Bereitstellung genetischer Ressourcen seit der Jahrtausendwende eine Vielzahl an Erkenntnissen über die Biologie von Organismen geliefert. Vor allem im Bereich der Genetik haben Sequenziertechnologien der nächsten Generation Anzahl, Umfang und Verlässlichkeit genetischer Studien auf eine neue Ebene gehoben. Dabei wurde unter anderem deutlich, dass die genetische Vielfalt innerhalb einer Art als Erklärungsansatz einer umfangreichen Palette an phenologischen und morphologischen Merkmalen dienen kann. Durch die Entwicklung und Verbesserung von chromatographie- und massenspektrometriebasierter Analytik haben wissenschaftliche Studien in letzter Zeit auch immer mehr den molekularen Phenotyp zu erfassen versucht. Speziell Metabolite und Proteine sind dabei von großem Interesse, da sie wichtige Teile von biochemischen Netzwerken sind. Durch die immer bessere Skalierbarkeit der involvierten Analytik etwa bezüglich der Anzahl an erfassten Variablen und Probendurchsatz, sind Metabolomics und Proteomics ein wichtiger Teil im Bereich der funktionellen Genomik geworden und liegen den modernen Ansätzen der Systembiologie zu Grunde. Dabei wird versucht, durch die Beschreibung und Vernetzung möglichst vieler Aspekte eines biologischen Systems das Verhalten des jeweiligen Systems mit Bezug auf bestimmte Reize wie etwa Variation der Wachstumsbedingungen zu erfassen und im Idealfall vorauszusagen. In diesem Zusammenhang habe ich zwei *Arabidopsis thaliana* Akzessionen mit variabler Frosttoleranz in ihrer Kälteakklimatisierung verglichen. Dabei konnte ich Unterschiede zwischen der kältetoleranten und der kältesensitiven Linie etwa in der Stärkemobilisierung aufzeigen. Um die Interpretierbarkeit solcher Studien zu erleichtern wurde ein graphentheoretischer Ansatz entwickelt, der die Änderung im Verhältnis der zweiten und ersten Ableitung nach der Zeit von interpolierten Splines der Metabolitzeitserien als Indikator für eine Änderung der Stoffwechselregulation heranzieht. Das Problem bei Studien, die in kontrollierten Wachstumsbedingungen durchgeführt werden, ist jedoch die Tatsache, dass eine Überlagerung von verschiedenen Stressfaktoren, wie sie im natürlichen Habitat regelmäßig vorkommen, oft zu einem unerwarteten Verhalten des untersuchten Systems führen kann. Daher habe ich einen Ansatz entwickelt, der die molekulare Charakterisierung von Pflanzenproben aus dem natürlichen Habitat ermöglicht um dann gezielte Folgeuntersuchungen anstellen zu können.

Introduction

Changing climates are imposing selective pressure on organisms. Previous studies have gone beyond merely acquiring meteorological data and found ways to quantify the impact of climate change on the biosphere (Gottfried et al., 2012; Pauli et al., 2012). These studies provide impressive information on how fluctuations in environmental parameters affect ecosystem composition. Because of rising minimum temperatures that prolong the vegetation period at high altitudes, thermophilic plant species (i.e. plant species adapted to warmer climates that usually are less frost tolerant but have higher growth rates) are able to invade and colonize elevated habitats in European mountain ranges, outcompeting native cold adapted alpine and nival species because of quicker biomass accumulation. Without the protective low temperatures that have prevented growth of thermophilic species, these more cold tolerant plant species have to shift their habitats to even higher elevations. When they reach the mountain peak and cannot climb any further, they become extinct and biodiversity is reduced. That process is not only visible in species confined to high altitudes but also in species growing in Northern latitudes (Tchebakova et al., 2009). Consequently, it threatens the present species inventory on a global scale. Unfortunately, these intriguing results do not elucidate underlying molecular processes resulting in an adapted stable metabolic homeostasis in a new environmental setup.

To remedy this drawback, biological research seeks to combine ecological research with molecular biological approaches to allow for an investigation of genotype-environment interactions (GEI). Fundamentally, each genotypical response has to be considered in context of an environment in order to produce a molecular phenotype (Weckwerth, 2003, 2011). Recent work has dealt with the impact of selection on genetic covariance data, highlighting the importance of pleiotropy (Blows and McGuigan, 2015). An approach to comprehensively analyze these molecular phenotypes was demonstrated in *Arabidopsis thaliana* by measuring and integrating metabolomic and proteomic phenotypes in dependence of different genotypes and different environmental conditions (Weckwerth et al., 2004; Morgenthal et al., 2005; Wienkoop et al., 2008; Wienkoop et al., 2010; Kleessen et al., 2012; Doerfler et al., 2013). Measuring the proteome (comprising as many proteins as possible) and metabolome (comprising as many metabolites as possible) allows for the estimation of dynamic reactions of genotypes to environments. These molecular levels are well suited for functional molecular phenotyping as abundance and posttranslational modifications cannot be predicted from genome or expression data (Weckwerth, 2011). Generally, proteome and metabolome together are key players of metabolic networks. The general topology of these biochemical reaction networks can be predicted from static genome information for model organisms (Poolman et al., 2009; Williams et al., 2010) and the translated set of proteins can be identified by shotgun-proteomics (Wienkoop and Weckwerth, 2006) and quantified in a targeted (Lehmann et al., 2008) as well as in an untargeted

approach (Hoehenwarter et al., 2011), trading off accuracy for comprehensiveness. These molecular data provide information on how biochemical pathways are regulated in a specific environmental setting. Additionally, dynamics of metabolite concentrations have to be recorded in order to estimate flow through biochemical pathways, which also is considered a property of molecular plant phenotypes (Weckwerth, 2008). Metabolomic covariance data can be fed to kinetic and structural mathematical models of metabolism (Steuer et al., 2003; Weckwerth, 2003, 2011; Sun and Weckwerth, 2012) which enables the identification of differentially regulated pathways under varying experimental conditions (Nägele et al., 2011; Doerfler et al., 2013), which is essential for the analysis and dissection of complex physiological adaptive processes.

The Species *Arabidopsis thaliana*

Arabidopsis thaliana has been established as a model organism for flowering plants since the 1940s by Friedrich Laibach, who collected plants in their natural habitats and analyzed phenological traits in common environments. As the last common ancestor of flowering plants dates back approximately 150 million years, this taxon provides a meaningful resource for elucidating fundamental plant biological processes in flowering plants (Somerville and Koornneef, 2002).

Phylogenetically, *Arabidopsis thaliana* is part of the Brassicaceae family. As such, it is closely related to genus *Arabis*, but also to crops like *Brassica napus* (rape), *Brassica oleracea* (cabbage, varieties like cauliflower and broccoli), *Amoracia rusticana* (horseradish) or *Sinapis hirta* (mustard). A compound class specific for the whole order of Brassicales are glucosinolates. These compounds are related to cyanogenic glycosides and the main aglycon groups (aliphatic, indolyl, aromatic) are derived from the seven amino acids alanine, leucine, isoleucine, valine, tyrosine, phenylalanine and tryptophan (Rask et al., 2000). They have a genotype-specific, age-specific and density-specific effect on plant fitness (Burow et al., 2010). A pivotal role of glucosinolates is herbivore repulsion. Upon tissue damage, myrosinases hydrolyze glucosinolates and cleave the glucose moiety, paving the way for conversion to toxic compounds such as thiocyanates (Rask et al., 2000). Recently, researchers have also elucidated regulatory effects of glucosinolates on root growth in multiple plant species (Malinovsky et al., 2017).

Arabidopsis thaliana is a ruderal plant and as such its habitats are often closely linked to human activity because it needs disturbed open soil to avoid being outcompeted by bigger, perennial plants. As in most winter annuals, flowering times of *Arabidopsis thaliana* is regulated by photoperiod and vernalization (Michaels et al., 2005). It is native to Eurasia, and it recolonized Europe after the last glaciation period from Northern Africa (The 1001 Genomes Consortium, 2016). Recent phylogenetic research has suggested, that the taxon has primarily originated from Southern Africa (Durvasula et al.,

2017). Climatic modeling of the biogeographical range of *Arabidopsis thaliana* has shown the suitability of most of the Northern hemisphere for supporting growth (Hoffmann, 2002). The success story of this taxon as a model organism is rooted in the phenology suitable for lab cultivation because of the small size of a maximum of 40 centimeters, unpretentious habitat demands and a short life cycle down to eight weeks (Somerville and Koornneef, 2002). It has become one of the biggest genomic mapping resources in a non-human species (Horton et al., 2012) because it is well applicable for genetic experiments (Koornneef and Meinke, 2010). Advantageous are on the one hand the relatively small genome size of approximately 135 million basepairs arranged in five chromosomes comprising around 27000 genes transcribed to some 33000 transcripts that are translated to some 35000 proteins (https://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp) and on the other hand the comparatively easy generation of mutants simply by spraying it with plasmid-containing bacteria. Indeed, novel genome editing technology has already and will continue to enhance these efforts (Doudna and Charpentier, 2014; Jia et al., 2016; Zhao et al., 2016). From an evolutionary perspective, the high selfing rate of 97% in *Arabidopsis thaliana* (Platt et al., 2010) is convenient as it ensures that natural populations are in fact mostly recombinant inbred lines.

Natural Variation in *Arabidopsis thaliana*

Natural variation in a species describes the differences in traits between a species' individuals. It is the basis for adaptive evolutionary change by natural selection. Natural variation can be observed both on the genotypic and phenotypic level. Because all species are exposed to temporally and spatially varying habitat parameters along their distribution range, natural selection acts in different ways upon the species and leads to genotypic divergence. The set of genetic differences of a species' individuals is called natural genetic variation and harbors ecologically relevant adaptive information (Weigel, 2012).

Natural Genetic Variation in *Arabidopsis thaliana*

An individual genotype is defined by the genome sequence of an individual. Naturally occurring variation of *Arabidopsis thaliana* genotypes has been studied for decades (Koornneef et al., 2004; Alonso-Blanco et al., 2009). Based on the complete sequence information of the Columbia (Col-0) accession (The Arabidopsis Genome Initiative, 2000) numerous studies on natural genetic variation in *Arabidopsis thaliana* have been undertaken, elucidating, for example, a correlation between light response of hypocotyl length and latitude of origin in 149 accessions (Maloof et al., 2001). It was also shown that the occurrence of summer annual accessions relies on mutations in only two genes, Flowering Locus C (FLC) and Frigida (FRI) (Michaels et al., 2003). Biogeographical analysis of the current distribution range of *Arabidopsis thaliana* has proven its suitability for studying adaptations to a variety of climatic conditions (Hoffmann, 2002). Population genetic studies have shown the genetic variation among populations to be higher than within populations (Bergelson et al., 1998) and demonstrated

the advantages of genome wide association mapping for analyzing *Arabidopsis thaliana* single nucleotide polymorphism (SNP) data to find candidate genes putatively causing phenotypic differences (Sharbel et al., 2000; Nordborg et al., 2002; Aranzana et al., 2005; Weigel and Nordborg, 2005; Bergelson and Roux, 2010). Natural genetic variation data has been correlated with geographic origin (Nordborg et al., 2005) and isolation by distance was found in Eurasia but not in more recently colonized North America (Platt et al., 2010). However, it was also demonstrated that climate variation among sites explained more genomic variation than mere geographic distance for 1003 accessions (Lasky et al., 2012). Based on genome-wide linkage-disequilibrium studies (Nordborg et al., 2002) that paved the way to a 250k SNP microarray (Kim et al., 2007), data provided evidence for a naturally varying trade-off between biomass production and pathogen resistance (Todesco et al., 2010). An *Arabidopsis thaliana* haplotype map was introduced (Clark et al., 2007) promoting the understanding of genetic architecture and mechanisms of adaptation (Buckler and Gore, 2007). Additionally, candidate genes for adaptations to climate could be identified (Hancock et al., 2011). Recently, high quality genomes of 1135 natural accessions of *Arabidopsis thaliana* have been published and relict populations where discovered in Southern Europe (The 1001 Genomes Consortium, 2016). This study also revealed Northern latitude to have a higher impact on natural genetic variation of *Arabidopsis thaliana* in Europe, most likely because of the orientation of European mountain ranges. Additionally, it is a formidable resource for genome-wide association studies (GWAS). In GWAS, genotypes are grouped by the nucleotide present at a specific locus and scored phenotypes are checked for significance in a statistical model reflecting the genotypic groups, for instance an ANOVA. The result is an estimated correlation of genomic regions with a given phenotype. For instance, GWAS identified variation in disease resistance protein 1 (RPM1) to be highly significantly correlated with pathogen resistance in a study scoring a total of 107 phenotypic variables and associating them with sequence polymorphisms in 199 *Arabidopsis thaliana* lines (Atwell et al., 2010). However, this technique does not test causality and is affected by linkage disequilibrium (Boyle et al., 2017), which decays in a distance of about 250 kilobases in *Arabidopsis thaliana* (Nordborg et al., 2002). Advances in epigenetic studies (Lister et al., 2008; Becker et al., 2011; Schmitz and Ecker, 2012; Schmitz et al., 2013) have culminated in the publication of 1107 methylomes and 1203 transcriptomes (Kawakatsu et al., 2016). This effort revealed geography and climate of origin to predict DNA methylation patterns, pointing towards a role of the epigenetic level in adaptive processes (Kawakatsu et al., 2016).

These studies provide the basis for further molecular and environmental research efforts to track down eco-physiological and subsequent morphological changes that in the long run are the results of adaptive processes (Malosetti et al., 2013). The difficulty in predicting phenotypes from genotype-phenotype correlations lies in the static nature of the genomic code (Weckwerth, 2011), as downstream information processing regulating biochemical activity during transcription, translation

and post-translational modifications cannot be predicted from the genome sequence. Also Crick's initial central dogma of molecular biology (Crick, 1970), has been continuously expanded to account for this reality, as prions (proteinaceous infectious particle) are able to mediate inheritance, especially in combination with genetic variation (Koonin, 2012).

Several authors demonstrated that there generally are non-linear dependencies between different levels of molecular information. For instance, freezing tolerance of nine *Arabidopsis thaliana* accessions has been extensively discussed in the context of effects on transcriptome and metabolome (Hannah et al., 2006). Two years later, it was demonstrated that the inter-experimental differences of the detected cold inducible transcripts significantly depend on the time of day the experiment was performed and moreover, that this effect cannot be cleared out by including a control group at normal temperature (Bieniawska et al., 2008). Furthermore, it was shown that there is no consistent transcriptional regulation which would allow the prediction of cold induced metabolite accumulation (Espinoza et al., 2010).

Systems Biology and OMICS-Technologies for Phenomics

Following the central dogma of molecular biology, biological systems operate in a complex and interdependently regulated network of multiple molecular levels. It is not possible to predict all traits in the life cycle only from the genome sequence (Pigliucci, 2010). Because predicting the behavior of the complex interconnected architecture of molecular levels is non-intuitive, we have to describe biological systems on as many molecular levels as possible with regard to experimental parameters. The acquired molecular phenotypes potentially elucidate the reactivity of the biochemical system to external stimuli. Consequently, a systemic view on the complex interactions of molecular levels provides interpretable insight in GEI (Kitano, 2002; Weckwerth, 2003). Phenomic research comprises analysis of these molecular phenotypes intersected with morphology and ecology.

Metabolomic technologies have become core technologies for functional genomics and molecular plant physiology (Trethewey et al., 1999; Fiehn, 2002; Weckwerth, 2003; Morgenthal et al., 2005; Tohge et al., 2005; Weckwerth, 2008; Stitt et al., 2010; Weckwerth, 2011). Metabolic steady states and transient dynamics of biological systems have been elucidated by quantification of metabolic components in response to specific environmental perturbations indicating the relation to the genotype (Weckwerth et al., 2004; Saito and Matsuda, 2010). To unravel the relation of different molecular levels experimentally, protocols for the integrative extraction, identification and quantification of metabolites, proteins, RNA and DNA from the same sample have been established (Valledor et al., 2014). Recently, tremendous effort was undertaken to analyze the reception and

molecular regulation of energy deprivation in a systemic approach combining metabolomics, proteomics and phosphoproteomics, which proved the pivotal role of SnRK1 in repressing energy intensive cellular processes like protein synthesis and even elucidated formerly unknown links of SnRK1 to the phosphorylation state of chloroplastic proteins (Nukarinen et al., 2016). Evidently, tackling specific biological research hypotheses in experiments comprising the analysis of multiple molecular levels potentially reveals multiple, previously unexpected systemic effects that broaden our horizon and lead to completely new scientific challenges.

Systems Biology in Cold Acclimation Research

Temperature and precipitation prominently shape the distribution range of *Arabidopsis thaliana* (Hoffmann, 2002). As expected, freezing tolerance therefore shows significant natural variation (Hannah et al., 2006). Most known *Arabidopsis* accessions are winter annuals (Michaels et al., 2005) which germinate in autumn and have to survive as leaf rosettes until bolting in the next spring. In parts of the distribution range, *Arabidopsis thaliana* plants have to sustain freezing temperatures during the winter months. Biogeographic analysis of the species' ecological amplitude revealed range limitation by cold winter but also cold autumn and spring temperatures (Hoffmann, 2002). Accordingly, freezing tolerance in *Arabidopsis thaliana* positively correlates with latitude of origin (Zhen and Ungerer, 2008). Environmental stimuli lead to a reprogramming of metabolism that increases stress tolerance (Kosova et al., 2011). Low but non-freezing temperatures increase the frost resistance in several temperate plant species in a complex process called cold acclimation (Thomashow, 1999) for which light is essential (Wanner and Junttila, 1999). Cold temperatures are presumably perceived via Ca^{2+} signals (Knight and Knight, 2012), changes in membrane fluidity (Los and Murata, 2004) and reorganization of the cytoskeleton (Orvar et al., 2000). The stimulus is further conveyed by numerous signaling cascades (Teige et al., 2004; Cramer et al., 2011).

Plant response on the transcript level is measurable after 15 minutes of cold treatment (Gilmour et al., 1998). Effects on the transcript level are diverse and reach a maximum after around 24 hours of cold exposure (Fowler and Thomashow, 2002). A prominently discussed set of transcription factors with regard to cold acclimation are the C-repeat binding factors (CBFs: CBF1, CBF2, CBF3) that activate transcription of the so-called CBF regulon, overexpression of which significantly increases cold tolerance (Chinnusamy et al., 2010; Thomashow, 2010). All three *Arabidopsis thaliana* CBFs are located on chromosome 4 and are in linkage disequilibrium (Zhao et al., 2016). CBFs bind to a cis-acting C-repeat/dehydration responsive element (CRT/DRE) (Stockinger et al., 1997) present in cold-regulated (COR) genes encoding proteins that regulate transcription, photosynthesis, protein metabolism, primary metabolism and stress response (Maruyama et al., 2004; Hannah et al., 2005) and increase freezing tolerance in non-acclimated as well as acclimated plants (Gilmour et al., 2000). This induction

is modulated by the Circadian Clock (Mikkelsen and Thomashow, 2009), CBF transcripts show highest levels after 4 hours of light and lowest levels after 16 hours of light (Fowler et al., 2005). Alternative splicing of the clock component Circadian Clock-Associated 1 (CCA1) contributes to freezing tolerance (Park et al., 2012; Seo et al., 2012). Generally, the circadian gating of cold response is thought to efficiently maintain growth in low temperatures (Greenham and McClung, 2015). Further, CBF transcription is positively regulated by Inducer of CBF Expression 1 (ICE1) (Chinnusamy et al., 2003) and inhibited by Myeloblastosis 15 (MYB15) (Agarwal et al., 2006). The gene regulons of CBF1, CBF2 and CBF3 share an overlapping set of 133 upregulated loci enriched in CRT/DRE and 39 downregulated genes that do not show this enrichment (Park et al., 2015). Together with targets of 11 other transcription factors that are upregulated in the first wave of cold acclimation, the set of COR genes comprises at least 248 upregulated and 155 downregulated genes (Park et al., 2015). Upregulated genes encode enzymes of carbohydrate and lipid metabolism, cell wall modification, plastid related proteins, transporters, kinases, Ca^{2+} and hormone signaling components, transcriptional regulators, proteins involved in cellular biogenesis and stress response (Gilmour et al., 2004; Zhao et al., 2016). The advent of the CRISPR/Cas9 technology enabled the creation of *cbf* single, double and triple mutants (Zhao and Zhu, 2016). Experiments have provided evidence for the importance of all CBFs in freezing tolerance but CBF2 has a prominent role in cold acclimation (Jia et al., 2016; Zhao et al., 2016). CBF2 expression is negatively regulated by CBF1 and CBF3 (Zhao and Zhu, 2016). In accordance with these results, varying freezing tolerance between accessions from Sweden and Italy could be partly explained by a non-functional CBF2 variant in the more sensitive Italian accession (Gehan et al., 2015). Considerable natural genetic variation of freezing tolerance has been described in *Arabidopsis thaliana* (Hannah et al., 2006).

The induction of COR gene transcription by CBF and other transcription factors upon cold exposure induces a comprehensive and naturally varying shift in the cellular homeostasis of *Arabidopsis thaliana* plants (Nagler et al., 2015). This shift is reflected by changes on the proteomic and metabolomic states and also includes differential posttranslational modification of proteins (Nagler et al., 2015). Generally, more proteins are upregulated than downregulated in stress conditions (Kosova et al., 2011). In the case of cold response, the Arrhenius equation gives one possible direct explanation for this as enzyme activity is negatively correlated to temperature and positively correlated to catalyst abundance. In order to maintain a stable reaction rate, the amount of protein has to increase if temperature decreases given all other parameters do not change.

A prominently discussed plant response during cold acclimation is the inhibition of photosynthesis and repression of genes encoding photosynthetic light reaction proteins (Strand et al., 1997), probably to diminish oxidative damage of the photosynthetic apparatus from reactive oxygen species (ROS) (Huner

et al., 1998) for instance via Mehler's reaction in the chloroplasts and thus prevent photoinhibition and other potentially harmful effects as membrane damage and protein oxidation (Janmohammadi et al., 2015). Accordingly, Calvin cycle protein activity is increased by upregulation of protein abundance to further process photochemical energy into the metabolic system (Strand et al., 1999).

Still, to maintain a proper cellular redox homeostasis, the ROS scavenging machinery is induced (Fanucchi et al., 2012). Electrons originating from the photosynthetic light reactions reduce O_2 to superoxide which has to be detoxified via the water-water cycle (Asada, 1999). The first line of defense against ROS are dismutases and catalases (Asada, 1999). The major form of superoxide dismutases (SD) in chloroplasts are Cu-Zn-SDs (CSD) reducing superoxide radicals to hydrogen peroxide (Asada, 1999). Catalases (CAT) reduce hydrogen peroxide to water and oxygen, some of which need the tetrapyrrole heme as cofactor. Peroxidases are another protein group that mitigates oxidative stress by reducing hydrogen peroxide to water and O_2. A prominent role of ascorbate in peroxidase activity has been identified (Smirnoff, 2000). Electrons are transferred either to ascorbate, which is oxidized to monodehydroascorbate (Asada, 1999), or thioredoxin, which in turn reduces oxidized glutathione disulfide (Noctor et al., 2011; Noctor et al., 2012). Monodehydroascorbate can spontaneously disproportionate to ascorbate and dehydroascorbate (Asada, 1999). Monodehydroascorbate is further reduced to ascorbate by monodehydroascorbate reductases (MDAR) with electrons from NADPH and dehydroascorbate is reduced to ascorbate by dehydroascorbate reductases (DHAR) while oxidizing glutathione to glutathione disulfide (Foyer and Noctor, 2011). Glutathione is synthesized from glutamate, cysteine and glycine in two steps catalyzed by the enzymes γ-glutamylcysteine synthase and glutathione synthase (Noctor et al., 2012). Ascorbate biosynthesis is more complex and part of galactose metabolism (Linster and Clarke, 2008) and regulated at multiple steps (Linster and Clarke, 2008; Wang et al., 2013; Wang et al., 2013). Thioredoxins serve as reductants for multiple peroxidases and peroxiredoxins (Noctor et al., 2012). Thioredoxins are known to regulate enzyme activities by disulfide bond reduction and accompanying conformational changes (Meyer et al., 2008). This way, four out of eight Calvin cycle enzymes (GAPDH, FBPASE, SBPASE, PRK) are regulated (Buchanan et al., 2002) and even enzymes involved in starch synthesis (Geigenberger et al., 2005).

Further, methionine sulfoxide reductases (MSRs) are also regenerated by thioredoxin activity (Meyer et al., 2008). Free and protein-bound methionine, levels of which are responsive to multiple stress situations (Obata and Fernie, 2012), can be oxidized by ROS to methionine sulfoxide. Proteins reducing methionine sulfoxide to methionine are called MSRs. MSRAs accept both free and protein-bound Methionine sulfoxide as substrate whereas MSRBs have low affinity towards free methionine sulfoxide (Le et al., 2013) but a good ability of repairing methionine containing proteins. MSRs are reponsive to photooxidative stress (Vieira Dos Santos et al., 2005).

Secondary metabolism is knowingly responding to environmental stimuli (Winkel-Shirley, 2002) and especially flavonoids have been discussed for their role in cold acclimation (Doerfler et al., 2013; Schulz et al., 2016) and UVb protection (Li et al., 1993; Agati and Tattini, 2010; Hectors et al., 2014). There also exists significant natural variation of flavonoid biosynthesis during cold acclimation (Schulz et al., 2015). Isoprenoid metabolism, especially tocopherol synthesis (Lange and Ghassemian, 2003), is also involved in adaptation to low temperatures (Maeda et al., 2006). Tocopherol is a strong antioxidant produced from tyrosine in the plastids which protects the photosynthetic apparatus, pigments and thylakoid lipids from oxidative degradation by ROS (Kanwischer et al., 2005). A compound class closely connected to isoprenoids are tetrapyrroles. Tetrapyrroles are classified in four groups: chlorophyll, heme, siroheme and phytochromobilin (Tanaka et al., 2011). Besides the obvious importance of chlorophyll synthesis for light harvesting in photoautotroph organisms, tetrapyrroles are also known to be responsive to drought stress because of their reducing properties on ROS (Nagahatenna et al., 2015). Consequently, transcription of many genes involved in tetrapyrrole synthesis is induced by ROS signaling (Nagai et al., 2007). Interestingly, transcripts of genes involved in tetrapyrrole synthesis have been identified to be negatively correlated to cold acclimation (Hannah et al., 2006). Glucosinolate metabolism is receiving more and more attention for the role in environmental adaptation and field fitness (Kerwin et al., 2015). In this context, epistatic effects have been shown to impact on glucosinolate accumulation and fitness in the field (Kerwin et al., 2017).

As extracellular ice formation induces a drop in the osmotic potential in the apoplast which results in cellular water loss, among other COR proteins like COR15a (Wang and Hua, 2009), dehydrins increase in abundance during cold acclimation (Amme et al., 2006) as well as other compatible solutes like proline (Cook et al., 2004; Kaplan et al., 2004) which also plays a role in redox buffering (Verslues and Sharma, 2010). Interestingly, a CBF triple knockout mutant does not accumulate proline during cold acclimation (Jia et al., 2016). The change in temperature and loss of water can also lead to protein denaturation which explains the induction of chaperone abundance as for instance heat-shock proteins (Rocco et al., 2013).

Carbohydrate metabolism is regulated during cold acclimation on the transcript (Hannah et al., 2006), protein (Nagler et al., 2015) and metabolite level (Cook et al., 2004; Nägele et al., 2011; Doerfler et al., 2013; Nagler et al., 2015). Fixed carbon is used for sucrose synthesis rather than starch accumulation (Strand et al., 1999; Nägele et al., 2011). Sucrose, as well as its building blocks glucose and fructose, have been discussed to increase during cold acclimation (Cook et al., 2004). Sucrose is the most important transport sugar in plants (Winter and Huber, 2000), whereas starch is an important carbon storage compound in plants (Strand et al., 2000) and both are hubs in plant carbohydrate metabolism. Besides regulating developmental effects, sucrose is known to increase in heat and cold stress (Kaplan

et al., 2004). Particularly, transcript level of cystathionine γ-synthase, a regulated enzyme in methionine biosynthesis (Galili et al., 2016), is upregulated by sucrose (Hacham et al., 2013). Sucrose cleavage is important for cellular hexose signals (Koch, 2004). *In vivo*, there are two of enzymatic groups involved in sucrose degradation: invertases irreversibly release fructose and glucose whereas sucrose synthases (SUS) reversibly release UDP-glucose and fructose (Koch, 2004). UDP-glucose is a substrate for starch synthesis via granule-bound starch synthases releasing amylose. Together with fructose-6-phosphate, UDP-glucose also participates in reversible sucrose synthesis via sucrose-phosphate-synthase (SPS), activity of which was found to positively correlate with cold tolerance (Nägele et al., 2011) and has been discussed as being involved in the natural variation of freezing tolerance (Nägele et al., 2012). UDP-glucose-1-phosphate uridylyltransferases reversibly catalyze the conversion of glucose-1-phosphate to UDP-glucose. Glucose-1-phosphate is reversibly convertible to glucose-6-phosphate by phosphoglucomutase which can be processed to starch for storage and via phosphoglucoisomerase also to fructose-6-phosphate (Zeeman et al., 2007).

Although starch anabolism is of less importance during cold acclimation, starch catabolism has been identified as a crucial part of the cold acclimation process also underlying significant natural variation (Espinoza et al., 2010; Nagler et al., 2015). The cold tolerant *Arabidopsis thaliana* accession Rschew (Rsch; Origin: Russia) is able to mobilize starch more efficiently than cold sensitive Cvi (Origin: Cape Verde) which results in the higher abundance of soluble sugars like sucrose, glucose, fructose and raffinose and osmoprotectants like proline in the cold acclimated state (Nagler et al., 2015). In this context, starch degradation has been linked to proline accumulation (Zanella et al., 2016). Sucrose also is a substrate for the synthesis of many cryoprotective substances like raffinose and raffinose family oligosaccharides (RFOs) (Peterbauer and Richter, 2001). Together with raffinose, which besides galactinol also increases during cold acclimation (Kaplan et al., 2004), sucrose stabilizes membrane integrity and fluidity which also contributes to temperature sensing (Hincha et al., 2003; Los and Murata, 2004; Knight and Knight, 2012). Similar mechanisms have been proposed for COR genes like COR15a and COR15b (Thalhammer et al., 2014). During cold acclimation, raffinose synthesized in the cytosol is transported to the plastids to protect chloroplast membranes, a trait that also exhibits natural variation (Nägele and Heyer, 2013). Raffinose and other RFOs have also been discussed to be involved in abiotic stress response and sugar sensing (Valluru and Van den Ende, 2011) as well as signaling compound in biotic stress response and priming (Kim et al., 2008). Raffinose, as sucrose, is also responsive to many other environmental stimuli (ElSayed et al., 2014).

The subcellular redistribution of metabolites and proteins has been recognized as crucial for cold acclimation (Hurry, 2017). Besides an increase in plastidic raffinose abundance, many aspartate derived amino acids increase in the cytosol during cold acclimation, for instance branched-chain amino

acids (Hoermiller et al., 2016). This is an interesting property, as amino acid catabolism is an important alternative energy source in plants, especially under stress conditions. Specifically the oxidation of branched-chain amino acids yields considerable amounts of ATP and reduction equivalents (Hildebrandt et al., 2015). The degradation is connected to autophagy, deficiency of which restricts alternative respiration in mitochondria (Barros et al., 2017). Different from the increase in proline, which results from increased synthesis, branched-chain amino acid concentrations have been shown to accumulate because of an upregulation of protein degradation under osmotic stress (Huang and Jander, 2017).

Once temperatures rise again in spring, acclimated freezing tolerance is lost in a process called deacclimation. This process is quicker than cold acclimation as transcripts and metabolome approach non-acclimated patterns of Col-0 after 24 hours of higher temperature which revert the biomolecular reprogramming during cold acclimation (Pagter et al., 2017).

Phenotypic Plasticity

Every genetic locus has to be considered in the context of an environment in order to give rise to a phenotype (Weckwerth, 2003; Dawkins, 2004; Weckwerth, 2011). Hence, GEI are defined by a set of genotypes which are mapped to a set of phenotypes by a specific environment. In Figure 1, this relationship is schematically drawn. An individual's genotype is called a genophene. Similar genophenes are grouped to genotypes, which constitute the genospecies in genotypic space. A specific environmental cue induces a corresponding phenotype, an individual ecophene. Similar ecophenes are combined to ecotypes and similar ecotypes constitute an ecospecies. The set of all ecospecies is the coenospecies comprising the species' ecological amplitude. A specific genotype can interact with different environments in variable ways, thus producing different (molecular) phenotypes depending on the environment they were exposed to. This phenomenon is called phenotypic plasticity and has puzzled biologists for decades (Turesson, 1922; Fusco and Minelli, 2010). Indeed, some genotypes do not grow in specific environments, thus they are schematically shown here with only one outgoing environmental arrow. Developing an explicit idea on how these mappings can be defined in a molecular and ecophysiological way is one of the main goals in phenotypic plasticity research (Pigliucci, 2001).

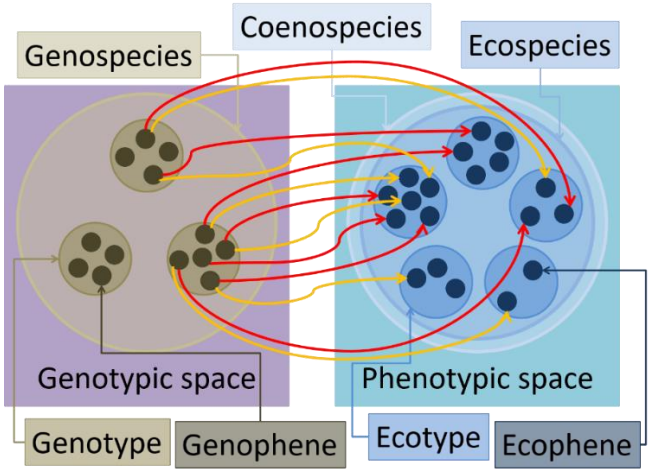


Figure 1: Genotype-Environment Interactions (GEI) as a function mapping from Genotypic to Phenotypic space. Different environmental settings are indicated by differently colored arrows – modified from (Turesson, 1922)

The analysis of a certain phenotype by a set of molecular profiles on different molecular levels, so called molecular ecophenes, has been shown to be a promising approach to promote the understanding of how a genotype shapes a molecular phenotype (Weckwerth et al., 2004; Morgenthal et al., 2005; Wienkoop et al., 2008; Wienkoop et al., 2010; Kleessen et al., 2012; Doerfler et al., 2013). Because proteins and metabolites are the prominent components of biochemical pathways, these molecular levels include most of the processed information encoded by GEI and are well suited for the definition of molecular phenotypes. In summary, research on phenotypic plasticity is key to understand, predict and anticipate the effects of changing climates on plant performance (Nicotra et al., 2010). Additionally, phenotypic plasticity has been discussed as a mechanism to preserve genetic diversity under directed natural selection (Gillespie and Turelli, 1989).

