Spatially resolved transcriptomics reveals plant host responses to the aphid pest *Myzus persicae*

Michael Giolai (M.Sc., B.Sc.)

Doctor of Philosophy (PhD)

University of East Anglia

School of Environmental Sciences

September 2019

Abstract

Phloem sap feeding insect pests cause devastating agricultural losses with poorly understood mechanisms of plant defence responses to these insects leaving potentially environmentally damaging pesticides as the only protection. A striking feature of these pests is the characteristic pattern of how plants are colonised; i.e. by continuous manipulation of the hosts immune system until the point of successful phloem sap feeding. Plants respond to aphid feeding via a complex network of defence processes locally, i.e. in an entire leaf and systemically throughout the plant. How can we gain insights into these complex spatial and temporal processes to gain a better understanding of how plants respond to phloem sap feeding insects? With my work I show that by linking electrophysiological insect assays with recent progress in spatial transcriptomics, it is possible to unravel some of the features of *Arabidopsis thaliana* responses to *Myzus persicae* and, more generally, plant interactions with other invertebrate pests and microbial plant pathogens.

Table of Contents

List of tables

List of figures

List of accompanying material

A CD containing the material listed in the **Appendices** section.

Acknowledgements

I want to thank my supervisory team (Prof. Dr. Saskia Hogenhout, Dr. Matthew Clark, Dr. Iain Macaulay, Dr. Sam Mugford and Dr. Walter Verweij) for the suggestions, guidance, advise and mentoring that I experienced throughout the process of my PhD studies and that led to the here presented work.

I especially thank Saskia Hogenhout and Sam Mugford for the introduction to the world of plant insect interactions and their constructive advice in studying plant insect interactions.

Special thanks also to Matthew Clark and Walter Verweij who greatly helped and advised on the establishment of genomic technologies and who introduced me to the field of genomics and data driven research.

Further special thanks to Iain Macaulay for contributing his expertise in low-input transcriptome sequencing technologies, transcriptome experiment design and analysis to my supervision.

Thanks to my colleagues in the Hogenhout lab at the John Innes Centre and the Clark and Macaulay labs at the Earlham Institute, with which I had the chance to work together. Particularly to Ashleigh Lister for her helping hands while I have been on my PIPS internship and to Dr. Darren Heavens for his helpful suggestions when I faced genomics wet-lab related challenges.

Thanks also to the Entomology and Insectary team, as well as the Horticultural staff at the John Innes Centre.

I also thank Dr. Freddy Tjallingii (EPG Systems, Wageningen, Netherlands) for the helpful suggestions in applying insect EPG techniques.

Lastly, I thank Dr. Talia Karasov, Prof. Dr. Detlef Weigel and Dr. Patrice Wegener (Max Planck Campus, Tübingen, Germany) for helping me to organise and conduct a productive and joyful PIPS internship.

Chapter-1 Introduction

The progression of spatial transcriptomics technologies

Most multicellular model organisms are composed of different cell types aggregating to tissues and organs in a well-defined and spatially heterogeneous body plan [1]. Knowledge about the processes determined by this spatial heterogeneity is important to understand a multicellular system [2]. In the past this has led to the development of a rich repertoire of *in situ* and *in vivo* methods that assess spatial biological information in a functional context [3]. Many of these methods label DNA, RNA or proteins on a subcellular scale using fluorescently or enzymatically linked complementary nucleotide probes [4] or antibodies [5]. These methods greatly contributed to our functional understanding of fine scale cellular events and are still of extraordinary importance in resolving biological processes [3–5]. A disadvantage of labelling and imaging based *in situ* methods is the limited throughput, which -in comparison to sequencing based technologies- is compensated by the great resolution of these methods that can indicate the position of a target at a subcellular scale[3]. Many *in-situ* technologies however require specialised tissue preparation and only a small number of markers can be assayed at a time [3–5]. Although techniques such as seqFISH+ [6] have been developed to image transcripts for up 10,000 genes in a single sample by continuous washing and re-labelling of tissues, especially advances in next-generation sequencing (NGS) made testing of entire genomes, transcriptomes [7, 8] or epigenomes [9, 10] possible.

Over the last years these NGS based spatial transcriptomics technologies experienced great development [3]. Earlier methods used photoactivatable fluorescent markers to label cells for extraction and sequencing [3, 11, 12]. Bulk photoactivation of cells in tissues and subsequent sample dissociation however

does not retain the spatial information of cells in a tissue. The precise position of a cell within the photoactivated area remains therefore unknown and the spatial information sparse [12]. Also not every organism is accessible for the transformation with photoactivatable markers and so the use of photoactivatable markers is limited to model organisms or culture based cell aggregates [3, 11]. Retainment of spatial information at single-cell level resolution was made possible by combining laser capture microdissection (LCM) with low-input NGS methods [13, 14]. This was achieved at the cost of a lowered throughput and the restriction to a very small and thin area [13, 14]. LCM technologies however require specialised equipment and training for precise, laborious excision of tissue elements [15].In comparison to LCM assisted technologies, recent advances of spatial transcriptome sequencing allow to probe a larger area of 6.2 mm x 6.6 mm at a lower resolution of 100 µm (on average 3 – 30 cells) using an array of solid-surface immobilised and positionally barcoded reverse transcription primers [16]. Messenger RNA (mRNA) molecules diffuse from of thin, permeabilised histological sections to the array and hybridise to the reverse transcription oligonucleotides for reverse transcription and sequencing library preparation [16]. The Spatial Transcriptomics method was recently improved using high density bead arrays with position index barcoded beads instead of oligonucleotides [17, 18]: Rodriques et al. [17] report for the Slide-seq method a spatial resolution of ~ 10 µm, Vickovic et al. report for the High Density Spatial Transcriptomics (HDST) protocol a resolution of ~ 2 µm [18].

Although the advances presented by Slide-seq [17] and HDST [18] substantially add to NGS based spatial transcriptome sequencing, the need for specialised tissue and sample preparation limits the applicability for many laboratories, as both methods require specific tissue preparations (e.g. cryo-sectioning, permeabilization or fixation) and specialised protocols to assess transcriptome levels in thinly sectioned, permeabilised samples [16–18].

In cases where an easier, and in comparison to Spatial Transcriptomics non-commercial [19] solution is desired, a protocol based on sample cryosectioning to ~ 18 µm thin sections and computational reconstruction of complex 3D gene

expression maps has been published by Junker et al. [20]. Despite the lower resolution in comparison to HDST, the method allows to process morphologically complex samples (i.e. an entire zebrafish embryo) and library preparation with common wet-lab methods at a high spatial accuracy [20].

Table 1 Methods for spatially resolved genomics and transcriptomics: A rich palette of spatial and low-input omics methods is available to profile transcripts or single nucleotide variants (SNVs) from a multi- to a subcellular resolution. Highest resolutions (i.e. to the subcellular level) are achieved by probe-based, fluorescent *in-situ* hybridisation (FISH) technologies. FISH methods however rely on complementary binding of short fragments. In contrast laser capture microscopy (LCM), microtome sequencing or slide sequencing based protocols which rely on tissue dissection and / or lysis also enable profiling of full length transcripts at the cost of a decreased resolution (besides of new slide sequencing technologies). Table modified after Crosetto et al. **[3]**.

Method	Sample	Target	Spatial Resolution	Sensitivity	Read length
smFISH	Fixed tissues or cells	RNA	Subcellular	Abundance; SNVs; isoforms	Short fragments
Branched FISH	Fixed tissues or cells	RNA	Subcellular	Abundance	Short fragments
LCM sequencing	Fixed tissues or cells	DNA, RNA	Multicellular to cellular	Abundance; SNVs; isoforms	Short fragments
Microtomy sequencing	Fixed and fresh tissues	RNA, DNA	Multicellular	Abundance; SNVs; isoforms	Short fragments to full length isoforms
FISSEQ	Fixed tissues or cells	RNA	Cellular	Abundance; SNVs; isoforms	Short fragments
Slide technology sequencing	Fixed tissues	RNA	Multicellular to subcellular	Abundance; SNVs; isoforms	Short reads to full length isoforms

The progression of spatial transcriptomics technologies for plant model systems

The development of plant spatial transcriptomics methods followed a similar timeline as for animal organisms [3, 21]: plant researchers greatly benefited and still benefit from imaging based *in situ* methods [22, 23], plant tissues are accessible for LCM methods [24, 25] and fluorescence activate cell sorting (FACS) technologies allow studies of single plant protoplasts [26–28].

The introduction of recent high resolution spatial NGS technologies however posed some difficulties. While methods based on the dissection of large tissue areas or entire plant organs have been in use [29, 30], high resolution spatial NGS methods require robust protocols to extract and process low RNA amounts from a few or single cells. However, the rigid plant cell wall [31] and plant tissues rich in secondary metabolites [32] make the extraction of low RNA quantities and downstream reactions inefficient [15, 21]. Giacomello et al. recently applied the spatial transcriptomics [16] technology to *Arabidopsis thaliana* inflorescence meristem, *Populus tremula* dormant and developing leaf buds and *Picea abies* female cones [15]. To transfer this technology from mammalian to plant tissues, Giacomello et al. adjusted multiple steps of the workflow i.e. by using a milder fixation method, adapted enzymatic cocktails to permeabilise the tissue sections and additional protocol steps to capture secondary metabolites during tissue permeabilization and cDNA synthesis [11]. These improvements make technologies such as Slide-seq [17] or HDST [18] compatible with plant tissues, but the requirement to optimise the tissue preparation for different tissue types limits the applicability of the methods.

The application of spatial transcriptomics in plant-pathogen interaction studies

To date many spatial transcriptomics studies characterising plant pathogen interactions rely on LCM; e.g. [24, 33–36]. The published studies report of attack-site specific gene expression profiles, higher expression magnitudes by testing exclusively stressed cells, novel detected plant response genes and spatially distinct expression patterns of plant defence genes [24, 33, 36–38].

In a recent manuscript Coker et al. [37] aimed to develop a simpler and less hands-on intensive method than LCM to isolate attacked plant cells. To achieve this, the authors used FACS of *Hyaloperonospora arabidopsidis* haustoriated *A. thaliana leaf* cells expressing a haustoriation sensitive fluorescent cell marker. Coker et al. identified 267 differentially expressed (DE) genes of which 128 were shared with published bulk tissue (i.e. entire leaves with infected and uninfected cells) sequencing experiments [39–41]. For the 128 DE genes the authors describe stronger log2 fold-changes in comparison to the bulk studies. Coker et al. also found 139 DE novel genes for *A. thaliana-H. arabidopsidis* interactions. However, the manuscript also describes introduction of experimental noise by FACS pooling of 20,000 – 100,000 cells to collect enough material for transcriptome profiling. An additional described technical difficulty was the sparse availability of cell markers - the promoter driven green fluorescent protein system produced detectable fluorescence earliest 5 days post-inoculation [37].

A recent publication by Mulema et al. [38] describes the spatial and temporal transcriptome response of *A. thaliana* to *Botrytis cinerea* using mechanical leaf dissection. The authors study two zones at 0 – 6 mm and 6 – 12 mm distance from the infection site. Mulema et al. show differential gene expression in both zones at 12- and 24 hours post-inoculation and highlight the potential role of certain transcription factor (TF) families in regulating spatiotemporal gene expression [38].

Zierold et al. [42] and Bruggmann et al. [43] also apply mechanical sampling to study transcriptome responses of barley to the obligate biotrophic fungus *Blumeria graminis*. As the fungus colonises epidermal cells the authors use tissue peels to access epidermis and mesophyll plant tissues for fine-scale tissue sampling and microarray analysis. Bruggmann et al. characterise 44 epidermis specific and 76 mesophyll specific transcripts [43] and Zierold et al. describe 293 plant DE genes in the plants epidermis with an additional 18 transcripts of fungal origin [42].

Although the published manuscripts highlight the potential for the analysis of spatiotemporal plant transcriptome responses in plant pathogen interactions, the studies are limited to strongly localised features such as fungal infection sites [24, 34, 35] or large nematode infested cells [33]. Such samples are compatible with fine-scale dissection of tissues, but other important pests such as herbivorous insects show mobile patterns of plant colonisation and actively search for suitable attack sites on plants [44]. Spatial transcriptomics methods versatile enough to robustly study such interactions at a millimetre-scale resolution and a high throughput would have the potential to deepen our understanding of plant immunity to insects. However, such methods have not been described yet.

Plant defences to insect herbivores

Plants developed sophisticated defence strategies against pathogens and pests [45, 46]. To defend themselves from herbivore attack plants use a series of defences such as the deployment of physical barriers (e.g. hairs, trichomes and waxes) and metabolic as well as chemical cues [46–50]. In contrast to the adaptive immune system of animals [51, 52] plants detect a wide range of pathogens with a large repertoire of immune recognition receptors [53, 54]. The two main components of this recognition system [51, 55] are cell surface localised pattern recognition receptors (PRRs) that detect microbe-associated molecular patterns (MAMPs) [54] and intracellular receptors that recognise pathogen or pest virulence molecules named effectors [53]. The downstream processes triggered

by MAMP or effector recognition are named pattern triggered immunity (PTI) and effector triggered immunity (ETI) [51, 55]. The induced defences of PTI and ETI cannot be strictly distinguished [56]. In a simplified model [57] ETI leads to an effective hypersensitive response (HR) often resulting in cell death [51] to an adapted pathogen able to overcome a plants PTI defence [56, 57]. In contrast, PTI shapes a first and effective barrier against non-adapted pathogens attacking a plant [51, 56, 57].

This wide concept of plant pathogen recognition also applies to plant herbivore interactions [46, 48–50]. So far only a few herbivory PRRs recognising insect herbivore associated molecular patterns (HAMPs) or damage associated molecular patterns (DAMPs) are known, but more and more PRRs for insect pests are being described [48, 58, 59]. Among the first defence and defence mediating processes downstream of DAMP or HAMP perception are rapid membrane potential changes [60], cytosolic calcium fluxes [49, 60, 61], the production of reactive oxygen species (ROS) [46, 49] and mitogen-activated protein kinase (MAPK) signalling cascades [46, 48].

Important elements of plant immunity against insect herbivores are activated hormone signalling pathways and the wide range of synthesised secondary metabolite processes [32, 62, 63]. Especially the plant hormone pathways of jasmonic acid (JA), salicylic acid (SA) and ethylene (ET) have been studied extensively in the past [46, 48, 50, 62]. Studies showed that JA signalling is the most important pathway in mediating resistance to chewing insects [46, 62]. SA signalling has been described to be effective in the defence against phloem feeding insects [63]. ET signalling is described to mediate responses to a broad spectrum of insect pests [62, 64]. Interesting, but less well understood are the effects of cross talk between hormone signalling pathways such as potential antagonistic roles between SA and JA, as well as agonistic and antagonistic processes between ET and JA signalling in insect-plant interactions [46, 48, 65].

Next to hormone signalling, plant secondary metabolites play an important role in defending plants from herbivores [62]. Secondary metabolites are synthesised for

multiple purposes and act as toxic feeding compounds, aid in attraction of predatory enemies to insects or repel and distract insect pests [46, 62, 66].

Such metabolites can be stored as inactive compounds by the plant and induced in response to an attack. Compounds which plants use to defend themselves against insect pests are manifold with thousands of derivatives; among the best characterised are glucosinolates, benzoxazinoids, isoflavonoids, terpenoids, alkaloids, phenolic compounds, tannins, etc. [67]. Many compounds directly alter insect fitness. In Arabidopsis for example glucosinolate synthesis mutants have been described to increase susceptibility to chewing herbivores. Tobacco with a decreased nicotine content is more susceptible to herbivores. Transgenic tomato plants with an increased terpene production show increased insect resistance. Maize benzoxazinoids are directly toxic to chewing insects. However, also indirect functions for secondary metabolites are known. In maize for example indole has been described to reduce a plants attractiveness for caterpillars and to attract parasitoids. Also in wheat overexpression of a terpene synthase was linked with repelling activity aphids and augmented parasitoids recrution [46, 48, 67].

Knowledge about plant resistance mechanisms becomes increasingly important as the restrictive pesticide legislation in agriculturally important nations [68] and the evolution of pesticide resistances by pathogens and pests threatens agricultural systems [69, 70]. Especially recent advances in reducing crop generation times for breeding [71], strategies to identify plant resistance mediating ETI receptors using NGS [72–74], the possibility to transfer resistant alleles between plant species (e.g. from wild, resistant varieties to crop plants) [75, 76] and tailor-made genome editing techniques (e.g. to target susceptibility genes) [77, 78] build a new and effective pest control repertoire.

One major insect pest, where application of this repertoire will be useful, is the green peach aphid (GPA) pest *Myzus persicae*. *M. persicae* infestations are difficult to control due to the rapid generation times and clonal reproduction of the aphid, the potential of the insect to colonise a variety of plant families [79, 80], transmit agriculturally important viruses [81] and quickly develop pesticide resistances [82].

The aphid (stylets) feeding pathway

Aphids are a particularly damaging clade of insect pathogens in temperate agricultural systems [80, 83]. Aphids belong to the order Hemiptera and feed by using specialised, piercing-sucking mouthparts composed of stylets [84]. Stylets are used to pierce plant cells without leaving excessive damage [85, 86], to actively deposit effector molecules that act on plant defences [87] and to ingest cell sap from within the plant [81]. Hemipterans are also important vectors for plant pathogenic viruses and bacteria, which in addition to insect related damages expose the plant to a second pathogenic threat [81, 88].

Among aphids *M. persicae* is the economically most important pest worldwide [83]. Characteristics of the insect species are a host range of more than 400 plant species in 40 plant families, short generation times, clonal reproduction, strong dispersal rates and the ability to transmit over 100 different plant viruses contribute to the importance of *M. persicae* as pest [80]. The pest is also highly pesticide resistant [82] and has the ability to colonise and adapt to new hosts quickly [79]. A repertoire of efficient virulence factors (e.g. the effectors Mp10 [89], MpC002 [90], Mp42 [91]) help *M. persicae* to overcome a host defence responses [87, 92, 93].

To feed, *M. persicae* ingests nutrients from the photo-assimilate rich phloem [94, 95]. Nutrient uptake from phloem elements is achieved by puncturing phloem bundles with the stylet and active uptake of phloem sap from within the plant [96]. On the stylet path to the phloem *M. persicae* tests (i.e. probes or punctures) multiple cells [44]. During this process [96] small amounts of cell sap are ingested and saliva rich in effector molecules [87, 93] is deposited in the host cells [86, 87, 93]. Potential physical but also chemical and molecular cues that the insect uses to find suitable feeding spots is not well understood yet. Plant resistance screens described higher probing frequencies and longer walking paths of *M. persicae* on resistant plant varieties [44, 97]; this suggests potential dissemination of

intracellular signals retrieved from probing (e.g. by sensing the molecular state of a cell upon cell sap ingestion) [98].

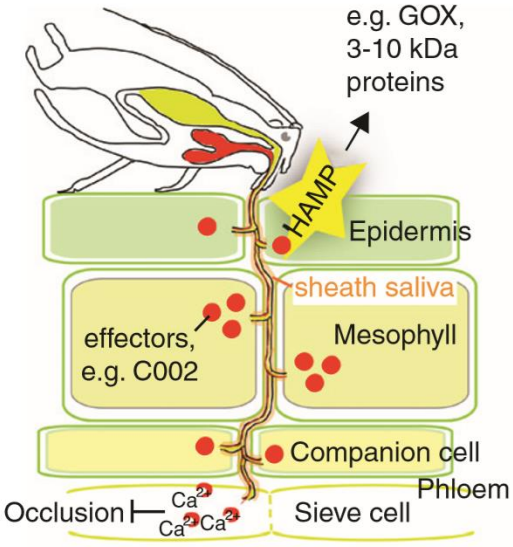

Figure 1 Infection process of a plant by a phloem feeding aphid: The insect uses stylets to navigate through a plants apoplast the nutrient rich phloem. Along the stylet path cells are punctured and effector molecules (e.g. here C002 as an example for the pea aphid) are actively deposited in the plant cells. Once the stylets reach the phloem, calcium binding proteins secreted by the aphid help to sequester calcium ions and so inhibit sieve element occlusion. Plants recognise the attacking pest by sensing HAMPs; the consequences of HAMP perception are downstream defence responses. Figure extracted from Hogenhout and Bos [87].

Plant defences to *Myzus persicae*

To date, the majority of our understanding of *A. thaliana* research promoted our understanding of molecular plant responses to *M. persicae* interactions [95, 99]: in a recent review article Louis and Shah [95] comprehend genetic factors involved in plant immunity to *M. persicae* [95]. The authors summarise key-genes in promoting plant resistance or susceptibility and the roles of JA, SA as well as ET hormone pathways [46, 65, 95]. A great role in promoting resistance to *M. persicae* can be attributed the accumulation of components that are toxic upon ingestion such as glucosinolates [100–102] and camalexin [66]. Toxic effects have also been observed for phloem localised lectin ingestion (e.g. *pp2-A1*) [103, 104]. Another important mechanism relies on sucrose sequestration (e.g. by

polymerisation to starch) and so the removal of a feeding stimulant promoted by the trehalose pathway enzyme *TPS11* [105]. The trehalose metabolic pathway also plays a role in upregulating important defence regulatory components (e.g. *PAD4* [105–107]).