Phenotypic Plasticity in *Arabidopsis thaliana*

Phenotypic plasticity has a long tradition in *Arabidopsis thaliana* research and has buried the genes-as-blueprint metaphor (Alberch, 1991; Pigliucci, 2010). Basically, all environmental response experiments belong to this category. In *Arabidopsis thaliana*, genetic variation of plasticity in light and nutrient response but not in response to gradients of water availability has been detected (Pigliucci et al., 1995). Stressful conditions significantly affect phenotypic plasticity of flowering time, life span, number of leaves, leaf weight, height of first flower, total plant height, number of branches, inflorescence weight and number of fruits (Pigliucci et al., 1995) and also a relationship between phenotypic plasticity and fitness has been discussed (Pigliucci and Schlichting, 1996).

Plasticity genes are genes that mediate environmental changes to a phenotypic output (Pigliucci, 1998). A prominent example of plasticity genes in *Arabidopsis thaliana* are phytochrome genes (Pigliucci, 1996) conveying shade-avoidance phenotypes in this annual weed. It was shown, that plants

deficient in all functional phytochromes are less plastic in their response to changes in light parameters, in fact they always show a shade-avoidance phenotype (Pigliucci and Schmitt, 1999). Light response shows high inter-population variation but low variation for phenotypic plasticity (Pigliucci and Kolodynska, 2002). In a reciprocal-transplant field trial, shading induced flowering at an earlier developmental stage with fewer rosette leaves and also at an earlier time point but this effect disappeared in a greenhouse environment, presumably because these two trends cancelled each other (Callahan and Pigliucci, 2002). From this study it was also concluded that the greenhouse experiment provided stronger evidence for variation between populations suggesting that in the field, environmental parameters have a higher impact on the variance of fitness than genetic background, as the correlation of date of bolting and number of leaves disappeared in the field (Callahan and Pigliucci, 2002). In an experiment testing phenotypic plasticity in *Arabidopsis thaliana* with regard to photoperiodic regimes in growth chambers reflecting a latitudinal light gradient, it turned out that leaf number is directly and leaf size indirectly correlated to Northern latitude whereas the effects on bolting time were not so much influenced by the light regime but by the genotype (Banta et al., 2007). Strikingly, considering local adaptation, the reproductive fitness estimated by the number of fruits was not highest in the populations originating from one of the three photoperiodic regimes (Banta et al., 2007), contrasted by other studies proving the better performance of local populations in their native habitats (Agren and Schemske, 2012; Wilczek et al., 2014). Other authors have determined the influence of climatic variables on alleles linked to fitness variation and found that local selection acts upon different genetic loci with diverse molecular functions (Fournier-Level et al., 2011). In this context, necrotroph responsive laccase 1 (LAC1), drought responsive senescence-associated gene 21 (SAG21) and DNA repair related chromatin remodeling 8 (CHR8) were identified to impact on fitness in multiple environments (Fournier-Level et al., 2011). It was also found that local adaptation is lagging behind changing environments employing the product of silique length and silique number as fitness proxy (Wilczek et al., 2014). In a common garden site in Finland, the native accession was not as fit as an accession from Germany whereas in common garden sites in Spain, the United Kingdom and Germany, the native accessions were outperforming or at least performing equally well (Wilczek et al., 2014).

Another study investigating phenotypic plasticity of 47 *Arabidopsis thaliana* accessions in flooding conditions highlighted a trade-off between root and stem biomass accumulation (Pigliucci and Kolodynska, 2002). Wind as a proxy for mechanic stimulation was found to impact on the number of basal branches and thus fecundity in many populations (Pigliucci, 2002). Besides genotypic variation, epigenetic variation has also been shown to dramatically impact on plant fitness and phenotypic plasticity (Bossdorf et al., 2010). However, the impact of epigenetic variation was only weakly correlated to phylogeny (Bossdorf et al., 2010).

Ecophenes

Compared to data on plant genomes, proteomes and metabolomes, environmental information is sparse on the accessions available in seed banks. Although habitat metadata is available to some extent, microclimatic conditions can vary on small spatial scales independent from latitudinal, longitudinal and altitudinal gradients. In other words, *Arabidopsis thaliana* perceives environments at the same degree of complexity at all spatial scales (Pigliucci, 1998). Therefore, data sets are not detailed enough to allow in-depth molecular ecological research. Although there may be enough information to justify studies of microevolution among populations, the available data is not suitable for investigating microevolution within populations as variable environmental conditions during development are long known to influence the plasticity of phenotypes (Gilbert, 1991). Additionally, natural metabolic variation was shown to be only weakly correlated to genetic diversity (Houshyani et al., 2012).

For example, consider two topographically and climatologically similar populations at a given geographic distance, each one comprising a ridge situation and a lower slope situation. Although genetic variability may be lower among populations than between them, because of phenotypic plasticity there could of course be significant intrapopulation molecular phenotypic variation and, because of interpopulation similarity of specific ecological niches (ridge and lower slope in this fictive example, respectively), these phenotypes could show a similarity pattern not reflecting geographic or genetic distance. Therefore, we have to employ techniques to characterize the microenvironments of natural populations on a spatial and temporal scale in connection with molecular profiles of proteome and metabolome and genotyping via microarrays or next-generation deep sequencing for a detailed description of GEIs to identify physiologically adaptive traits and patterns of molecular phenotypic plasticity as well as their morphological and phenological outputs. Experiments in controlled environments with collected seeds allow a check for heritability and plasticity of discovered traits.

In this context, Turesson's ecotype concept (Turesson, 1922) seems to provide the theoretical basis to define GEI outputs (Figure 1) in an ecological meaningful way. Investigating this frame work with modern technologies potentially enables us to investigate the physiological and molecular basis for environmentally induced natural phenotypic variation (Pigliucci, 1998; Alonso-Blanco and Koornneef, 2000). Consider one of Turesson's examples in his monography on phenotypic plasticity in the plant kingdom (Turesson, 1922), the species *Polygonum amphibium*. This member of the Polygonaceae family can be modified into a land form, a water form and a dune form simply by exposition of plants to specific environments. In context of the ecotype concept, we could assume the coenospecies *Polygonum amphibium* to consist of three ecospecies (land, water and dune forms, respectively). Each of these ecospecies can be divided in ecotypes, which show a similar phenotype but because of

differing reasons. For instance, we could think of the land ecotype in the example to consist of different drought responses of individuals exposed to different environments, which are nevertheless perceived by the plant through one and the same stress signaling cascade and thus result in related ecophenes. On the one hand, we could have drought because of soil grain size and as a consequence of substrate chemistry on the other hand. Thirdly, there could be variation in precipitation or temperature data. It is also known that drought response varies with development and stress severity (Skirycz et al., 2010). Theoretically, it is even possible to unleash defined ecotypes from the genetic species concept, and thus comprising ecophenes of multiple species. For example, if more than one species are under scrutiny, those ecotypes would then represent ecological units which show similar reaction types to specific environmental signals. This could facilitate the identification of functional niches in ecosystems and thus keystone species and ecosystem engineers which are important in conservation and restoration biology. Therefore, a given ecotype is closely related on the spatial scale to a given (micro)ecosystem and on a chronological scale to a specific development of the same ecosystem throughout a given time period. Among major questions arising is how to sufficiently characterize an ecosystem to provide enough environmental resolution to successfully distinguish and associate ecophenes. Investigating molecular ecophysiological properties of individuals in extensively monitored natural habitats is the logical next step in the quest for the better understanding of the evolutionarily shaped genetic architecture of the metabolome (Chan et al., 2010) as well as community and ecosystem dynamics. As indicated classically in Ellenberg's experiment on the performance of grass species in monoculture and mixed cultures (Ellenberg, 1953), the realized ecological niche is depending on many parameters that have to be considered in an *in vitro* growth setting to reflect *in situ* plant performance, as for instance phytosociology. Additionally, choosing the ecologically relevant levels of abiotic parameters is important to derive relevant conclusions for *in situ* performance. In drought stress response, the investigation of lethal soil water deficits is explicitly not connected to increased survival or yield gain in more realistic, milder drought conditions in *Arabidopsis thaliana* (Skirycz et al., 2011).

Technological possibilities for continuously measuring abiotic habitat parameters, as for example soil properties, such as temperature, absolute water matrix potential or electric conductivity, and atmospheric parameters, such as air temperature, relative humidity and precipitation, exist and are longing for application in molecular ecological research. Methods for inferring biotic habitat parameters such as phytosociology are available since decades (Braun-Blanquet, 1964) and statistical classification methods have been developed (Hill, 1979; Tichy, 2002). Additionally, ecosystem properties such as land use classes, hemeroby maps and electronic soil classifications can potentially bridge the gap from model to ecosystem research.

Biochemical modeling

The parts of a (biological) system interact with each other in complex and sometimes non-intuitive ways. In order to provide an idea of how a system reacts if it's constituents are partly altered, the whole network of responses has to be considered. An established way for describing systemic behavior is modeling the dependencies of a system's variables in a set of coupled ordinary differential equations (ODEs) that describe the temporal fluctuation of system variables that reflect the system's state (Nägele, 2014). For instance, in metabolomics data, these equations describe the concentration changes of the modelled metabolites as the sum of anabolic and catabolic reactions, in more elaborate models considering the subcellular compartmentation even transport processes. The ODE for a given metabolite x_i is:

$$\dot{x_\iota}(t) = \frac{d\, x_i(t)}{d\, t} = anabolism - catabolism \pm transport = f_i(x_1, x_2, \dots, x_n, t)$$

Numerical integration of ODEs within biochemically and physiologically relevant boundaries yields an explicit mathematical expression that allows the time-dependent estimation of systemic variables. Metabolite data enable this approach as they provide the basis for solving this initial value problem. Indeed, the major problem arising is to gain enough knowledge of relevant system parameters to allow for the exact formulation of the ODE. For instance, the temporal change of fructose could be considered to depend on the anabolic reaction catalyzed by invertase and the catabolic reaction catalyzed by fructokinase, both depending on substrate and product concentrations. In order to exactly define the differential equation, all enzyme kinetic parameters have to be known, an endeavor that is further complicated by the existence of multiple isoforms with varying optimal conditions. In the last decade, a work-around has been proposed with the introduction of the inverse calculation of the biochemical Jacobian from metabolic covariance data (Steuer et al., 2003).

The Biochemical Jacobian

The concept of the Jacobian matrix is long known in systems theory as the matrix containing all first-order partial derivatives of the functions of systemic variables and thus describing the reactivity of the system to changes in those variables:

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Systems biologists have incorporated this concept into their considerations and have connected it to covariance data with the van Kampen equation (Steuer et al., 2003):

$$\boldsymbol{JC} + \boldsymbol{CJ}^T = -2\boldsymbol{D}$$

Here, **J** is the n x n Jacobian matrix (each system variable is defined by one function) constituted by the first order partial derivatives of the n functions with respect to all of the n variables. **C** is the n x n covariance matrix of the n system variables and **D** is an n x n diagonal fluctuation matrix of white noise modeled by a Langevin-type equation (Steuer et al., 2003). The solution of this inverse problem after reformation yields the Jacobian of the biochemical system. The metabolite functions combine multiple non-linear biochemical dependencies as for instance Michaelis-Menten enzyme kinetics and are linearized around a metabolic steady state in which the prerequisite $\dot{x}_i(t_0) = \frac{d\, x_i(t_0)}{d\, t} = 0$ is met, assuming there is no net change in a metabolite's concentration and the biochemical system is well adapted and in homeostasis. This assumption is of course an idealization that might not always be justified as it ignores for instance diurnal rhythms in metabolite pools. Consequently, the steady state assumption can be alleviated to not require an exact equilibrium of reaction rates but rather do not show a statistically significant change in metabolite concentrations. Additionally, the linearization is only valid in a narrow temporal interval, i.e. at the specific steady state time point t_0. Analytically, the metabolite functions, which are characterized by time-dependent metabolite concentrations and time-dependent reaction parameters, can be described as the product of a n x m stoichiometric matrix **N** containing the stoichiometries of n metabolites in m biochemical reactions and the non-linear flux vector v of length m at the n metabolite concentrations **X** (Steuer et al., 2003):

$$\frac{d\, \boldsymbol{X}(t)}{d\, t} = \boldsymbol{f}(\boldsymbol{X}(t), \boldsymbol{p}(t)) = \boldsymbol{N}\boldsymbol{v}(\boldsymbol{X}(t), \boldsymbol{p}(t))$$

At the steady state $\mathbf{S^0}$ another equivalency can be formulated and the above mentioned formulae equal the product of the stoichiometric matrix, the partial derivatives of reaction rates with respect to all metabolites at steady state time t_0 and metabolite concentrations **X**:

$$\frac{d\, \boldsymbol{X}(t_0)}{d\, t} = \boldsymbol{f}(\boldsymbol{X}(t_0), \boldsymbol{p}(t_0)) = \boldsymbol{N}\boldsymbol{v}(\boldsymbol{X}(t_0), \boldsymbol{p}(t_0)) = \boldsymbol{N}\frac{\partial \boldsymbol{v}}{\partial \boldsymbol{X}}\Big|_{t_0} \boldsymbol{X} = \boldsymbol{J}\boldsymbol{X} = 0$$

This equation also implies, that the entries of the biochemical Jacobian can be interpreted as the product of the stoichiometric matrix and the partial derivatives of reaction rates with respect to all metabolites.

Hence, the entries quantify the varying reactivity of the biochemical system to changes in metabolite concentrations at the steady state. For the first time this approach allows the inverse calculation of the biochemical system directly from the metabolomics covariance matrix and this was proven in 2012

(Sun and Weckwerth, 2012). As in general systems theory, the Jacobian eigenvalues' real parts provide information on the stability of the biochemical system (Nägele and Weckwerth, 2013). By considering the medians of 1000s of inverse numerically estimated biochemical Jacobians from the van Kampen equation, the role of pyruvate dehydrogenase in regulating the varying metabolic homeostasis in light and extended darkness has been demonstrated without exact knowledge of all involved reactions' parameters (Nägele et al., 2014). This proves that the inverse calculation of the biochemical Jacobian is in principal a possibility to assess the dynamic behavior of a biochemical system without the laborious experimental determination of kinetic parameters in all situations of interest (Doerfler et al., 2013; Nägele and Weckwerth, 2013; Nukarinen et al., 2016; Wang et al., 2016). We have also applied this approach for the first time to *in situ* metabolomics data from natural *Arabidopsis thaliana* populations (Nagler et al., manuscript in preparation, see third manuscript of cumulative thesis). This approach has also been complemented via structural kinetic modeling, which surpasses some disadvantages of the discussed approach of deriving the biochemical Jacobian from a linear approximation of metabolic functions at a steady state by considering local linear models at each point in parameter space which allows for the exact estimation of the biochemical Jacobian at a given point in parameter space (Steuer et al., 2006).

Dynamic responses of biochemical systems to experimental conditions can also be investigated in time series experiments that provide a view on the evolution of corresponding molecular homeostases. A possibility are regression approaches to determine the time-shifted correlation in variable abundancies as in Granger causality (Granger, 1969). This was also recently applied to metabolomics time series in Arabidopsis cold stress data (Doerfler et al., 2013).

Publications featured in this Cumulative Thesis

The results of the experiments conducted during this doctoral thesis have been published or will be submitted for publication in international and peer-reviewed journals. The relevant publications are:

- **Nagler M***, Nukarinen E*, Weckwerth W, Nägele T (2015) Integrative molecular profiling indicates a central role of transitory starch breakdown in establishing a stable C/N homeostasis during cold acclimation in two natural accessions of *Arabidopsis thaliana*. Bmc Plant Biology 15

  * contributed equally

  We examined the natural variation of cold acclimation on the metabolome, proteome and phosphoproteome with regard to two geographically well separated *Arabidopsis thaliana* accessions Rsch and Cvi. We presented evidence that starch degradation is differentially regulated between these cold-tolerant and cold-sensitive genotypes. Further, an interaction network of the cold acclimated accession revealed more comprehensive reprogramming of the molecular plant system during cold acclimation resulting in a higher degree of connectivity of the cold-induced protein-protein interaction network in the cold tolerant accession pointing towards higher resilience of the cold reprogrammed metabolic homeostasis.

  To this experiment, Ella Nukarinen and I contributed equally. I measured the proteomics samples and analyzed the data. Together, we wrote the manuscript for the publication.

Nagler et al. BMC Plant Biology (2015) 15:284
DOI 10.1186/s12870-015-0668-1

BMC Plant Biology

RESEARCH ARTICLE Open Access

Integrative molecular profiling indicates a central role of transitory starch breakdown in establishing a stable C/N homeostasis during cold acclimation in two natural accessions of *Arabidopsis thaliana*

Matthias Nagler[1†], Ella Nukarinen[1†], Wolfram Weckwerth[1,2] and Thomas Nägele[1,2*]

Abstract

Background: The variation of growth and cold tolerance of two natural *Arabidopsis* accessions, Cvi (cold sensitive) and Rschew (cold tolerant), was analysed on a proteomic, phosphoproteomic and metabolomic level to derive characteristic information about genotypically distinct strategies of metabolic reprogramming and growth maintenance during cold acclimation.

Results: Growth regulation before and after a cold acclimation period was monitored by recording fresh weight of leaf rosettes. Significant differences in the shoot fresh weight of Cvi and Rschew were detected both before and after acclimation to low temperature. During cold acclimation, starch levels were found to accumulate to a significantly higher level in Cvi compared to Rschew. Concomitantly, statistical analysis revealed a cold-induced decrease of beta-amylase 3 (BAM3; AT4G17090) in Cvi but not in Rschew. Further, only in Rschew we observed an increase of the protein level of the debranching enzyme isoamylase 3 (ISA3; AT4G09020). Additionally, the cold response of both accessions was observed to severely affect ribosomal complexes, but only Rschew showed a pronounced accumulation of carbon and nitrogen compounds. The abundance of the Cold Regulated (COR) protein COR78 (AT5G52310) as well as its phosphorylation was observed to be positively correlated with the acclimation state of both accessions. In addition, transcription factors being involved in growth and developmental regulation were found to characteristically separate the cold sensitive from the cold tolerant accession. Predicted protein-protein interaction networks (PPIN) of significantly changed proteins during cold acclimation allowed for a differentiation between both accessions. The PPIN revealed the central role of carbon/nitrogen allocation and ribosomal complex formation to establish a new cold-induced metabolic homeostasis as also observed on the level of the metabolome and proteome.

Conclusion: Our results provide evidence for a comprehensive multi-functional molecular interaction network orchestrating growth regulation and cold acclimation in two natural accessions of *Arabidopsis thaliana*. The differential abundance of beta-amylase 3 and isoamylase 3 indicates a central role of transitory starch degradation in the coordination of growth regulation and the development of stress tolerance. Finally, our study indicates naturally occurring differential patterns of C/N balance and protein synthesis during cold acclimation.

Keywords: Cold acclimation, *Arabidopsis thaliana*, Natural variation, Starch metabolism, Amylases, Systems biology, Metabolomics, Proteomics, Phosphoproteomics, Growth regulation

* Correspondence: Thomas.Naegele@univie.ac.at
†Equal contributors
[1]Department of Ecogenomics and Systems Biology, University of Vienna, Althanstr. 14, 1090 Vienna, Austria
[2]Vienna Metabolomics Center (VIME), University of Vienna, Althanstr. 14, 1090 Vienna, Austria

Background

Plant growth together with stress tolerance and flowering traits are known to be orchestrated in a complex and interdependent molecular manner. Water supply, temperature and soil quality have been shown to be the most relevant abiotic factors which significantly affect these traits [1]. During the last decade, naturally occurring genetic and phenotypic variation of *Arabidopsis thaliana* has been shown to be a promising tool for studying the molecular architecture of such physiological traits. On the cellular level, abiotic stress affects the integrity of membrane systems, transport proteins, metabolic enzymes and signalling compounds, ultimately leading to disfunctions in cellular metabolism which directly impair plant growth and development. Previous studies have shown and discussed significant differences in naturally occurring stress tolerance, morphology, developmental programming and flowering of *Arabidopsis thaliana* [2–9].

Low temperature belongs to one of the most important abiotic factors limiting the geographic distribution of plants. In many temperate species, the exposure of plants to low but non-freezing temperatures initiates a process termed cold acclimation resulting in increased freezing tolerance [10]. The process of cold acclimation is a multigenic trait being characterized by a comprehensive reprogramming of the transcriptome, proteome and the metabolome, but also of enzyme activities and the composition of membranes [3, 11–17]. Particularly, reprogramming of primary metabolism plays a crucial role during cold acclimation leading to a changed photosynthetic activity and the accumulation of soluble sugars, amino acids and polyamines. Concentrations of the di- and trisaccharide sucrose and raffinose, respectively, have been shown to correlate well with winter hardiness in several plant species [18, 19]. Further, several roles for sugars in protecting cells from freezing injury have been proposed [10]. Yet, soluble carbohydrates have been shown to be insufficient to fully describe the development of freezing tolerance [20]. While sugar levels are often found to positively correlate with freezing tolerance, the underlying regulatory mechanisms are poorly understood. On a whole plant level, it remains elusive whether sugar accumulation may result from reduced sink activity, because growth retardation at low temperatures is stronger than the reduction of photosynthetic activity [21]. Additionally, it is not clear whether sugars function as cryoprotective substances or because they are substrates for the cryoprotectant synthesis [19].

Together with sugars, also pools of organic and amino acids are significantly affected during cold-induced metabolic reprogramming. Aspartate, ornithine and citrulline were found to increase during cold exposure of *Arabidopsis thaliana* indicating the reprogramming of the urea cycle [14]. Beyond, the authors observed a cold-induced increase in levels of alpha-ketoglutarate, fumarate, malate and citrate which they suggested to result from an up-regulation of the citric acid cycle. Although many observations revealed an increase of metabolite levels to be characteristic for cold acclimation, the magnitude of changes in the metabolome does not necessarily indicate the capacity of *Arabidopsis* to increase its freezing tolerance [12]. A prominent example which shows the possible discrepancy between metabolic reprogramming and gain of freezing tolerance is the comparison of the freezing sensitive natural accessions Cvi, which originates from Cape Verde Islands, and C24, originating from the Iberian Peninsula. Both accessions similarly increase their freezing tolerance during cold acclimation while concomitant metabolome changes were found to differ dramatically [3]. It might not be surprising that the coordination of a complex trait like freezing tolerance cannot be directly related to one certain metabolic output, but, simultaneously, this observation indicates a high level of plasticity which is characteristic for intraspecific molecular responses to environmental cues. In this context, most of the naturally occurring biochemical mechanisms and metabolic regulatory strategies to acclimate to low temperature still remain elusive.

Plant growth is significantly reduced due to cold exposure. Although low temperature significantly affects metabolic processes and resource allocation, growth is not necessarily limited by photosynthetic activity. Following a period of 1 to 3 days after exposure to low temperature, during which cold stress is sensed and acclimation is initiated, rates of photosynthetic carbon assimilation can be almost fully recovered [22]. Together with the finding that growth is affected more significantly than photosynthesis during exposure to water deficit [23], this indicates that growth during stress exposure might rather be limited by sinks than sources. Such a cold-induced sink limitation has been discussed to be the reason for the characteristic accumulation of sugars during cold exposure. Although high levels of sugars have been shown to potentially repress the expression of photosynthetic genes [24, 25], cold acclimation and development at low temperature was found to reduce or even fully revert this effect [26–28]. Additionally, cold acclimation was found to have a significant effect on leaf respiration of *Arabidopsis thaliana* [29]. Both respiration rates in the light and in the dark were described to increase significantly during cold acclimation, while the more pronounced effect was found for respiration in darkness. Moreover, although cytosolic hexose phosphate concentrations increased dramatically, there was no significant correlation observed with respiration in the light indicating that respiration is not limited by substrate availability under low temperature stress [29].

Although the above-mentioned findings only represent an excerpt from current findings about growth regulation

and cold acclimation strategies in *Arabidopsis*, it clearly indicates a highly complex and interlaced relationship between metabolic and physiological consequences of low temperature. Systems biology focuses on such complex questions and has become a rapidly expanding and attractive research area during the last decade [30]. In a systems biology approach, elements of an interaction network, e.g. a metabolic map, are rather analysed and discussed as interacting components than isolated parts in order to improve the understanding of how a complex biological system is organized and regulated [31].

Research on plant freezing tolerance, growth regulation and plant systems biology has largely been driven by studies in *Arabidopsis thaliana*. The species is native to Europe and central Asia, its biogeography was described in detail, and it was shown that climate on a global scale is sufficient for shaping the range boundaries [32]. When compared to other *Brassicaceae* species, *Arabidopsis* has a wide climatic amplitude and shows a latitudinal range from 68 to 0°N, which makes it suitable for the analysis of variation in adaptive traits [33]. *Arabidopsis* represents a predominantly selfing species, and, hence, most of the individual *Arabidopsis* plants collected in nature represent homozygous inbred lines [34]. These homozygous lines are commonly referred to as accessions, representing genetically distinct natural populations that are specialized to particular sets of environmental conditions. The variation of morphological and physiological phenotypes enables the differentiation of most of the collected *Arabidopsis* accessions from others. In particular, considering the tolerance to abiotic factors, e.g. low temperature, a large variation has been reported (e.g. [33]), making *Arabidopsis* an attractive system to study plant-environment interactions.

In the present study, two of these *Arabidopsis* accessions were analysed with respect to naturally occurring variation in the traits of growth regulation and freezing tolerance. The selection of the two accessions, Cvi (origin: Cape Verde Islands) and Rschew (origin: Western Russia), was based on findings of previous studies which have shown that Cvi represents a freezing sensitive accession while Rsch is freezing tolerant (e.g. [35]). Based on this finding and due to their large distance with respect to geographical origin, cold acclimation capacity and cold-induced gene regulation [3], the molecular and biochemical study of both accessions can be expected to provide a suitable approach to quantify strategies of growth maintenance during environmental fluctuations. As previous work has already indicated, a multi-layered design of molecular physiological studies was necessary in order to derive coherent conclusions on a genome-wide level [11, 36]. Thus, the present study aimed at a comprehensive characterization of metabolomic, proteomic and phosphoproteomic levels of both natural accessions to unravel differential strategies of growth regulation in a changing environment.

Results

Differential growth of Cvi and Rsch during cold acclimation

Growth behaviour of both accessions was characterized by recording the total fresh weight of leaf rosettes from 15 independently grown plants for each acclimation state, i.e. the non-acclimated (na) and acclimated (acc) state (Fig. 1a). Analysis of variance (ANOVA) revealed a significantly higher fresh weight of Rsch plants before (na) and after (acc) cold acclimation compared to Cvi (Fig. 1b). Additionally, plants of the accession Rsch were found to increase their fresh weight significantly (~1.6-fold) during cold acclimation while this was not observed for Cvi (Fig. 1b; Remark: when applying Student's t-test, the increase in fresh weight of Cvi was detected to be significant; $p = 0.018$). Furthermore, cold acclimated plants of Cvi did not differ in their fresh weight compared to non-acclimated plants of Rsch. Most distinct differences in fresh weight, which we interpreted in terms of an average growth rate [37], were observed between cold acclimated plants of Rsch and Cvi (Ratio >2).

Integrative profiling of metabolites, proteins and phosphoproteins during cold acclimation

For a comprehensive molecular characterization of both accessions, the metabolome, proteome and the phosphoproteome, i.e. phosphopeptide abundance, was analysed applying an integrative analytical GC-MS and LC-MS platform [38–43]. Statistical dimensionality reduction by Principal Component Analysis (PCA) revealed a clear separation of both accessions and acclimation states on all levels of molecular organization (Fig. 2). In the non-acclimated state, the accessions were not separated by metabolite profiling including the main components of C/N leaf metabolism. (Fig. 2a). In contrast, after cold-acclimation both accessions were significantly separated (Fig. 2a). Levels of soluble sugars, threonic acid, citrate, succinate, malate, fumarate, glutamate, proline and aspartate were found to be significantly higher in Rsch, while a high level of transitory starch was found to be characteristic for Cvi (Fig. 3a, b; Additional file 1: Table S1; Additional file 2: Figure S1).

On the proteome level, PCA revealed a clear separation of both accessions and conditions (Fig. 2b). Accessions were separated on PC1 while the acclimation process became visible on PC2. Although the explanatory power of PC1 was only about 8 % higher than that of PC2 (Additional file 3: Figure S2), this indicated that the strongest observable effect in the proteome was due to accession-specific differences followed by changes induced by the cold acclimation process. The strongest observed accession-specific separation in the proteome

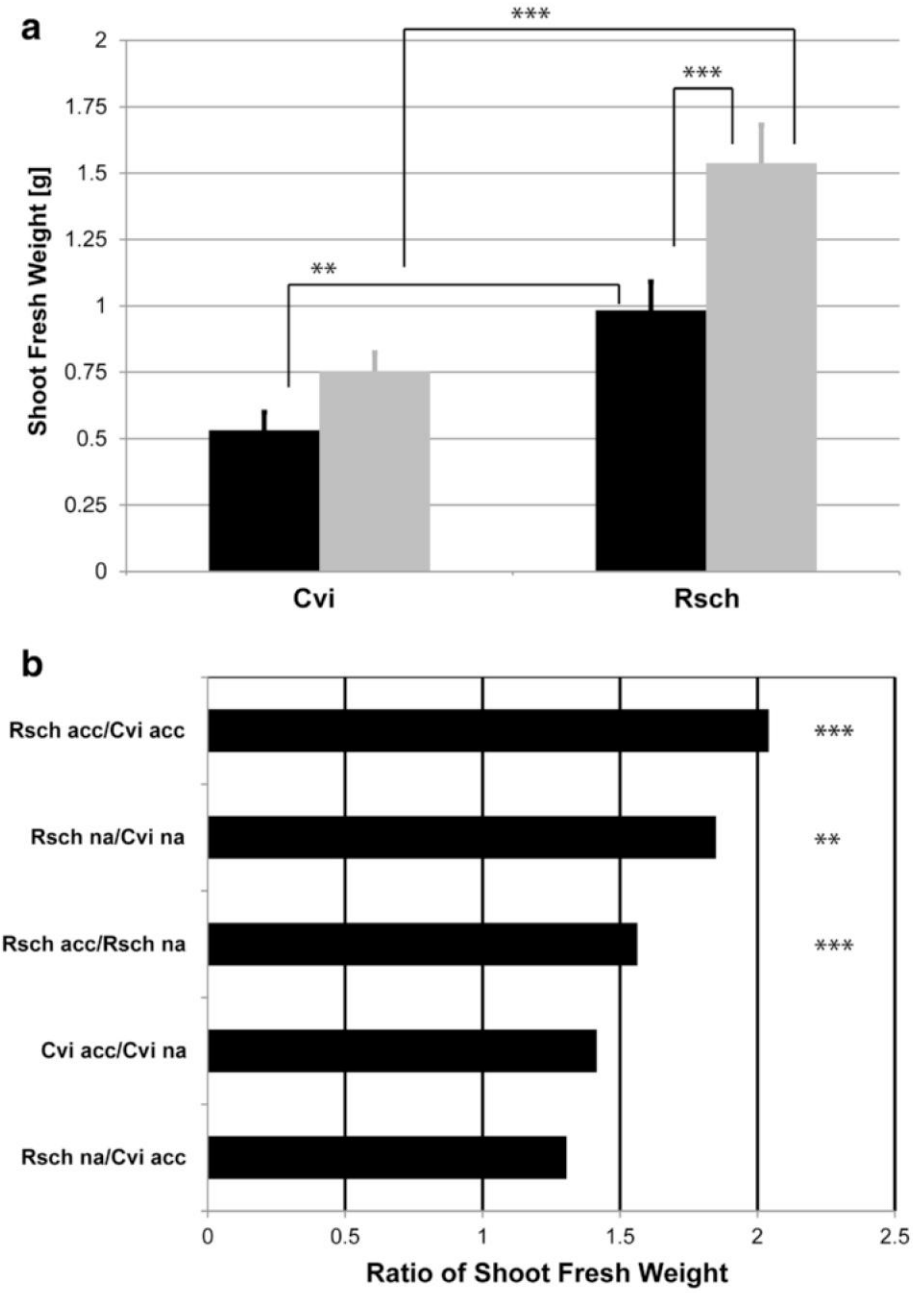


Fig. 1 Comparison of shoot fresh weight. **a** Absolute shoot fresh weight of accessions Cvi and Rsch before (na, black bars) and after (acc, grey bars) cold acclimation. Error bars represent means ± SE ($n = 15$). **b** Ratios of mean shoot fresh weights. Asterisks indicate significance tested in an ANOVA (** $p < 0.01$; *** $p < 0.001$)

appeared due to differences in carbohydrate metabolism, amino acid metabolism, abiotic stress-related proteins, protein synthesis and degradation, sulphur assimilation (ATP-sulfurylase, ATP-S), glucosinolate biosynthesis, and redox regulation (Additional file 4: Table S2). Particularly, relative alpha- and beta-amylase enzyme levels, i.e. alpha-amylase-like 3 (AMY3; AT1G69830) and chloroplast beta-amylase (BAM3; AT4G17090), showed a differential pattern in both accessions (Fig. 4). While AMY3-levels were found to be constitutively higher in Rsch (Fig. 4a), levels of BAM3 showed an acclimation-dependent decrease in Cvi (Fig. 4b). Levels of isoamylase 3 (ISA3; AT4G09020) were found to significantly increase during cold acclimation in Rsch while no significant change in ISA3-levels was observed for Cvi (Fig. 4c).