Less clear is the role for oxylipin pathways [107] (i.e. oxylipins are derived from oxidized plant lipids [108]). The resistance associated JA hormone is synthesised by the 13-lipoxygenase (13-LOX) pathway [109, 110]. In contrast 9-lipoxygenase (9-LOX) pathway oxylipins are described to stimulate *M. persicae* feeding and so act as susceptibility factors [111]. However the 9-LOX pathway enzyme *LOX5* is involved in promoting *PAD4* expression and therefore also in mediating resistance to *M. persicae* [109].

Studies about leaf senescence also indicate the need for better understanding of plant responses to *M. persicae*: *M. persicae* feeding increases *PAD4* promoted expression of senescence genes, which leads to reduced aphid performance [112, 113]. Pegadaraju et al. hypothesise that induction of senescence mechanisms by the plant serves as defence mechanism as premature senescence could limit nutrient flow to insect infested leaves and so lead to less insect colonisation [112]. However, also studies describing increased aphid performance under induced senescence are published [114, 115]. Machado-Assefh et al. for example describe of increased phloem sap ingestion by aphids, potentially due to the increased nutrient mobilisation in the senescent organs (i.e. by compounds degradation) after triggering senescence [114].

Described results for the role of the SA pathway in promoting successful defences to *M. persicae* also diverge in literature: It is known that members of the JA, SA and ET pathways are induced upon *M. persicae* attack [65, 99, 107]. Whereas increased JA pathway activity promotes resistance against *M. persicae* [110], studies measuring insect performance on SA knock-out mutants did not show an increased effect on aphid fitness (i.e. ICS1 [113], EDS5 [116]). For other SA genes such as the signalling regulator NPR1 [117] experiments are less conclusive showing unaltered [116] or increased [118] aphid fitness on knock-out mutants. The specific role of SA signalling in *M. persicae* interactions therefore remains to

be elucidated. A hypothesis of Walling et al. describes *M. persicae* driven induction of SA responses to negatively act on the resistance promoting JA signalling pathway [119].

The specific role of ET components is less precisely described as well [95]; an example of resistance promoting ET signalling is JA induction [64], however also reports about susceptibility factors in ET responses exist (e.g. *MYB102* [120]). The regulation of the ET signalling pathway by some MYB transcription factors makes these genes to potential susceptibility factors in plant aphid interactions [95]. For example, Zhu et al. described increased MYB102 expression upon aphid perception and a positive feedback loop of MYB102 on enhancing ET levels and therefore increasing susceptibility [120].

The study of *Arabidopsis thaliana* - *Myzus persicae* interactions with spatial transcriptomics

The broad host range, clonal preproduction and rapid generation time [80, 83], the availability of genomic resources [79] and transgenic techniques [90] make *M. persicae* an excellent and well used model in understanding plant responses to phloem feeding insects. Over the last decades many studies of plant interactions with *M. persicae* explaining plant defence mechanisms have been published – e.g. [59, 61, 66, 89, 91–93, 95, 106, 109, 121]. More recent studies indicate consecutive and strongly localised layers of defence responses such as calcium bursts [61], reactive oxygen bursts [122], gene expression changes [66, 123] at and close to sites of *M. persicae* attack. Especially fine-scale gene expression changes have so far not been characterised well. A previous study by Kettles et al. showed highly variable spatial gene expression patterns and magnitudes for the defence marker PAD3 at local *M. persicae* attack sites [66]. The pattern of this localised immune response likely depends on the direct interaction between plant defences and the applied countermeasures by the pest [57].

Motivated by these recent studies, I am interested in characterising plant transcriptome responses to *M. persicae* at and near to sites of insect attack. This requires the combination of fine-scale transcriptome sequencing technologies with a suitable method to determine insect activities on a plant. Local insect attacks can be studied by measuring the number of cell punctures using the electrical penetration graph (EPG) technology [124]. However, as phloem feeding insects actively search for nutrient rich phloem bundles, stylets routes extend for a few hundred microns through the apoplast with numerous attacked cells along the path [96, 125]; therefore the described LCM [24], FACS [37], array [15] and macro-scale dissection [38] based NGS methods are not easily applied.

Here I therefore present the development of a novel and robust fine-scale spatial RNA-seq technology that enables the dissemination of transcriptome level maps from very small amounts of any eukaryotic tissue. I show the potential of the method by analysing spatial responses to bacterial (the bacterial peptide flagellin-22) and insect elicitors (crude aphid extract) and I present the application of EPG coupled spatiotemporal transcriptomics in an *A. thaliana – M. persicae* interaction study. This allows me to disseminate strongly localised spatial plant transcriptome responses to live insects and provide novel insights in the complex interaction of plants with phloem feeding insects.

Chapter-2 Spatially resolved transcriptomics reveals plant host responses to MAMPs

Introduction

Most model plants and all crops species are multicellular, consisting of multiple organs and cell types with a multitude of physiological states [126]. A thorough understanding of these complex systems requires the ability to dissect and characterise processes in the different organs and cell types. This is challenging, though recently multi-omics single-cell studies have been flourishing [7], but high-throughput, high-resolution methodologies that assess molecular conditions with spatial resolution are sparsely available [3, 15, 16, 20].

Although some spatial and low-input transcriptome profiling methods have been developed for animal model organisms [3, 16, 20, 127], these methods are difficult to transfer to plants [15, 38]. In comparison to animal cells, plant tissues hold a series of additional challenges: the robust plant cell wall requires specialised sample preparation (which makes reproducible, high-throughput sample preparation more difficult) and some plant secondary metabolites e.g. polyphenols can inhibit downstream enzymatic processes [128]. For plants, single plant cells (protoplasts) can be obtained by enzymatic removal of plant-cell walls and subsequent FACS assays [129]. At the sub-cellular scale plant nuclei can be isolated within minutes by cell lysis and FACS [129–131]. However, 'stimulus and response' assays, such as differential gene-expression experiments or the characterisation of cell-type transcripts could be affected by these additional experimental procedures before RNA-extraction. Another important factor is the loss of spatial information when nuclei or protoplasts are extracted from a tissue. Thus methods such as fluorescent *in situ* hybridisation (FISH) [23], LCM [25, 132, 133] or the Spatial Transcriptomics [15, 16] workflow are better suited to understand spatial transcription changes. However, all three methods need specific tissue preparations (e.g. cryo-sectioning, permeabilization or fixation) and

specialised protocols to assess transcriptome levels: FISH methods require imaging of transcripts and are restricted to multiplexing a few fluorescent anti-sense probes at a time [3], LCM requires specialised equipment and training for precise, laborious excision of specific tissue elements [15] and the Spatial Transcriptomics protocol requires preparation of thinly sectioned, permeabilised samples and custom made DNA arrays [15, 16].

Despite the high level of resolution that can be achieved with all these methods, they are not easily applied in most laboratories. I aimed to overcome this with my spatial transcriptomics workflow – hereafter referred to under the working name spatial-transcriptome sequencing (ST-seq). ST-seq is designed to quickly process mechanically dissected millimetre sized samples into sequencing libraries using standard laboratory equipment and can be used in most modern laboratories. ST-seq is based on three consecutive steps: (1) rapid, mechanical sample dissection of small e.g. 1 mm^2 leaf areas, (2) a high-throughput method for high quality mRNA extraction of difficult to lyse plant tissues and (3) NGS library construction (**Figure 2**).

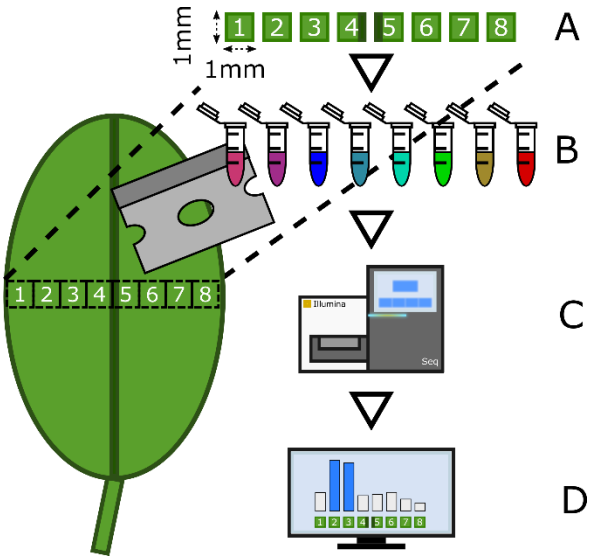

Figure 2 Overview of the ST-seq workflow: (A) Tissue sections of approximately 1 mm^2 size are mechanically dissected (e.g. a cross-section of a leaf) and after mRNA extraction (B) prepared into uniquely barcoded Illumina sequencing libraries. After (C) Illumina sequencing (D) transcript specific, spatial expression data can be assessed and analysed.

In a series of experiments, I compare the performance of ST-seq with standard RNA-seq (Illumina TruSeq) experiments. Using Illumina sequencing I identify DE genes in 1D, 1 mm wide lateral leaf sections. This shows how transcript and expression levels change across the tissues that make up the leaf. I compare a large-scale vs. fine-scale transcriptome experiments ability to detect plant responses induced by the bacterial peptide flagellin-22 (flg22), a well-described MAMP that triggers plant immune responses [55]. By comparing my data with published datasets for 'flagellin rapidly elicited' (FLARE) genes [134] I identify 143 of 253 described FLARE genes that overlap with my data, and a further 428 genes with similar expression patterns to FLARE genes. I show that the detected 428 transcripts, are enriched for plant defence responses and that spatial transcriptome data can be used to reconstruct the spatial expression of pathway components across leaves. Importantly, I demonstrate that these findings are independent of potentially wounding induced genes that could be responding to the mechanical dissection.

Results

Does leaf dissection induce wounding response gene expression profiles?

Physical wounding of plants is known to induce wounding related gene expression [135]. This is an important point to consider as the ST-seq workflow dissects tissue into ~ 1 mm^2 squares followed by immediate snap freezing on dry ice. Yet dissection could potentially lead to activation of wounding related gene expression and dissection takes longer as the resolution increases (grid size). As any wounding effect could form a technical limitation to ST-seq I measured the number of DE genes found after tissue dissection. For this I tested ~1 mm^2 leaf squares (3 biological replicates per time-point) prepared at the time-points: 0-minutes, 2.5-minutes, 5-minutes and 10-minutes between cutting and freezing on dry ice (when all enzymatic reactions cease) for DE genes (**Table 3**). To determine the number of DE genes at each time-point I compared the 2.5-minute, 5-minute

and 10-minute samples with the 0-minute samples as an unwounded reference. This analysis showed just 1 DE-gene (AT2G37130) at the 2.5-minute time-point (which was not significant at later time-points) there were no DE genes at the 5-minute time-point and 13 genes at the 10-minute time-point (see: **Chapter2_additional_file1.pdf: 'Wounding time-series experiment'** and **Chapter2_additional_file2.xlsx: 'Wounding_DEgenes'**) suggesting that the transcriptional response to wounding starts between 5 and 10 minutes. I looked for enriched biological processes in the combined set of 14 genes and detected three genes at the 10-minute time-point being associated with the GO-term 'response to wounding': *TPS04, TAT3* and AT1G62660 (see: **Chapter2_additional_file1.pdf: 'Wounding time-series experiment'** and **Chapter2_additional_file2.xlsx: 'Wounding_GOterms'**). This indicates that it is highly desirable to cut and snap freeze sample material within 10 minutes to avoid perturbation of results – a time window in which I find that sample dissection is easily achievable.

Spatially resolved transcriptomics data reveals leaf tissue specific gene expression

I assayed ST-seq's ability to detect known gene expression differences between tissue types in untreated leaves. Briefly, I dissected a lateral cross-section of an *A. thaliana* leaf (3 biological replicates) into a 1-dimensional (1D) expression map of eight circa 1 mm^2 squares (**Figure 3A and Table 3**). Each cross-section was sampled according to the same pattern: the leaf margins were located at square-1 and square-8 and the midvein at square-5. I then identified DE genes by comparing the midvein with the lamina and the leaf margins with the lamina. This resulted in 393 DE genes for the midvein and 686 DE genes for the leaf margins comparison (**Figure 3** and **Chapter2_additional_file2.xlsx: 'Leaf-untreated_spatial-DEgenes'**).

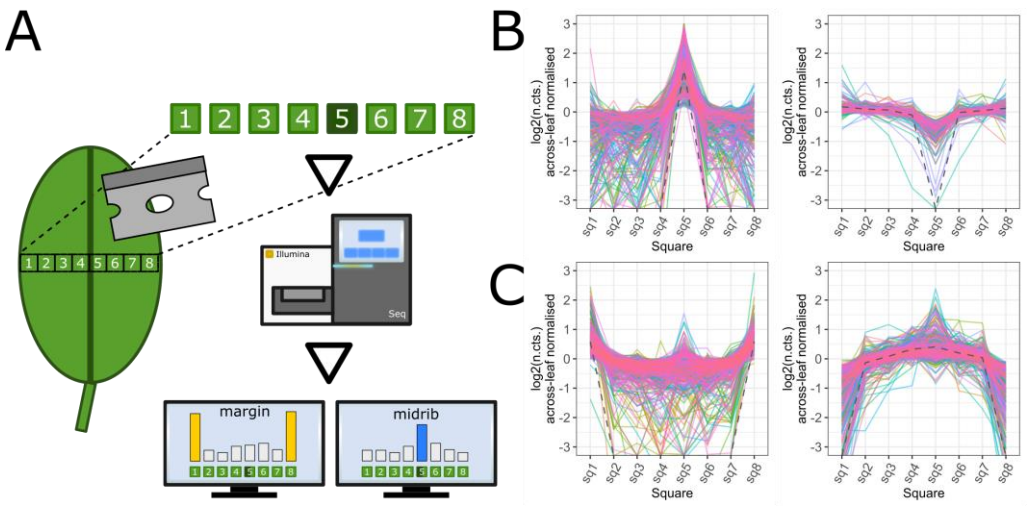

Figure 3 Identification of midvein and edge DE genes in a lateral 1D leaf cross-section: (A) DE-gene analysis of a 1D *A. thaliana* leaf lateral cross sections by comparing the midvein (square-5) or the margin squares (square-1 and square-8) with the 'bulk' (remaining) leaf sections. The two images in (B) show 393 DE genes with higher (left, 256 DE genes) or lower (right, 137 DE genes) expression values in the midvein. The images in (C) show 686 DE genes higher (left, 403 DE genes) or lower (right, 283 DE genes) expression in the leaf margins. The grey dashed line in each plot (B and C) represents a trend-line for the average log2(normalised counts) of all genes normalised across the leaf squares.

Comparison of spatial and bulk transcriptomics after localised flg22 stimulation

To compare spatial (only treated areas) with bulk (large leaf areas with treated and untreated areas) MAMP immune responses I used a flg-22 syringe infiltration assay. For this I produced small, local infiltration spots on the abaxial, left-hand side of a leaf of 6 biological *A. thaliana* replicates using either 500 nM flg22 or water (**Figure 4A and Table 3**). I incubated the plants for 1 hour and sampled by dissecting leaf samples with a 1D system as above, briefly: square-1 (in the middle of the left half of a leaf) as infiltration spot and then laterally towards the midvein square-2 as non-vascular leaf tissue, square-3 as the midvein and square-4 as non-vascular leaf tissue.

In my analysis I wanted to measure how bulk RNA-seq datasets compared to spatially collected ones by using the number of detectable DE genes. I

hypothesised that the spatial analysis of the small, local treatment spot (square-1) and its surroundings (square-2, square-3 and square-4) would reveal more and distinct types (or waves) of flg22 responsive DE genes than a bulk analysis would – especially of rarer transcripts. To measure the effect of spatial information alone I simulated an *in silico* flg22 bulk experiment by combining the data from flg22 or water treated square-1 with the other untreated squares-2, 3 and 4 of the same leaf. I then called the treatment responsive DE genes from the bulk files, detecting 65 DE genes (39 higher expressed, 26 lower expressed) 1 hour after flg22 infiltration. I detected 887 more DE genes (952 in total) by comparing the single squares of the flg22 and water infiltration dataset: 646 DE genes for square-1 (416 higher, 230 lower expressed), 401 DE genes for square-2 (306 higher, 95 lower expressed), 9 DE genes for square-3 (8 higher, 1 lower expressed) and any DE genes for square-4 (see: **Chapter2_additional_file2.xlsx: 'Infilt_spatial_vs_bulk_DEgenes'** and **'Infilt_spatial_vs_bulk_DEgenes'**). In contrast comparing the gene lists of the *in-silico* bulk and spatial analysis I detected that 4 DE genes were exclusively called from the bulk dataset and 64 genes were shared by both datasets.

To identify the biological processes uncovered by my transcriptomics experiments I performed a GO-term enrichment analysis on the spatial flg22 related DE-gene datasets. From all (952) DE genes I obtained 168 enriched GO-terms. Among them I observed many biological processes, which grouped under the GO parent terms: 'response to organonitrogen compound', 'jasmonic acid metabolism', 'regulation of reactive oxygen species metabolism', 'respiratory burst', etc. (**Figure 4B**) and so can be associated with stress and defence responses.

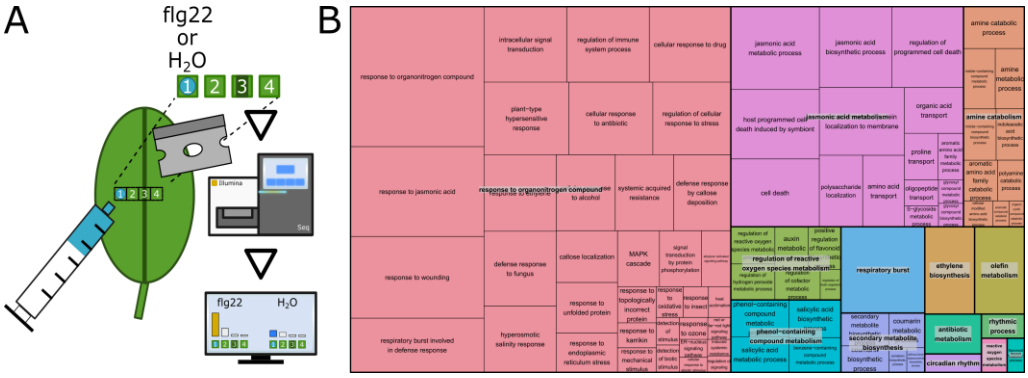

Figure 4 Elicitation of early *A. thaliana* defence response genes by infiltrating the bacterial peptide flg22: (A) To provide a strong stress stimulus I used syringe infiltration of either 500 nM flg22 or, as a control, water on a small area of the abaxial side of a leaf (square-1). 1 hour after infiltration I dissected 4 squares of a lateral leaf section with square-1 being the infiltration spot, square-2 and square-4 untreated, non-vascular leaf tissue and square-3 as midvein. The figure (B) shows the REVIGO **[136]** treemap of the detected 168 GO-terms grouped under parent terms such as: 'response to organonitrogen compound', 'jasmonic acid metabolism', 'amine catabolism', 'regulation of reactive oxygen species metabolism', 'phenol-containing compound metabolism', 'respiratory burst', 'secondary metabolite synthesis', 'ethylene biosynthesis', 'oleofin metabolism', 'antibiotic metabolism', 'circadian rhythm', 'reactive oxygen metabolism' and 'flavonoid metabolism'. The size of each rectangle relates to the absolute log10(q-value) – the larger the more significant.

Early elicited flg22 response genes of local, fine-scale stimulation

To study an initial pathogen encounter I analysed data from a milder stimulus method than the above described syringe infiltration: 6 biological *A. thaliana* replicates were exposed to deposition of 1 µl of 500 nM flg22 on square-3 of the abaxial side of a leaf and 1 µl of water (internal control) on square-6 (equivalent locations due to leaf bilateral symmetry). After one hour the treated leaf area was extracted as a 1D lateral cross-section containing 8 separate 1 mm^2 squares (**Figure 5A** and **Table 3**). I was interested in DE genes at the site of flg22 spotting (square-3) and in adjacent sections (square-2 and square-4) as I reasoned that the plant would respond to the MAMP locally at first and then responses via signalling to adjacent tissues and the rest of the plant. I called DE genes by comparing the flg22 with the water droplet spots (square-3 vs square-6) and the adjacent

sections with their corresponding bilateral equivalents (square-2 vs square-7 and square-4 vs square-5). Due to the milder stimulus in comparison to the flg22 infiltration dataset I expected the number of DE genes could be lower than in the syringe infiltration experiment where I detected 952 DE genes. Indeed, I identified a lower number of 523 DE genes (491 higher expressed, 32 lower expressed) for the droplet spot, and 5 DE genes in the adjacent sections (1 higher expressed DE-gene in the square-4 square-5 comparison and 4 higher expressed DE genes in the square-2 vs square 7 comparison). Thus, in total I detected 526 individual DE genes.

I compared both droplet spotting and syringe infiltration datasets (each dataset was collected 1 hour after flg22 exposure) for biological processes using GO-term enrichment. Both experiments produced a similar number of enriched GO-terms with 159 biological processes enriched in the droplet spotting dataset and 168 biological processes enriched in the infiltration dataset, with an overlap of 132 biological processes (83.0 % of the spotting dataset and 78.5 % of the infiltration dataset) between both datasets (**Chapter2_additional_file2.xlsx: 'Spotting_DE genes'** and **'Spotting_GOterms'**). The percentage of shared, enriched GO-terms indicated the presence of a similar plant response to flg22 in both experiments despite the difference in stimulus strength.