In addition to this accession-specific effect, the cold acclimation process most significantly affected proteins related to processes involved in photosynthetic light reactions and

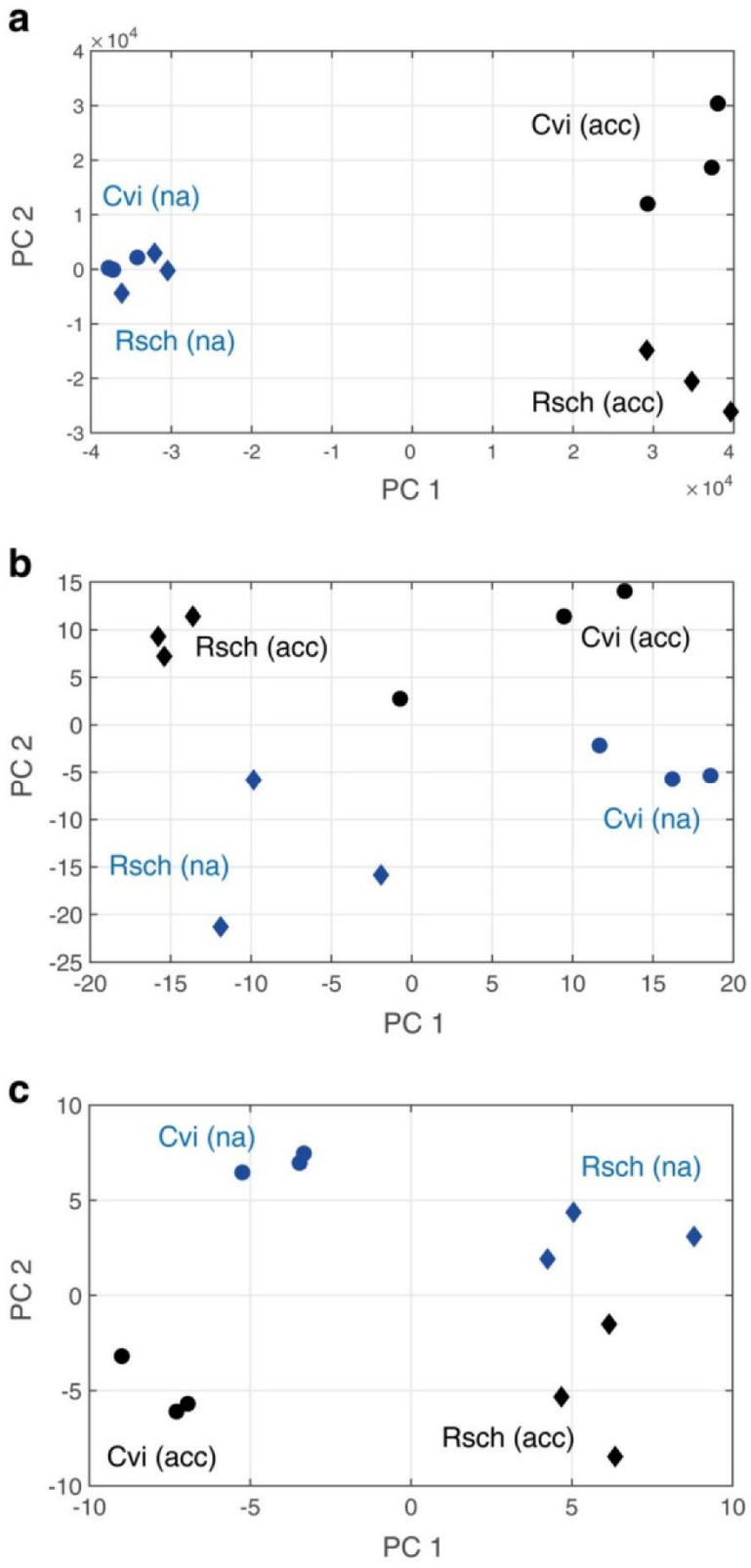


Fig. 2 (See legend on next page.)

(See figure on previous page.)
Fig. 2 Principal component analysis (PCA) on levels of (**a**) the primary C/N-metabolome, (**b**) protein abundance, and (**c**) phosphopeptide abundance. Accession samples are represented by filled circles (Cvi) and filled diamonds (Rsch). Blue colour indicates non-acclimated samples, black colour indicates acclimated samples. Detailed information about loadings and explained variances of the PCA as well as absolute levels of metabolites, relative levels of proteins and phosphopeptides are provided in the supplements

the Calvin cycle (Additional file 4: Table S2). PCA revealed a very pronounced cold acclimation-induced effect for levels of the ribosomal 40 and 60S subunit (see Additional file 4: Table S2) indicating a systematic reprogramming of the translational machinery in both accessions (Fig. 5). A detailed list of ribosomal components is provided in the supplements (Additional file 5: Table S3). In both accessions, levels of several ribosomal protein components were significantly increased after cold acclimation, and this effect was found to be even more pronounced in Rsch than in Cvi (see Additional file 5: Table S3).

A full and detailed list of all functional categories of the proteome and their hierarchy concerning the accession- and acclimation-specific separation is provided in the supplements (Additional file 4: Table S2).

Changes in the phosphoproteome of Cvi and Rsch during cold acclimation

Similar to the proteome, also the phosphoproteome, i.e. the detected and quantified phosphopeptide abundances, revealed a stronger separation of accessions compared to acclimation states (Fig. 2c, Additional file 3: Figure S2). Yet, also in this context the explained variances by PC1 (accession) and PC2 (acclimation) only differed by ~6 % indicating a similar contribution to the separation. The most dominating accession-specific effects in the phosphoproteome were found to comprise processes of membrane transport and trafficking, modulation of transcription factors and ubiquitination (Additional file 6: Table S4). In particular, one of the most characteristic and significant differences between Cvi and Rsch could be observed for the phosphorylation levels of BASIC PENTACYSTEINE 6 (BPC6; AT5G42520; Fig. 6a), a member of a plant-specific transcription factor family. The phosphorylation level was found to be constitutively higher in Rsch compared to Cvi ($p < 0.01$). In contrast, phosphorylation levels of the plasma membrane intrinsic protein PIP2;3 (AT2G37180) were found to be constitutively higher in Cvi (Fig. 6b; $p < 0.001$).

Detected cold acclimation-induced changes in the phosphoproteome, which were displayed on PC2 (Fig. 2c), revealed a complex pattern of *in vivo* phosphorylation affecting various transcription factors, photosynthetic electron carriers, ribosomal subunits, processes of protein assembly and the cytoskeleton (Additional file 6: Tables S4 and Additional file 7: Table S5). The most significant cold acclimation-induced effect on phosphopeptide levels was detected for the protein Cold Regulated 78, COR78 (AT5G52310). In both accessions, relative levels of phosphorylated COR78 peptides were found to be significantly increased after cold acclimation ($p < 0.001$; Fig. 7a). Further, a significantly higher phosphorylation level was detected in cold acclimated samples of Rsch compared to acclimated samples of Cvi ($p < 0.05$). The same pattern was observed for the relative protein abundance of COR78 which was also significantly higher in non-acclimated samples of Rsch ($p < 0.05$; Fig. 7b).

Integrative analysis of metabolism and predicted protein-protein-interaction networks (PPIN) during cold acclimation

To derive a comprehensive overview of accession-specific and cold acclimation-induced molecular processes, collected experimental information about metabolite, protein and phosphopeptide levels was clustered according to their Euclidean distance after standardization (zero mean & unit variance; Fig. 8a). While for both Cvi and Rsch clusters could be identified which were not affected by the cold acclimation process (Additional file 8: Table S8), cold affected proteins were analysed in protein interaction networks predicted by the STRING database (see Methods) (Fig. 8b, c). Both created interaction networks differed clearly in their size. While the cold-response network of the cold-tolerant accession Rsch comprised almost 4000 protein interactions (Additional file 9: Table S6), the Cvi network only comprised about 500 interactions (Additional file 10: Table S7). A predominant and common effect of cold acclimation in both accessions was the reprogramming of protein synthesis, i.e. of ribosomal subunits (Table 1). About 65–80 % of all cold-affected protein interactions were found to be related to this functional category. In a more specific context, this finding is also displayed in Fig. 5 showing the cold-induced reprogramming of the ribosomal 40 and 60S subunit. A more contrasting picture between both accessions was observed for proteins and phosphorylation levels associated with processes of protein degradation, Calvin Cycle, photosynthetic light reactions, TCA cycle, amino acid synthesis, photorespiration, redox metabolism, protein folding, glycolysis, and lipid metabolism (Table 1). These processes were found to be involved much stronger in the cold acclimation responsenetwork of Rsch compared to Cvi.

Discussion

Cold acclimation of plants represents a multifaceted and multigenic process affecting various levels of molecular organisation, e.g. gene expression, RNA processing or post-

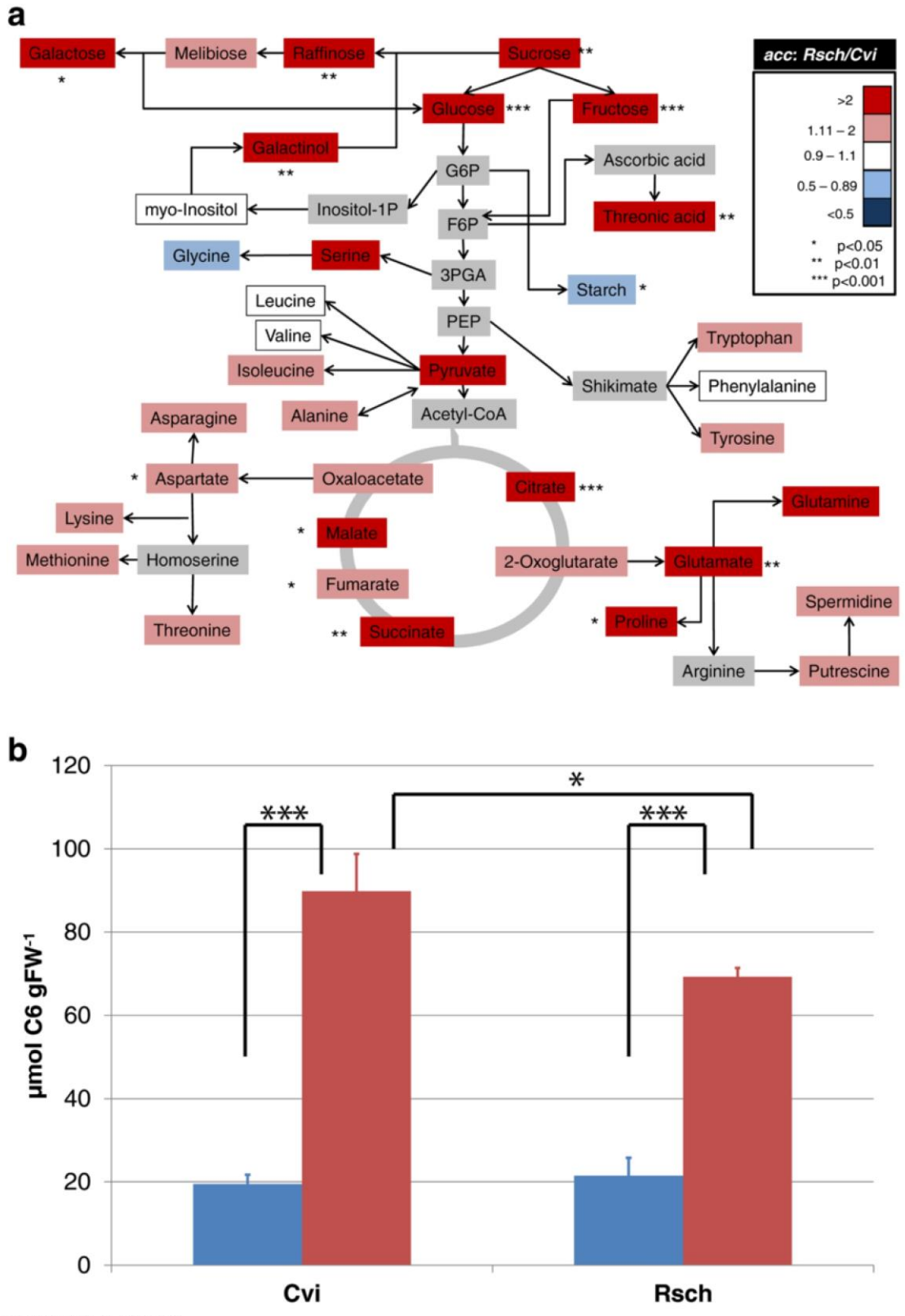


Fig. 3 (See legend on next page.)

(See figure on previous page.)
Fig. 3 The primary metabolome in cold-acclimated leaf samples of accessions Rsch and Cvi. **a** Ratios of metabolite levels which were built by dividing the absolute mean values of metabolite levels of Rsch by levels of Cvi which were assessed by a GC-TOF/MS measurement (see Methods - GC-MS Metabolite Analysis; $n = 3$). Asterisks indicate significant differences as described in the figure. Grey-coloured metabolites were not experimentally analysed. **b** Absolute starch levels in non cold-acclimated (blue bars) and cold acclimated (red bars) leaf samples of Cvi and Rsch ($n = 3$). Asterisks indicate significant differences (* $p < 0.05$; ** $p<0.01$; *** $p < 0.001$)

translational regulation [44, 45]. Hence, although numerous comprehensive studies have unravelled many crucial processes being involved in the acclimation process (for an overview see e.g. [46]), it is not surprising that many gaps still exist in our understanding of how metabolism is reprogrammed, and how the metabolic output is linked to the observed physiological output, e.g. changes in growth and yield. In general, plant growth requires a sufficient supply with energy, water and nutrients and is regulated in response to environmental changes. These environmental cues are sensed and integrated by a highly complex and conserved signalling network [47].

An efficient balancing of photosynthesis and respiration was shown to be a prerequisite for plant growth [48] and cold acclimation [29]. With regard to these two central processes, our findings revealed a more complex cold-induced metabolic reprogramming in the cold tolerant *Arabidopsis* accession Rsch which also showed a significantly higher shoot fresh weight than both non-acclimated and acclimated plants of Cvi (see Fig. 1). In addition, also glycolysis, TCA cycle and pathways of amino acid biosynthesis were found to be differentially affected by low temperature in both accessions. Together with the observed levels of sugars, organic and amino acids, which were, on an average, significantly higher in acclimated plants of Rsch, this points to a differential cold-induced redirection of carbon equivalents in both accessions. While we cannot experimentally exclude a limitation of CO_2 uptake as a reason for the lower metabolite levels in cold-acclimated plants of Cvi, there are several indications which rather suggest a differential regulation of carbon allocation to be the reason for the observed phenotype. First, on the level of the total proteome, we could observe a separation of acclimation states but not of accessions by cold-induced protein dynamics related to photosynthetic dark and light reactions (Additional file 11: Table S9). Second, in a former study, the analysis of the photosynthetic carbon uptake was found to be similar in cold-acclimated plants of cold sensitive and tolerant accessions [49]. While Nägele and colleagues did not analyse the Cape Verde accession Cvi but the cold-sensitive accession C24 originating from the Iberian Peninsula, further support of this hypothesis is provided by another study in which photosynthetic acclimation of Cvi was compared to the Finnish accession Hel-1, originating from Helsinki [50]. There, the author found that both accessions, originating from contrasting climates, showed a highly similar capability to acclimate to a broad regime of temperature and irradiance. Another indication for a non-limited CO_2-uptake is provided by the starch levels which were found to increase to a significantly higher level in Cvi than in Rsch (see Fig. 3). This agrees with the findings of Guy and co-workers who also described a significantly higher starch level in Cvi compared to Rsch after cold acclimation [12]. Based on this observation, Guy and co-workers suggested that, following a sufficiently long acclimation period, even in poorly acclimating accessions like Cvi energy constraints do not seem to limit the acquisition of freezing tolerance. Although our growth conditions (5 °C/7d of acclimation/125 μmol m^{-2} s^{-1}) do not exactly reflect the growth conditions applied in the study of Guy and co-workers (4 °C/14d acclimation/ 90 μmol m^{-2} s^{-1}), we still observed a similar output of starch metabolism.

To derive an explanation for the observed differences in starch metabolism, which has previously been suggested to be a major regulator of plant growth [51], the regulation of both starch synthesis and degradation has to be considered. While our study does not account for enzymatic activity, our proteomic results provide evidence for a different regulation of starch metabolism in cold acclimated plants of Cvi and Rsch. While, independently from cold exposure, levels of alpha-amylase AMY3 were found to be constitutively higher in Rsch than in Cvi, a cold-induced significant reduction in the level of beta-amylase BAM3 could only be observed for Cvi, while isoamylase 3, ISA3, was significantly increased only in cold-acclimated plants of Rsch. Alpha-, beta- and isoamylases play crucial roles in starch degradation [52–54], and, hence, these findings hint towards a distinct regulation of starch degradation which was previously discussed to play a decisive role in the process of cold acclimation [55, 56]. Starch molecules consist of mostly unbranched amylose (alpha-1,4-linked glucosyl moieties) and branched amylopectin (alpha-1,6-linked moieties). While alpha-amylase, hydrolysing the alpha-1,4-glucosidic linkages of starch, plays a central role in the degradation of storage starch in endosperm of germinating cereal seeds [57], a disruption of AtAMY3 by insertional mutagenesis did not affect starch degradation in *Arabidopsis* leaves [58]. However, removal of AMY3 in addition to the debranching, alpha-1,6-linkage hydrolysing, enzyme ISA3 was shown to lead to a strong starch excess phenotype [54]. A triple mutant with

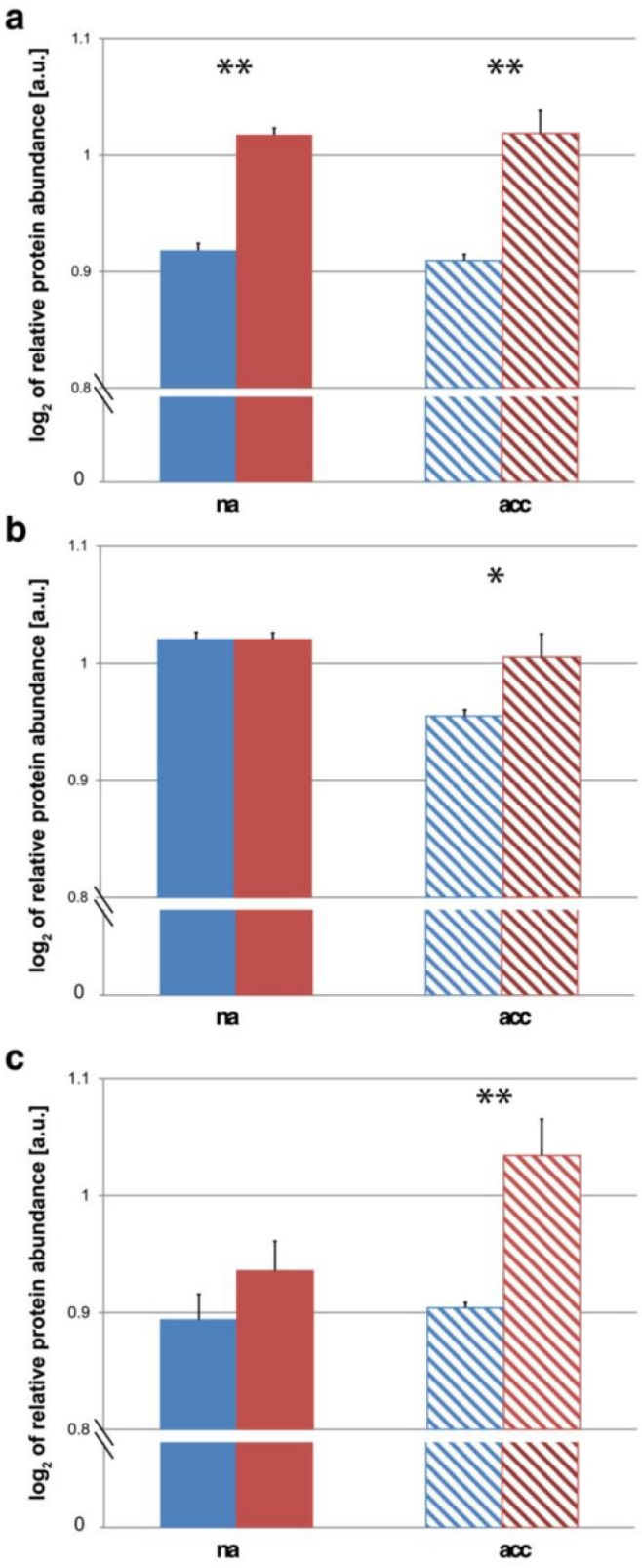


Fig. 4 (See legend on next page.)

(See figure on previous page.)
Fig. 4 Relative protein levels of amylase enzymes in non cold-acclimated (na) and cold-acclimated (acc) leaf samples. **a** Levels of alpha-amylase-like 3 (AMY3; AT1G69830), and (**b**) Levels of chloroplast beta-amylase (BAM3; AT4G17090), and (**c**) Levels of isoamylase 3 (ISA3; AT4G09020). Blue colour indicates the accession Cvi, red colour indicates the accession Rsch ($n = 3$). Filled bars represent means ± SD of na samples, hatched bars represent means ± SD of acc samples. Asterisks indicate significant differences between accessions (* $p < 0.05$; ** $p < 0.01$). Abundances were normalised to total protein content of the sample

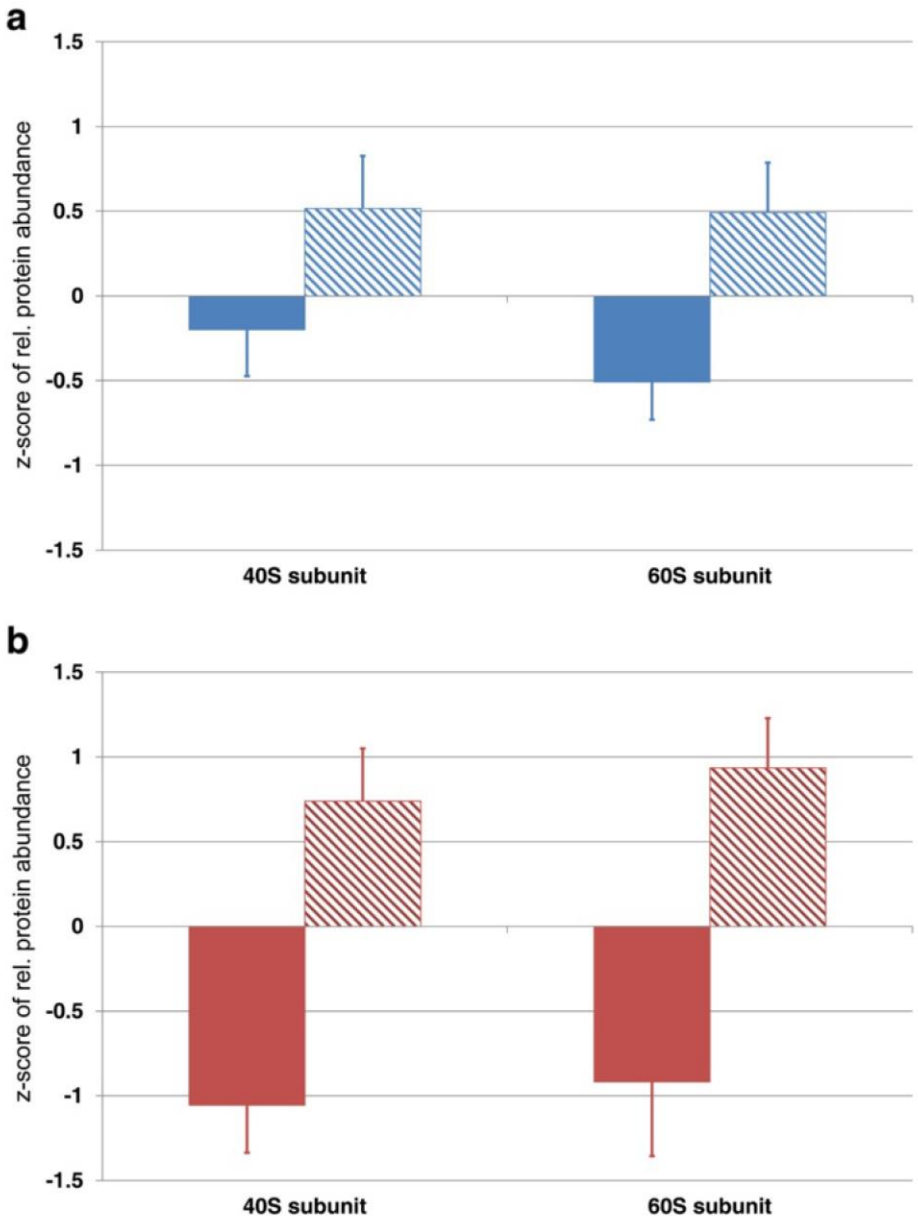


Fig. 5 Cold-induced increase of the ribosomal 40S and 60S subunit in the *Arabidopsis* accessions (**a**) Cvi and (**b**) Rsch. Colours indicate the different accessions (blue: Cvi; red: Rsch), filled and hatched bars differentiate cold acclimation states (filled: na; hatched: acc). Bars and error bars represent the mean ± SD of relative protein abundance after standardization (zero mean & unit variance, *z-score*). Means ± SD were built from those ribosomal protein compounds which were identified to contribute strongest to the separation of na and acc samples (see PCA in Fig. 2b and Additional file 4: Table S2; 60S subunit $n = 11$; 40S subunit $n = 12$)

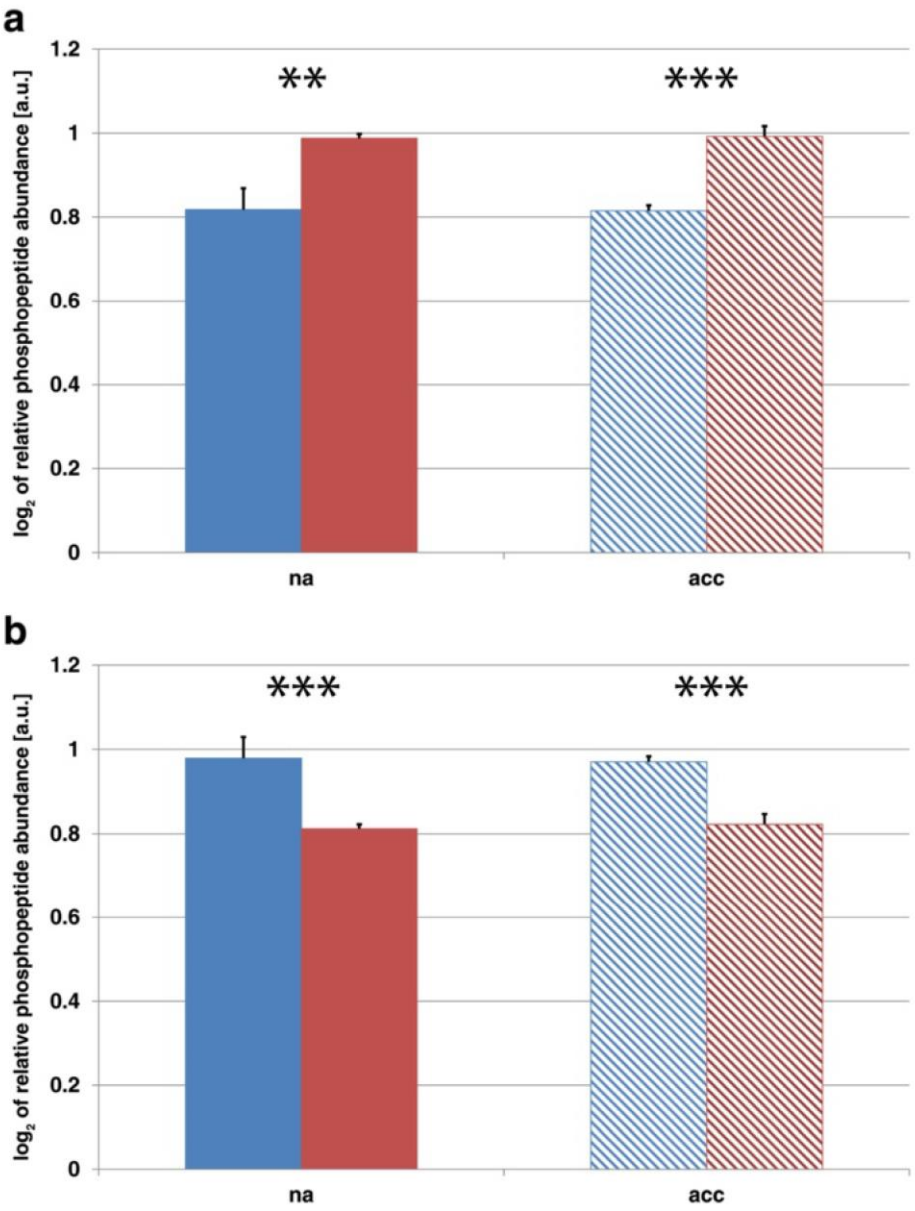


Fig. 6 Relative abundance of phosphorylated peptides of (**a**) BASIC PENTACYSTEINE 6, and (**b**) plasma membrane intrinsic protein PIP2;3. Colours indicate samples of the two different accessions Cvi (blue) and Rsch (red) before (na; filled bars) and after (acc; hatched bars) cold acclimation. Bars and error bars represent the mean ± SD of relative phosphopeptide abundance ($n = 3$). Asterisks indicate significant differences (** $p < 0.01$; *** $p < 0.001$)

an additional removal of limit dextrinase, LDA, which represents another debranching enzyme, was finally shown to result in an effective block of starch breakdown accumulating even higher levels of starch than observed before in the double mutant [54]. While our presented shotgun proteomics approach could not resolve the cold-induced effect on LDA in either of both accessions, our findings indicate that the combination of constantly lower AMY3-levels in Cvi and a cold-induced increase in ISA3-levels in Rsch might provide an explanation for the higher starch levels observed in cold-acclimated plants of Cvi.

The complete process of (transitory) starch breakdown from the insoluble granule to the soluble compounds maltose and glucose comprises numerous additional steps and classes of enzymes, finally resulting in a complex and tightly (redox) regulated pathway [52, 59]. Beta-amylases (BAMs) primarily hydrolyse glucan chains, which have been previously released and linearized, liberating maltose [52]. The multigene family of BAMs in *Arabidopsis thaliana* comprises nine genes, and *BAM3* was shown to encode a catalytically active plastidial enzyme playing a central role in leaf starch degradation at night in mesophyll

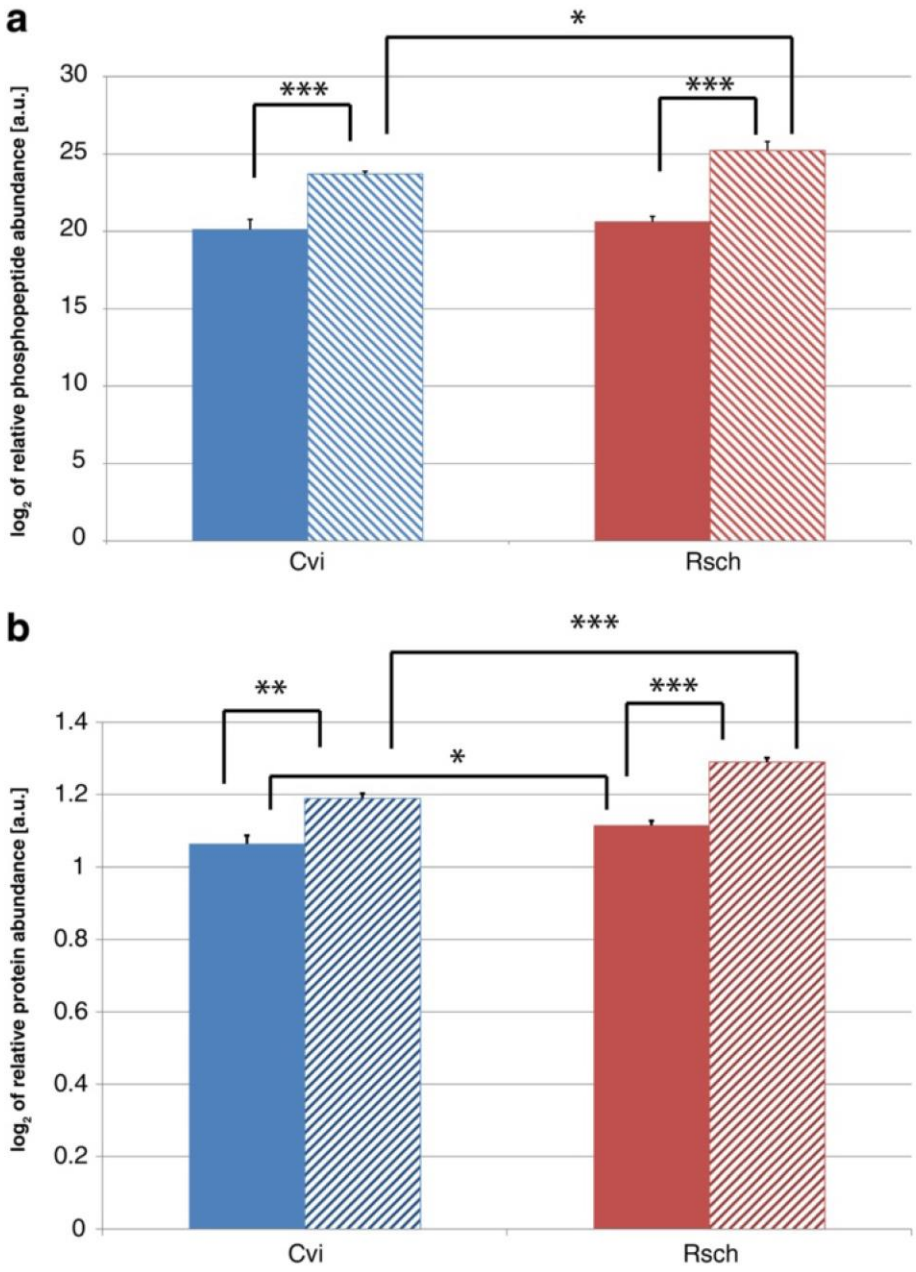


Fig. 7 Relative phosphorylation and protein levels of COR78. **a** Bars represent mean values (±SD, $n = 3$) of relative COR78 (AT5G52310) phosphopeptide abundance. **b** Bars represent mean values (±SD, $n = 3$) of relative COR78 (AT5G52310) protein abundance. Colours indicate the accessions (Cvi: blue; Rsch: red). Filled bars indicate values of non-cold acclimated samples, hatched bars indicate values of cold acclimated samples. Asterisks indicate significant differences (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

cells [60, 61]. Hence, our finding of a significant decrease of BAM3 protein levels in cold acclimated plants of Cvi provides a further explanation for the strong increase of starch levels. The observation of a decrease in BAM3 protein levels contrasts the finding of a cold induced increase of *BAM3* expression [56]. However, in a recent publication Monroe and co-workers derived a more complex picture in which the authors observed a decline in BAM3 activity after 2d and 4d of cold stress while *BAM3* mRNA levels clearly increased [62]. Although these results were derived from studies within the genetic background of the Arabidopsis accession Columbia-0, and, hence, might not directly be comparable to the background of Cvi, they indicate the complex interplay of molecular levels of organization during exposure to a fluctuating environment. Such an adaptive and differential regulation of starch metabolism in response to cold was also exemplified in a previous study on the starchless *Arabidopsis thaliana pgm* mutant being

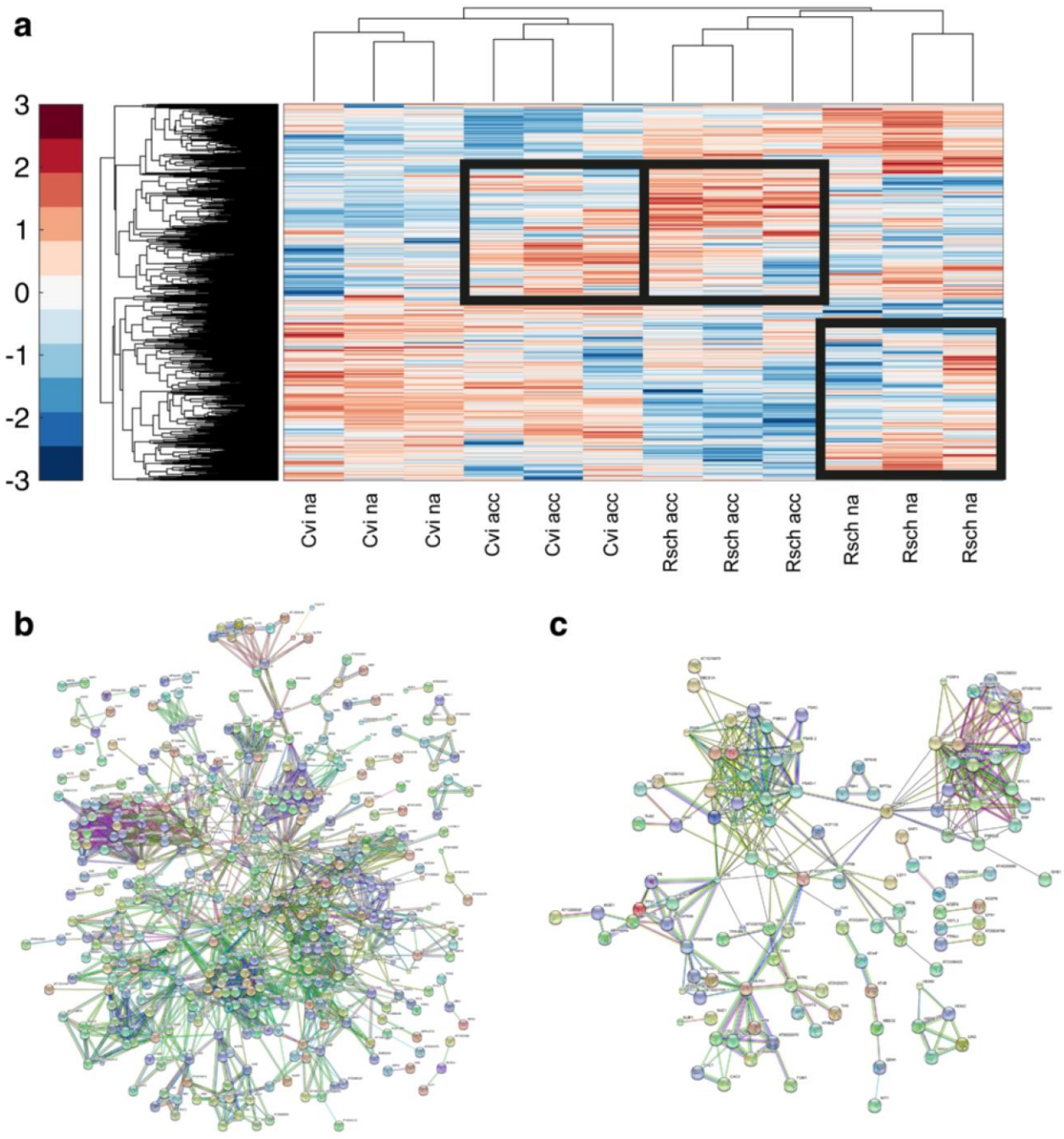


Fig. 8 Hierarchical cluster analysis and functional protein interaction networks of cold acclimation-induced reprogramming. **a** Hierarchical clustering of *Arabidopsis* accessions, acclimation states, and metabolite, protein and phosphopeptide abundances based on Euclidean distances. Columns represent non-cold acclimated (na) and cold-acclimated (acc) samples of Rsch and Cvi. Rows represent metabolites, proteins and phosphopeptides. Blue rectangles indicate the characteristic compounds which were chosen for reconstruction of the cold-acclimation induced interaction networks (part (**b**) and (**c**)). **b** Protein-protein interaction network of all proteins and phosphoproteins which were found to be involved in the cold acclimation-induced reprogramming of Rsch. **c** Protein-protein interaction network of all proteins and phosphoproteins which were found to be involved in the cold acclimation-induced reprogramming of Cvi. Interaction networks were created using the STRING database for known and predicted protein-protein interactions (setting: highest confidence (0.9); http://string-db.org/) [85]. A detailed list of protein-protein interactions for both accessions is provided in the supplement (Additional file 9: Table S6 and Additional file 10: Table S7)

Table 1 Proteomic adjustment and functions in the cold acclimation of Rsch and Cvi

Functional Category	*Arabidopsis* Accession	Relative contribution to accession-specific cold response [%]
Protein synthesis	Rsch	65.59
	Cvi	79.47
Protein degradation	Rsch	5.81
	Cvi	1.83
Calvin cycle	Rsch	4.54
	Cvi	0.34
Light reactions	Rsch	4.01
	Cvi	0.34
TCA	Rsch	2.76
	Cvi	0.57
Amino acid synthesis	Rsch	2.23
	Cvi	1.72
Not assigned	Rsch	1.61
	Cvi	0.92
Photorespiration	Rsch	1.40
	Cvi	-
Redox	Rsch	1.23
	Cvi	0.46
Protein folding	Rsch	1.17
	Cvi	0.57
Glycolysis	Rsch	1.00
	Cvi	0.23
Lipid metabolism	Rsch	1.00
	Cvi	0.23
Nucleotide metabolism	Rsch	0.95
	Cvi	-
OPP	Rsch	0.91
	Cvi	-
Stress	Rsch	0.72
	Cvi	5.50
Gluconeogenesis	Rsch	0.65
	Cvi	-
Secondary metabolism	Rsch	0.61
	Cvi	0.23
S-assimilation	Rsch	0.40
	Cvi	-
Protein targeting	Rsch	0.33
	Cvi	0.46
mETC	Rsch	0.32
	Cvi	0.46
DNA synthesis	Rsch	0.27
	Cvi	-
C1 metabolism	Rsch	0.26
	Cvi	-
Major CHO metabolism	Rsch	0.26
	Cvi	-
Tetrapyrrole synthesis	Rsch	0.26
	Cvi	1.61
Transport	Rsch	0.23
	Cvi	-
Misc.	Rsch	0.20
	Cvi	-
N-Metabolism	Rsch	0.20
	Cvi	-
Co-factor and vitamin metabolism	Rsch	0.13
	Cvi	0.69
Signalling	Rsch	0.13
	Cvi	0.34
Amino acid degradation	Rsch	0.12
	Cvi	-
Minor CHO metabolism	Rsch	0.12
	Cvi	-
Cell cycle	Rsch	0.09
	Cvi	-
Protein assembly/ Cofactor ligation	Rsch	0.09
	Cvi	0.11
RNA	Rsch	0.09
	Cvi	3.56
Protein PTM	Rsch	0.07
	Cvi	-
Biodegradation of xenobiotics	Rsch	0.06
	Cvi	-
Protein activation	Rsch	0.06
	Cvi	0.23
Hormone metabolism	Rsch	0.04
	Cvi	-
Cell organisation	Rsch	0.03
	Cvi	-
Metal handling	Rsch	0.03
	Cvi	-
Cell division	Rsch	0.01
	Cvi	-

All listed categories represent protein functions and their relative portion to all interactions which were identified by the STRING database analysis and which were found to be affected after the cold acclimation process (see Fig. 8)

deficient in a phosho-glucomutase activity [63]. In this study, the cold/heat-stress-induced increase of raffinose-family-oligosaccharide levels in the *pgm* mutant plants revealed an unexpected flexibility to adjust central metabolism to temperature stress in the absence of transitory starch.

Based on our investigation of the two natural accessions Rsch and Cvi during cold-acclimation, we suggest that the orchestration of growth and cold acclimation differs significantly in the redirection of photoassimilates between soluble metabolic compounds and the insoluble storage compound starch. In addition, the observation described in a previous study, that the biomass formation in the starchless *pgm* mutant is restricted by high respiratory losses in the root [48], allows us to hypothesise that the differences we observed in the fresh weight of Cvi and Rsch might also be due to a differential regulation of sink-source interaction both before and after the cold acclimation period. In future studies it would be interesting to analyse whether the observed differences in starch degradation are somehow related to resource allocation and root respiration in both accessions.

In context of *Arabidopsis* cold acclimation, the C-repeat binding factor (CBF) pathway belongs to one of the most intensively studied pathways which has a crucial role in the development of freezing tolerance [64]. Within minutes after transfer to low temperature, the CBF1-3 [65], i.e. DREB1a-c [66], expression is induced. They encode members of the AP2/ERF family of transcription factors recognizing the C-repeat (CRT)/dehydration-responsive element (DRE) being present in the promotors of CBF-targeted genes [66]. The constitutive overexpression of either CBF1, 2 or 3 alters the expression of cold-regulated (COR) genes resulting in an increase of freezing tolerance without exposure to low temperature [67, 68]. In the present study, the level of COR78 (AT5G52310) and its phosphorylation were observed to be positively correlated with the acclimation state of both accessions. Further, independent from the acclimation state, protein levels were found to be constitutively higher in Rsch than in Cvi. Interestingly, COR78 transcript abundance was previously discussed to be regulated by sucrose [69] which would explain our findings of higher protein abundance and sucrose levels in Rsch (see Figs. 3 and 7). In addition, these observations allow for the speculation about a link between sugar signalling networks and the cold responsive gene regulation which could probably comprise central conserved signalling compounds like the complex and antagonistic interaction network spanned by the kinases Sucrose-non-fermenting-1-Related Protein Kinase 1(SnRK1) and Target Of Rapamycin (TOR) [70].

Finally, the observation of differentially phosphorylated transcription factors, like the BASIC PENTACYSTEINE (BPC), but also membrane proteins, e.g. PIP2;3 aquaporins which are involved in numerous developmental and growth-regulatory processes [71, 72], clearly shows the wide range of cellular processes which might contribute to a systematic and differential stress acclimation output in naturally occurring accessions of Arabidopsis. Our results indicate that a comprehensive reprogramming not only of the process of protein synthesis, but also of metabolic pathways regulating the flux of photoassimilates to the TCA cycle and to pathways of amino acid biosynthesis, contributes to the stabilization of a metabolic homeostasis during cold acclimation. Together with previous studies on the stress-induced dynamics of protein phosphorylation patterns, which have, for example, revealed the central role of protein phosphorylation in cold-induced subcellular sugar allocation [73], and its applicability to crop science [74], this clearly indicates the necessity for integrative molecular profiling approaches to unravel a comprehensive picture of complex plant acclimation strategies.

Conclusions

The findings presented in this study provide evidence for a central role of the starch degradation pathway in the molecular orchestration of plant growth and abiotic plant-environment interactions in different natural Arabidopsis accessions. We conclude that manipulation of the starch degradation pathway represents a promising target for improving plant yield and stress tolerance. We hypothesise that stress-induced reprogramming of starch degradation plays a central role in the orchestration of photosynthetic metabolism rather than being a pure consequence from cold-induced metabolic changes. Together with reprogramming of translational regulation and protein synthesis it seems to differentially affect the cold-induced metabolic homeostasis which finally contributes to the observed acclimation output.

Methods

Plant cultivation and sampling strategy

Plants of *Arabidopsis thaliana* natural accessions Cvi-0 (NASC ID: N1097) and Rsch-0 (NASC ID: N1490; both accessions donated by: Albert Kranz Institute for Molecular Biosciences, Department of Biological Sciences, Johann Wolfgang Goethe-Universität Frankfurt am Main) were cultivated in a growth chamber under controlled conditions. The substrate for plant growth was composed of Einheitserde® ED63 and perlite. Plants were watered daily and fertilized once with NPK fertilization solution (WUXAL®Super; MANNA®-Dünger, Ammerbuch). Light intensity was 75 μmol m^{-2} s^{-1} in a 8/16 h day/night cycle with a relative humidity of 70 % and a temperature of 22 °C/16 °C. 28 days after sowing, light intensity was increased to 125 μmol m^{-2} s^{-1} in a 16/8 h day/night cycle. At bolting stage, which was 43 days after sowing, samples of non-

acclimated plants were collected from both accessions at midday, i.e. 8 h after light on. One sample consisted of 3 leaf rosettes. Non-sampled plants were transferred to 5 °C at 125 μmol m^{-2} s^{-1} in a 16/8 h day/night cycle with 70 % humidity. After 7 days at 5 °C, leaf rosettes were sampled as described for non-acclimated plants, i.e. each sample consisted of 3 leaf rosettes. At this growth stage, both accessions had induced inflorescence which was slightly higher (<1 cm) in Cvi than in Rsch. All samples were immediately quenched in liquid nitrogen. Sample material was stored at −80 °C until use.

GC-MS metabolite analysis

Frozen sample rosettes were ground to a fine powder with pestle and mortar under frequent cooling with liquid nitrogen. Polar metabolites were extracted and chemically derivatized as described previously [75, 76]. Gas chromatography coupled to mass spectrometry (GC-MS) analysis was performed on an Agilent 6890 gas chromatograph (Agilent Technologies®, Santa Clara, CA, USA) coupled to a LECO Pegasus® 4D GCxGC-TOF mass spectrometer (LECO Corporation, St. Joseph, MI, USA). Compounds were separated on an Agilent HP5MS column (length: 30 m, diameter: 0.25 mm, film: 0.25 μm). Deconvolution of the total ion chromatograms was performed using the LECO Chromatof® software. For absolute quantification of metabolites, peak areas were compared to calibration curves within a linear range of detection. Compound names, retention indices and mass-charge (m/z)-ratios which were used for peak quantification are provided in the supplements (Additional file 12: Table S10).

Protein extraction, phosphopeptide enrichment and LC-MS analysis

Total protein was extracted from 1 g of ground plant material as previously described [77]. Protein pellets were dissolved in 8 M urea/100 mM ammonium bicarbonate (AmBic) and protein concentration was determined with the Bio-Rad Bradford Assay using BSA as a standard. 1050 μg of total protein per sample were first reduced with dithiothreitol (DTT) at concentration of 5 mM at 37 °C for 45 min. Cysteine residues were alkylated with 10 mM iodoacetamide (IAA) in darkness at room temperature (RT) for 60 min. Alkylation was stopped by increasing DTT concentration to 10 mM and incubating in the dark at RT for 15 min. Proteins were first pre-digested with Lys-C (1:1000 w:w) at 30 °C for 5 h. Then the urea concentration was diluted to 2 M with 50 mM AmBic/10 % acetonitrile (ACN).$CaCl_2$ was added to a final concentration of 2 mM. Trypsin digestion (Poroszyme immobilized trypsin; 1:100 v:w) was performed at 37 °C overnight. Protein digests were desalted with C18 extraction materials (Agilent Technologies, Santa Clara, USA) and carbon graphite solid phase extraction (SPE) materials as described elsewhere [78]. After both SPEs, corresponding eluates were pooled, split in two tubes (50 μg for total proteomics and 1000 μg for phosphopeptide enrichment) and dried in a vacuum concentrator. Phosphopeptide enrichment was performed using 10 mg of TiO_2 (Glygen Corp.) as described previously [40, 79].

One microgram of total protein was separated on a PepMap RSLC 75 μm × 50 cm column (Thermo Fisher Scientific Inc., Waltham, USA) using a 120 min linear gradient from 2 to 40 % of mobile phase B (mobile phase A: 0.1 % [v/v] formic acid (FA) in water; mobile phase B: 0.1 % [v/v] FA in 90 % [v/v] ACN) with 300 nL/min flow rate. MS analysis was done with an Orbitrap Elite instrument (Thermo Fisher Scientific Inc., Waltham, USA) using a data-dependent acquisition method. Precursor masses at range 350–1800 Th were measured in the Orbitrap mass analyser with a resolution of 120 000, 1×10^6 ion population, and 200 ms injection time. MS/MS analysis was done in the linear ion trap with CID fragmentation and rapid scan mode for the 20 most intense ions. Prediction of ion injection time was enabled and the trap was set to gather 5×10^3 ions for up to 50 ms. Dynamic exclusion was enabled with repeat duration of 30 s, exclusion list size was set to 500 and exclusion duration to 60 s.

Phosphopeptides were dissolved in 10 μL of 5 % ACN/0.5 % FA and 5 μL were loaded on the column. The LC-MS analysis was done as the analysis of total protein digest with a few modifications. The gradient was 150 min from 2 to 40 % of mobile phase B and multistage activation was enabled with neural losses of 24.49, 32.66, 48.999, 97.97, 195.94, and 293.91 Da for the 10 most intense precursor ions. Further information about LC-MS analysis for reproducibility of experiments is provided in the supplements (Additional file 13: Table S11).

Data analysis and statistics

Peptide identification, phosphosite mapping as well as protein and phosphopeptide quantification were performed with MaxQuant 1.4 (http://www.maxquant.org) [80] and the Andromeda search algorithm [81] against the TAIR10 protein database. Total proteomics analysis was done with the following settings: maximum 2 missed cleavages, methionine oxidation, and protein N-terminal acetylation as dynamic modifications were allowed. Mass tolerance for precursors was set to 5 ppm and for fragment masses to 0.8 Da. The maximum FDR was set to 1 % for both peptide and protein levels. Protein quantification was done with a peptide ratio count of, at least, 2. Phosphopeptide identification was performed applying similar settings as in the total protein analysis. Phosphorylation of serine, threonine and tyrosine residues were additionally allowed to occur as dynamic modifications. Because the phosphorylation near a tryptic site could hinder digestion, 3 missed cleavages

were allowed. Quantification was done at peptide level. Further data processing was done with the Perseus 1.5 software. Total proteomics data was $\log_2$ transformed and filtered so that at least in one of the four conditions all values were present. Data was normalized to median of each sample and missing values were replaced with random numbers drawn from normal distribution of each sample. Phosphoproteomics data was handled similarly but additional filtering steps were applied: only phosphopeptides belonging to category I (localization probability >0.75 and score difference >5) [82] were considered for further analyses.

Data evaluation, normalisation and transformation was performed in Microsoft Excel® (http://www.microsoft.com). For Principal Component Analysis (PCA) and hierarchical cluster analysis, z-scores (zero mean, unit variance) were calculated for relative protein and phosphopeptide abundance. Metabolite PCA was performed on absolute levels. Analysis of variance (ANOVA) and Student's t-test were performed with the R software (The R Project for Statistical Computing; http://www.r-project.org/) (R Core [83]). PCA and hierarchical cluster analysis was performed within the numerical software environment Matlab® (V8.4.0 R2014b; www.mathworks.com) and the toolbox COVAIN [84]. Protein-protein interaction networks were created using the STRING database for Known and Predicted Protein-Protein Interactions (setting: highest confidence, 0.900; http://string-db.org/) (von Mering et al. [85]).

Availability of supporting data

All supporting data are included as additional files.

Additional files

Additional file 1: Table S1. Loadings of the metabolite PCA shown in Fig. 2a of the main document. (XLS 38 kb)

Additional file 2: Figure S1. Comparison of metabolite levels between non-acclimated and acclimated plants. Ratios were built by dividing the absolute mean values of metabolite levels of Rsch by levels of Cvi, or by dividing absolute mean values of metabolites of acc by na plants. Asterisks indicate significant differences as described in the figure. Grey-coloured metabolites were not experimentally analysed. (TIF 1649 kb)

Additional file 3: Figure S2. Explained variances of principal components describing the separation by metabolites, protein and phosphoproteins in Fig. 2 of the main document. (TIF 692 kb)

Additional file 4: Table S2. Loadings of the protein PCA shown in Fig. 2b of the main document. (XLS 241 kb)

Additional file 5: Table S3. Cold-induced effects on protein levels of components of the 40S and 60S ribosomal subunits. (XLSX 9 kb)

Additional file 6: Table S4. Loadings of the phosphoprotein PCA shown in Fig. 2c of the main document. (XLS 66 kb)

Additional file 7: Table S5. Normalised experimental data of metabolite, protein and phosphopeptide levels. (XLSX 1210 kb)

Additional file 8: Table S8. Proteomic and phosphoproteomic components of Cvi and Rsch which were not included in the cold-induced protein-protein interaction network analysis of Fig. 8. (XLSX 34 kb)

Additional file 9: Table S6. Components of the acclimation-induced protein-protein interaction network of Rsch shown in Fig. 8b of the main document. (XLSX 392 kb)

Additional file 10: Table S7. Components of the acclimation-induced protein-protein interaction network of Cvi shown in Fig. 8c of the main document. (XLSX 57 kb)

Additional file 11: Table S9. Protein dynamics related to photosynthetic dark and light reactions. (XLSX 34 kb)

Additional file 12: Table S10. List of metabolic components, retentions indices and mass-to-charge ratios which were applied for absolute quantification of metabolites. (XLSX 11 kb)

Additional file 13: Table S11. Proteomics minimal information for reproducibility of experiments. (XLSX 11 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MN and EN performed proteomic measurements, data analysis, statistics and wrote the paper. WW conceived the study and wrote the paper, TN conceived the study, performed metabolic measurements, performed statistical analysis and wrote the paper. All authors read and approved the final version of the manuscript.

Acknowledgements

We would like to thank the gardeners Andreas Schröfl and Thomas Joch for excellent plant cultivation in the department-associated greenhouse facility. We also thank the whole team of the MoSys Department for all the support, suggestions and constructive discussions. We thank the Vienna Metabolomics Center (ViMe) at the University of Vienna for support and advice. This work was supported by the EU-Marie-Curie ITN MERIT (GA 2010–264474) and the Austrian Science Fund (FWF, P 26342-B21 and P 25488).

Received: 16 October 2015 Accepted: 23 November 2015
Published online: 01 December 2015

References

1. Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K. Effects of abiotic stress on plants: a systems biology perspective. BMC Plant Biol. 2011;11(1):163.
2. Weigel D. Natural variation in Arabidopsis: from molecular genetics to ecological genomics. Plant Physiol. 2012;158(1):2–22.
3. Hannah MA, Wiese D, Freund S, Fiehn O, Heyer AG, Hincha DK. Natural genetic variation of freezing tolerance in Arabidopsis. Plant Physiol. 2006; 142:98–112.
4. Hannah MA, Heyer AG, Hincha DK. A global survey of gene regulation during cold acclimation in Arabidopsis thaliana. PLoS Genet. 2005;1:179–96.
5. Juenger TE. Natural variation and genetic constraints on drought tolerance. Curr Opin Plant Biol. 2013.
6. Ikram S, Bedu M, Daniel-Vedele F, Chaillou S, Chardon F. Natural variation of Arabidopsis response to nitrogen availability. J Exp Bot. 2012;63(1):91–105.
7. Clauw P, Coppens F, De Beuf K, Dhondt S, Van Daele T, Maleux K, et al. Leaf Responses to Mild Drought Stress in Natural Variants of Arabidopsis thaliana. Plant Physiol. 2015.
8. Nägele T, Heyer AG. Approximating subcellular organisation of carbohydrate metabolism during cold acclimation in different natural accessions of Arabidopsis thaliana. New Phytol. 2013;198(3):777–87.
9. Samis KE, Murren CJ, Bossdorf O, Donohue K, Fenster CB, Malmberg RL, et al. Longitudinal trends in climate drive flowering time clines in North American Arabidopsis thaliana. Ecol Evol. 2012;2(6):1162–80.
10. Xin Z, Browse J. Cold comfort farm: the acclimation of plants to freezing temperatures. Plant Cell Environ. 2000;23:893–902.
11. Kosova K, Vitamvas P, Prasil IT, Renaut J. Plant proteome changes under abiotic stress–contribution of proteomics studies to understanding plant stress response. J Proteome. 2011;74(8):1301–22.
12. Guy C, Kaplan F, Kopka J, Selbig J, Hincha DK. Metabolomics of temperature stress. Physiol Plant. 2008;132(2):220–35.
13. Stitt M, Hurry V. A plant for all seasons: alterations in photosynthetic carbon metabolism during cold acclimation in Arabidopsis. Curr Opin Plant Biol. 2002;5:199–206.

14. Cook D, Fowler S, Fiehn O, Thomashow MF. A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc Natl Acad Sci U S A. 2004;101(42):15243–8.
15. Maruyama K, Takeda M, Kidokoro S, Yamada K, Sakuma Y, Urano K, et al. Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. Plant Physiol. 2009;150(4):1972–80.
16. Mikkelsen MD, Thomashow MF. A role for circadian evening elements in cold-regulated gene expression in Arabidopsis. Plant J. 2009.
17. Davey MP, Woodward FI, Quick WP. Intraspecfic variation in cold-temperature metabolic phenotypes of Arabidopsis lyrata ssp. petraea. Metabolomics. 2009; 5(1):138–49.
18. Scarth GW, Levitt J. The frost-hardening mechanism of plant cells. Plant Physiol. 1937;12(1):51–78.
19. Klotke J, Kopka J, Gatzke N, Heyer AG. Impact of soluble sugar concentrations on the acquisition of freezing tolerance in accessions of Arabidopsis thaliana with contrasting cold adaptation - evidence for a role of raffinose in cold acclimation. Plant Cell Environ. 2004;27:1395–404.
20. Hincha DK, Sonnewald U, Willmitzer L, Schmitt JM. The role of sugar accumulation in leaf frost hardiness: investigations with transgenic tobacco expressing a bacterial pyrophosphatase or a yeast invertase gene. J Plant Physiol. 1996;147:604–10.
21. Huner NPA, Öquist G, Sarhan F. Energy balance and acclimation to light and cold. Trends Plant Sci. 1998;3(6):224–30.
22. Nägele T, Kandel BA, Frana S, Meissner M, Heyer AG. A systems biology approach for the analysis of carbohydrate dynamics during acclimation to low temperature in Arabidopsis thaliana. FEBS J. 2011;278(3):506–18.
23. Muller B, Pantin F, Genard M, Turc O, Freixes S, Piques M, et al. Water deficits uncouple growth from photosynthesis, increase C content, and modify the relationships between C and growth in sink organs. J Exp Bot. 2011;62(6):1715–29.
24. Goldschmidt EE, Huber SC. Regulation of photosynthesis by End-product accumulation in leaves of plants storing starch, sucrose, and hexose sugars. Plant Physiol. 1992;99(4):1443–8.
25. Sheen J. Feedback control of gene expression. Photosynth Res. 1994;39(3): 427–38.
26. Savitch LV, Barker-Astrom J, Ivanov AG, Hurry V, Oquist G, Huner NP, et al. Cold acclimation of Arabidopsis thaliana results in incomplete recovery of photosynthetic capacity, associated with an increased reduction of the chloroplast stroma. Planta. 2001;214(2):295–303.
27. Strand A, Hurry V, Gustafsson P, Gardeström P. Development of Arabidopsis thaliana leaves at low temperatures releases the suppression of photosynthesis and photosynthetic gene expression despite the accumulation of soluble carbohydrates. Plant J. 1997;12(3):605–14.
28. Masclaux-Daubresse C, Purdy S, Lemaitre T, Pourtau N, Taconnat L, Renou J-P, et al. Genetic variation suggests interaction between cold acclimation and metabolic regulation of leaf senescence. Plant Physiol. 2007;143(1): 434–46.
29. Talts P, Parnik T, Gardestrom P, Keerberg O. Respiratory acclimation in Arabidopsis thaliana leaves at low temperature. J Plant Physiol. 2004; 161(5):573–9.
30. Weckwerth W. Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. J Proteomics. 2011;75(1):284–305.
31. van Norman JM, Benfey PN. Arabidopsis thaliana as a model organism in systems biology. Wiley Interdiscip Rev Syst Biol Med. 2009;1(3):372–9.
32. Hoffmann MH. Biogeography of Arabidopsis thaliana (L.) heynh. (brassicaceae). J Biogeography. 2002;29:125–34.
33. Koornneef M, Alonso-Blanco C, Vreugdenhil D. Naturally occurring genetic variation in Arabidopsis thaliana. Annu Rev Plant Biol. 2004;55:141–72.
34. Lawrence MJ. Variations in natural populations of Arabidopsis thaliana (L.) Heynh. 1976. p. 167–90.
35. Mishra A, Mishra KB, Höermiller II, Heyer AG, Nedbal L. Chlorophyll fluorescence emission as a reporter on cold tolerance in Arabidopsis thaliana accessions. Plant Signal Behav. 2011;6(2):301–10.
36. Brunetti C, George RM, Tattini M, Field K, Davey MP. Metabolomics in plant environmental physiology. J Exp Bot. 2013;64(13):4011–20.
37. Chen D, Neumann K, Friedel S, Kilian B, Chen M, Altmann T, et al. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. Plant Cell Online. 2014;26(12): 4636–55.
38. Weckwerth W. Green systems biology—from single genomes, proteomes and metabolomes to ecosystems research and biotechnology. J Proteome. 2011;75(1):284–305.
39. Beckers GJ, Hoehenwarter W, Rohrig H, Conrath U, Weckwerth W. Tandem metal-oxide affinity chromatography for enhanced depth of phosphoproteome analysis. Methods Mol Biol. 2014;1072:621–32.
40. Chen Y, Hoehenwarter W, Weckwerth W. Comparative analysis of phytohormone responsive phosphoproteins in Arabidopsis thaliana using TiO2phosphopeptide enrichment and mass accuracy precursor alignment. Plant J. 2010;63(1):1–17.
41. Weckwerth W. Unpredictability of metabolism—the key role of metabolomics science in combination with next-generation genome sequencing. Anal Bioanal Chem. 2011;400(7):1967–78.
42. Weckwerth W, Wienkoop S, Hoehenwarter W, Egelhofer V, Sun X. From proteomics to systems biology: MAPA, MASS WESTERN, PROMEX, and COVAIN as a user-oriented platform, Plant Proteomics. New York: Humana Press; 2014. p. 15–27.
43. Morgenthal K, Wienkoop S, Scholz M, Selbig J, Weckwerth W. Correlative GC-TOF-MS-based metabolite profiling and LC-MS-based protein profiling reveal time-related systemic regulation of metabolite–protein networks and improve pattern recognition for multiple biomarker selection. Metabolomics. 2005;1(2):109–21.
44. Barrero-Gil J, Salinas J. Post-translational regulation of cold acclimation response. Plant Sci. 2013;205–206:48–54.
45. Zhu J, Dong CH, Zhu JK. Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. Curr Opin Plant Biol. 2007;10(3):290–5.
46. Hincha DK, Zuther E. Plant cold acclimation and freezing tolerance. Methods Mol Biol. 2014;1166:1–6.
47. Tomé FS, Nägele T, Adamo M, Garg A, Marco-llorca C, Nukarinen E, et al. The low energy signaling network. Front Plant Sci. 2014;5:353.
48. Brauner K, Hormiller I, Nägele T, Heyer AG. Exaggerated root respiration accounts for growth retardation in a starchless mutant of Arabidopsis thaliana. Plant J. 2014;79(1):82–91.
49. Nägele T, Stutz S, Hörmiller II, Heyer AG. Identification of a metabolic bottleneck for cold acclimation in Arabidopsis thaliana. Plant J. 2012;72(1): 102–14.
50. Pons T. Interaction of temperature and irradiance effects on photosynthetic acclimation in two accessions of Arabidopsis thaliana. Photosynth Res. 2012; 113(1–3):207–19.
51. Sulpice R, Pyl E-T, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, et al. Starch as a major integrator in the regulation of plant growth. Proc Natl Acad Sci U S A. 2009;106(25):10348–53.
52. Smith AM, Zeeman SC, Smith SM. Starch degradation. Annu Rev Plant Biol. 2005;56:73–98.
53. Zeeman SC, Smith SM, Smith AM. The diurnal metabolism of leaf starch. Biochem J. 2007;401(1):13–28.
54. Streb S, Eicke S, Zeeman SC. The simultaneous abolition of three starch hydrolases blocks transient starch breakdown in Arabidopsis. J Biol Chem. 2012;287(50):41745–56.
55. Yano R, Nakamura M, Yoneyama T, Nishida I. Starch-related alpha-glucan/ water dikinase is involved in the cold-induced development of freezing tolerance in Arabidopsis. Plant Physiol. 2005;138(2):837–46.
56. Sicher R. Carbon partitioning and the impact of starch deficiency on the initial response of Arabidopsis to chilling temperatures. Plant Sci. 2011; 181(2):167–76.
57. Fincher GB. Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. Annu Rev Plant Physiol Plant Mol Biol. 1989;40(1):305–46.
58. Yu TS, Zeeman SC, Thorneycroft D, Fulton DC, Dunstan H, Lue WL, et al. alpha-Amylase is not required for breakdown of transitory starch in Arabidopsis leaves. J Biol Chem. 2005;280(11):9773–9.
59. Santelia D, Trost P, Sparla F. New insights into redox control of starch degradation. Curr Opin Plant Biol. 2015;25:1–9.
60. Lao NT, Schoneveld O, Mould RM, Hibberd JM, Gray JC, Kavanagh TA. An Arabidopsis gene encoding a chloroplast-targeted beta-amylase. Plant J. 1999;20(5):519–27.
61. Kaplan F, Guy CL. RNA interference of Arabidopsis beta-amylase8 prevents maltose accumulation upon cold shock and increases sensitivity of PSII photochemical efficiency to freezing stress. Plant J. 2005;44(5):730–43.

62. Monroe JD, Storm AR, Badley EM, Lehman MD, Platt SM, Saunders LK, et al. Beta-Amylase1 and beta-amylase3 are plastidic starch hydrolases in Arabidopsis that seem to be adapted for different thermal, pH, and stress conditions. Plant Physiol. 2014;166(4):1748–63.
63. Wienkoop S, Morgenthal K, Wolschin F, Scholz M, Selbig J, Weckwerth W. Integration of metabolomic and proteomic phenotypes analysis of data covariance dissects starch and RFO metabolism from Low and high temperature compensation response in Arabidopsis thaliana. Mol Cell Proteomics. 2008;7(9):1725–36.
64. Thomashow MF. Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. Plant Physiol. 2010;154(2):571–7.
65. Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF. Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. Plant J. 1998;16(4):433–42.
66. Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, et al. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. Plant Cell. 1998;10(8):1391–406.
67. Gilmour SJ, Fowler SG, Thomashow MF. Arabidopsis transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. Plant Mol Biol. 2004;54(5):767–81.
68. Jaglo-Ottosen KR, Gilmour SJ, Zarka DK, Schabenberger O, Thomashow MF. Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. Science. 1998;280:104–6.
69. Rekarte-Cowie I, Ebshish OS, Mohamed KS, Pearce RS. Sucrose helps regulate cold acclimation of Arabidopsis thaliana. J Exp Bot. 2008;59(15): 4205–17.
70. Lastdrager J, Hanson J, Smeekens S. Sugar signals and the control of plant growth and development. J Exp Bot. 2014;65(3):799–807.
71. Monfared MM, Simon MK, Meister RJ, Roig-Villanova I, Kooiker M, Colombo L, et al. Overlapping and antagonistic activities of BASIC PENTACYSTEINE genes affect a range of developmental processes in Arabidopsis. Plant J. 2011;66(6):1020–31.
72. Lee SH, Chung GC, Jang JY, Ahn SJ, Zwiazek JJ. Overexpression of PIP2;5 aquaporin alleviates effects of Low root temperature on cell hydraulic conductivity and growth in Arabidopsis. Plant Physiol. 2012;159(1):479–88.
73. Schulze WX, Schneider T, Starck S, Martinoia E, Trentmann O. Cold acclimation induces changes in Arabidopsis tonoplast protein abundance and activity and alters phosphorylation of tonoplast monosaccharide transporters. Plant J. 2012;69(3):529–41.
74. Rampitsch C, Bykova NV. The beginnings of crop phosphoproteomics: exploring early warning systems of stress. Front Plant Sci. 2012;3:144.
75. Doerfler H, Lyon D, Nägele T, Sun X, Fragner L, Hadacek F, et al. Granger causality in integrated GC-MS and LC-MS metabolomics data reveals the interface of primary and secondary metabolism. Metabolomics. 2013;9(3): 564–74.
76. Weckwerth W, Wenzel K, Fiehn O. Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. Proteomics. 2004;4(1):78–83.
77. Colby T, Röhrig H, Harzen A, Schmidt J. Modified metal-oxide affinity enrichment combined with 2D-PAGE and analysis of phosphoproteomes. Methods Mol Biol. 2011;779:273–86.
78. Furuhashi T, Nukarinen E, Ota S, Weckwerth W. Boron nitride as desalting material in combination with phosphopeptide enrichment in shotgun proteomics. Anal Biochem. 2014;452:16–8.
79. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat Methods. 2007;4(3):231–7.
80. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367–72.
81. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011;10(4):1794–805.
82. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell. 2006;127(3):635–48.
83. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013.
84. Sun X, Weckwerth W. COVAIN: a toolbox for uni-and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. Metabolomics. 2012;8(1):81–93.
85. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, et al. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. 2005;33(Database issue):D433–7.