I measured the DE genes with the 253 FLARE genes described by seedling and cell culture flg22 exposure experiment of Navarro et al. [134] (**Chapter2_additional_file2.xlsx: 'FLARE_Navarro_reference'**). I found that the DE genes of the droplet spotting experiment contained 24.90 % FLARE genes (63 of 253 genes). These consisted of 32 FLARE genes associated with signal transduction, 11 genes associated with roles in signal perception, 14 with known or putative roles as effector proteins and 9 FLARE genes identified by Navarro et al. as 'other' FLAREs (**Chapter2_additional_file2.xlsx: 'Infilt_spot_FLARE-overlap'**). I found a slightly higher number in the infiltration experiment: 80 DE genes were shared with the 253 FLARE genes (31.62 %) with 39 genes associated with signal transduction, 16 genes in signal perception, 15 genes with known or putative roles as effector proteins and 11 genes with other functions

(**Chapter2_additional_file2.xlsx: 'Infilt_spot_FLARE-overlap'**). Interestingly the percentage of known FLARE and the number of enriched biological processes was higher in the infiltration than the spotting experiment (possibly suggesting other processes are triggered by infiltration).

I was interested in the spatial expression patterns of the 63 shared FLARE genes between my droplet spotting experiment and the Navarro et al. dataset. For this I visualised the expression patterns of the FLAREs across the studied leaf area. All 63 FLARE genes showed high expression levels at the area of flg22 exposure in comparison to adjacent leaf squares (**Figure 6B**). To study the expression profiles of the remaining 460 DE genes identified in the flg22 droplet spotting experiment, I affinity propagation clustered [137] these DE genes based on their spatial expression patterns and visualised the expression profile of each cluster (**Figure 5C**). I identified three gene clusters: two of the three clusters contained genes with higher expression levels at or adjacent to the area of flg22 treatment and one cluster contained a group of lower expressed genes at the flg22 treated area (**Figure 5C**).

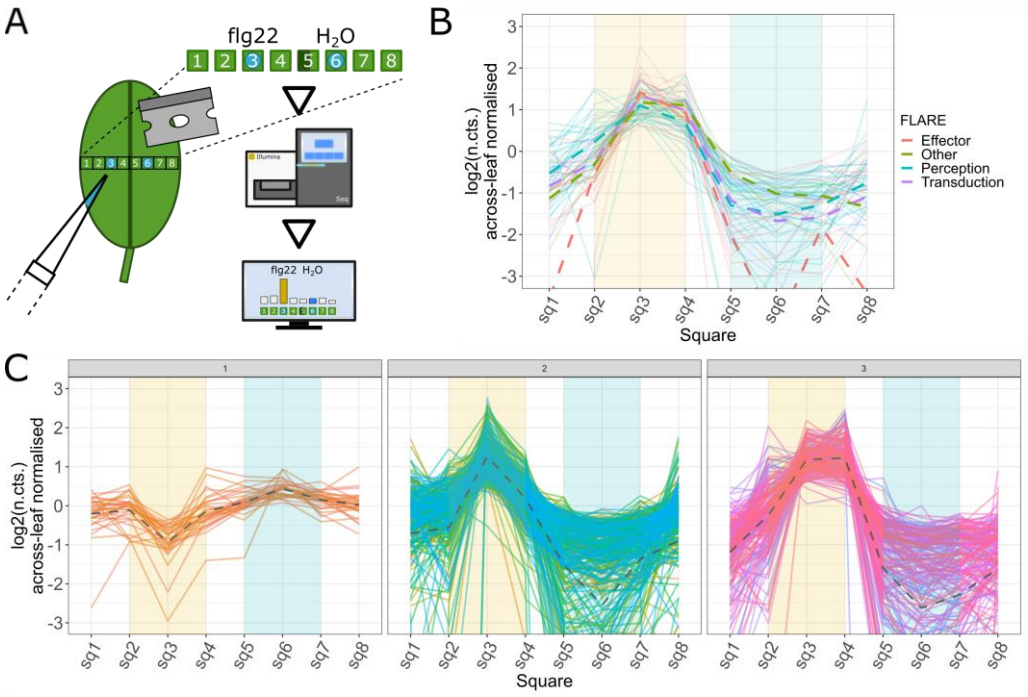

Figure 5 Elicitation of early *A. thaliana* defence response genes by droplet depositing the bacterial peptide flg22: (A) As a milder stress stimulus than flg22 syringe infiltration a 1 µl droplet of 500 nM flg22 and, as an internal control, water was pipetted on the abaxial surface of a leaf. 1 hour after droplet deposition a lateral section was dissected into 8 squares with square-1 and square-8 as leaf margins, square-3 as flg22 treated spot, square-6 as water treated spot and square-5 as midvein. Image (B) shows an overlay of the spatial expression patterns of the 63 FLARE genes characterised by Navarro et al. **[134]** present in our dataset. Each group is coloured separately, the average expression of each FLARE group is shown as the dashed line. Image (C) shows the spatial expression of all 523 detected DE genes grouped in three different clusters. From left to right: One cluster (1) contains genes which are lower expressed at the flg22 treatment area, two clusters contain genes with higher expression at the flg22 treatment spot in comparison to adjacent areas but with narrower (2) and broader (3) spatial expression. The yellow background in the plot indicates the flg22 treated area, the blue background indicates the water treated control area.

I analysed all clusters for enriched biological processes using GO-term analysis. I could not detect any enrichment in biological processes for the cluster containing the DE genes which were lower expressed at the flg22 site. The two clusters with expression peaks at the site of flg22 stimulation however enriched 135 and 122 biological processes. Of all biological processes 100 were shared between both clusters and 35 as well as 22 biological processes unique for each cluster respectively (**Chapter2_additional_file1.pdf: 'Early plant response of local, fine-scale flg22 stimulation'** and **Chapter2_additional_file2.xlsx: 'Spotting_DE genes_3-clusters'**). This suggests that biological processes (host responses) could be associated with the spatial expression profiles of their corresponding genes. To study this further I used the affinity propagation clustering algorithm [137] to determine the number of clusters without a specified cluster number preference value. This grouped the 523 DE genes into 36 more tightly resolved spatial expression clusters, comprising 35 clusters with between three and 57 genes and a single cluster containing only one gene (**Figure 6A**). 28 of 35 clusters were enriched for biological processes (**Chapter2_additional_file3.csv**). To test if different spatial expression patterns enrich different biological processes, I correlated all multi-gene clusters based on the presence / absence of all enriched GO-terms. I saw little overlap in biological processes between clusters, indicating that each spatial expression cluster enriched slightly different GO-terms. (**Figure 6B** and **Chapter2_Additional_file3.csv**).

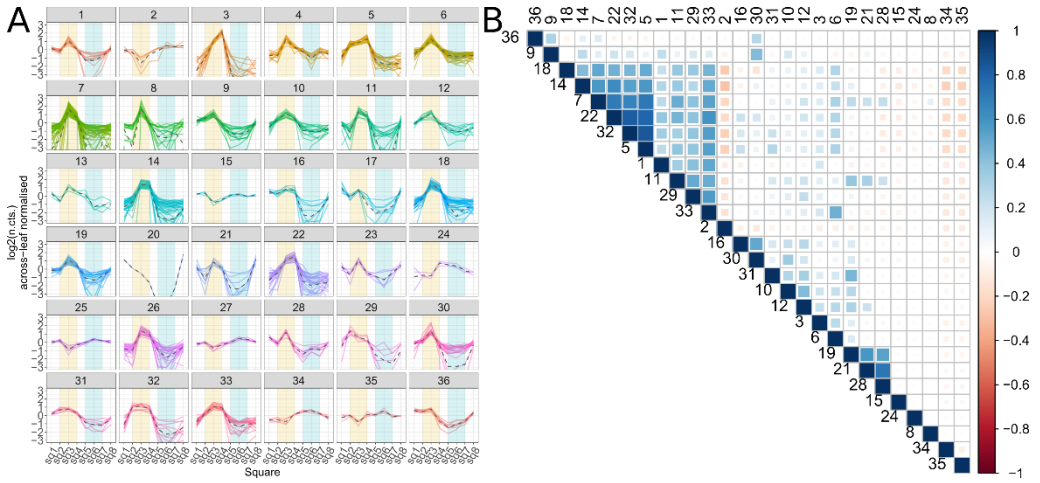

Figure 6 Unsupervised clustering of flg22 elicited DE genes and GO-term correlation matrix of the predicted clusters: (A) shows the expression profiles of the 523 flg22 elicited DE genes grouped to 36 clusters precisely clustered according to their spatial expression pattern across the tested leaf area. Many of the clusters show differences in their induction profile at the site of flg22 deposition (yellow background) but also differences in expression at the water treated area (blue background) or the expression at the leaf boarders. (B) shows the correlation analysis of the enriched GO-terms from the genes of the spatial clusters shown in (A) – 28 clusters grouped with hierarchical clustering for enriched GO-terms.

Characterisation of spatial regulatory elements

I characterised the expression patterns of the 36 obtained clusters in **Figure 6**. 11 clusters (1, 2, 6, 8, 12, 15, 23, 25, 28, 30, 35) showed a peak of higher expression at the site of flg22 perception. 14 clusters (3, 5, 7, 10, 11, 14, 18, 19, 21, 22, 26, 29, 32, 33) indicated spatially elevated gene expression patterns with higher expression also at sites adjacent to the area of flg22 perception. The remaining 11 clusters showed less clear expression profiles (**Chapter2_additional_file2.xlsx: 'Spotting_DE-genes-affinity-prop'**).

To identify plant regulatory elements that are potentially involved in MAMP perception and signal propagation to adjacent areas, I selected DE genes belonging to the flg22 locally and adjacently elevated clusters, and then filtered the genes for the TAIR-10 [138] GO-terms 'receptor' and 'transcription'. This included the leucine-rich repeat receptor like kinase (LRR-RLK) *RLK7* (cluster 1)

which was locally elevated, whereas the LRR-RLK *CERK1* and the serine/threonine-protein kinase *PSKR1* while strongly elevated at the area of flg22 perception were also broadly expressed throughout all sites (cluster 7). I detected a larger set of 48 DE genes associated with transcriptional processes. WRKY (15 genes, **Figure 7**), ERF (8 genes) and MYB (4 genes) transcription factor family members [139–141] were the most abundant in my dataset.