- Nägele T*, Fürtauer L*, **Nagler M***, Weiszmann J*, Weckwerth W (2016) A Strategy for Functional Interpretation of Metabolomic Time Series Data in Context of Metabolic Network Information. Front Mol Biosci 3: 6

  * contributed equally

  We have established a graph theoretical approach which connects metabolic time series information to the underlying biochemical network to identify regulation in a specific temporal frame with regard to the interaction of derivatives of specific metabolic functions indicating a switch in the dependency of metabolites resulting in an altered homeostasis of the biochemical system. Our approach is strictly analytically derivable and averts numerical estimations. We also provided a graphical user interface that facilitates application of this approach based on the Matlab® software.

  In this publication, to which Thomas Nägele, Lisa Fürtauer, Jakob Weiszmann and myself contributed equally, I was involved in elaborating the algorithm and performed unit testing. I was also involved in writing the manuscript for the publication.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 07 March 2016
doi: 10.3389/fmolb.2016.00006

A Strategy for Functional Interpretation of Metabolomic Time Series Data in Context of Metabolic Network Information

Thomas Nägele[1,2*†], *Lisa Fürtauer*[1†], *Matthias Nagler*[1†], *Jakob Weiszmann*[1†] and *Wolfram Weckwerth*[1,2]

[1] Department of Ecogenomics and Systems Biology, University of Vienna, Vienna, Austria, [2] Vienna Metabolomics Center, University of Vienna, Vienna, Austria

The functional connection of experimental metabolic time series data with biochemical network information is an important, yet complex, issue in systems biology. Frequently, experimental analysis of diurnal, circadian, or developmental dynamics of metabolism results in a comprehensive and multidimensional data matrix comprising information about metabolite concentrations, protein levels, and/or enzyme activities. While, irrespective of the type of organism, the experimental high-throughput analysis of the transcriptome, proteome, and metabolome has become a common part of many systems biological studies, functional data integration in a biochemical and physiological context is still challenging. Here, an approach is presented which addresses the functional connection of experimental time series data with biochemical network information which can be inferred, for example, from a metabolic network reconstruction. Based on a time-continuous and variance-weighted regression analysis of experimental data, metabolic functions, i.e., first-order derivatives of metabolite concentrations, were related to time-dependent changes in other biochemically relevant metabolic functions, i.e., second-order derivatives of metabolite concentrations. This finally revealed time points of perturbed dependencies in metabolic functions indicating a modified biochemical interaction. The approach was validated using previously published experimental data on a diurnal time course of metabolite levels, enzyme activities, and metabolic flux simulations. To support and ease the presented approach of functional time series analysis, a graphical user interface including a test data set and a manual is provided which can be run within the numerical software environment Matlab®.

Keywords: metabolic network, data integration, metabolomics, time series analysis, systems biology, network dynamics

OPEN ACCESS

Edited by:
Guowang Xu,
Chinese Academy of Sciences, China

Reviewed by:
Michal Jan Markuszewski,
Medical University of Gdansk, Poland
Huub C. J. Hoefsloot,
University of Amsterdam, Netherlands

****Correspondence:***
Thomas Nägele
thomas.naegele@univie.ac.at

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

Received: *11 December 2015*
Accepted: *19 February 2016*
Published: *07 March 2016*

Citation:
Nägele T, Fürtauer L, Nagler M, Weiszmann J and Weckwerth W (2016) A Strategy for Functional Interpretation of Metabolomic Time Series Data in Context of Metabolic Network Information. Front. Mol. Biosci. 3:6. doi: 10.3389/fmolb.2016.00006

INTRODUCTION

The functional interpretation of experimental data in context of biochemical network information represents one of the central challenges in current biological research. While genome sequencing projects have enabled the reconstruction of genome-scale metabolic networks, their high dimensionality precludes a direct and intuitive application to interpret experimental data. Hence, although genome sequence information and metabolic networks have become available for

numerous organisms, tissues, or cell types (Herrgard et al., 2008; Chang et al., 2011; De Oliveira Dal'Molin and Nielsen, 2013; Thiele et al., 2013), functional metabolic data interpretation still represents a major obstacle in systems biology. Various mathematical and computational strategies from the fields of multivariate statistics, ordinary, and partial differential equations (ODEs/PDEs), optimization or statistical time series analysis have been developed and applied to reveal a biologically meaningful interpretation of comprehensive and multidimensional experimental data sets. For example, a computational model of starch metabolism in plants enabled the analysis of starch kinetics through diurnal metabolic and circadian sensors (Pokhilko et al., 2014). The authors developed a model of 28 ODEs which were numerically simulated in order to analyze diurnal kinetics of carbon metabolism *in silico*. In another study, the response of *Escherichia coli* to varying oxygen concentrations was analyzed applying a mathematical model of the central metabolism (Ederer et al., 2014). Here, the authors derived a prediction about the impact of product formation on biomass concentration using steady state simulations at varying environmental conditions.

Both examples for mathematical modeling differ in organism and application. Besides, the dynamic approach can be distinguished from the steady state approach. However, in both approaches, dynamics of metabolic systems can be described by sets of ODEs. If sufficient kinetic information is available, such ODEs can be numerically integrated revealing simulated metabolic concentrations depending on time, enzyme parameters, thermodynamic constraints, etc. Yet, statistically robust experimental enzyme kinetic information often limits the applicability of such modeling approaches. Particularly, the resolution of enzyme activities, substrate affinities, or inhibitory constants is very laborious and only possible if well-established experimental assays and sufficient biochemical knowledge are available. Additionally, uncertainties about model structures and reaction kinetics complicate the interpretation of a numerically simulated output (Schaber et al., 2009). Such limitations have been addressed by different theoretical approaches, for example by structural kinetic modeling, SKM (Steuer et al., 2006). In the SKM approach, local linear models are applied to explore and statistically analyze a given parameter space without the need for explicit information about functional forms of enzyme kinetics and rate equations. Finally, a Jacobian matrix is derived which characterizes the dynamic capabilities of a metabolic system at a certain steady state. In previous publications, we have developed a procedure to determine Jacobian matrices directly from experimental metabolomics data (Nägele, 2014; Nägele et al., 2014). Based on experimental metabolic (co)variance information a metabolic regulator was identified indicating a strategy how plant metabolism is reprogrammed during exposure to energy limiting conditions. In a different context, other studies have also shown that it is possible to infer regulatory information about metabolic steady states from experimental data with such approaches (see e.g., Steuer et al., 2003; Sun and Weckwerth, 2012; Kügler and Yang, 2014).

Beyond these approaches of dynamic and steady state modeling, time series analysis and related regression models offer another mathematical strategy to reveal information about molecular system dynamics (Schelter, 2006). For example, Dutta and co-workers developed an algorithm for identification of differentially expressed genes in a time series experiment (Dutta et al., 2007), which they also applied to integrate transcriptome and metabolome data (Dutta et al., 2009). In another study, statistical modeling and regression analysis revealed a nitrogen-dependent modulation of root system architecture in the genetic model plant *Arabidopsis thaliana* (Araya et al., 2015). While these exemplarily mentioned studies present only a very small fraction of possible statistical applications, it already becomes evident that these are promising and necessary mathematical approaches to reveal biologically meaningful information from comprehensive experimental data sets being preliminary for hypothesis generation and experimental validation. However, a common problem of regression and correlation approaches in a biochemical context is a missing functional linkage of the results to causal biochemical interrelations, i.e., enzymatically driven reactions. To overcome this limitation and to facilitate the biochemical interpretation of the statistical results, the present study derives a theoretical connection between mathematical approaches of ODE-based dynamic modeling and statistical time series analysis. Based on the stoichiometric matrix information of a metabolic network, ratios of time-dependent derivative functions were built providing an estimate for the strength and probability of a metabolic interaction during the time course. The suggested strategy was tested using previously published experimental data sets on diurnal and stress-induced dynamics of metabolite concentrations and related enzyme kinetic information. Finally, a graphical user interface for Matlab is provided which intends to facilitate the application of the presented strategy.

RESULTS

Deriving Metabolic Functions by Inverse Variance-Weighted Regression Analysis

Time-dependent dynamics of metabolite concentrations in a biochemical network can be described by a set of ODEs:

$$\frac{d}{dt}\boldsymbol{M}(t) = \boldsymbol{N}\boldsymbol{v}\left(\boldsymbol{M},\boldsymbol{p},t\right) = f(\boldsymbol{M},\boldsymbol{p},t) \tag{1}$$

Here, $\boldsymbol{M}$ represents an n-dimensional vector of mean metabolite concentrations (c_n), $\boldsymbol{N}$ is the $n \times k$ stoichiometric matrix and $\boldsymbol{v}$ describes the k-dimensional vector of reaction rates which depend on metabolite concentrations $\boldsymbol{M}$, enzyme kinetic parameters $\boldsymbol{p}$ and time t. The right side of the ODE system can be summarized by metabolic functions, $f(\boldsymbol{M},\boldsymbol{p},t)$. Hence, these metabolic functions define the time-dependent changes in metabolite concentration as a sum of all biochemical reactions either consuming or producing a metabolite. A metabolic steady state is described by ODEs which equal zero:

$$\frac{d}{dt}\boldsymbol{M}(t) = 0 \tag{2}$$

Linearization enables the investigation of the system behavior close to a steady state. The linearization process results in the so-called Jacobian matrix $\boldsymbol{J}$ which characterizes the dynamic properties of the system at a steady state:

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial c_1} & \cdots & \frac{\partial f_1}{\partial c_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial c_1} & \cdots & \frac{\partial f_n}{\partial c_n} \end{pmatrix} \quad (3)$$

Hence, in a biochemical context, the Jacobian matrix $\boldsymbol{J}$ describes the behavior of metabolic functions f_i (for $i = 1,...,n$) of a metabolic network with regard to small changes of variables c_i (for $i = 1,...,n$), i.e., metabolite concentrations at the considered steady state. The information if a metabolic function f_i biochemically depends on the concentration of a metabolite c_i is provided by the stoichiometric matrix $\boldsymbol{N}$ of a metabolic network (see Equation 1).

To derive, i.e., predict, time continuous information from time discrete experimental observations, interpolation methods can be applied. To prevent unrealistic oscillations of high-degree polynomial interpolation, intervals of approximation can be partitioned in subintervals which can be approximated, for example, by cubic polynomials which form a cubic spline $S_{c_i}(t)$ (see e.g., Bronstein et al., 2008):

$$S_{c_i}(t) = \alpha_{ij} + \beta_{ij}(t - t_j) + \gamma_{ij}(t - t_j)^2 + \delta_{ij}(t - t_j)^3 \quad (4)$$

Here, it is $t \in [t_j, t_{j+1}]$ with $(j = 1, 2,..., z-1)$, where z represents the number of interpolation nodes (t_j, g_{ij}), and it is $S_{c_i}(t_j) = g_{ij}$. Interpolation coefficients are represented by α, β, γ, and δ. Due to the occurrence of experimental errors, the requirement of $S_{c_i}(t_j) = g_{ij}$ is not fulfilled which prevents the suitability of such a type of interpolation. Instead, a smoothing element can be introduced accounting for those experimental errors:

$$min\left(\sum_{j=1}^{z} w_{ij}[g_{ij} - S_{c_i}(t_j)]^2 + \lambda \int_{t=t_1}^{t_z} [S_{c_i}''(t)]^2 dt\right) \quad (5)$$

Here, w_{ij} represents a weighting factor, $S_{c_i}''(t)$ is the second derivative of $S_{c_i}(t)$, and λ (with $\lambda \geq 0$) represents a smoothing factor. For $\lambda = 0$, one obtains the cubic spline interpolation, while the degree of smoothing increases with the value of λ. To connect the smoothing spline generation to experimentally observed errors we defined the weighting factor w_{ij} to equal the inverse variance information, i.e., the inverse squared standard deviation:

$$w_{ij} = \sigma_{ij}^{-2} = \left(\frac{1}{r-1}\sum_{k=1}^{r}(c_{ij,k} - \bar{c}_{ij})^2\right)^{-1} \quad (6)$$

Here, r represents the number of experimental and independent replicates.

Merging Equations (1), (5) and (6) and replacing g_{ij} by the mean concentration of metabolite i at time point j, $\bar{c}_{ij}$, reveals a description of metabolic functions by inverse variance-weighted regression analysis:

$$\begin{aligned} \frac{d}{dt}M_i(t) &= f_i(\boldsymbol{M},\boldsymbol{p},t) \\ &= \frac{d}{dt}\left[min\left(\sum_{j=1}^{z}\left[\left(\frac{1}{r-1}\sum_{k=1}^{r}(c_{ij,k} - \bar{c}_{ij})^2\right)^{-1}\right.\right.\right. \\ &\quad \left.\left.\left.[\bar{c}_{ij} - S_{c_i}(t_j)]^2\right] + \lambda\int_{t=t_1}^{t_z}[S_{c_i}''(t)]^2 dt\right)\right] \quad (7) \end{aligned}$$

Hence, building the first derivative of the smoothed interpolation of experimental time-course data reveals information about the connected metabolic function. In the present study, this approach was applied to evaluate a diurnal time course of previously published metabolite concentrations (Nägele et al., 2012) belonging to the central carbohydrate metabolism in leaves of the genetic model plant *A. thaliana*. Diurnal dynamics of metabolic functions are shown exemplarily (**Figure 1**) for the metabolite pools of sucrose (Suc) and sugar phosphates (SP) in a control experiment (non-cold acclimated, na) and after exposure to low temperature (acc).

To characterize time-dependent changes in metabolic functions, the second time-dependent derivative was built from the approximated diurnal time course of metabolite concentrations:

$$\begin{aligned} \frac{d^2}{dt^2}M_i(t) &= \frac{d}{dt}f_i(\boldsymbol{M},\boldsymbol{p},t) \\ &= \frac{d^2}{dt^2}\left[min\left(\sum_{j=1}^{z}\left[\left(\frac{1}{r-1}\sum_{k=1}^{r}(c_{ij,k} - \bar{c}_{ij})^2\right)^{-1}\right.\right.\right. \\ &\quad \left.\left.\left.[\bar{c}_{ij} - S_{c_i}(t_j)]^2\right] + \lambda\int_{t=t_1}^{t_z}[S_{c_i}''(t)]^2 dt\right)\right] \quad (8) \end{aligned}$$

As described for **Figure 1**, diurnal dynamics of those time-dependent changes of metabolic functions are also shown exemplarily for Suc and SP (**Figure 2**).

Connecting Metabolic Functions Based on Biochemical Network Information

While the metabolic time-course information derived before characterizes time-dependent rates of changes in each considered metabolite concentration separately (see Equations 7 and 8), information of biochemical interdependencies, i.e., the information about a substrate - product relationship between two or more metabolites, is only contained implicitly. To explicitly analyze and visualize these biochemical interdependencies with regard to the time-dependent rates of concentrations changes, a metabolic $n \times n$ interaction matrix, $\boldsymbol{Y}$, was derived where n represents the number of metabolites comprised by the model. In $\boldsymbol{Y}$, each entry indicates whether two metabolites are biochemically connected (entry: 1) or not (entry: 0). The interaction is characterized analogous to entries of the Jacobian

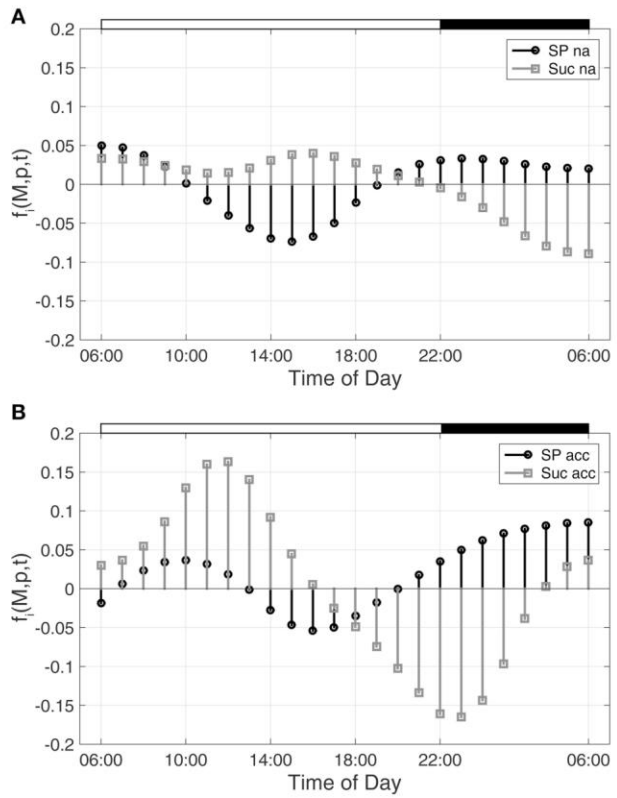


FIGURE 1 | Metabolic functions derived from inverse variance-weighted regression analysis for (A) non-cold acclimated and (B) cold acclimated plants. Metabolic functions $f_i(\boldsymbol{M},\boldsymbol{p},t)$ were derived for sugar phosphates (SP) and sucrose (Suc) as described in the text (see Equation 7). Experimental data were derived from a previous study comprising metabolite levels of non-cold acclimated (na) and cold acclimated (acc) leaf material of *Arabidopsis thaliana*, accession Col-0 (Nägele et al., 2012). White and black bars on the top indicate light and dark phase of a diurnal cycle.

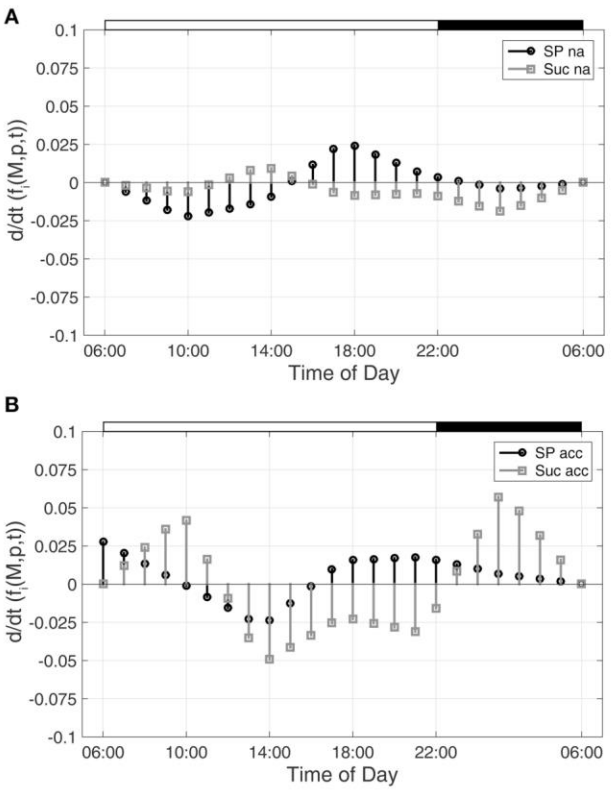


FIGURE 2 | Time-dependent dynamics of metabolic functions derived from inverse variance-weighted regression analysis for (A) non-cold acclimated and (B) cold acclimated plants. Function dynamics were derived for sugar phosphates (SP) and sucrose (Suc) as described in the text (see Equation 8). Experimental data were derived from a previous study comprising metabolite levels of non-cold acclimated (na) and cold acclimated (acc) leaf material of *A. thaliana*, accession Col-0 (Nägele et al., 2012). White and black bars on the top indicate light and dark phase of a diurnal cycle.

matrix (Equation 3): if the metabolic function of metabolite A is biochemically connected to changes in concentrations of metabolite B, the corresponding entry in $\boldsymbol{Y}$ is 1. Information about metabolic functions is given row-wise, while biochemically connected metabolites are indicated column-wise for each function. In a simple example, containing three reactions (r_1–r_3) and four metabolites (A–D), the construction and content of $\boldsymbol{Y}$ is exemplified (**Figure 3**). The diagonal entries indicate the biochemical dependencies of metabolic functions on substrate concentrations. For example, $\boldsymbol{Y}_{11} = 1$ indicates that metabolic function f(A,t) depends on the concentration of A(t). The non-diagonal entries describe interdependencies between different metabolite pools. For example, $\boldsymbol{Y}_{21} = 1$ indicates that metabolic function f(B,t) depends on the concentration of A(t).

Based on a previously published metabolic model (Nägele et al., 2012), an interaction matrix $\boldsymbol{Y}$ was derived for the central carbohydrate metabolism in leaves of *A. thaliana*. The metabolic functions (Equation 7) and their time-dependent derivatives (Equation 8) were related to each other according to the entries of $\boldsymbol{Y}$. This finally resulted in functions $\omega(a \rightarrow b, t)$ indicating changes in metabolic functions of b in context of concentration changes of a which might represent substrates, inhibitors or activators:

$$\omega\left(a \rightarrow b, t\right) = \frac{\frac{d}{dt} f_b(\boldsymbol{M},\boldsymbol{p}, t)}{f_a(\boldsymbol{M},\boldsymbol{p}, t)},\ D = \left\{\mathbb{R} \setminus f_a\left(\boldsymbol{M},\boldsymbol{p}, t\right) = 0\right\} \quad (9)$$

With regard to the analyzed time-course of sugar phosphate (SP) and sucrose (Suc) concentrations (see **Figures 1**, **2**), ω(SP→Suc, t) revealed information about the reaction of sucrose biosynthesis, catalyzed by the enzyme sucrose phosphate synthase (SPS):

$$SugarPhosphates\ (SP) \overset{SPS}{\rightarrow} Sucrose \quad (10)$$

In detail, ω(SP→Suc, t) described changes in the metabolic function of sucrose in context of concentration changes of its biochemical substrate sugar phosphates:

$$\omega\left(SP \rightarrow Suc, t\right) = \frac{\frac{d}{dt} f_{Suc}(\boldsymbol{M},\boldsymbol{p}, t)}{f_{SP}(\boldsymbol{M},\boldsymbol{p}, t)} \quad (11)$$

Comparing ω(SP→Suc, t) for na and acc plants revealed a noticeable difference between both conditions within the first

4 h of the day (**Figure 4**). Interestingly, in the same time period, simulations of sucrose biosynthesis, based on a system of ordinary differential equations (ODEs), revealed a similar picture in which rates of sucrose biosynthesis were decreased only in acc plants (Nägele et al., 2012).

Characterization of ω(t)

$\omega(t)$ represents a ratio of metabolic functions and related derivatives. Hence, the unit of ω(t) is derived from the flux unit of a metabolic function, [mM s^{-1}]:

$$\left[\omega\left(a \rightarrow b, t\right)\right] = \frac{\left[\frac{d}{dt} f_b(\boldsymbol{M},\boldsymbol{p}, t)\right]}{\left[f_a(\boldsymbol{M},\boldsymbol{p}, t)\right]} = \frac{\left[mM\ s^{-2}\right]}{\left[mM\ s^{-1}\right]} = \left[\frac{1}{s}\right] \quad (12)$$

Here, concentrations are given in mM (mmol l^{-1}) and the time unit is seconds (s).

This results in the unit of a rate or frequency. Hence, $|\omega(a \rightarrow b,t)|$ was interpreted as oscillations of a metabolic function per time-period with reference to a biochemical effector.

In the case of $|\omega(a \rightarrow b,\ t)| \rightarrow \infty$ for $t \rightarrow \tau$, the influence of the biochemical effector on a metabolic function was defined to be strong, while $|\omega(a \rightarrow b,\ t)| \rightarrow 0$ for $t \rightarrow \tau$ indicated a weak effect. In detail, $|\omega(a \rightarrow b,\ t)| \rightarrow \infty$ for $t \rightarrow \tau$ indicates that it is $|d/dt\ (f_b(\boldsymbol{M},\boldsymbol{p},t))| >> |f_a(\boldsymbol{M},\boldsymbol{p},t)|$. *Vice versa*, $|\omega(a \rightarrow b,\ t)| \rightarrow 0$ for $t \rightarrow \tau$ indicates that $|d/dt\ (f_b(\boldsymbol{M},\boldsymbol{p},t))| << |f_a(\boldsymbol{M},\boldsymbol{p},t)|$.

Application Example: Stress-Induced Metabolic Reprogramming in *Arabidopsis thaliana*

While in the above mentioned example, the calculation and interpretation of ω(t) was demonstrated in context of a previously published kinetic ODE model, another published data set was analyzed by this strategy comprising metabolite levels of the primary and secondary metabolism in *A. thaliana* (Doerfler et al., 2013). In the experiment performed by Doerfler and co-workers, a combined strategy of gas chromatography and liquid chromatography coupled to mass spectrometry was applied in order to reveal a comprehensive picture of metabolic reprogramming during exposure to low temperature and high light intensity. The time period of stress exposure comprised more than 2 weeks which allowed for the analysis of a short- and long-term acclimation response in the metabolome of *A. thaliana,* accession Col-0. A central output of the study was the characterization of metabolomic and regulatory dynamics at the interface of primary and secondary metabolism. The authors observed a fast increase of stress-responsive compounds, e.g., sucrose, which became significant already after 2 days of stress exposure, while the interaction with the secondary metabolism, resulting in biosynthesis of flavonoids, became most significant after 8 days of stress exposure.

To prove the suitability of deriving the absolute value function $|\omega(a \rightarrow b,\ t)|$ in order to reveal steps of metabolic regulation within a considered time interval, regression analysis and metabolic functions were calculated for the dataset of Doerfler et al. (2013) and compared to their observations. The metabolic interaction matrix $\boldsymbol{Y}$ was derived from the metabolic network model which was previously suggested and applied for inverse approximations of the Jacobian matrix (Doerfler et al., 2013). For regression analysis and for integration of metabolic network information we developed and applied a graphical user interface (*FEMTO*, ***F***unctional ***E***valuation of ***M***etabolic ***T***ime series ***O***bservations) which is based on the numerical software environment Matlab® (http://www.mathworks.com), and which is provided in the supplements together with a user manual (**Supplementary Files S1**, **S2**).

To characterize sucrose metabolism, changes of the metabolic function of sucrose were related to changes in sucrose concentrations:

$$|\omega\ (Suc \rightarrow Suc,\ t)| = \left|\frac{\frac{d}{dt} f_{Suc}(\boldsymbol{M},\boldsymbol{p},\ t)}{f_{Suc}(\boldsymbol{M},\boldsymbol{p},\ t)}\right| \quad (13)$$

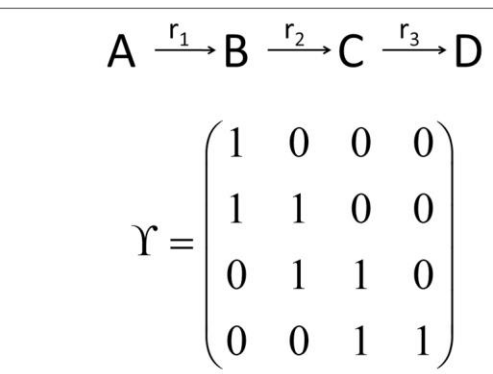


FIGURE 3 | Schematic reaction chain and the derived interaction matrix Y. Rows (metabolic functions) and columns (metabolites) of **Y** describe biochemical interactions of metabolites A (first row/column), B (second row/column), C (third row/column), and D (fourth row/column). Entries, i.e., 0 and 1, indicate if two metabolites interact (entry 1) or not (entry 0).

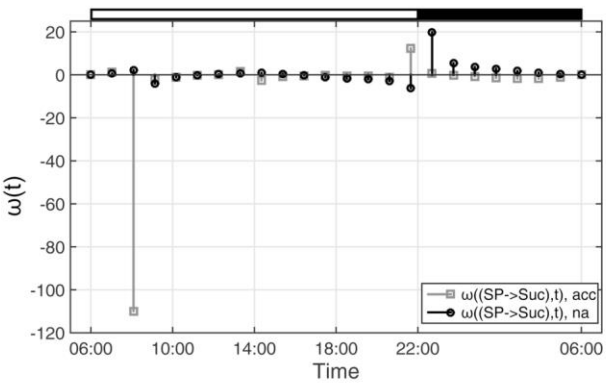


FIGURE 4 | Functions ω(t) indicating changes in metabolic functions in context of concentration changes of biochemical interaction partners. ω(t) was calculated as described in the main text (see Equation 9). Results of ω(t) are shown for the reaction of sucrose biosynthesis for non-cold acclimated (na) and cold acclimated (acc) leaf material of *A. thaliana,* accession Col-0 (Nägele et al., 2012). White and black bars on the top indicate light and dark phase of a diurnal cycle.

For time-dependent characterization of flavonoid dynamics, changes in the flavonoid (Flav) metabolic function were related to substrate concentration changes, i.e., phenylalanine (Phe) dynamics:

$$\left|\omega\left(Phe \rightarrow Flav, t\right)\right| = \left|\frac{\frac{d}{dt} f_{Flav}(\boldsymbol{M},\boldsymbol{p}, t)}{f_{Phe}(\boldsymbol{M},\boldsymbol{p}, t)}\right| \tag{14}$$

Results of metabolic function analysis and the resulting time course of $|\omega(t)|$ revealed an early de-regulation of sucrose metabolism during the first 2 days of stress exposure (**Figure 5A**), while the peak value of $|\omega(t)|$ for flavonoid biosynthesis occurred delayed after 8 days (**Figure 5B**).

These findings coincide with the previous findings described by Doerfler and colleagues who applied the method of Granger causality time-series correlation and a covariance-based inverse approximation of Jacobian matrices to reveal strategies of metabolic regulation (Doerfler et al., 2013). Conclusions which have been drawn from the $|\omega(t)|$ calculation were found to be highly similar to the output of other statistical methods, finally substantiating the validity of the suggested workflow and the derived method to unravel time points of regulatory perturbation in a biochemical system.

DISCUSSION

Mathematical analysis of biochemical system dynamics represents a central focus of current biomathematical, biochemical and biotechnological research due to the need for methods and algorithms enabling a functional interpretation of experimental data in context of a biochemical network. Particularly, system dynamics which arise due to circadian regulation (Harmer, 2009; Kumar Jha et al., 2015), diurnal metabolic adjustment (Geiger and Servaites, 1994; Pokhilko et al., 2014) or stress-induced metabolic reprogramming (Jozefczuk et al., 2010; Kanshin et al., 2015) are hardly traceable by intuition. Hence, this indicates a strong need for suitable theoretical approaches being capable of resolving and functionally connecting molecular moieties with underlying biochemical regulation.

Various theoretical strategies have addressed this complex issue, providing a comprehensive methodological platform for time-series analysis, dynamic flux balance analysis, kinetic and Boolean modeling (see e.g., Mahadevan et al., 2002; Schelter, 2006; Rohwer, 2012; Steinway et al., 2015). In a recent approach, Willemsen and colleagues have modified the approach of dynamic flux balance analysis by incorporating time-resolved metabolomics measurements (Willemsen et al., 2015). With their extended method, the authors derived an estimate of dynamic flux profiles which allowed them to generate and test hypotheses related to environmentally induced molecular dynamics. In another recent study, a computational approach was suggested to translate metabolomics data into flux information (Cortassa et al., 2015). One main methodological difference between the studies of Willemsen et al. and Cortassa et al. was the extent of kinetic information which was needed to estimate cellular behavior and metabolic fluxes. While Willemsen et al. focused on minimalistic kinetic information, the study of Cortassa and co-workers used a detailed kinetic model of glucose catabolic pathways to derive flux information.

In our presented approach, flux information, which was implicitly derived from spline interpolation, was interpreted only indirectly by comparing time-dependent changes in metabolic functions to concentration changes of biochemical reaction partners. This procedure revealed information about a rate which was interpreted in terms of metabolic functions related to concentration changes in a substrate or co-substrate. Comparing derived results to other methods, it was shown that changes in ratios of second- to first-order derivatives between functionally connected variables potentially reveal time points of regulatory perturbation within a biochemical interaction. Hence, these observed perturbations might indicate a change in enzymatic activity, protein abundance, or allosteric regulation ultimately leading to a change in the metabolic functions.

The information content of the introduced time-dependent functions $\omega(t)$ is related to entries of the Jacobian matrix $\boldsymbol{J}$ (see Equation 3) indicating the dynamics of metabolic functions with respect to (small) concentration changes at a certain steady state. This theoretical connection of $\boldsymbol{J}$ and $\omega(t)$ at a considered time point t_0 might be illustrated in a simple first-order reaction scheme.

$$A \xrightarrow{k} B \tag{15}$$

Here, substance A is interconverted into substance B, and the reaction velocity is characterized by the rate constant k. The time-dependent change in concentration of A equals $dA/dt = -k{\cdot}A$. Hence, a general solution of this ODE is given by $A(t) = A_0 e^{-\mathrm{kt}}$ which finally yields $J_{11}(t_0) = \omega(A \rightarrow A, t_0) = -\mathrm{k}$.

With this, the information of $\omega(t)$ becomes comparable to entries of the Jacobian matrix $\boldsymbol{J}$. Yet, in contrast to entries of $\boldsymbol{J}$, characterizing dynamic properties of a metabolic steady state ($d/dt\ \boldsymbol{M}(t) = 0$), functions $\omega(t)$ were derived from a time series of experimental data and might rather be valid for a non-infinitesimal than for an infinitesimal time frame. While for $\lim_{t \rightarrow t_0} \left|\omega(t)\right|$, $|\omega(t)|$ might be assumed to approach entries of $\boldsymbol{J}$, this was not tested in the present study and would need experimental validation. In addition, while a connection, and probable correlation, to other molecular levels, such as the proteome or transcriptome, was not experimentally analyzed, this might be a promising target for analysis in future studies. However, the incorporation of an interaction matrix, which, in the present study, was derived from a previously published reaction network, and which might be derived from genome-scale metabolic reconstruction works in future studies (Weckwerth, 2011; King et al., 2015), provides direct evidence for the biochemical and physiological relevance of the performed theoretical analysis.

While our results indicate a realistic and biochemically interpretable output of the presented method, limitations of application might occur due to several reasons. First, the presented method significantly depends on the knowledge about the biochemical network structure and involved regulatory interactions, e.g., feedback inhibition or feedforward activation. Although regression analysis of time series data might be

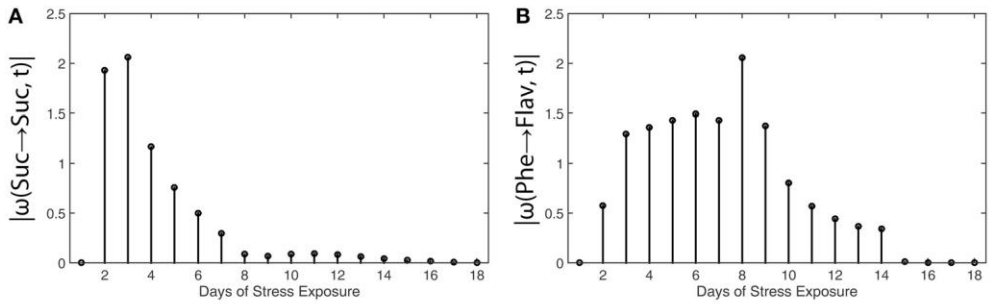


FIGURE 5 | Absolute value functions of ω(t) for (A) sucrose metabolism and (B) flavonoid biosynthesis in leaves of *Arabidopsis thaliana*. Abscissae indicate days of exposure to low temperature and high-light stress conditions. Detailed information about the calculation is provided in the main text (see Equations 13 and 14). Experimental data were derived from a previous study (Doerfler et al., 2013).

performed for all network components independently, deriving a reliable biochemical interaction matrix $\boldsymbol{Y}$ is essential to reveal realistic information about time-dependent changes in metabolic interactions. A second central prerequisite for a meaningful regression analysis is the design of an adequate experimental setup. This comprises the number of biological (independent) replicates as well as the number and interval of sampling points. It is hardly possible to generalize a number of replicates or sampling points due to heterogeneous technical or environmental fluctuations which are introduced by different analytical techniques, growth conditions or sample types. Yet, spanning various experimental scenarios, it might be generalized that the interval of sampling points is crucial to be able to discriminate between metabolic fast or short-term responses and slow or long-term responses. Particularly to resolve fast metabolic regulation, a narrow sampling interval is needed in order to prevent any over-interpretation of regression analysis and related derivatives. Comparing the presented approach to methods of metabolic modeling, a third major limitation is the missing predictive output by model simulations. For example, enzyme kinetic models of metabolism aim at going beyond the time interval of measured rate constants or metabolite concentrations to predict changes in system dynamics under changing environmental conditions or due to a mutated gene. However, although our presented method cannot afford this simulation output, time-dependent changes within the considered time interval might indicate regulatory bottlenecks and kinetic information supporting the numerical solution and simulation of metabolic ODE models.

In summary, the suggested approach intends to promote the functional interpretability of metabolic time series data in context of metabolic network information. Particularly with regard to multidimensional metabolomics data sets, this might unravel strategies of complex biochemical regulation and might overcome some limitations in the generation of testable hypotheses as we have discussed previously (Nägele and Weckwerth, 2012). Finally, the direct integration of biochemical network information with experimental data promises to enable the functional interpretation and the causal connection of various levels of molecular organization.

MATERIALS AND METHODS

The described procedure of data analysis, spline interpolation and graphical representation was performed within the numerical software environment Matlab®. A Matlab-based graphical user interface (FEMTO, *F*unctional *E*valuation of *M*etabolic *T*ime series *O*bservations) was developed and is provided, together with a user manual, in the supplements (**Supplementary Files S1**, **S2**).

AUTHOR CONTRIBUTIONS

TN, LF, MN, and JW performed data analysis, statistics, wrote the source code of the graphical user interface and wrote the paper. TN and WW conceived the study and wrote the paper. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the Austrian Science Fund, FWF, Project P 26342 and Project I 2071.

ACKNOWLEDGMENTS

We would like to thank the members of the MoSys Department for constructive advice and fruitful discussions. Further, we thank Arnd G. Heyer from the Department of Plant Biotechnology at the University of Stuttgart, Germany, for providing experimental data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmolb.2016.00006

Supplementary File S1 | Matlab code-file and test data file.

Supplementary File S2 | Manual for the graphical user interface FEMTO 1.0.

REFERENCES

Araya, T., Kubo, T., von Wirén, N., and Takahashi, H. (2015). Statistical modeling of nitrogen-dependent modulation of root system architecture in *Arabidopsis thaliana*. *J. Integr. Plant. Biol.* doi: 10.1111/jipb.12433. [Epub ahead of print].

Bronstein, I. J. N., Semendjajew, K. A., Musiol, G., and Mühlig, H. (2008). *Taschenbuch der Mathematik*. Frankfurt am Main: Wissenschaftlicher Verlag Harri Deutsch GmbH.

Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F., Balaji, S., Fu, W., et al. (2011). Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* 7, 518. doi: 10.1038/msb.2011.52

Cortassa, S., Caceres, V., Bell, L. N., O'Rourke, B., Paolocci, N., and Aon, M. A. (2015). From metabolomics to fluxomics: a computational procedure to translate metabolite profiles into metabolic fluxes. *Biophys. J.* 108, 163–172. doi: 10.1016/j.bpj.2014.11.1857

De Oliveira Dal'Molin, C. G., and Nielsen, L. K. (2013). Plant genome-scale metabolic reconstruction and modelling. *Curr. Opin. Biotechnol.* 24, 271–277. doi: 10.1016/j.copbio.2012.08.007

Doerfler, H., Lyon, D., Nägele, T., Sun, X., Fragner, L., Hadacek, F., et al. (2013). Granger causality in integrated GC-MS and LC-MS metabolomics data reveals the interface of primary and secondary metabolism. *Metabolomics* 9, 564–574. doi: 10.1007/s11306-012-0470-0

Dutta, B., Kanani, H., Quackenbush, J., and Klapa, M. I. (2009). Time-series integrated "omic" analyses to elucidate short-term stress-induced responses in plant liquid cultures. *Biotechnol. Bioeng.* 102, 264–279. doi: 10.1002/bit.22036

Dutta, B., Snyder, R., and Klapa, M. I. (2007). Significance analysis of time-series transcriptomic data: a methodology that enables the identification and further exploration of the differentially expressed genes at each time-point. *Biotechnol. Bioeng.* 98, 668–678. doi: 10.1002/bit.21432

Ederer, M., Steinsiek, S., Stagge, S., Rolfe, M. D., Ter Beek, A., Knies, D., et al. (2014). A mathematical model of metabolism and regulation provides a systems-level view of how *Escherichia coli* responds to oxygen. *Front. Microbiol.* 5:124. doi: 10.3389/fmicb.2014.00124

Geiger, D. R., and Servaites, J. C. (1994). Diurnal regulation of photosynthetic carbon metabolism in C3 plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 45, 235–256. doi: 10.1146/annurev.pp.45.060194.001315

Harmer, S. L. (2009). The circadian system in higher plants. *Annu. Rev. Plant Biol.* 60, 357–377. doi: 10.1146/annurev.arplant.043008.092054

Herrgard, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., et al. (2008). A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* 26, 1155–1160. doi: 10.1038/nbt1492

Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhauser, D., et al. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* 6, 364. doi: 10.1038/msb.2010.18

Kanshin, E., Kubiniok, P., Thattikota, Y., D'Amours, D., and Thibault, P. (2015). Phosphoproteome dynamics of *Saccharomyces cerevisiae* under heat shock and cold stress. *Mol. Syst. Biol.* 11, 813. doi: 10.15252/msb.20156170

King, Z. A., Lloyd, C. J., Feist, A. M., and Palsson, B. O. (2015). Next-generation genome-scale models for metabolic engineering. *Curr. Opin. Biotechnol.* 35, 23–29. doi: 10.1016/j.copbio.2014.12.016

Kügler, P., and Yang, W. (2014). Identification of alterations in the Jacobian of biochemical reaction networks from steady state covariance data at two conditions. *J. Math. Biol.* 68, 1757–1783. doi: 10.1007/s00285-013-0685-3

Kumar Jha, P., Challet, E., and Kalsbeek, A. (2015). Circadian rhythms in glucose and lipid metabolism in nocturnal and diurnal mammals. *Mol. Cell Endocrinol.* *418*(Pt 1), 74–88. doi: 10.1016/j.mce.2015.01.024

Mahadevan, R., Edwards, J. S., and Doyle, F. J. III (2002). Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* 83, 1331–1340. doi: 10.1016/S0006-3495(02)73903-9

Nägele, T. (2014). Linking metabolomics data to underlying metabolic regulation. *Front. Mol. Biosci.* 1:22. doi: 10.3389/fmolb.2014.00022

Nägele, T., Mair, A., Sun, X., Fragner, L., Teige, M., and Weckwerth, W. (2014). Solving the differential biochemical Jacobian from metabolomics covariance data. *PLoS ONE* 9:e92299. doi: 10.1371/journal.pone.0092299

Nägele, T., Stutz, S., Hörmiller, I., and Heyer, A. G. (2012). Identification of a metabolic bottleneck for cold acclimation in *Arabidopsis thaliana*. *Plant J.* 72, 102–114. doi: 10.1111/j.1365-313X.2012.05064.x

Nägele, T., and Weckwerth, W. (2012). Mathematical modeling of plant metabolism—from reconstruction to prediction. *Metabolites* 2, 553–566. doi: 10.3390/metabo2030553

Pokhilko, A., Flis, A., Sulpice, R., Stitt, M., and Ebenhöh, O. (2014). Adjustment of carbon fluxes to light conditions regulates the daily turnover of starch in plants: a computational model. *Mol. Biosyst.* 10, 613–627. doi: 10.1039/c3mb70459a

Rohwer, J. M. (2012). Kinetic modelling of plant metabolic pathways. *J. Exp. Bot.* 63, 2275–2292. doi: 10.1093/jxb/ers080

Schaber, J., Liebermeister, W., and Klipp, E. (2009). Nested uncertainties in biochemical models. *IET Syst. Biol.* 3, 1–9. doi: 10.1049/iet-syb:20070042

Schelter, B. (2006). *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. Weinheim: WILEY-VCH.

Steinway, S. N., Biggs, M. B., Loughran, T. P. Jr., Papin, J. A., and Albert, R. (2015). Inference of network dynamics and metabolic interactions in the gut microbiome. *PLoS Comput. Biol.* 11:e1004338. doi: 10.1371/journal.pcbi.1004338

Steuer, R., Gross, T., Selbig, J., and Blasius, B. (2006). Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11868–11873. doi: 10.1073/pnas.0600013103

Steuer, R., Kurths, J., Fiehn, O., and Weckwerth, W. (2003). Observing and interpreting correlations in metabolomic networks. *Bioinformatics* 19, 1019–1026. doi: 10.1093/bioinformatics/btg120

Sun, X. L., and Weckwerth, W. (2012). COVAIN: a toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. *Metabolomics* 8, S81–S93. doi: 10.1007/s11306-012-0399-3

Thiele, I., Swainston, N., Fleming, R. M., Hoppe, A., Sahoo, S., Aurich, M. K., et al. (2013). A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* 31, 419–425. doi: 10.1038/nbt.2488

Weckwerth, W. (2011). Unpredictability of metabolism—the key role of metabolomics science in combination with next-generation genome sequencing. *Anal. Bioanal. Chem.* 400, 1967–1978. doi: 10.1007/s00216-011-4948-9

Willemsen, A. M., Hendrickx, D. M., Hoefsloot, H. C., Hendriks, M. M., Wahl, S. A., Teusink, B., et al. (2015). MetDFBA: incorporating time-resolved metabolomics measurements into dynamic flux balance analysis. *Mol. Biosyst.* 11, 137–145. doi: 10.1039/C4MB00510D

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

- **Nagler M**[*], Nägele T[*], Gilli C, Fragner L, Korte A, Platzer A, Farlow A, Nordborg M, Weckwerth W: Manuscript in preparation, running title: Metabolic signatures in ecology: making the leap from model systems in the lab to native populations in the field

  * contributed equally

  We investigated the applicability of GC- and LC-MS based primary and secondary metabolite profiling for the characterization of *in situ* Arabidopsis thaliana populations. We contributed seeds of the scrutinized populations to the 1001 genomes project. Almost all analyzed substance classes showed significant variation. Prominent differences were found in flavonoid content, polyamine metabolism and TCA cycle intermediates, most of all fumarate. Although the environmental regulation remains elusive, we provide evidence for the suitability of metabolomics in analyzing genotype-environment interactions *in situ*.

  To this manuscript, I identified and collected the populations, performed extractions, measurements and statistical analysis of results and wrote the manuscript.

Metabolic signatures in ecology: making the leap from model systems in the lab to native populations in the field

Matthias NAGLER[§,1], Thomas NÄGELE[§,1], Christian GILLI[1], Lena FRAGNER[1], Arthur KORTE[2], Alexander PLATZER[2], Ashley FARLOW[2], Magnus NORDBORG[2] and Wolfram WECKWERTH[*,1]

[§] **Authors contributed equally**

Author(s) Affiliation:

[1]Department of Ecogenomics and Systems Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

[2]Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Dr.Bohr Gasse 3, 1030 Vienna, Austria

Running title: In situ metabolic phenotypes of Arabidopsis

***Corresponding author:**

Wolfram Weckwerth

Department of Ecogenomics and Systems Biology

University of Vienna

Althanstrasse 14

1090 Vienna, Austria

E-mail address: wolfram.weckwerth@univie.ac.at

T.: +43-1-4277-76550

Abstract (max. 250)

Experimental high-throughput analysis of molecular networks is a central approach to characterize the adaptation of plant metabolism to the environment. However, recent studies have demonstrated that it is hardly possible to predict *in situ* metabolic phenotypes from experiments under controlled conditions, such as growth chambers or greenhouses. This is particularly due to the high molecular variance of *in situ* samples induced by environmental fluctuations. An approach of functional metabolome interpretation of in field samples would be desirable in order to be able to identify and trace back the impact of environmental changes on plant metabolism. To test the applicability of metabolomics studies for a characterization of plant populations in the field, we have identified and analyzed *in situ* samples of nearby grown natural populations of *Arabidopsis thaliana* in Austria. *Arabidopsis thaliana* is the primary molecular biological model system in plant biology with one of the best functionally annotated genomes representing a reference system for all other plant genome projects. The genomes of these natural populations were sequenced exemplarily and phylogenetically compared to a comprehensive genome database. Experimental results on the primary and secondary metabolomic constitution were functionally integrated by a data mining strategy which combines statistical output of metabolomics data with genome-derived biochemical pathway information. This approach indicated different strategies of metabolic regulation on a population level which enabled the direct comparison, differentiation and prediction of metabolic phenotypes. Finally, hyphenated LC/MS-driven analysis of secondary metabolism allowed for the direct classification of *Arabidopsis* populations within geographically contiguous sampling sites.

Keywords: metabolic phenotype, *in situ* analysis, metabolomics, *Arabidopsis thaliana*, metabolic modeling, natural variation, Jacobian matrix, Green Systems Biology

Introduction

Natural variation, as first described by Darwin (Darwin 1859), is the ultimate point of attack for natural selection and still the only known process that is able to produce adaptive evolutionary change. *Arabidopsis thaliana* has become a powerful model organism for studying many aspects of plant biology (Somerville & Koornneef 2002). After the publication of a first complete reference genome sequence (AGI, 2000), it was discovered that it is inappropriate to think about 'the' genome of a species (Weigel & Mott 2009). In fact, all species are exposed to specific environmental clines differently affecting individual plants' phenotypic performance (Ellenberg 1953; Hoffmann 2002; Lasky *et al.* 2012; Turesson 1922; Weigel 2012). Therefore, they comprise different populations colonizing different habitats. These habitats may impose differing directions of natural selection upon the coenospecies, and thus - together with genetic drift - lead to diverging allele frequencies and to an inhomogeneous genetic structure. This inhomogeneity is called natural genetic variation and potentially provides insights in genome evolution, population structure and selective mechanisms (Mitchell-Olds & Schmitt 2006). However, the genetic side represents only one level in the complex molecular architecture which builds up the basis for physiological and morphological responses of plants to environmental stimuli (Pigliucci 2010). The experimental analysis and interpretation of these molecular architectures is non-intuitive, particularly because of the highly complex organization of plant molecular networks. Numerous studies have shown that a multitude of genes, proteins, metabolites and underlying regulatory processes are involved in plant-environment interactions (Chan *et al.* 2010; Keurentjes 2009; Koornneef *et al.* 2004; Lasky *et al.* 2012; Macel *et al.* 2010; Wienkoop *et al.* 2008). However, interpreting these findings in the context of environmental conditions and, particularly, in an ecological context is highly challenging. This is particularly due to a missing stringent definition of the genotype-phenotype relationship which can hardly be expected to be derivable from a single methodology but rather from a comprehensive platform of experimental and theoretical strategies (Diz *et al.* 2012; Weckwerth 2011). Recording environmentally induced fluctuations in a metabolic homeostasis has been shown to be a promising approach to unravel complex patterns of metabolic regulation and adaptation. For example, the metabolism of floral anthocyanins, which is a central group of secondary metabolites, was found to represent a suitable metabolic system to characterize the process of environmental regulation (Lu *et al.* 2009). The authors suggested that environmental regulation of the anthocyanin pathway is mainly affected by daily average temperature and UV light intensity modulating anthocyanin transcript levels at floral developmental stages. In another study, a metabolomics approach has been applied to elucidate *in situ* allelopathic relationships of individual species to phytosociological gradients (Scherling *et al.* 2010). The authors could show that metabolic signatures of five different plant species correlate with a biodiversity gradient. In a more general, metabolomics approaches can be expected to provide a detailed information about metabolic

processes in context of genomic signatures (Chae *et al.* 2014). Particularly in model systems with functionally annotated genomes this makes it the method of choice to unravel and interpret molecular ecological properties.

For the genetic and molecular biological model plant *Arabidopsis thaliana,* one of the best functionally annotated genomes (Baerenfaller *et al.* 2012; Lavagi *et al.* 2012) and a comprehensive catalogue of genetic variation is available (The 1001 Genomes Consortium 2016). Recently, an *in vitro* study of the physiological homeostasis of 92 *Arabidopsis thaliana* accessions in multiple growth settings has demonstrated the devastating impact of varying environmental conditions on the correlation of *in vitro* metabolism to geographic origin (Kleessen *et al.* 2012). Yet, as microhabitats may vary significantly on relatively small spatial scales and are not necessarily corresponding to geographic distance, the investigation of the molecular performance of plants *in situ* seems inevitable to get a realistic picture of plant-environment interactions and their ecophysiological consequences. A well-known example indicating the need of such *in situ* studies is Ellenberg's Hohenheimer groundwater table experiment (Ellenberg 1953; Hector *et al.* 2012). Here, it was shown that the phenotypic performance of plants *in vitro* significantly differ from their *in situ* physiological homeostasis, as important microhabitat parameters may not be included in the *in vitro* growth setting (Shulaev *et al.* 2008). Both plant communities and plant populations seem to be an appropriate target for the development and tuning of *in situ* methodologies due to their sessile nature and the availability of a large set of *in vitro* reference data for some species. This enables the intersection of individual molecular with environmental data, and even ecosystem properties can be accounted for via geographic information systems. Genotyping approaches in *Arabidopsis thaliana* have already been established (Atwell *et al.* 2010; Hancock *et al.* 2011; Horton *et al.* 2012; Long *et al.* 2013; Platt *et al.* 2010; Todesco *et al.* 2010) and are easily transferable to *in situ* samples (Hunter *et al.* 2013). Metabolomics and proteomics technologies provide the means for generating upstream molecular phenotypes (Doerfler *et al.* 2013; Hoehenwarter *et al.* 2008; Wienkoop *et al.* 2010). Thus, these techniques are suitable for experimental high-throughput analysis at the molecular level, representing the basis for strategies of multivariate statistics and mathematical modeling to identify biochemical perturbation sites and gain predictive power (Nägele 2014; Nägele & Weckwerth 2013). In this context, particularly metabolomic analysis has proven to be a suitable approach for the comprehensive and representative investigation of complex metabolic networks with respect to the underlying phenotypic diversity (Keurentjes 2009; Scherling *et al.* 2010).

In the present study, the genomes and metabolomes of *in situ* samples from three Austrian natural populations of *Arabidopsis thaliana* were experimentally characterized. Applying a combination of multivariate statistics and mathematical modeling based on genome-derived

biochemical pathway information, metabolomic signatures of *in situ Arabidopsis* populations could be identified. Different metabolic steady states on a population level and general patterns common to all populations were distinguished by this novel metabolic modeling which finally allowed the prediction of the secondary metabolite constitution, and, with this, characteristic processes of environmental regulation.

Materials & Methods

Plant material and sampling strategy

In situ sampling of *Arabidopsis thaliana* leaf rosettes in three Austrian locations (see Figure 1) was performed during one day in a rotatory way, thus accounting for diurnal changes in the physiological performance of sampling plants. The first location (OOE1) was a hay meadow, the second (OOE2) was a rocky spot with variable substrate thickness, and the third sampling site (OOE3) was an unused meadow with steep slope and a nearby well. All populations were located in close proximity to intensively used grassland. Each sample consisted of one whole leaf rosette. Samples for LC-MS analysis consisted of mature leaves of individual leaf rosettes. Global Positioning System (GPS) coordinates of the sampling sites were recorded using a Garmin Oregon300 handheld GPS receiver (Garmin®, Schaffhausen, Switzerland) with an accuracy of approximately 3 meters. Rosettes were cut and immediately frozen in liquid nitrogen. Samples were stored at -80°C until further processing. The waypoints were imported into Garmin Mapsource Version 6.15.6 (Garmin®, Schaffhausen, Switzerland) and projected on the OpenStreetMap (http://www.freizeitkarte-osm.de/de/oesterreich.html).

DNA Sequencing and SNP calling

Sequencing was performed on 1 plant from the OOE1, 3 plants from the OOE2 and 2 plants from the OOE3 population. Genomic DNA preparation, and SNP calling was performed as described previously in Long *et al.* (Long *et al.* 2013). The samples were sequenced using 100bp paired-end reads on an Illumina HiSeq platform. Pairwise genetic differences (θ_p) between these 6 accessions and a set of additional 24 accessions for which DNA sequence is publically available (The 1001 Genomes Consortium 2016) has been calculated by dividing the number of polymorphic sites by the number of informative sites. These values have been used to create a hierarchical clustering using the McQuitty method within the function *hclust* in R.

Gas chromatography coupled to time-of-flight mass spectrometry

Frozen sample rosettes were homogenized in a ball mill (Retsch®, Haan, Germany) under frequent cooling with liquid nitrogen for 3 minutes. Polar metabolites were extracted and derivatized as

described previously (Doerfler *et al.* 2013; Weckwerth *et al.* 2004). Gas chromatography coupled to mass spectrometry (GC-MS) analysis was performed on an Agilent 6890 gas chromatograph (Agilent Technologies®, Santa Clara, CA, USA) coupled to a LECO Pegasus® 4D GCxGC-TOF mass spectrometer (LECO Corporation, St. Joseph, MI, USA). Compounds were separated on an Agilent HP5MS column (length: 30m length, diameter: 0.25mm, film: 0.25 µm). Deconvolution of the total ion chromatograms was performed using the LECO Chromatof® software. A calibration curve was recorded for absolute quantification of central primary metabolites.

LC-MS analysis of polar compounds

The frozen plant leaf material was homogenized as the samples for the GC-MS analysis. The polar metabolites were extracted as described elsewhere (Doerfler *et al.* 2013). Extracts were weighed and dissolved in 5% Acetonitrile 0.5% Formic acid to an extract concentration of 0.5g/L. From these solutions, 3µL where injected to an Agilent Ultimate 3000 LC-system and separated on a reversed-phase column on a 60min effective gradient prior to data-dependent mass spectrometric analysis of *+1* - charged ions.

GC-MS data analysis and inverse approximation of Jacobian matrix entries

ANOVA and computation of p-values adjusted for sample size by Tukey Honest Significant Differences was performed on the GC-MS data using R (R Core Team 2013). For multivariate analysis, outliers (all values that were lower/higher than 1.5*interquartile range from the 25%/75% quantile) were removed from the dataset. Missing values of variables, which were missing in more than half of all measurements in a population were filled with half of the matrix minimum. The remaining missing values were imputed by random forest computation (Gromski *et al.* 2014; Stekhoven & Buhlmann 2012). This dataset was centered and scaled to unit variance prior to sPLS regression. Sparse partial least squares (sPLS) regression analysis was performed using the mixOmics package (Gonzalez *et al.* 2012; Gonzalez *et al.* 2011; Le Cao *et al.* 2009) for the statistical software environment R (R Core Team 2013).

The functional integration of GC-MS metabolomics data into a metabolic network was performed, as previously described (Nägele *et al.* 2014), by the approximation of the biochemical Jacobian matrix. This approximation directly connects the covariance matrix C, which was built from the experimental metabolomics data, with a metabolic network structure derived from Arabidopsis genome information. Linkage of covariance data with the network structure follows equation 1 (Steuer *et al.* 2003):

$$JC + CJ^T = -2D \quad (1)$$

Here, J represents the Jacobian matrix and D is a fluctuation matrix which integrates a Gaussian noise function simulating metabolic fluctuations around a steady state condition. In context of a metabolic network, entries of the Jacobian matrix J represent the elasticity of reaction rates to any change of metabolite concentrations which are characterized by equation 2:

$$J = N\frac{\partial r}{\partial M} \quad (2)$$

N is the stoichiometric matrix or a metabolic interaction matrix if reactions and reactants have been modified in the original network. r represents the rates for each reaction and M represents metabolite concentrations. As stated before, the Jacobian approximation comprises the stochastic term D. Therefore, we performed 10 x 10^5 inverse approximations for each population, finally resulting in 10 technical replicates of the Jacobian matrices.

All calculations of Jacobian matrices were performed based on a modified version of the toolbox COVAIN (Sun & Weckwerth 2012) within the numerical software environment Matlab® (V8.4.0 R2014b).

Data mining of LC-MS ions

Acquired LC-MS runs were converted to the open mzXML data format using the MassMatrix File Conversion Tools. Subsequently, MS1 intensities of all mass traces, that were fragmented at least once in a sample were summed over the whole runs with ProtMAX2012 (Egelhofer *et al.* 2013; Hoehenwarter *et al.* 2011), thus losing the chromatographic information but accounting for chemically closely related stereoisomers that exhibit deviating retention times. The data set was filtered for features that were measured in at least half of the replicates of one population and remaining variables were normalized to the sum of all variables of the respective sample. The resulting values were used to fit ANOVA models and Tukey Honest Significant Difference were used to estimate sample-size adjusted p-values in R (R Core Team 2013). VENNY was used to visualize the number of detected significant differences (Oliveros 2007).

For multivariate analysis, outliers (all values that were lower/higher than 1.5*interquartile range from the 25%/75% quantile) were removed from the dataset. Missing values of variables, which were missing in more than half of all measurements in a population were filled with half of the matrix

minimum. The remaining missing values were imputed by random forest computation (Gromski *et al.* 2014; Stekhoven & Buhlmann 2012). This dataset was centered and scaled to unit variance prior to sPLS regression.

Results

GC-MS analysis of *in situ* samples

In situ sampling of *Arabidopsis thaliana* leaf rosettes was performed on three nearby locations in Upper Austria (Oberoesterreich; OOE; see Figure 1 and *Materials and Methods*). All *Arabidopsis* rosettes were sampled at a developmental stage in which inflorescence and mature leaf rosettes had been established (example pictures are provided in Appendix S1). Metabolomic analysis of *in situ* samples comprised absolute levels of 39 central compounds of the C/N metabolism comprising sugars and sugar alcohols, organic acids, amino acids, and polyamines (Figure 2). Results of an ANOVA indicated that only levels of fumaric acid discriminated all three populations significantly (Fig. 2 b). Populations OOE1 and OOE3 could be discriminated significantly by the concentrations of galactose, melibiose, threitol, ascorbic acid, fumaric acid, gluconic acid, malic acid, threonic acid, alanine and proline ($p<0.05$; Figure 2). For populations OOE2 and OOE3, significant differences were found to exist for absolute levels of galactinol, raffinose, threitol, myo-inositol, ascorbic acid, fumaric acid, succinic and threonic acid as well as for the amino acids alanine, glutamic acid, lysine, methionine, and ornithine ($p<0.05$; Figure 2). Populations OOE1 and OOE2 could be discriminated by levels of citric acid, fumaric acid, gluconic acid and malic acid. To summarize these findings, most significant differences between absolute metabolite levels of populations OOE1, 2 and 3 were determined for the class of organic acids (13 out of 27, i.e. ~ 50%).

Multivariate analysis indicates a discrimination of *in situ* populations by metabolic phenotypes

Sparse partial least squares (sPLS) regression analysis of primary metabolites versus a response matrix comprising geographical coordinates and altitude above sea level indicated a separation of population OOE3 from populations OOE1 and OOE2 across latent variable 1 (Figure 3). The metabolite levels of fumaric acid, melibiose, alanine, putrescine, gluconic acid, threonic acid, myo-inositol, galactinol and succinic acid were identified to contribute most to this separation with elevated levels in OOE3 whereas mainly ascorbic acid and threitol were elevated in OOE1 and OOE2. Discrimination of populations OOE1 and OOE2 was indicated on latent variable 2 (Fig. 3). Here, a higher abundance of 2-oxoglutaric acid, glutamic acid, raffinose, glycine, succinic acid, serine and threonic acid in OOE1 and malic acid, gluconic acid and citric acid in OOE2 (see Appendix S2 for a complete list of loadings, and Appendix S3 for PCA of the primary metabolome).

Approximating entries of Jacobian matrices indicates a differential interplay between primary and secondary metabolism in natural populations

While absolute metabolite levels can provide a representative view on a metabolic homeostasis, it can hardly be directly interpreted in terms of biochemical regulation. Instead, strategies of multivariate statistics and modeling were shown to be essential to provide a comprehensive view on the biochemical regulation of a metabolic homeostasis. Based on a biomathematical strategy we have developed and applied in former studies, entries of Jacobian matrices were directly inferred from experimental metabolomic covariance data (Nägele *et al.* 2014). As described in our previous work, we derived a metabolic network model according to our experimental GC-MS data comprising reactants and reactions indicated in the supplement (Appendix S4). The metabolic covariance information was linked to a genome-information derived biochemical network structure, finally satisfying a Lyapunov equation (for more details about the method and the metabolic network model we refer to the section *Material and Methods* as well as to our previous work (Nägele *et al.* 2014)). The calculation procedure, i.e. solving the equation after stochastic perturbation, was performed 10 x 10^5 times and median values of all entries of the Jacobian matrices were determined. Principal Component Analysis (PCA) of the entries revealed a clear separation of the population-specific Jacobian information in which the technical variance was found to be significantly lower than the biological variance (Fig. 4). Loadings of the PCA revealed that the strong separation of population OOE1 from OOE2 and 3 on component 1 (PC1) was predominantly due to differences in amino acid, polyamine and raffinose metabolism (see also Appendix S4,S5). OOE2 was separated from OOE1 and OOE3 on PC2 which was found to be predominantly due to Jacobian entries being related to the biosynthesis of aromatic amino acids (Appendix S4,S5). This output indicated a potential difference in the regulation of secondary metabolism, or, at least, of the interface between primary and secondary metabolism. Hence, secondary metabolite abundance of the three Austrian Arabidopsis populations was recorded applying LC-MS analysis. The quantitative analysis of specific mass traces in the chromatograms showed that there was no feature separating all of the populations significantly (ANOVA, $p<0.05$). Yet, we were able to identify 70 features that discriminated at least two of the populations (Figure 3).

To statistically evaluate the separation of populations by secondary metabolites, LC-MS data were analyzed by sPLS regression analysis. The first latent variable was found to separate OOE1 from OOE2 and OOE3 (Figure 6; Loadings are provided in Appendix S2). The second latent variable indicated a separating effect of several putative anthocyanins attached to sinapoyl moieties (A6, A7/A17, A8, A10, A11 and 1329, respectively) in the OOE2 population by which it was discriminated from OOE1 and OOE3.

Austrian *in situ* population sequencing

A SNP-based genotyping approach was performed to unravel the genomic relationship of the three populations. Genotyping showed clear differences between the three populations (Figure 7). Plants of population OOE2 were found to be nearly identical (12, 23 and 13 SNPs, respectively). The population OOE2 was found to differ by approximately 300,000 SNPs from both populations OOE1 and OOE3, which were likewise separated by more than 300,000 SNPs. Interestingly, plants that have been sequenced from the OOE3 population were genetically different as well but to a minor extent (~260,000 SNPs).The comparison with genomic data from other ecotypes show the expected genetic differences not only between these populations but also with respect to global samples, in which accessions from Austria, Italy and the Czech Republic are most similar.

Discussion

The importance and central role of metabolomics in an ecological context has extensively been outlined in previous studies and overview articles (see e.g. (Jones *et al.* 2013; Sardans *et al.* 2011)). One of the central issues of eco-metabolomic approaches is the detection and characterization of environmentally induced phenotypic mechanisms in context of key metabolic processes and ecologically relevant parameters, i.e. all kinds of environmental cues. Yet, due to the non-linear relationship between single levels of molecular organization, the reliable interpretation of metabolomics results is highly challenging. The metabolic output or homeostasis of a biochemical system depends on numerous molecular parameters and variables, and, finally, a metabolic network sums up to a highly branched, interlaced and non-linearly behaving molecular system (Nägele 2014).

While under controlled conditions such plasticity of molecular systems already significantly limits our ability to intuitively draw conclusions about regulatory mechanisms, *in situ* data interpretation has to cope even more with a potential ambiguousness introduced by environmental dynamics and fluctuations (Macel *et al.* 2010). In the present study, such fluctuations were taken into account by considering (co)variance information of metabolite pools, and by a modeling approach which focuses on the characterization of dynamical behavior of metabolic systems around a metabolic homeostasis (Nägele *et al.* 2014; Steuer *et al.* 2003). In detail, data dimensionality reduction via PCA indicated a clear separation of all populations by Jacobian entries being related to the biochemistry at interface of primary and secondary metabolism as well as the metabolism of metabolic stress-markers, such as polyamines or raffinose, which have been discussed to be involved in the protection of the

photosynthetic apparatus against various stress types (Alcázar *et al.* 2011; Alcázar *et al.* 2006; Bouchereau *et al.* 1999; Knaupp *et al.* 2011).

Predictions about the differentiation via signatures in secondary metabolism could be validated by LC-MS metabolomics focusing a central set of secondary metabolite backbones with close similarity to previously identified anthocyanins attached to sinapoyl moieties (Tohge *et al.* 2005). Such metabolic differences are in line with previous findings reporting on metabolic signatures which are due to characteristic differences in specialized or secondary metabolism (Chae *et al.* 2014; Lu *et al.* 2009; Moore *et al.* 2014; Wink 2003). The accumulation of anthocyanin pigments in vegetative tissue of plants represents an approved metabolic stress and acclimation output (Winkel-Shirley 2002). Hence, while a characteristic differentiation of the three *in situ* populations by their anthocyanin-related leaf color was hardly possible (Appendix 1), the molecular analysis provided a more detailed view on the metabolic constitution of secondary metabolites. With this, evidence is provided for the suitability of metabolic phenotyping of *in situ* samples by a combined GC- MS and LC-MS platform. While, at this point, we can only speculate on the environmental cues which initiated the observed differences in secondary stress-associated metabolism, flavonoid metabolites in general are heavily discussed in context of their UV absorption and reactive oxygen species (ROS) scavenging properties (Agati & Tattini 2010; Hectors *et al.* 2014; Winkel-Shirley 2002). Together with the finding of a differentially regulated polyamine metabolism, which became visible rather by covariance information than by mean values, our results point towards a differential macro- or microclimatic environment at the three Austrian *in situ* sampling sites. To elucidate the possible roles and contributions of environmental factors to the establishment of the observed metabolic homeostases, future studies will have to correlate these metabolic markers to micro- and macroclimatic data which are derived from long lasting environmental data logging approaches.