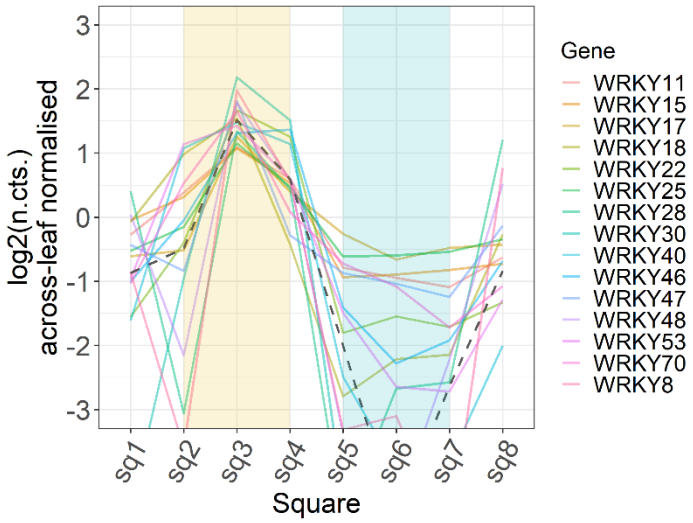

Figure 7 Spatial expression profile of all 15 detected DE WRKY transcription factors in my dataset: All detected DE WRKY transcription factors were higher expressed at the area of flg22 droplet spotting (yellow background) in comparison to the water exposed control area (blue background). The majority of WRKY transcription factors shows spatial expression profiles which indicate elevated expression in areas adjacent to the droplet deposition spot sq3 (i.e. area sq1, sq2 and sq4).

To start to understand the possible gene regulatory network controlling this spatial expression I used the TF2Network software [142] to search for putative regulatory interactions between these 48 DE genes i.e. by transcription factor binding. The resulting gene networks are built from genes with at least one target and a q-value < 0.01 (**Chapter2_additional_file1.pdf: 'Characterisation of spatial regulatory elements'**). These linked 4 transcription factors of which all belonged to the WRKY family (WRKY11, WRKY15, WRKY17 and WRKY47*)* to 388 other DE genes indicating a possible regulatory network (TF2Network authors suggest their tools has a very low false positive rate, whilst being sensitive enough to detect 75-92% of correct links). Of the detected transcription factors *WRKY17* (cluster 4) and

WRKY47 (cluster 8) were associated with local expression patterns, whereas *WRKY11* (cluster 10) and *WRKY15* (cluster 10) showed spatially wider expression.

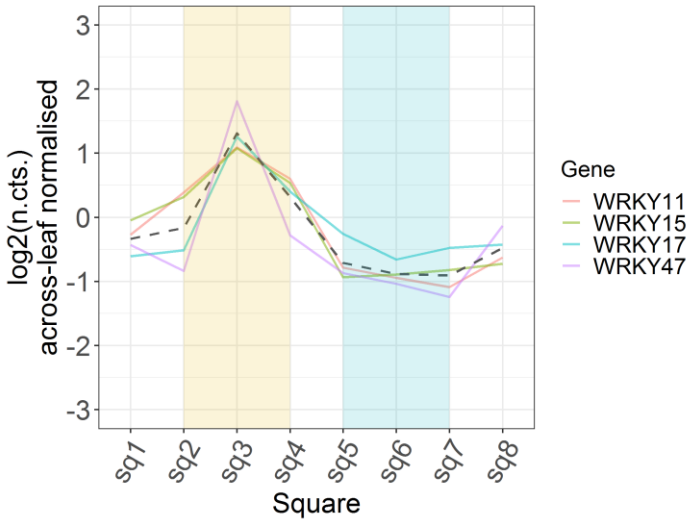

Figure 8 Spatial expression profile of WRKY11, WRKY15, WRKY17 and WRKY47: In my analysis the four transcription factors were associated with a potential role in contributing to the control of spatial expression. All four transcription factors are higher expressed at the area of flg22 exposure (yellow background), whereas at the area of water exposure (blue background) gene expression is not peaking. Especially WRKY11 and WRKY15 show a wider expression profile which is also elevated in adjacent squares to the sq3 flg22 droplet spot.

Discussion

The ability to profile gene expression patterns in small specific areas without bulk sequencing provides access to lower level transcripts, especially tissue and cell specific ones [3, 16, 20, 143–145]. Spatial, low RNA-input transcriptomics methods allow deeper insights in how an organism develops and reacts to its environment than conventional 'gross-scale ' RNA-seq methods [15, 16, 20]. By combining rapid dissection with ST-seq, I was able to reconstruct spatial transcriptional differences across organs and localised defence responses.

Although some specialised protocols are already available to profile transcriptomes from minute input amounts such as single-cells [7, 129], or even nuclei [130, 131], these detailed techniques do not retain the spatial information

of a starting tissue and any time-consuming experimental procedures to preserve spatial data could induce experimental bias by altering the transcriptome. For spatial analyses in plants LCM [25, 132] methods for fine-scale transcriptome analyses are available, however, these procedures are time-consuming and this limits the scale of the application. Large scale spatial analysis have been performed in the past [38], but still required bulk sampling of material by pooling multiple replicates. Methods relying on sample dissection and reaction-tube processing of tissue sections to sequencing libraries have already proven to be able to identify transcripts patterns in the zebrafish embryo [20] allowing to process multiple samples easily for modelling the transcriptome landscape of an entire organism. Recently Giacomello et al. [15] published a workflow to blot the transcriptome landscape from permeabilised plant tissues by vertical diffusion onto a slide containing an array of barcoded primers and on slide library construction which maintained the mRNA's location via the barcode. This method, for the first time in plants, allowed access to spatial transcriptome data in thin tissue slices of plant organs with a great level of resolution. However, gathering and optimisation of permeabilization conditions of thin tissue sections can be challenging especially if the tissue due to the volume and shape of the sample, are not suitable to be processed on an array and the workflow is only commercially accessible.

Plants grow in a microbiologically rich environment with their own microbiomes and even symbionts [146], but they are also attacked by pathogens, pests, herbivores and other biotic stresses [147]. As plants can't move away from attacks they defend themselves using molecular and cellular biology responses, however overstimulation of these processes leads to stunted development and lower fitness [55]. As plants must balance the need to defend themselves against constant plant-microbial interactions and attack [55] I hypothesised that local attacks might be integrated into a plant-wide defence response decision. Yet there are no appropriate assays to measure the molecular and cell biology changes at the required resolution. Here, I demonstrate a novel millimetre-scale method to pursue spatial transcriptomics experiments in plants in an easy manner based on

easily accessible methods in sequencing library construction and bioinformatics tools.

I developed a robust micro-spatial expression methodology that enables the creation of transcriptome level maps from very small amounts of any eukaryotic tissue. Key advantages of the method are (a) a 96-well format mRNA extraction protocol to rapidly lyse and extract mRNA from small leaf areas, (b) an optimised SMART [148] based reverse transcription protocol to generate full length ds-cDNA from leaf mRNA and (c) a cost optimised Illumina Nextera reaction to enzymatically fragment ds-cDNA to Illumina libraries. The ST-seq workflow evolved by transferring elements from existing single-cell RNA-seq methods [149, 150] from animal systems to plants and refining these methods for stable, low-cost generation of sequencing libraries from small amounts of RNA starting material. This allows the design of experiments in which spatial information is required but only small pieces of tissue can be obtained. In the process of method development, I tested and included several features to efficiently generate double stranded cDNA (ds-cDNA) with reduced PCR amplification in both ds-cDNA synthesis and subsequent amplification after Nextera tagmentation. I introduced sample specific barcodes in the Nextera amplification step to allow pooling of 2,304 of samples per sequencing run. This optimisation altogether allowed me to construct sequencing libraries by hand for just £ 6.00 per library (**Table 2** and **Chapter2_additional_file2.xlsx: 'Costs'**) in comparison to £ 65.56 for an Illumina TruSeq library (RS-122-2001, Illumina) or £ 62.60 for a SMARTer PCR cDNA Synthesis Kit library (634926, TaKaRa).

In my benchmarking experiments I compared the ST-seq workflow with the widely used Illumina TruSeq sequencing protocol and show that the ST-seq method compares well with this common commercial RNA-seq protocol; (see: **Chapter2_additional_file1.pdf: 'Comparison of RNA sequencing library methods'**). I also show that ST-seq can detect transcript level differences across 1D leaf sections in distinct leaf elements such as leaf margins or vascular tissues and that spatial mapping of transcript levels to specific sections of leaves is possible, which allows drawing of transcriptional expression profiles across

tissues. This easily used, low cost protocol makes feasible experiments that require spatial transcriptome analysis.

To apply the ST-seq method for studying biotic actions I challenged *A. thaliana* leaves with the bacterial peptide flg22, a conserved 22 amino acid sequence of the bacteria flagellin protein, which to the plant indicates an encounter with potentially pathogenic bacteria [151]. Plants recognize such potential threats as the pathogenic cell surface molecules, so called MAMPs, perceived by the plant Pattern Recognition Receptors (PRRs) on the plant cell surface [54]. This event initiates an intracellular plant signalling cascade leading eventually to immunity or disease [55, 76]. In my experiments I could detect the triggering of immune and defence response related biological processes and show that the results obtained by RNA-seq are independently reproducible using qRT-PCR (**Chapter2_additional_file1.pdf: 'Validation of flg22 induced local plant response genes using qRT-PCR'**). I was able to find overlap in my data with already described flg22 elicited (FLARE) genes from a gross-scale experiment using a strong stimulus [134]. In comparative analyses of my dataset with the spatial expression patterns of the described FLARE genes I was able to identify genes which share similar spatial expression and are potential novel FLARE genes. Cluster based analysis of spatial expression data revealed sets of genes with highly similar expression profiles enriched in distinct biological processes; including FLARE genes to which I add new and increased expression resolution. Characterisation of spatial cluster expression profiles highlighted plant regulatory elements with local or spatially elevated expression levels and so potential short distance signal propagators upon flg22 stimulus.

Material and Methods

Plant growing conditions

I used 4 – 6 week old *A. thaliana* Col-0 plants that were grown in a controlled environment room with an 8 hours light, 16-hour dark cycle at a constant temperature of 22 °C and 70 % humidity.

Flg22 exposure experiments

Before flg22 treatment experiments, I transferred the plants from the controlled environment room to a laboratory working bench (room with constant light exposure and temperature). To elicit plant responses with flg22 I either syringe infiltrated [152] the peptide or spotted a droplet of flg22 on a leaf using a pipette.

To produce small, local infiltration spots I used a 1 ml syringe (BS01T, R&L Slaughter Ltd, Basildon, UK) loaded with 500 nM flg22 peptide solution. By application of mild pressure on the plunger of the syringe when infiltrating I produced an approximately 2 - 3 mm diameter infiltration spot on the left-hand side of a leaf. In parallel to flg22 infiltration I produced an infiltration series with DNase/RNase-free water as control. The plants were subsequently incubated on the laboratory working bench for 1 hour until sampling.

For the flg22 spotting experiment 1 µl droplet of 500 nM flg22 was loaded on the abaxial surface of a leaf using a pipette (diameter approximately 1 mm). The flg22 was pipetted onto the left half of the leaf and a 1 µl droplet of the water control droplet spotted on the right half of the leaf. After spotting the plants were incubated for 1 hour on the laboratory bench before sampling. The concentration of flg22 was used as described by Zhang et al. [153].