In the context of ecological and ecophysiological research, the approach of a metabolomic *in situ* classification promotes the detection and subsequent understanding of plant-environment interactions by revealing metabolic markers being characteristic for intraspecific variation and/or environmental adaptation. While we cannot directly relate the metabolic signatures to explicit environmental data, e.g. light intensity or humidity, compared population samples were collected during consistent meteorological conditions. In addition, results of SNP-based genotyping revealed three genetically different populations, which are, however, closer related to each other than to other European accessions (Fig. 7). In terms of temperature regimes and low temperature tolerance, which can be expected to have major influence on the geographic range of *Arabidopsis thaliana* (Hoffmann 2002), the genetic distance between the Austrian populations can be regarded as relatively small when compared to sensitive (Cvi,Co-1) and tolerant accessions (Rsch-4) (Hannah *et al.* 2006). Based on this

observation, we hypothesize that the variance in observed metabolic phenotypes are rather plasticity effects than conceptual differences in the acquisition of (abiotic) stress tolerance. This again might indicate a high intraspecific metabolic variation which would affect the evolutionary capacity of Arabidopsis in context of adaptation to macro- and micro-environmental fluctuation (Moore *et al.* 2014). Yet, the resilience and interpretability of the presented findings in such a context remains limited and indicates the need for long lasting experiments to characterize environmental fluctuations. Nevertheless, due to the fact that characteristic metabolomic signatures could be identified for the genetically closely related - but still different - populations, the presented approach can be summarized as a suitable molecular monitoring strategy. The findings contribute to the comprehensive understanding of ecological processes and may contribute to the design of future studies focusing the estimation of the impact of climate change on plant societies and evolution (Ward & Kelly 2004).

Statistics on absolute primary metabolite levels revealed major differences between natural *in situ* Austrian *Arabidopsis thaliana* populations. Almost all classes of analyzed substances, comprising sugars, carboxylic and amino acids displayed significant differences indicating different homeostases in primary metabolism of all three populations. Only levels of metabolite, the TCA intermediate fumaric acid, were found to significantly differ between all *in situ* samples indicating suitability to classify these populations. While it has been shown that fumaric acid metabolism plays a central role in diurnal carbon allocation (Pracharoenwattana *et al.* 2010), and, hence, indirectly affects the orchestration of photosynthesis in *Arabidopsis* leaves, it remains elusive if it can directly report on changes in plant-environment interactions. In addition, due to the complex regulation of plant primary metabolism it can hardly be assumed that one metabolite level provides representative information for robust metabolic *in situ* classification. Yet, together with the finding of a significant difference in potential photosystem-protective substances, e.g. polyamines and flavonoids, it can be hypothesized that differential metabolic homeostases evolved due to differences in the microenvironment of the three populations being characteristic enough to separate them according to the resulting metabolic signatures.

In summary, it was demonstrated that intraspecific metabolic phenotypes of geographically nearby-grown Arabidopsis plants can be characterized and differentiated by their primary-secondary metabolic signature. In future studies, monitoring of (micro)climatic properties will enable the characterization of sampling sites by continuous quantitative environmental data and thus improve the understanding of the ecological context of *in situ* molecular profiles. Additionally, biotic and abiotic habitat parameters, such as soil properties and phytosociological association, might even promote our current understanding of individual plants' physiology. Finally, our study points to the importance of considering variance and covariance information in biological data sets (Violle *et al.* 2012) which,

together with genome-derived pathway information, potentially provide information about environmental fluctuations and associated biochemical system properties.

Acknowledgements

The authors would like to thank the members of the Department Ecogenomics and Systems Biology for valuable discussions and advice. Additionally, the authors thank the gardeners and the whole team of the department-associated greenhouse facility for their support and advice. The study was funded by the Austrian Science Fund (FWF, Project P 26342 and I 1022).

Author contributions

MNA collected natural populations of Arabidopsis thaliana, performed measurements and statistical analysis and wrote the manuscript; TN performed measurements, statistical analysis, metabolic modeling and wrote the manuscript; CG identified and collected natural populations of *Arabidopsis thaliana*, LF harvested sample material, AK, AP, AF, MNO performed SNP calling, population analysis and wrote the manuscript; WW conceived the study, performed statistical analysis and wrote the manuscript.

References

Agati G, Tattini M (2010) Multiple functional roles of flavonoids in photoprotection. *New Phytologist* **186**, 786-793.

Alcázar R, Cuevas JC, Planas J*, et al.* (2011) Integration of polyamines in the cold acclimation response. *Plant Science* **180**, 31–38.

Alcázar R, Marco F, Cuevas JC*, et al.* (2006) Involvement of polyamines in plant response to abiotic stress. *Biotechnology Letters* **28**, 1867–1876.

Atwell S, Huang YS, Vilhjalmsson BJ*, et al.* (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**, 627-631.

Baerenfaller K, Bastow R, Beynon J*, et al.* (2012) Taking the Next Step: Building an Arabidopsis Information Portal. *Plant Cell* **24**, 2248-2256.

Bouchereau A, Aziz A, Larher F, Martin-Tanguy J (1999) Polyamines and environmental challenges: recent development. *Plant Science* **140**, 103–125.

Chae L, Kim T, Nilo-Poyanco R, Rhee SY (2014) Genomic signatures of specialized metabolism in plants. *Science* **344**, 510-513.

Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The complex genetic architecture of the metabolome. *PLoS genetics* **6**, e1001198.

Darwin C (1859) The Origin of Species. *London: John Murray*.

Diz AP, MartÍNez-FernÁNdez M, RolÁN-Alvarez E (2012) Proteomics in evolutionary ecology: linking the genotype with the phenotype. *Mol Ecol* **21**, 1060-1080.

412 Doerfler H, Lyon D, Nägele T*, et al.* (2013) Granger causality in integrated GC-MS and LC-MS
413 metabolomics data reveals the interface of primary and secondary metabolism.
414 *Metabolomics : Official journal of the Metabolomic Society* **9**, 564-574.
415 Egelhofer V, Hoehenwarter W, Lyon D, Weckwerth W, Wienkoop S (2013) Using ProtMAX to create
416 high-mass-accuracy precursor alignments from label-free quantitative mass spectrometry
417 data generated in shotgun proteomics experiments. *Nature Protocols* **8**, 595-601.
418 Ellenberg H (1953) Physiologisches und ökologisches Verhalten derselben Pflanzenarten. *Ber.*
419 *Deutsch. Bot. Ges* **65**, 351-362.
420 Gonzalez I, Cao KA, Davis MJ, Dejean S (2012) Visualising associations between paired 'omics' data
421 sets. *BioData mining* **5**, 19.
422 Gonzalez I, Lê Cao K, Déjean S (2011) MixOmics: Omics data integration project. *URL http://www.*
423 *math. univ-toulouse. fr/~ biostat/mixOmics*.
424 Gromski PS, Xu Y, Kotze HL*, et al.* (2014) Influence of missing values substitutes on multivariate
425 analysis of metabolomics data. *Metabolites* **4**, 433-452.
426 Hancock AM, Brachi B, Faure N*, et al.* (2011) Adaptation to Climate Across the Arabidopsis thaliana
427 Genome. *Science* **334**, 83-86.
428 Hannah MA, Wiese D, Freund S*, et al.* (2006) Natural genetic variation of freezing tolerance in
429 Arabidopsis. *Plant Physiology* **142**, 98–112.
430 Hector A, von Felten S, Hautier Y, Weilenmann M, Bruelheide H (2012) Effects of Dominance and
431 Diversity on Productivity along Ellenberg's Experimental Water Table Gradients. *Plos One* **7**.
432 Hectors K, Van Oevelen S, Geuns J*, et al.* (2014) Dynamic changes in plant secondary metabolites
433 during UV acclimation in Arabidopsis thaliana. *Physiologia Plantarum* **152**, 219-230.
434 Hoehenwarter W, Larhlimi A, Hummel J*, et al.* (2011) MAPA Distinguishes Genotype-Specific
435 Variability of Highly Similar Regulatory Protein Isoforms in Potato Tuber. *Journal of Proteome*
436 *Research* **10**, 2979-2991.
437 Hoehenwarter W, van Dongen JT, Wienkoop S*, et al.* (2008) A rapid approach for phenotype-
438 screening and database independent detection of cSNP/protein polymorphism using mass
439 accuracy precursor alignment. *Proteomics* **8**, 4214-4225.
440 Hoffmann MH (2002) Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). *Journal of*
441 *Biogeography* **29**, 125-134.
442 Horton MW, Hancock AM, Huang YS*, et al.* (2012) Genome-wide patterns of genetic variation in
443 worldwide Arabidopsis thaliana accessions from the RegMap panel. *Nature genetics* **44**, 212-
444 216.
445 Hunter B, Wright KM, Bomblies K (2013) Short read sequencing in studies of natural variation and
446 adaptation. *Current Opinion in Plant Biology* **16**, 85-91.
447 Jones OAH, Maguire ML, Griffin JL*, et al.* (2013) Metabolomics and its use in ecology. *Austral Ecology*
448 **38**, 713-720.
449 Keurentjes JJB (2009) Genetical metabolomics: closing in on phenotypes. *Current Opinion in Plant*
450 *Biology* **12**, 223-230.
451 Kleessen S, Antonio C, Sulpice R*, et al.* (2012) Structured patterns in geographic variability of
452 metabolic phenotypes in Arabidopsis thaliana. *Nat Commun* **3**, 1319.
453 Knaupp M, Mishra KB, Nedbal L, Heyer AG (2011) Evidence for a role of raffinose in stabilizing
454 photosystem II during freeze-thaw cycles. *Planta* **234**, 477-486.
455 Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in
456 Arabidopsis thaliana. *Annual Review of Plant Biology* **55**, 141–172.
457 Lasky JR, Des Marais DL, McKay JK*, et al.* (2012) Characterizing genomic variation of Arabidopsis
458 thaliana: the roles of geography and climate. *Molecular Ecology* **21**, 5512-5529.
459 Lavagi I, Estelle M, Weckwerth W, Beynon J, Bastow RM (2012) From bench to bountiful harvests: a
460 road map for the next decade of Arabidopsis research. *Plant Cell* **24**, 2240-2247.
461 Le Cao KA, Gonzalez I, Dejean S (2009) integrOmics: an R package to unravel relationships between
462 two omics datasets. *Bioinformatics* **25**, 2855-2856.

Long Q, Rabanal FA, Meng D*, et al.* (2013) Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. *Nature Genetics* **45**, 884-890.
Lu Y, Du JIN, Tang J*, et al.* (2009) Environmental regulation of floral anthocyanin synthesis in Ipomoea purpurea. *Mol Ecol* **18**, 3857-3871.
Macel M, Van Dam NM, Keurentjes JJB (2010) Metabolomics: the chemistry between ecology and genetics. *Molecular Ecology Resources* **10**, 583-593.
Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* **441**, 947-952.
Moore BD, Andrew RL, Külheim C, Foley WJ (2014) Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist* **201**, 733-750.
Nägele T (2014) Linking metabolomics data to underlying metabolic regulation. *Frontiers in Molecular Biosciences* **1**.
Nägele T, Mair A, Sun X*, et al.* (2014) Solving the differential biochemical Jacobian from metabolomics covariance data. *PLoS ONE* **9**, e92299.
Nägele T, Weckwerth W (2013) Eigenvalues of Jacobian matrices report on steps of metabolic reprogramming in a complex plant-environment interaction. *Applied Mathematics* **4**, 44-49.
Oliveros J (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams. *http://bioinfogp.cnb.csic.es/tools/venny/index.html.*
Pigliucci M (2010) Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 557-566.
Platt A, Horton M, Huang YS*, et al.* (2010) The scale of population structure in Arabidopsis thaliana. *PLoS genetics* **6**, e1000843.
Pracharoenwattana I, Zhou W, Keech O*, et al.* (2010) Arabidopsis has a cytosolic fumarase required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. *Plant Journal* **62**, 785-795.
Sardans J, Peñuelas J, Rivas-Ubach A (2011) Ecological metabolomics: overview of current developments and future challenges. *Chemoecology* **21**, 191-225.
Scherling C, Roscher C, Giavalisco P, Schulze ED, Weckwerth W (2010) Metabolomics unravel contrasting effects of biodiversity on the performance of individual plant species. *PLoS One* **5**, e12569.
Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. *Physiologia Plantarum* **132**, 199-208.
Somerville C, Koornneef M (2002) Timeline - A fortunate choice: the history of Arabidopsis as a model plant. *Nature Reviews Genetics* **3**, 883-889.
Stekhoven DJ, Buhlmann P (2012) MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112-118.
Steuer R, Kurths J, Fiehn O, Weckwerth W (2003) Observing and interpreting correlations in metabolomic networks. *Bioinformatics* **19**, 1019-1026.
Sun X, Weckwerth W (2012) COVAIN: a toolbox for uni-and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. *Metabolomics* **8**, 81-93.
Team RC (2013) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.
The 1001 Genomes Consortium (2016) 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*.
The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**, 796-815.
Todesco M, Balasubramanian S, Hu TT*, et al.* (2010) Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. *Nature* **465**, 632-U129.
Tohge T, Nishiyama Y, Hirai MY*, et al.* (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *The Plant journal : for cell and molecular biology* **42**, 218-235.

Turesson G (1922) The genotypical response of the plant species to the habitat. *Hereditas* **3**, 211-350.
Violle C, Enquist BJ, McGill BJ*, et al.* (2012) The return of the variance: intraspecific variability in community ecology. *Trends in Ecology & Evolution* **27**, 244-252.
Ward JK, Kelly JK (2004) Scaling up evolutionary responses to elevated CO2: lessons from Arabidopsis. *Ecology Letters* **7**, 427-440.
Weckwerth W (2011) Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *J Proteomics* **75**, 284-305.
Weckwerth W, Wenzel K, Fiehn O (2004) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* **4**, 78-83.
Weigel D (2012) Natural variation in Arabidopsis: from molecular genetics to ecological genomics. *Plant physiology* **158**, 2-22.
Weigel D, Mott R (2009) The 1001 genomes project for Arabidopsis thaliana. *Genome biology* **10**, 107.
Wienkoop S, Baginsky S, Weckwerth W (2010) Arabidopsis thaliana as a model organism for plant proteome research. *Journal of Proteomics* **73**, 2239-2248.
Wienkoop S, Morgenthal K, Wolschin F*, et al.* (2008) Integration of metabolomic and proteomic phenotypes: analysis of data covariance dissects starch and RFO metabolism from low and high temperature compensation response in Arabidopsis thaliana. *Molecular & cellular proteomics : MCP* **7**, 1725-1736.
Wink M (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3-19.
Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. *Current Opinion in Plant Biology* **5**, 218–223.

Figure Legends

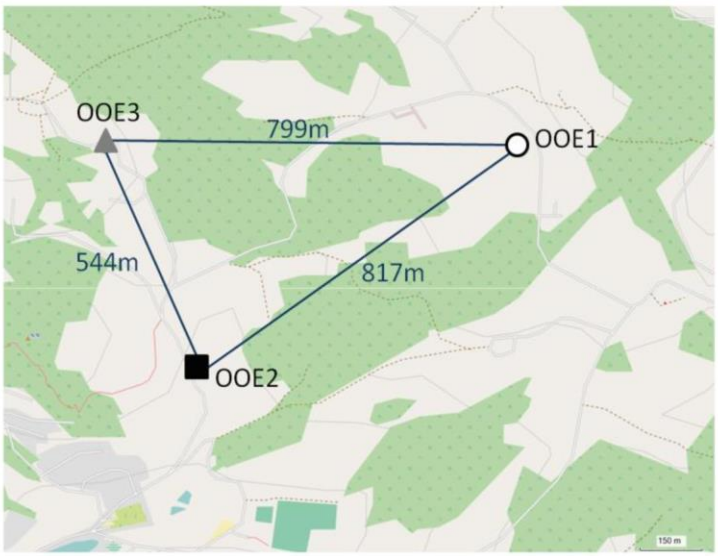


Figure 1: Projection of sample coordinates within OpenStreetMap® (http://www.openstreetmap.org). The air-line distance between populations is given in meters**.**

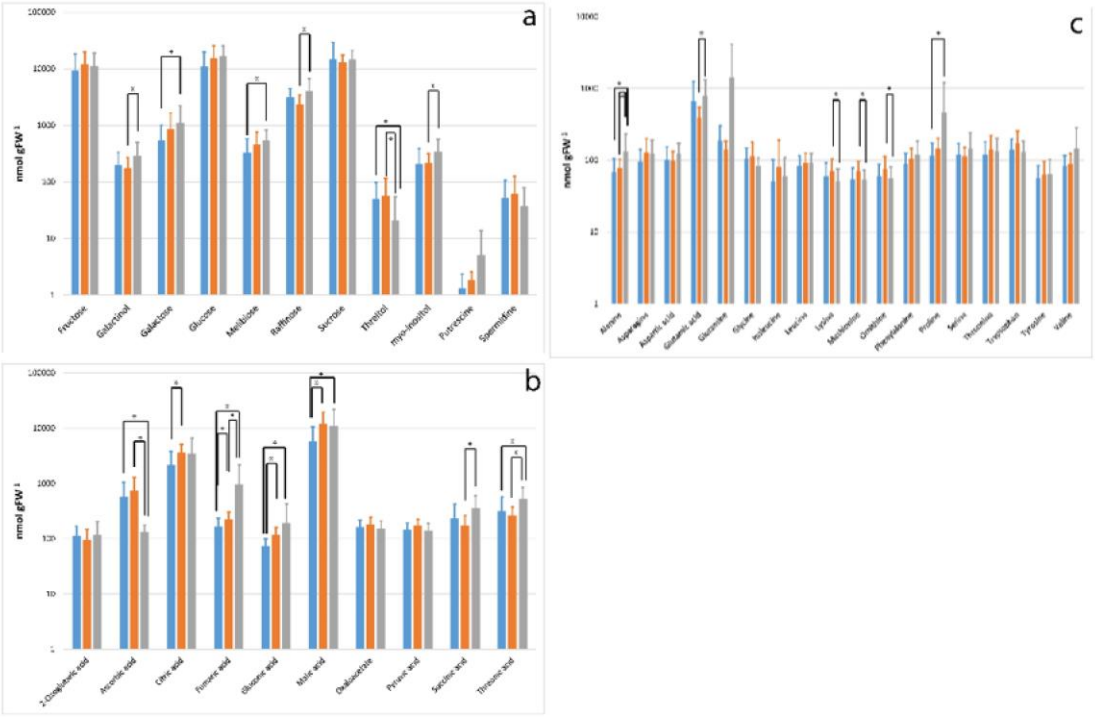
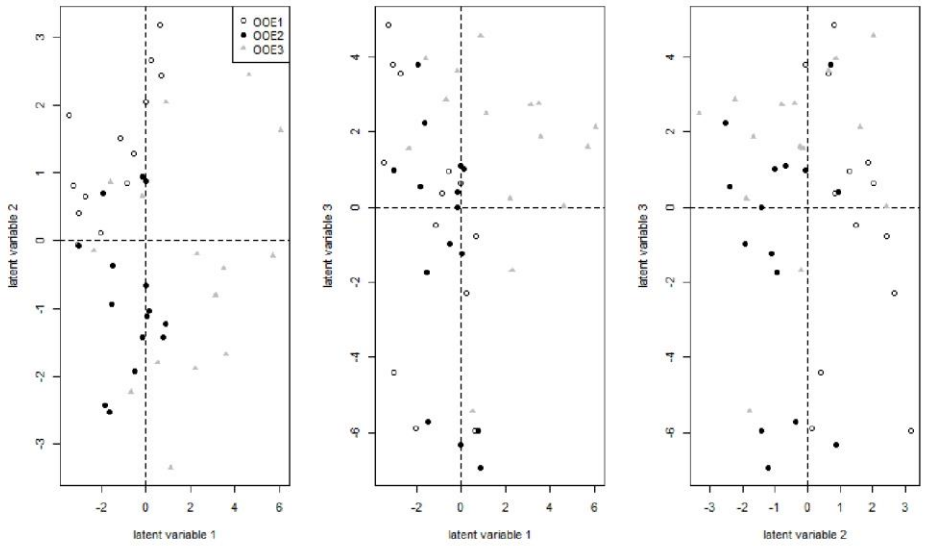
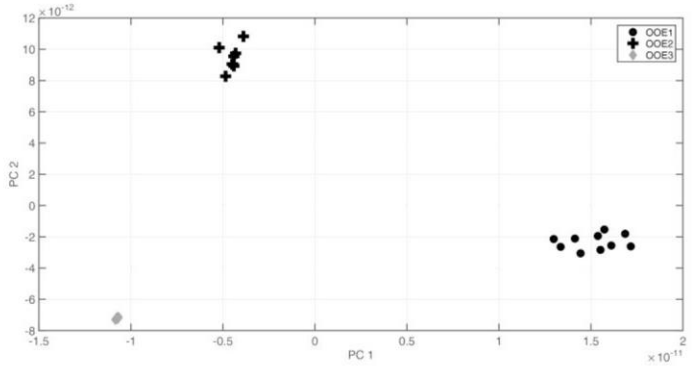


Figure 2: Absolute levels of primary metabolites. Metabolites are grouped according to the substance classes of (**a**) sugars, sugar alcohols and polyamines, (**b**) organic acids, (**c**) amino acids. Significant differences evaluated by ANOVA are indicated by asterisks (* p<0.05; ** p<0.01; *** p<0.001). Metabolite levels from samples of OOE1 are indicated by blue bars, OOE2 by orange bars, and OOE3 by grey bars.



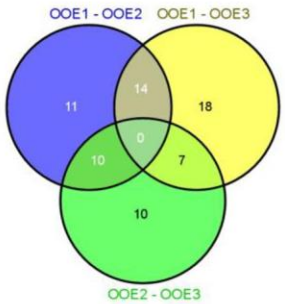
551

552 **Figure 3:** Projection of samples on latent variables of the primary metabolite matrix (GC-MS data)
553 after sPLS regression. Detailed information on the loadings are provided in the supplement
554 (Appendix S2).



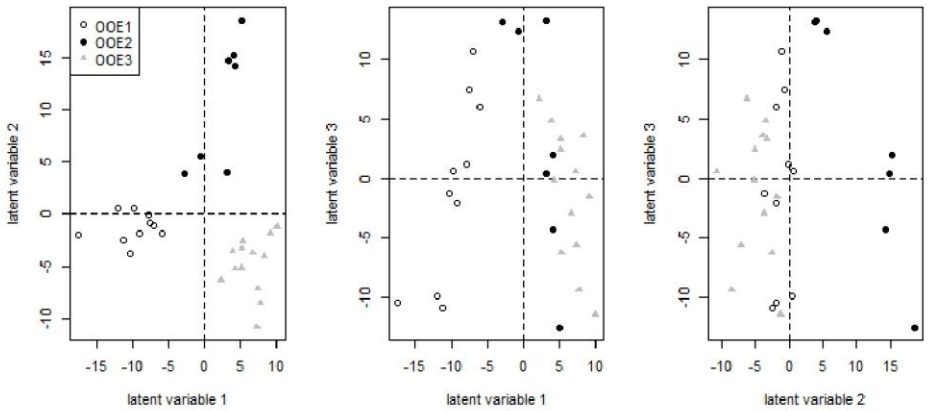
555

556 **Figure 4:** Principal component analysis (PCA) of Jacobian matrix entries for populations OOE1, OOE2
557 and OOE3. PC1 strongly separates OOE1 (black filled circles) from OOE2 (black filled crosses) and
558 OOE3 (grey filled diamonds). PC2 separates OOE2 from OOE1 and 3 most signficantly.



559

560 **Figure 5:** Venn diagram showing the number of LC-MS features which significantly discriminated the
561 three Arabidopsis populations OOE1, 2 and 3.



562

563 **Figure 6:** Projection of samples on latent variables of the secondary metabolite matric (LC-MS data)
564 after sPLS regression. Detailed information on the loadings are provided in the supplement
565 (Appendix S2).

566

567 **Figure 7:** SNP genotyping of three Austrian Arabidopsis populations. All three plants that have been
568 sequenced from the population OOE2 are nearly genetically identical. OOE2 differs by nearly 300,000
569 SNPs from both the OOE1 and OOE3 population, which are likewise separated by more than 300,000
570 SNPs. The Comparison with genomic data from other ecotypes show the expected genetic
571 differences not only within these populations but also to global samples, in which accessions from
572 Austria, Italy and the Czech Republic are most related. The genome information of the additional 24
573 accessions is publically available (The 1001 Genomes Consortium 2016)

Conclusive Comments and Future Perspectives

In the last decade, molecular phenotyping by measuring metabolite and protein profiles has been established as primary approaches in functional genomics. It promotes the understanding of the biochemical implications of genetic variants. As metabolites and proteins are central parts of any biochemical systems, knowledge of these molecular variables is crucial for understanding and subsequently predicting systemic behavior in context of key cellular processes. These reactions cannot be forecasted by only considering the genomic sequence of a particular organism because of the phenotypically plastic characteristics of morphologic and molecular phenotypes. In my thesis, I have been able to identify core components of *Arabidopsis thaliana*'s metabolism in context of acclimation and adaptive potential to cold environments that underlie phenotypic plasticity with regard to cold-tolerant and cold-sensitive accessions. I have produced evidence that the mobilization of starch connected to the accumulation of sucrose is of particular interest. Concerning agronomical importance, it will be interesting to determine the impact of my results in crop plants.

With the manuscript in preparation, I have been able to suggest the suitability of metabolic profiling to characterize geographically contiguous *Arabidopsis* populations. Modeling of metabolic regulations is even more powerful when it comes to defining potentially important adaptive markers because it does not only consider mean values but finds regulatory differences based on metabolic covariance information. For instance, if there are no significant differences in a metabolite pool, because of the varying covariance of that pool with other metabolites, we can estimate regulatory sites in the biochemical system that can be efficiently addressed in follow up experiments, for instance on the proteomic level or with enzyme activities. Novel genome editing techniques enable the targeted creation of specific knock-out mutants, even if the alleles of interest are in linkage disequilibrium. This approach will emerge with increasing importance in the quest for finding adaptively important genomic variants.

As the interpretation of metabolite data is often not straightforward due to the complexity of metabolic systems, tools that facilitate understanding the interrelations of molecular variables are needed. With the established graph theoretical approach to analyze metabolic time series data in context of a specific biochemical network, I have contributed to the endeavor of understanding the reaction of metabolism to user-defined stimuli. As the topology of metabolic systems can be predicted from static genome information, it is readily transferable also to studies in organisms other than *Arabidopsis thaliana*.

Future studies should address the question how genotypic diversity is related to biogeographical clines and phenotypic plasticity (Pigliucci, 2009). For this purpose, *Arabidopsis thaliana* is an appropriate

model system representing variable phenotypes and a complex biogeographical gradient (Hoffmann, 2005). *In situ* habitats should be described by time-continuous recording of environmental parameters to allow for the identification of the most prominent environmental differences between populations and subsequently targeted monitoring of plant performance with regard to those parameters.

Further, the concept should be extended to elucidate for instance life-form specific differences in observed molecular adaptive patterns or divergence caused by ecological strategies. *Arabidopsis thaliana* is an ephemeral ruderal plant so it is not straightforward to draw conclusions for K-strategists like phanerophytes or perennial plants. Genetic analysis of phylogenetically close relatives has already started (Koenig and Weigel, 2015) and will shed light on the variation of reaction types of different life-forms classified in the Raunkiær system. A specifically powerful synergy will be the combination of molecular profiling and GIS enabling the concatenation of biochemical properties with all kinds of ecological information on all possible spatial scales like phytosociological associations, electronic soil classifications, hemeroby maps and other landscape ecological datasets. Further, I suggest that characterizing molecular ecophenes by metabolome and proteome profiling and functionally integrating these profiles with genotypic and continuous environmental data by mathematical modeling will significantly advance our understanding of phenotypic plasticity, adaptive processes as well as their regulatory dynamics. Common garden and transplantation experiments will increase the comparability of acquired *in situ* data to also identify genotypically variable features encoded in *in situ* data. Finally, this will promote the molecular characterization of ecophysiological adaptation which is central to the improvement of plant performance in changing environments by marker assisted plant breeding.

Other Publications

During my thesis, I had multiple opportunities to contribute my practical and theoretical proficiencies in a diverse panel of scientific endeavors.

- Kawakatsu, T., Huang, S. S., Jupe, F., Sasaki, E., Schmitz, R. J., Urich, M. A., Castanon, R., Nery, J. R., Barragan, C., He, Y., Chen, H., Dubin, M., Lee, C. R., Wang, C., Bemm, F., Becker, C., O'Neil, R., O'Malley, R. C., Quarless, D. X., **1001 Genomes Consortium**, Schork, N. J., Weigel, D., Nordborg, M., and Ecker, J. R. (2016) Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. *Cell* 166, 492-505

- **1001 Genomes Consortium** (2016) 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*

- Ghatak, A., Chaturvedi, P., **Nagler, M.**, Roustan, V., Lyon, D., Bachmann, G., Postl, W., Schrofl, A., Desai, N., Varshney, R. K., and Weckwerth, W. (2016) Comprehensive tissue-specific proteome analysis of drought stress responses in Pennisetum glaucum (L.) R. Br. (Pearl millet). *J Proteomics* 143, 122-135

- Pascual, J., Alegre, S., **Nagler, M.**, Escandon, M., Annacondia, M. L., Weckwerth, W., Valledor, L., and Canal, M. J. (2016) The variations in the nuclear proteome reveal new transcription factors and mechanisms involved in UV stress response in Pinus radiata. *J Proteomics* 143, 390-400

- Alegre, S., Pascual, J., **Nagler, M.**, Weckwerth, W., Canal, M. J., and Valledor, L. (2016) Dataset of UV induced changes in nuclear proteome obtained by GeLC-Orbitrap/MS in Pinus radiata needles. *Data Brief* 7, 1477-1482

- Kerou, M., Offre, P., Valledor, L., Abby, S. S., Melcher, M., **Nagler, M.**, Weckwerth, W., and Schleper, C. (2016) Proteomics and comparative genomics of Nitrososphaera viennensis reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci U S A*

- Hanak, A. M., **Nagler, M.**, Weinmaier, T., Sun, X., Fragner, L., Schwab, C., Rattei, T., Ulrich, K., Ewald, D., Engel, M., Schloter, M., Bittner, R., Schleper, C., and Weckwerth, W. (2014) Draft Genome Sequence of the Growth-Promoting Endophyte Paenibacillus sp. P22, Isolated from Populus. *Genome Announc* 2

- Ischebeck, T., Valledor, L., Lyon, D., Gingl, S., **Nagler, M.**, Meijon, M., Egelhofer, V., and Weckwerth, W. (2014) Comprehensive Cell-specific Protein Analysis in Early and Late Pollen Development from Diploid Microsporocytes to Pollen Tube Growth. *Mol Cell Proteomics* 13, 295-310

- Kitner, M., Majesky, L., Gillova, L., Vymyslicky, T., and **Nagler, M**. (2012) Genetic structure of Artemisia pancicii populations inferred from AFLP and cpDNA data. *Preslia* 84, 97-120

- **Nagler, M.**, Grünweis, F. M., and Frank, B. (2012) Artemisia pancicii in Austria - Habitats, Community Association, Nature Conservation. *Verh. Zool.-Bot. Ges. Österreich* 148/149, 1-21

References

Agarwal M, Hao Y, Kapoor A, Dong CH, Fujii H, Zheng X, Zhu JK (2006) A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. J Biol Chem **281:** 37636-37645

Agati G, Tattini M (2010) Multiple functional roles of flavonoids in photoprotection. New Phytol **186:** 786-793

Agren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. New Phytologist **194:** 1112-1122

Alberch P (1991) From Genes to Phenotype - Dynamic-Systems and Evolvability. Genetica **84:** 5-11

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M (2009) What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? Plant Cell **21:** 1877-1896

Alonso-Blanco C, Koornneef M (2000) Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. Trends in Plant Science **5:** 22-29

Amme S, Matros A, Schlesier B, Mock HP (2006) Proteome analysis of cold stress response in Arabidopsis thaliana using DIGE-technology. J Exp Bot **57:** 1537-1546

Aranzana MJ, Kim S, Zhao KY, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang CL, Toomajian C, Traw B, Zheng HG, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. Plos Genetics **1:** 531-539

Asada K (1999) THE WATER-WATER CYCLE IN CHLOROPLASTS: Scavenging of Active Oxygens and Dissipation of Excess Photons. Annu Rev Plant Physiol Plant Mol Biol **50:** 601-639

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng DZ, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang CL, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature **465:** 627-631

Banta JA, Dole J, Cruzan MB, Pigliucci M (2007) Evidence of local adaptation to coarse-grained environmental variation in Arabidopsis thaliana. Evolution **61:** 2419-2432

Barros JAS, Cavalcanti JHF, Medeiros DB, Nunes-Nesi A, Avin-Wittenberg T, Fernie AR, Araujo WL (2017) Autophagy Deficiency Compromises Alternative Pathways of Respiration following Energy Deprivation in Arabidopsis thaliana. Plant Physiol **175:** 62-76

Becker C, Hagmann J, Muller J, Koenig D, Stegle O, Borgwardt K, Weigel D (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature **480:** 245-U127

Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. Nature Reviews Genetics **11:** 867-879

Bergelson J, Stahl E, Dudek S, Kreitman M (1998) Genetic variation within and among populations of Arabidopsis thaliana. Genetics **148:** 1311-1323

Bieniawska Z, Espinoza C, Schlereth A, Sulpice R, Hincha DK, Hannah MA (2008) Disruption of the Arabidopsis circadian clock is responsible for extensive variation in the cold-responsive transcriptome. Plant Physiology **147:** 263-279

Blows MW, McGuigan K (2015) The distribution of genetic variance across phenotypic space and the response to selection. Mol Ecol **24:** 2056-2072

Bossdorf O, Arcuri D, Richards CL, Pigliucci M (2010) Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in Arabidopsis thaliana. Evolutionary Ecology **24:** 541-553

Boyle EA, Li YI, Pritchard JK (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell **169:** 1177-1186

Braun-Blanquet J (1964) Pflanzensoziologie. Plant sociology. Springer-Verlag, New York

Buchanan BB, Schurmann P, Wolosiuk RA, Jacquot JP (2002) The ferredoxin/thioredoxin system: from discovery to molecular structures and beyond. Photosynth Res **73:** 215-222
Buckler E, Gore M (2007) An Arabidopsis haplotype map takes root. Nature Genetics **39:** 1056-1057
Burow M, Halkier BA, Kliebenstein DJ (2010) Regulatory networks of glucosinolates shape Arabidopsis thaliana fitness. Current Opinion in Plant Biology **13:** 348-353
Callahan HS, Pigliucci M (2002) Shade-induced plasticity and its ecological significance in wild populations of Arabidopsis thaliana. Ecology **83:** 1965-1980
Chan EKF, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The Complex Genetic Architecture of the Metabolome. Plos Genetics **6**
Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev **17:** 1043-1054
Chinnusamy V, Zhu JK, Sunkar R (2010) Gene regulation during cold stress acclimation in plants. Methods Mol Biol **639:** 39-55
Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen HM, Frazer KA, Huson DH, Schoelkopf B, Nordborg M, Raetsch G, Ecker JR, Weigel D (2007) Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science **317:** 338-342
Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **101:** 15243-15248
Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K (2011) Effects of abiotic stress on plants: a systems biology perspective. Bmc Plant Biology **11**
Crick F (1970) Central dogma of molecular biology. Nature **227:** 561-563
Dawkins R (2004) Extended phenotype - But not too extended. A reply to Laland, Turner and Jjablonka. Biology & Philosophy **19:** 377-396
Doerfler H, Lyon D, Nägele T, Sun XL, Fragner L, Hadacek F, Egelhofer V, Weckwerth W (2013) Granger causality in integrated GC-MS and LC-MS metabolomics data reveals the interface of primary and secondary metabolism. Metabolomics **9:** 564-574
Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science **346:** 1258096
Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Picó FX, Alonso-Blanco C, Hancock AM (2017) African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. Proceedings of the National Academy of Sciences**:** 201616736
Ellenberg H (1953) Physiologisches und ökologisches Verhalten derselben Pflanzenarten. Ber. Deutsch. Bot. Ges **65:** 351-362
ElSayed AI, Rafudeen MS, Golldack D (2014) Physiological aspects of raffinose family oligosaccharides in plants: protection against abiotic stress. Plant Biol (Stuttg) **16:** 1-8
Espinoza C, Degenkolbe T, Caldana C, Zuther E, Leisse A, Willmitzer L, Hincha DK, Hannah MA (2010) Interaction with Diurnal and Circadian Regulation Results in Dynamic Metabolic and Transcriptional Changes during Cold Acclimation in Arabidopsis. Plos One **5**
Fanucchi F, Alpi E, Olivieri S, Cannistraci CV, Bachi A, Alpi A, Alessio M (2012) Acclimation increases freezing stress response of Arabidopsis thaliana at proteome level. Biochim Biophys Acta **1824:** 813-825
Fiehn O (2002) Metabolomics - the link between genotypes and phenotypes. Plant Molecular Biology **48:** 155-171
Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A Map of Local Adaptation in Arabidopsis thaliana. Science **334:** 86-89
Fowler S, Thomashow MF (2002) Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell **14:** 1675-1690

Fowler SG, Cook D, Thomashow MF (2005) Low temperature induction of Arabidopsis CBF1, 2, and 3 is gated by the circadian clock. Plant Physiol **137:** 961-968

Foyer CH, Noctor G (2011) Ascorbate and glutathione: the heart of the redox hub. Plant Physiol **155:** 2-18

Fusco G, Minelli A (2010) Phenotypic plasticity in development and evolution: facts and concepts. Philosophical Transactions of the Royal Society B-Biological Sciences **365:** 547-556

Galili G, Amir R, Fernie AR (2016) The Regulation of Essential Amino Acid Synthesis and Accumulation in Plants. Annu Rev Plant Biol **67:** 153-178

Gehan MA, Park S, Gilmour SJ, An CF, Lee CM, Thomashow MF (2015) Natural variation in the C-repeat binding factor cold response pathway correlates with local adaptation of Arabidopsis ecotypes. Plant Journal **84:** 682-693

Geigenberger P, Kolbe A, Tiessen A (2005) Redox regulation of carbon storage and partitioning in response to light and sugars. J Exp Bot **56:** 1469-1479

Gilbert SF (1991) Epigenetic Landscaping - Waddington Use of Cell Fate Bifurcation Diagrams. Biology & Philosophy **6:** 135-154

Gillespie JH, Turelli M (1989) Genotype-Environment Interactions and the Maintenance of Polygenic Variation. Genetics **121:** 129-138

Gilmour SJ, Fowler SG, Thomashow MF (2004) Arabidopsis transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. Plant Mol Biol **54:** 767-781

Gilmour SJ, Sebolt AM, Salazar MP, Everard JD, Thomashow MF (2000) Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant Physiol **124:** 1854-1865

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. Plant J **16:** 433-442

Gottfried M, Pauli H, Futschik A, Akhalkatsi M, Barancok P, Alonso JLB, Coldea G, Dick J, Erschbamer B, Calzado MRF, Kazakis G, Krajci J, Larsson P, Mallaun M, Michelsen O, Moiseev D, Moiseev P, Molau U, Merzouki A, Nagy L, Nakhutsrishvili G, Pedersen B, Pelino G, Puscas M, Rossi G, Stanisci A, Theurillat JP, Tomaselli M, Villar L, Vittoz P, Vogiatzakis I, Grabherr G (2012) Continent-wide response of mountain vegetation to climate change. Nature Climate Change **2:** 111-115

Granger CWJ (1969) Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. Econometrica **37:** 424-438

Greenham K, McClung CR (2015) Integrating circadian dynamics with physiological processes in plants. Nat Rev Genet **16:** 598-610

Hacham Y, Matityahu I, Amir R (2013) Light and sucrose up-regulate the expression level of Arabidopsis cystathionine gamma-synthase, the key enzyme of methionine biosynthesis pathway. Amino Acids **45:** 1179-1190

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J (2011) Adaptation to Climate Across the Arabidopsis thaliana Genome. Science **334:** 83-86

Hannah MA, Heyer AG, Hincha DK (2005) A global survey of gene regulation during cold acclimation in Arabidopsis thaliana. Plos Genetics **1:** 179-196

Hannah MA, Wiese D, Freund S, Fiehn O, Heyer AG, Hincha DK (2006) Natural genetic variation of freezing tolerance in arabidopsis. Plant Physiology **142:** 98-112

Hectors K, Van Oevelen S, Geuns J, Guisez Y, Jansen MA, Prinsen E (2014) Dynamic changes in plant secondary metabolites during UV acclimation in Arabidopsis thaliana. Physiol Plant **152:** 219-230

Hildebrandt TM, Nunes Nesi A, Araujo WL, Braun HP (2015) Amino Acid Catabolism in Plants. Mol Plant **8:** 1563-1579

Hill M (1979) TWINSPAN - A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Cornell University: New York.

Hincha DK, Zuther E, Heyer AG (2003) The preservation of liposomes by raffinose family oligosaccharides during drying is mediated by effects on fusion and lipid phase transitions. Biochim Biophys Acta **1612:** 172-177

Hoehenwarter W, Larhlimi A, Hummel J, Egelhofer V, Selbig J, van Dongen JT, Wienkoop S, Weckwerth W (2011) MAPA Distinguishes Genotype-Specific Variability of Highly Similar Regulatory Protein Isoforms in Potato Tuber. Journal of Proteome Research **10:** 2979-2991

Hoermiller, II, Naegele T, Augustin H, Stutz S, Weckwerth W, Heyer AG (2016) Subcellular reprogramming of metabolism during cold acclimation in Arabidopsis thaliana. Plant Cell Environ

Hoffmann MH (2002) Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). Journal of Biogeography **29:** 125-134

Hoffmann MH (2005) Evolution of the realized climatic niche in the genus Arabidopsis (Brassicaceae). Evolution **59:** 1425-1436

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjalmsson BJ, Nordborg M, Borevitz JO, Bergelson J (2012) Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nature Genetics **44:** 212-216

Houshyani B, Kabouw P, Muth D, de Vos RCH, Bino RJ, Bouwmeester HJ (2012) Characterization of the natural variation in Arabidopsis thaliana metabolome by the analysis of metabolic distance. Metabolomics **8:** S131-S145

Huang T, Jander G (2017) Abscisic acid-regulated protein degradation causes osmotic stress-induced accumulation of branched-chain amino acids in Arabidopsis thaliana. Planta

Huner NPA, Oquist G, Sarhan F (1998) Energy balance and acclimation to light and cold. Trends in Plant Science **3:** 224-230

Hurry V (2017) Metabolic reprogramming in response to cold stress is like real estate, it's all about location. Plant Cell Environ **40:** 599-601

Janmohammadi M, Zolla L, Rinalducci S (2015) Low temperature tolerance in plants: Changes at the protein level. Phytochemistry **117:** 76-89

Jia Y, Ding Y, Shi Y, Zhang X, Gong Z, Yang S (2016) The cbfs triple mutants reveal the essential functions of CBFs in cold acclimation and allow the definition of CBF regulons in Arabidopsis. New Phytol **212:** 345-353

Kanwischer M, Porfirova S, Bergmuller E, Dormann P (2005) Alterations in tocopherol cyclase activity in transgenic and mutant plants of Arabidopsis affect tocopherol content, tocopherol composition, and oxidative stress. Plant Physiol **137:** 713-723

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of Arabidopsis. Plant Physiol **136:** 4159-4168

Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, Chen H, Dubin M, Lee CR, Wang C, Bemm F, Becker C, O'Neil R, O'Malley RC, Quarless DX, Genomes C, Schork NJ, Weigel D, Nordborg M, Ecker JR (2016) Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell **166:** 492-505

Kerwin R, Feusier J, Corwin J, Rubin M, Lin C, Muok A, Larson B, Li BH, Joseph B, Francisco M, Copeland D, Weinig C, Kliebenstein DJ (2015) Natural genetic variation in Arabidopsis thaliana defense metabolism genes modulates field fitness. Elife **4**

Kerwin RE, Feusier J, Muok A, Lin C, Larson B, Copeland D, Corwin JA, Rubin MJ, Francisco M, Li B, Joseph B, Weinig C, Kliebenstein DJ (2017) Epistasis x environment interactions among Arabidopsis thaliana glucosinolate genes impact complex traits and fitness in the field. New Phytol **215:** 1249-1263

Kim MS, Cho SM, Kang EY, Im YJ, Hwangbo H, Kim YC, Ryu CM, Yang KY, Chung GC, Cho BH (2008) Galactinol is a signaling component of the induced systemic resistance caused by Pseudomonas chlororaphis O6 root colonization. Mol Plant Microbe Interact **21:** 1643-1653

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in Arabidopsis thaliana. Nature Genetics **39:** 1151-1155

Kitano H (2002) Systems biology: a brief overview. Science **295:** 1662-1664

Kleessen S, Antonio C, Sulpice R, Laitinen R, Fernie AR, Stitt M, Nikoloski Z (2012) Structured patterns in geographic variability of metabolic phenotypes in Arabidopsis thaliana. Nature Communications **3**

Knight MR, Knight H (2012) Low-temperature perception leading to gene expression and cold tolerance in higher plants. New Phytol **195:** 737-751

Koch K (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. Curr Opin Plant Biol **7:** 235-246

Koenig D, Weigel D (2015) Beyond the thale: comparative genomics and genetics of Arabidopsis relatives. Nat Rev Genet **16:** 285-298

Koonin EV (2012) Does the central dogma still stand? Biol Direct **7:** 27

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in Arabidopsis thaliana. Annual Review of Plant Biology **55:** 141-172

Koornneef M, Meinke D (2010) The development of Arabidopsis as a model plant. Plant J **61:** 909-921

Kosova K, Vitamvas P, Prasil IT, Renaut J (2011) Plant proteome changes under abiotic stress--contribution of proteomics studies to understanding plant stress response. J Proteomics **74:** 1301-1322

Lange BM, Ghassemian M (2003) Genome organization in Arabidopsis thaliana: a survey for genes involved in isoprenoid and chlorophyll metabolism. Plant Mol Biol **51:** 925-948

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH (2012) Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. Molecular Ecology **21:** 5512-5529

Le DT, Tarrago L, Watanabe Y, Kaya A, Lee BC, Tran U, Nishiyama R, Fomenko DE, Gladyshev VN, Tran LS (2013) Diversity of plant methionine sulfoxide reductases B and evolution of a form specific for free methionine sulfoxide. PLoS One **8:** e65637

Lehmann U, Wienkoop S, Tschoep H, Weckwerth W (2008) If the antibody fails - a mass Western approach. Plant Journal **55:** 1039-1046

Li JY, Oulee TM, Raba R, Amundson RG, Last RL (1993) Arabidopsis Flavonoid Mutants Are Hypersensitive to Uv-B Irradiation. Plant Cell **5:** 171-179

Linster CL, Clarke SG (2008) L-Ascorbate biosynthesis in higher plants: the role of VTC2. Trends Plant Sci **13:** 567-573

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133:** 523-536

Los DA, Murata N (2004) Membrane fluidity and its roles in the perception of environmental signals. Biochim Biophys Acta **1666:** 142-157

Maeda H, Song W, Sage TL, DellaPenna D (2006) Tocopherols play a crucial role in low-temperature adaptation and Phloem loading in Arabidopsis. Plant Cell **18:** 2710-2732

Malinovsky FG, Thomsen MF, Nintemann SJ, Jagd LM, Bourgine B, Burow M, Kliebenstein DJ (2017) An evolutionarily young defense metabolite influences the root growth of plants via the ancient TOR signaling pathway. Elife **6**

Maloof JN, Borevitz JO, Dabi T, Lutes J, Nehring RB, Redfern JL, Trainer GT, Wilson JM, Asami T, Berry CC, Weigel D, Chory J (2001) Natural variation in light sensitivity of Arabidopsis. Nature Genetics **29:** 441-446

Malosetti M, Ribaut JM, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Frontiers in Physiology **4**

Maruyama K, Sakuma Y, Kasuga M, Ito Y, Seki M, Goda H, Shimada Y, Yoshida S, Shinozaki K, Yamaguchi-Shinozaki K (2004) Identification of cold-inducible downstream genes of the

Arabidopsis DREB1A/CBF3 transcriptional factor using two microarray systems. Plant Journal **38:** 982-993

Meyer Y, Siala W, Bashandy T, Riondet C, Vignols F, Reichheld JP (2008) Glutaredoxins and thioredoxins in plants. Biochim Biophys Acta **1783:** 589-600

Michaels SD, He Y, Scortecci KC, Amasino RM (2003) Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. Proc Natl Acad Sci U S A **100:** 10102-10107

Michaels SD, Himelblau E, Kim SY, Schomburg FM, Amasino RM (2005) Integration of flowering signals in winter-annual Arabidopsis. Plant Physiol **137:** 149-156

Mikkelsen MD, Thomashow MF (2009) A role for circadian evening elements in cold-regulated gene expression in Arabidopsis. Plant J **60:** 328-339

Morgenthal K, Wienkoop S, Scholz M, Selbig J, Weckwerth W (2005) Correlative GC-TOF-MS-based metabolite profiling and LC-MS-based protein profiling reveal time-related systemic regulation of metabolite-protein networks and improve pattern recognition for multiple biomarker selection. Metabolomics **1:** 109-121

Nagahatenna DS, Langridge P, Whitford R (2015) Tetrapyrrole-based drought stress signalling. Plant Biotechnol J **13:** 447-459

Nagai S, Koide M, Takahashi S, Kikuta A, Aono M, Sasaki-Sekimoto Y, Ohta H, Takamiya K, Masuda T (2007) Induction of isoforms of tetrapyrrole biosynthetic enzymes, AtHEMA2 and AtFC1, under stress conditions and their physiological functions in Arabidopsis. Plant Physiol **144:** 1039-1051

Nägele T (2014) Linking metabolomics data to underlying metabolic regulation. Front Mol Biosci **1:** 22

Nägele T, Heyer AG (2013) Approximating subcellular organisation of carbohydrate metabolism during cold acclimation in different natural accessions of Arabidopsis thaliana. New Phytologist **198:** 777-787

Nägele T, Kandel BA, Frana S, Meissner M, Heyer AG (2011) A systems biology approach for the analysis of carbohydrate dynamics during acclimation to low temperature in Arabidopsis thaliana. Febs Journal **278:** 506-518

Nägele T, Mair A, Sun XL, Fragner L, Teige M, Weckwerth W (2014) Solving the Differential Biochemical Jacobian from Metabolomics Covariance Data. Plos One **9**

Nägele T, Stutz S, Hormiller II, Heyer AG (2012) Identification of a metabolic bottleneck for cold acclimation in Arabidopsis thaliana. Plant Journal **72:** 102-114

Nägele T, Weckwerth W (2013) Eigenvalues of Jacobian Matrices Report on Steps of Metabolic Reprogramming in a Complex Plant-Environment Interaction. Applied Mathematics **04:** 44-49

Nagler M, Nukarinen E, Weckwerth W, Nagele T (2015) Integrative molecular profiling indicates a central role of transitory starch breakdown in establishing a stable C/N homeostasis during cold acclimation in two natural accessions of Arabidopsis thaliana. Bmc Plant Biology **15**

Nicotra AB, Atkin OK, Bonser SP, Davidson AM, Finnegan EJ, Mathesius U, Poot P, Purugganan MD, Richards CL, Valladares F, van Kleunen M (2010) Plant phenotypic plasticity in a changing climate. Trends Plant Sci **15:** 684-692

Noctor G, Mhamdi A, Chaouch S, Han Y, Neukermans J, Marquez-Garcia B, Queval G, Foyer CH (2012) Glutathione in plants: an integrated overview. Plant Cell Environ **35:** 454-484

Noctor G, Queval G, Mhamdi A, Chaouch S, Foyer CH (2011) Glutathione. Arabidopsis Book **9:** e0142

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in Arabidopsis thaliana. Nature Genetics **30:** 190-193

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng HG, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao KY, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J (2005) The pattern of polymorphism in Arabidopsis thaliana. Plos Biology **3:** 1289-1299

Nukarinen E, Nägele T, Pedrotti L, Wurzinger B, Mair A, Landgraf R, Bornke F, Hanson J, Teige M, Baena-Gonzalez E, Droge-Laser W, Weckwerth W (2016) Quantitative phosphoproteomics

reveals the role of the AMPK plant ortholog SnRK1 as a metabolic master regulator under energy deprivation. Sci Rep **6:** 31697

Obata T, Fernie AR (2012) The use of metabolomics to dissect plant responses to abiotic stresses. Cell Mol Life Sci **69:** 3225-3243

Orvar BL, Sangwan V, Omann F, Dhindsa RS (2000) Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity. Plant J **23:** 785-794

Pagter M, Alpers J, Erban A, Kopka J, Zuther E, Hincha DK (2017) Rapid transcriptional and metabolic regulation of the deacclimation process in cold acclimated Arabidopsis thaliana. BMC Genomics **18:** 731

Park MJ, Seo PJ, Park CM (2012) CCA1 alternative splicing as a way of linking the circadian clock to temperature response in Arabidopsis. Plant Signal Behav **7:** 1194-1196

Park S, Lee CM, Doherty CJ, Gilmour SJ, Kim Y, Thomashow MF (2015) Regulation of the Arabidopsis CBF regulon by a complex low-temperature regulatory network. Plant J **82:** 193-207

Pauli H, Gottfried M, Dullinger S, Abdaladze O, Akhalkatsi M, Alonso JLB, Coldea G, Dick J, Erschbamer B, Calzado RF, Ghosn D, Holten JI, Kanka R, Kazakis G, Kollar J, Larsson P, Moiseev P, Moiseev D, Molau U, Mesa JM, Nagy L, Pelino G, Puscas M, Rossi G, Stanisci A, Syverhuset AO, Theurillat JP, Tomaselli M, Unterluggauer P, Villar L, Vittoz P, Grabherr G (2012) Recent Plant Diversity Changes on Europe's Mountain Summits. Science **336:** 353-355

Peterbauer T, Richter A (2001) Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Science Research **11:** 185-197

Pigliucci M (1996) How organisms respond to environmental changes: From phenotypes to molecules (and vice versa). Trends in Ecology & Evolution **11:** 168-173

Pigliucci M (1998) Ecological and evolutionary genetics of Arabidopsis. Trends in Plant Science **3:** 485-489

Pigliucci M (2001) Syntheses in Ecology and Evolution. Phenotypic plasticity: Beyond nature and nuture. Johns Hopkins University Press; Johns Hopkins University Press

Pigliucci M (2002) Touchy and bushy: Phenotypic plasticity and integration in response to wind stimulation in Arabidopsis thaliana. International Journal of Plant Sciences **163:** 399-408

Pigliucci M (2009) An extended synthesis for evolutionary biology. Annals of the New York Academy of Sciences **1168:** 218-228

Pigliucci M (2010) Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. Philosophical Transactions of the Royal Society B-Biological Sciences **365:** 557-566

Pigliucci M, Kolodynska A (2002) Phenotypic plasticity and integration in response to flooded conditions in natural accessions of Arabidopsis thaliana (L.) Heynh (Brassicaceae). Annals of Botany **90:** 199-207

Pigliucci M, Kolodynska A (2002) Phenotypic plasticity to light intensity in Arabidopsis thaliana: invariance of reaction norms and phenotypic integration. Evolutionary Ecology **16:** 27-47

Pigliucci M, Schlichting CD (1996) Reaction norms of Arabidopsis .4. Relationships between plasticity and fitness. Heredity **76:** 427-436

Pigliucci M, Schlichting CD, Whitton J (1995) Reaction Norms of Arabidopsis .2. Response to Stress and Unordered Environmental Variation. Functional Ecology **9:** 537-547

Pigliucci M, Schmitt J (1999) Genes affecting phenotypic plasticity in Arabidopsis: pleiotropic effects and reproductive fitness of photomorphogenic mutants. Journal of Evolutionary Biology **12:** 551-562

Pigliucci M, Whitton J, Schlichting CD (1995) Reaction Norms of Arabidopsis .1. Plasticity of Characters and Correlations across Water, Nutrient and Light Gradients. Journal of Evolutionary Biology **8:** 421-438

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Agren J, Bossdorf O, Byers D, Donohue K, Dunning M, Holub EB, Hudson A, Le Corre V, Loudet O, Roux F, Warthmann N, Weigel D, Rivero L, Scholl R, Nordborg M, Bergelson J, Borevitz JO (2010) The Scale of Population Structure in Arabidopsis thaliana. Plos Genetics **6**

Poolman MG, Miguet L, Sweetlove LJ, Fell DA (2009) A Genome-Scale Metabolic Model of Arabidopsis and Some of Its Properties. Plant Physiology **151:** 1570-1581
Rask L, Andreasson E, Ekbom B, Eriksson S, Pontoppidan B, Meijer J (2000) Myrosinase: gene family evolution and herbivore defense in Brassicaceae. Plant Mol Biol **42:** 93-113
Rocco M, Arena S, Renzone G, Scippa GS, Lomaglio T, Verrillo F, Scaloni A, Marra M (2013) Proteomic analysis of temperature stress-responsive proteins in Arabidopsis thaliana rosette leaves. Mol Biosyst **9:** 1257-1267
Saito K, Matsuda F (2010) Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. Annual Review of Plant Biology, Vol 61 **61:** 463-489
Schmitz RJ, Ecker JR (2012) Epigenetic and epigenomic variation in Arabidopsis thaliana. Trends in Plant Science **17:** 149-154
Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen HM, Schork NJ, Ecker JR (2013) Patterns of population epigenomic diversity. Nature **495:** 193-198
Schulz E, Tohge T, Zuther E, Fernie AR, Hincha DK (2015) Natural variation in flavonol and anthocyanin metabolism during cold acclimation in Arabidopsis thaliana accessions. Plant Cell and Environment **38:** 1658-1672
Schulz E, Tohge T, Zuther E, Fernie AR, Hincha DK (2016) Flavonoids are determinants of freezing tolerance and cold acclimation in Arabidopsis thaliana. Scientific Reports **6**
Seo PJ, Park MJ, Lim MH, Kim SG, Lee M, Baldwin IT, Park CM (2012) A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in Arabidopsis. Plant Cell **24:** 2427-2442
Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in Arabidopsis thaliana: biogeography and postglacial colonization of Europe. Molecular Ecology **9:** 2109-2118
Skirycz A, De Bodt S, Obata T, De Clercq I, Claeys H, De Rycke R, Andriankaja M, Van Aken O, Van Breusegem F, Fernie AR, Inze D (2010) Developmental stage specificity and the role of mitochondrial metabolism in the response of Arabidopsis leaves to prolonged mild osmotic stress. Plant Physiol **152:** 226-244
Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, Dhondt S, Pucci A, Gonzalez N, Hoeberichts F, Tognetti VB, Galbiati M, Tonelli C, Van Breusegem F, Vuylsteke M, Inze D (2011) Survival and growth of Arabidopsis plants given limited water are not equal. Nat Biotechnol **29:** 212-214
Smirnoff N (2000) Ascorbate biosynthesis and function in photoprotection. Philos Trans R Soc Lond B Biol Sci **355:** 1455-1464
Somerville C, Koornneef M (2002) Timeline - A fortunate choice: the history of Arabidopsis as a model plant. Nature Reviews Genetics **3:** 883-889
Steuer R, Gross T, Selbig J, Blasius B (2006) Structural kinetic modeling of metabolic networks. Proceedings of the National Academy of Sciences of the United States of America **103:** 11868-11873
Steuer R, Kurths J, Fiehn O, Weckwerth W (2003) Observing and interpreting correlations in metabolomic networks. Bioinformatics **19:** 1019-1026
Stitt M, Sulpice R, Keurentjes J (2010) Metabolic Networks: How to Identify Key Components in the Regulation of Metabolism and Growth. Plant Physiology **152:** 428-444
Stockinger EJ, Gilmour SJ, Thomashow MF (1997) Arabidopsis thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci U S A **94:** 1035-1040
Strand A, Hurry V, Gustafsson P, Gardestrom P (1997) Development of Arabidopsis thaliana leaves at low temperatures releases the suppression of photosynthesis and photosynthetic gene expression despite the accumulation of soluble carbohydrates. Plant J **12:** 605-614
Strand A, Hurry V, Henkes S, Huner N, Gustafsson P, Gardestrom P, Stitt M (1999) Acclimation of Arabidopsis leaves developing at low temperatures. Increasing cytoplasmic volume

accompanies increased activities of enzymes in the Calvin cycle and in the sucrose-biosynthesis pathway. Plant Physiol **119:** 1387-1398

Strand A, Zrenner R, Trevanion S, Stitt M, Gustafsson P, Gardestrom P (2000) Decreased expression of two key enzymes in the sucrose biosynthesis pathway, cytosolic fructose-1,6-bisphosphatase and sucrose phosphate synthase, has remarkably different consequences for photosynthetic carbon metabolism in transgenic Arabidopsis thaliana. Plant J **23:** 759-770

Sun XL, Weckwerth W (2012) COVAIN: a toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. Metabolomics **8:** S81-S93

Tanaka R, Kobayashi K, Masuda T (2011) Tetrapyrrole Metabolism in Arabidopsis thaliana. Arabidopsis Book **9:** e0145

Tchebakova NM, Parfenova E, Soja AJ (2009) The effects of climate, permafrost and fire on vegetation change in Siberia in a changing climate. Environmental Research Letters **4**

Teige M, Scheikl E, Eulgem T, Doczi R, Ichimura K, Shinozaki K, Dangl JL, Hirt H (2004) The MKK2 pathway mediates cold and salt stress signaling in Arabidopsis. Mol Cell **15:** 141-152

Thalhammer A, Bryant G, Sulpice R, Hincha DK (2014) Disordered cold regulated15 proteins protect chloroplast membranes during freezing through binding and folding, but do not stabilize chloroplast enzymes in vivo. Plant Physiol **166:** 190-201

The 1001 Genomes Consortium (2016) 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408:** 796-815

Thomashow MF (1999) Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology **50:** 571-599

Thomashow MF (2010) Molecular Basis of Plant Cold Acclimation: Insights Gained from Studying the CBF Cold Response Pathway. Plant Physiology **154:** 571-577

Tichy L (2002) JUICE, software for vegetation classification. Journal of Vegetation Science **13:** 451-453

Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, Laitinen RAE, Huang Y, Chory J, Lipka V, Borevitz JO, Dangl JL, Bergelson J, Nordborg M, Weigel D (2010) Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. Nature **465:** 632-U129

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant Journal **42:** 218-235

Trethewey RN, Krotzky AJ, Willmitzer L (1999) Metabolic profiling: a Rosetta Stone for genomics? Curr Opin Plant Biol **2:** 83-85

Turesson G (1922) The genotypical response of the plant species to the habitat. Hereditas **3:** 211-350

Valledor L, Escandon M, Meijon M, Nukarinen E, Canal MJ, Weckwerth W (2014) A universal protocol for the combined isolation of metabolites, DNA, long RNAs, small RNAs, and proteins from plants and microorganisms. Plant Journal **79:** 173-180

Valluru R, Van den Ende W (2011) Myo-inositol and beyond--emerging networks under stress. Plant Sci **181:** 387-400

Verslues PE, Sharma S (2010) Proline metabolism and its implications for plant-environment interaction. Arabidopsis Book **8:** e0140

Vieira Dos Santos C, Cuine S, Rouhier N, Rey P (2005) The Arabidopsis plastidic methionine sulfoxide reductase B proteins. Sequence and activity characteristics, comparison of the expression with plastidic methionine sulfoxide reductase A, and induction by photooxidative stress. Plant Physiol **138:** 909-922

Wang J, Yu Y, Zhang Z, Quan R, Zhang H, Ma L, Deng XW, Huang R (2013) Arabidopsis CSN5B interacts with VTC1 and modulates ascorbic acid synthesis. Plant Cell **25:** 625-636

Wang J, Zhang Z, Huang R (2013) Regulation of ascorbic acid synthesis in plants. Plant Signal Behav **8:** e24536

Wang L, Nägele T, Doerfler H, Fragner L, Chaturvedi P, Nukarinen E, Bellaire A, Huber W, Weiszmann J, Engelmeier D, Ramsak Z, Gruden K, Weckwerth W (2016) System level analysis of cacao seed ripening reveals a sequential interplay of primary and secondary metabolism leading to polyphenol accumulation and preparation of stress resistance. Plant Journal **87:** 318-332

Wang Y, Hua J (2009) A moderate decrease in temperature induces COR15a expression through the CBF signaling cascade and enhances freezing tolerance. Plant J **60:** 340-349

Wanner LA, Junttila O (1999) Cold-induced freezing tolerance in Arabidopsis. Plant Physiol **120:** 391-400

Weckwerth W (2003) Metabolomics in systems biology. Annual Review of Plant Biology **54:** 669-689

Weckwerth W (2008) Integration of metabolomics and proteomics in molecular plant physiology - coping with the complexity by data-dimensionality reduction. Physiologia Plantarum **132:** 176-189

Weckwerth W (2011) Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. Journal of Proteomics **75:** 284-305

Weckwerth W (2011) Unpredictability of metabolism-the key role of metabolomics science in combination with next-generation genome sequencing. Analytical and Bioanalytical Chemistry **400:** 1967-1978

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. Proceedings of the National Academy of Sciences of the United States of America **101:** 7809-7814

Weigel D (2012) Natural variation in Arabidopsis: from molecular genetics to ecological genomics. Plant Physiol **158:** 2-22

Weigel D, Nordborg M (2005) Natural variation in arabidopsis. How do we find the causal genes? Plant Physiology **138:** 567-568

Wienkoop S, Baginsky S, Weckwerth W (2010) Arabidopsis thaliana as a model organism for plant proteome research. Journal of Proteomics **73:** 2239-2248

Wienkoop S, Morgenthal K, Wolschin F, Scholz M, Selbig J, Weckwerth W (2008) Integration of metabolomic and proteomic phenotypes. Molecular & Cellular Proteomics **7:** 1725-1736

Wienkoop S, Weckwerth W (2006) Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in Arabidopsis thaliana. Journal of Experimental Botany **57:** 1529-1535

Wilczek AM, Cooper MD, Korves TM, Schmitt J (2014) Lagging adaptation to warming climate in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **111:** 7906-7913

Williams TCR, Poolman MG, Howden AJM, Schwarzlander M, Fell DA, Ratcliffe RG, Sweetlove LJ (2010) A Genome-Scale Metabolic Model Accurately Predicts Fluxes in Central Carbon Metabolism under Stress Conditions. Plant Physiology **154:** 311-323

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol **5:** 218-223

Winter H, Huber SC (2000) Regulation of sucrose metabolism in higher plants: Localization and regulation of activity of key enzymes. Critical Reviews in Biochemistry and Molecular Biology **35:** 253-289

Zanella M, Borghi GL, Pirone C, Thalmann M, Pazmino D, Costa A, Santelia D, Trost P, Sparla F (2016) beta-amylase 1 (BAM1) degrades transitory starch to sustain proline biosynthesis during drought stress. Journal of Experimental Botany **67:** 1819-1826

Zeeman SC, Smith SM, Smith AM (2007) The diurnal metabolism of leaf starch. Biochem J **401:** 13-28

Zhao C, Zhang Z, Xie S, Si T, Li Y, Zhu JK (2016) Mutational Evidence for the Critical Role of CBF Transcription Factors in Cold Acclimation in Arabidopsis. Plant Physiol **171:** 2744-2759

Zhao C, Zhu JK (2016) The broad roles of CBF genes: From development to abiotic stress. Plant Signal Behav **11:** e1215794

Zhen Y, Ungerer MC (2008) Clinal variation in freezing tolerance among natural accessions of Arabidopsis thaliana. New Phytologist **177:** 419-427

Acknowledgements

Wen die Dankbarkeit geniert,

Der ist übel dran;

Denke, wer dich einst geführt,

Wer für dich getan!

Goethe

Writing this thesis and conducting all contained and uncontained experiments took tremendous effort. It would not have been possible without the endurance and supervision of Wolfram Weckwerth. Likewise, my mentor and friend Thomas Nägele taught me so much about scientific method in countless discussions that I cannot possibly overrate his effort and impact on my work. I'm proud that you were part of this journey. Enjoy your professorship and go for the Hessian tensor! Ditt!!!

As a complex endeavor like a PhD thesis comes with boons and banes, I am endlessly grateful for the love, patience and motivating words from my partner Isabella Rieder and my son Jonas. You were keeping me going forward through good and bad and I'm happy to finally share this accomplishment with you. I want to express my deep gratitude to my parents Ilse and Rupert Nagler who always supported me both mentally and practically for instance by lending me their car so I could search and visit *in situ* populations. I also want to thank my sister Sophie and her family Christian, Laurenz and Valentin for support.

I am thankful to Christian Gilli who let his botanical knowledge spill over into my mind and reliably explained species inventories of various habitats. When robots finally take over all the lab work, your knowledge will be all that's left until machine learning algorithms can deal with the subtle morphological differences of *Arabidopsis*, *Arabis* and *Eriophila* seedlings.

Jakob Weiszmann, thank you for being a true friend and helping hand. Going for *in situ* sample harvest has been just a great experience, cooling our ice cream in liquid nitrogen and listening to great music I had almost forgotten.

Thanks to Ella Nukarinen for being such an awesome collaborator. Thanks for sharing your knowledge with me and so much nice time. Everybody should have a magic window like we did!

A big "Thank You" to the people who made the working experience in the office and lab a nice one, most of all Reinhard Turetschek, Lisa Fürtauer and Julian Preiner. I'm much obliged for the help and assistance of Wolfgang Höhenwarter, Christa Schleper, Luis Valledor and David Lyon. Thumbs up to the technicians Lena Fragner, Martin Brenner and Sonja Tischler who kept the lab running.

Finally, I have to apologize to lots of people who had to put up with my increased social incompatibility in the recent past. Especially, I want to thank my band colleagues who endured a hard time with me skipping rehearsals because I got obsessed with manuscripts. Still, you guys were a prominent reason that kept my mental health upright, for making music with you is definitely major psychic hygiene and a true source of power. We've gotta keep on movin'! And we will. Although we never made it to the Waterhole. At least not just yet.