Leaf sectioning and sample harvesting

I used single margin razor blades (T586, Agar Scientific Ltd., Stansted, UK) to cut leaves into approximately 1 mm^2 small leaf squares (**Figure 9**). To create a clean surface for cutting I used the pealed, non-sticky paper cover of a 96-well plate seal (AB0580, Thermo Fisher Scientific, Waltham, USA). With a previously in RNaseZAP (AM9780, Thermo Fisher Scientific, Waltham, USA) washed and air-dried forceps (T083, TAAB Laboratories Equipment Ltd, Berks, UK) I transferred each leaf square immediately after cutting into a well of a 96-well plate (E1403-0100-C, Starlab, Milton Keynes, UK) which I had pre-cooled on a 96-well metal block in dry ice (- 70 ^{o}C), or alternatively, a dry-ice cooled 1.5 ml tube (10051232, Fisher Scientific,

Loughborough, UK). The sample wells of 96-well plates were sealed using domed PCR cap strips (AB0602, Thermo Fisher Scientific, Waltham, USA). Post harvesting the samples were stored at -80 ^{o}C until use.

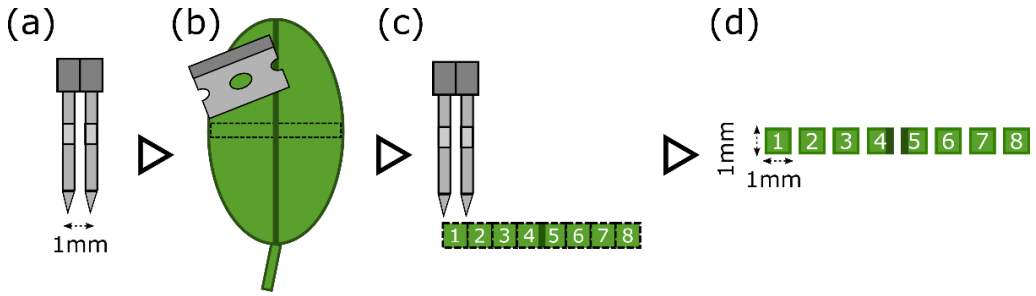

Figure 9 Cutting technique to obtain 1 mm^2 leaf squares: To dissect leaf strips at a width of 1 mm (a) two single margin razor blades were held together and (b) a cross section dissected from a leaf. With a fresh pair of razor blades (or alternatively a ruler), the so obtained leaf strip is dissected into small 1 mm^2 areas, which immediately upon dissection are transferred to a dry ice cooled 96-well plate or 1.5 ml tube.

Leaf sample lysis and preparation for mRNA extraction

To lyse the leaf samples stored in 1.5 ml tubes I first added 10 µl lysis buffer composed of 100 mM Tris-HCl pH 7.5 (BP1757, Fisher Scientific, Loughborough, UK), 500 mM LiCl (L7026, Sigma Aldrich, St. Louis, USA), 10 mM EDTA pH 8.0 (E7889, Sigma Aldrich, St. Louis, USA), 1 % LiDS (L4632, Sigma Aldrich, St. Louis, USA), 5 mM DTT (18064014, Thermo Fisher Scientific, Waltham, USA) to each sample immediately after removing the sample tube from the cold storage.

I subsequently ground the leaf sections in lysis buffer using polypropylene pestles (Z359947, Sigma Aldrich, St. Louis, USA), which, before use, were washed with RNaseZAP (R2020, Sigma Aldrich, St. Louis, USA, three times with 80% ethanol (32221, Sigma Aldrich, St. Louis, USA) rinsed with UltraPure DNase/RNase-free distilled Water (10977049, Thermo Fisher Scientific, Waltham, USA) and air-dried after washing. After sample lysis I transferred the lysate to an ice-cooled 96-well plate and continued with the mRNA extraction.

Samples stored in 96-well plates were lysed by using 1 mm diameter grade 1000 hardened 1010 carbon steel ball bearings (Simply Bearings Ltd, Leigh, UK). For this,

before use of the ball bearings, I treated a bulk batch sequentially with RNaseZAP and DNA AWAY, after this washed the ball bearings three times with 80% ethanol and transferred them to sterile screw-cap 2.0 ml tubes (E1420-2341, Starlab, Milton Keynes, UK) and heat dried with a slightly loosened lid on a 95 ^{o}C heating block (N2400-4001, Starlab, Milton Keynes, UK).

To lyse the collected leaf samples stored in a 96-well plate, I transferred the 96-well plate to a dry ice temperature cooled 96-well metal block. I carefully opened the domed PCR cap lids to avoid sample spillage and added 4 – 6 (room temperature) ball bearings to each sample well. After this I transferred 10 µl lysis buffer to each well and re-sealed the plate with new domed PCR cap lids, and immediately proceeded to the 2010 Geno/Grinder (SPEX SamplePrep, Stanmore, UK) disrupting the samples for 30 seconds at 1,750 rpm. I gathered the sampled using a centrifuge (Centrifuge 5910 R, Eppendorf UK Ltd, Stevenage, UK) for 10 seconds at 2,000 rcf. A strongly green-coloured solution without any remaining solid leaf material indicated good sample lysis. If satisfactory sample lysis was not achieved, I disrupted the samples again for another 10 seconds on the 2010 Geno/Grinder at 1,750 rpm and centrifuged for 30 seconds at 2,000 rcf. I immediately transferred the lysis solutions into a new 96-well plate using a 10 µl multichannel pipette. After transfer of the lysis solutions, I stored the new 96-well plate on ice, discarded the 96-well plate containing the ball bearings and proceeded immediately with mRNA extraction.

Leaf mRNA purification

The leaf tissue mRNA was purified using 1 µl NEBNext Poly(A) mRNA Magnetic Isolation Module oligo-dT(25) beads (E7490, New England Biolabs Ltd, Hitchin, UK) per extraction. Before the extraction the required volume of oligo-dT(25) magnetic beads was washed twice in 200 µl lysis buffer on a DynaMag-2 Magnet rack (12321D, Thermo Fisher Scientific, Waltham, USA) and resuspended in 10 µl lysis buffer for each 1 µl oligo-dT(25) beads input volume. The beads were mixed by a quick vortex and 10 µl of the resuspended beads were transferred to each well of the 96-well plate containing the lysis solutions. The wells were sealed with

domed PCR cap strips, the 96-well plate vortexed briefly and attached to a tube rotator (444-0502, VWR International Ltd, Luterworth, UK) with adhesive tape. After 10 minutes rotation on room temperature I collected the lysis solution at the bottom of the wells by spinning the plate for 10 seconds at 2,000 rcf and pelleted the oligo-dT(25) magnetic beads on a 96-ring magnetic plate (A001219, Alpaqua, Beverly, USA). Using a multichannel pipette I washed the oligo-dT(25) magnetic beads twice with 50 µl Wash Buffer A (10mM Tris-HCl pH7.5, 0.15M LiCl, 1mM EDTA, 0.1% LiDS) and once with Wash Buffer B (10mM Tris-HCl pH 7.5, 0.15M LiCl, 1mM EDTA). After washing I centrifuged the plate for 10 seconds at 2,000 x rcf to collect the remaining Wash Buffer B at the bottom of the tube, pelleted the oligo-dT(25) magnetic beads on a magnet and removed the remaining volume of Wash Buffer B with a multichannel pipette. The oligo-dT(25) beads were resuspended immediately in 8 µl DNase/RNase-free water, incubated for 2 minutes at 80 ^{o}C on a G-Storm GS1 thermal cycler (G-Storm, Somerton, UK), then immediately pelleted on a 96-ring magnetic plate to elute the mRNA off and separate from the oligo-dT(25) beads. The solutions containing the purified mRNA were immediately transferred to a new 96-well plate, which was placed in a - 80 ^{o}C freezer until needed. At this step the mRNA is not quality controlled, this is performed after the ds-cDNA synthesis step.

Double-stranded cDNA synthesis reaction

For ds-cDNA synthesis I used a protocol based on the template switching mechanism of the reverse transcriptase enzymes [148]. Briefly: 2.50 µl extracted mRNA was mixed (on ice) with 2 µl 5x First Strand buffer (18064014, Thermo Fisher Scientific, Waltham, USA), 1 µl 10 mM dNTPs (10297018, Thermo Fisher Scientific, Waltham, USA), 1 µl 5'-biotinylated 10 µM STRT-V3-T30-VN oligonucleotide: 5'-/5Biosg/TTAAGCAGTGGTATCAACGCAGAGTCGACTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3' (Integrated DNA Technologies, Leuven, BE), 1 µl 20 mM DTT (18064014, Thermo Fisher Scientific, Waltham, USA), 0.10 µl 40 U/µl RNase Inhibitor (M0314S, New England Biolabs Ltd, Hitchin, UK), 0.25 µl 10 µM template switching oligo 5'-

AAGCAGTGGTATCAACGCAGAGTGCAGUGCUTGATGATGGrGrGrG-3' (Integrated DNA Technologies, Leuven, BE), 0.30 µl 200 U/µl SuperScript II Reverse Transcriptase (18064014, Thermo Fisher Scientific, Waltham, USA), 0.30 µl 100 µM $MnCl_2$ (M1787, Sigma Aldrich, St. Louis, USA) and 1.55 µl DNase/RNase-free water to a total reaction volume of 10 µl. The reverse transcription reaction was run in a G-Storm GS1 thermal cycler for 90 minutes at 42 ^{o}C with additional 10 minutes at 72 ^{o}C to inactivate the reverse transcriptase. After reverse transcription I immediately added 2 µl RNase H (M0297S, New England Biolabs Ltd, Hitchin, UK) diluted to 0.5 U/µl (5 U/µl stock concentration) to the reaction and incubated the reaction in the GS1 thermal cycler for 30 minutes at 37 ^{o}C. The RNase H treated reactions were purified using a 0.83x (10 µl) AMPure XP bead ratio (Beckman Coulter, High Wycombe, UK) and eluted in 18 µl 1x TE buffer. After this step I added 5 µl 5x Kapa HiFi PCR buffer (KK2102, KAPA BioSystems, Wilmington, USA), 0.75 µl 10 mM dNTPs, 0.75 µl 10 µM PCR+G primer 5'-GAAGCAGTGGTATCAACGCAGAGT-3' (Integrated DNA Technologies, Leuven, BE) and 0.50 µl 1 U/µl Kapa HiFi polymerase (KK2102, KAPA BioSystems, Wilmington, USA) to the cleaned ds-cDNA resulting in a total reaction volume of 25 µl and amplified the ds-cDNA in a G-Storm GS1 thermal cycler according to the following programme: (1) 3 minutes at 94 ^{o}C, (2) 17 cycles with 30 seconds at 94 ^{o}C, 30 seconds at 63 ^{o}C and 1 minute 30 seconds at 72 ^{o}C, (3) a final elongation step for 5 minutes at 72 ^{o}C. The amplified libraries were purified using a 1x (25 µl) AMPure XP bead ratio and eluted in 20 µl 1x TE buffer. The ds-cDNA libraries could be stored at this point in a -20 ^{o}C freezer. Before continuing with Illumina sequencing library preparation, I measured the ds-cDNA library concentrations with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA) dsDNA HS Assay Kit reagents (Q32854, Thermo Fisher Scientific, Waltham, USA) and also assessed the size distributions of randomly picked libraries on an Agilent Bioanalyser (G2939BA, Agilent Technologies, Stockport, UK) using the Agilent High Sensitivity DNA Kit (5067-4626, Agilent Technologies, Stockport, UK).

At later stages I modified the ds-cDNA synthesis integrating elements of the Smart-seq2 protocol [154]. The reverse transcription reactions of ST-seq-1.0 (as

described above) and Smart-seq2 were already highly similar, but Smart-seq2 had proven to require less hands-on time than the ST-seq-1.0 reverse transcription workflow. ST-seq-1.1 integrates elements of the Smart-seq2 library preparation workflow with minor modifications: Smart-seq2 uses single-cells sorted into a 2 % v/v Triton-X100 buffer as reverse transcription template. Instead of single-cells I supply previously extracted mRNA in DNase/RNase-free water to the reaction. Smart-seq2 further uses Illumina Nextera XT kit reagents in half-volume reactions at a cost of £ 14.52 per sample. With an additional reduction of the reaction volumes and optimisation of the reaction conditions (see below) I reduced the costs of this step to £ 2.12 per reaction.

ST-seq-1.1 ds-cDNA synthesis is performed as follows (**Figure 10**):2.50 µl extracted mRNA were combined with 1 µl 10 µM Smart-seq2 Oligo-dT30VN (5'-AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3', Integrated DNA Technologies, Leuven, BE) and 1 µl 10 mM dNTPs to a total volume of 4.5 µl (on ice). To anneal the Smart-seq2 Oligo-dT30VN I incubated the library for 30 seconds at 72 ^{o}C and snap-cooled the mixture on ice. The reverse transcription was conducted by adding the following reagents (while keeping the reaction plate on ice) to the reaction with a final reaction volume of 10 µl: 2 µl 5x First Strand buffer, 2 µl 5 M betaine (B0300, Sigma Aldrich, St. Louis, USA), 0.06 µl 1 M $MgCl_2$ (AM9530G, Thermo Fisher Scientific, Waltham, USA), 0.5 µl 100 mM DTT, 0.25 µl 40 U/µl RNase Inhibitor (2313A, Takara Clontech, Mountain View, USA), 0.10 µl 10 µM Smart-seq2 template switching oligo (5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3', Exiqon, Vedbaek, DK), 0.50 µl 200 U/µl SuperScript II Reverse Transcriptase, and 0.09 µl DNase/RNase-free water. I performed the reverse transcription reaction for (1) 90 minutes at 42 ^{o}C, (2) 15 cycles with 2 minutes at 50 ^{o}C and 2 minutes at 42 ^{o}C and finally (3) 15 minutes at 70 ^{o}C. After reverse transcription I added 12.50 µl 2x Kapa HiFi HotStart ReadyMix (KK2601, KAPA BioSystems, Wilmington, USA), 0.25 µl 10 µM Smart-seq2 IS-PCR primers (5'-AAGCAGTGGTATCAACGCAGAGT-3', Integrated DNA Technologies, Leuven, BE) and 2.25 µl DNase/RNase-free water to the reaction resulting in a total volume of 15 µl per reaction. Amplification was performed in a

G-Storm GS1 cycler according to the programme: (1) 3 minutes at 98 °C, (2) 15 cycles with 20 seconds at 98 °C, 15 seconds at 67 °C and 6 minutes at 72 °C and a (3) final elongation step for 5 minutes at 72 °C. The PCR reactions were purified using a 0.65x (9.75 µl) AMPure XP clean-up and eluted in 20 µl 1x TE buffer. After clean-up I measured the ds-cDNA library concentrations with the Qubit 2.0 Fluorometer dsDNA HS Assay Kit reagents (yields between 10 ng/µl and 20 ng/µl per ds-cDNA amplification reaction can be expected) and loaded randomly selected libraries on the Agilent Bioanalyser using the Agilent High Sensitivity DNA Kit before continuing with Illumina sequencing library preparation (**Figure 11**).

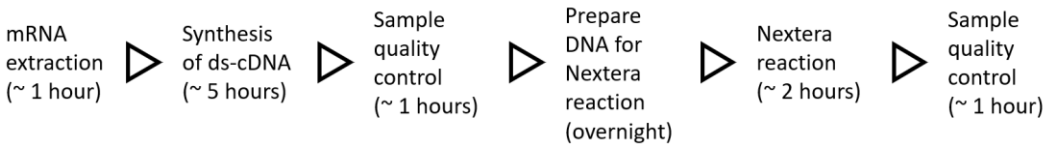

Figure 10 Diagram of ST-seq 1.1 working steps and timings: Starting from leaf squares the protocol can be conducted in approximately two days, performing the mRNA extraction and synthesis of ds-cDNA steps on day one, preparing the DNA for the Nextera reaction overnight and performing the Nextera reaction and Nextera reaction quality control on the second day.

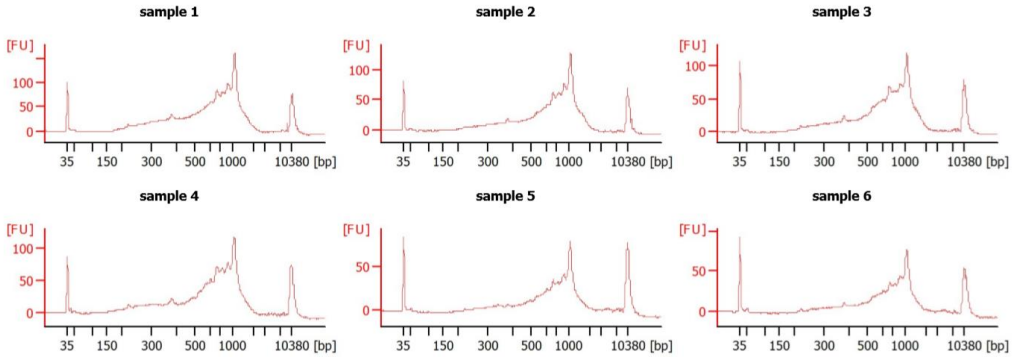

Figure 11 Double stranded cDNA traces of six different leaf areas amplified using the ST-seq-1.1 workflow: Sample 1 – 6 show High Sensitivity Bioanalyser ds-cDNA traces amplified using the ST-seq-1.1 workflow. The peaks at ~ 35 and ~ 13,380 bp are ladder peaks spiked into the reaction to calculate smear sizes. Successful amplification of ds-cDNA is indicated by a smear from 150 bp upwards peaking at ~ 1.0 – 1.5 kb, with few small amplified fragments and without amplified primer dimers. Similar to the data presented by Picelli et al. in [149].

Illumina library preparation from ds-cDNA

I prepared Illumina sequencing libraries using an Illumina Nextera (FC-121-1030, Illumina Cambridge, UK) based protocol with minor modifications: I exclusively used the Tagment DNA Enzyme 1 and the Tagment DNA Buffer and amplified the tagmented DNA with the Kapa 2G Robust Polymerase (KK5024, Sigma Aldrich, St. Louis, USA). I used custom Nextera barcodes that allow to multiplex hundreds of samples (**Chapter2_additional_file2.xlsx: 'Nextera_adapters'**) [155].

I reduced the costs of the library preparation by reducing the total tagmentation reaction volume to 5 µl (from 50 µl as recommended) with 1 ng ds-cDNA library input and using less enzyme. I performed a titration experiment of Tagment DNA Enzyme vs. 1 ng of selected ds-cDNA libraries aiming for Illumina sequencing libraries with a modal insert size distribution in the range of 400-500 bp (base pairs) with little short insert fragments and found that 0.1 µl Nextera enzyme was optimal.

The Nextera reactions were performed by combining 1 ng of ds-cDNA (air-dried over-night at room temperature in a drawer with the 96-well plate loosely covered to allow evaporation of liquid) with 2.5 µl 2 x Nextera buffer, 2.4 µl water and 0.1 µl Nextera enzyme on ice. The tagmentation plate was immediately transferred for 5 minutes at 55 ^{o}C on a G-Storm GS1 thermal cycler. Meanwhile I prepared a fresh 96-well plate with 2.0 µl 2.5 µM P5 and 2.0 µl 2.5 µM P7 custom multiplexing primers (**Chapter2_additional_file2.xlsx: 'Nextera_adapters'**). After tagmentation I transferred the tagmentation reactions to the previously prepared 96-well plate containing the sequencing adapters (see above) and added the following to each well: 5.00 µl 5 x Kapa 2G Robust Buffer, 0.50 µl 10 mM dNTPs, 0.10 µl 5 U / µl Kapa 2G Robust Polymerase, 10.4 µl water to a total final volume of 25 µl.

Amplification was performed on a GStorm GS-1 cycler using the following program: (1) 3 minutes at 72 ^{o}C, 1 minute at 95 ^{o}C (2) 11 cycles of 10 seconds at 95 ^{o}C, 30 seconds at 65 ^{o}C, 2 minutes 30 seconds at 72 ^{o}C (2) a final elongation step for 2 minutes 30 seconds at 72^{o}C. After amplification I purified the libraries

using a 0.64x ratio (16 µl) AMPure XP beads, eluted the libraries in 20 µl 1x TE buffer, measured the library yields with the Qubit 2.0 Fluorometer dsDNA HS Assay Kit reagents (yields between 5 ng/µl and 10 ng/µl per library can be expected) and assessed the size distributions of randomly selected libraries on the Agilent Bioanalyser with the Agilent High Sensitivity DNA Kit (**Figure 12**).

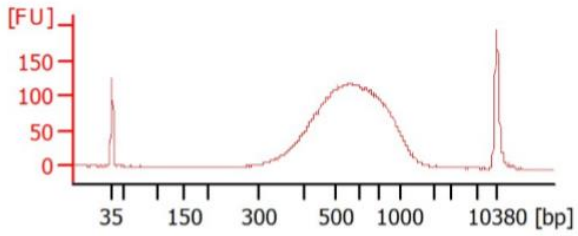

Figure 12 Illumina Nextera library pool constructed from double stranded cDNA: The image shows a High Sensitivity Bioanalyser electropherogram of a final Illumina Nextera library pool that was submitted for successful sequencing. The smear of the Illumina library starts at ~ 300bp and peaks between 500 – 600bp without containing primer dimers. Similar to the library smear presented by Picelli et al. in [149].

Sample pooling and sequencing

For sequencing, all library concentrations were determined using the Qubit 2.0 Fluorometer with dsDNA HS Assay kit reagents and pooled at equal molarity. The profile and concentration of the final library pool was assessed on the Agilent Bioanalyser using Agilent High Sensitivity DNA Kit Sequencing reagents. After this the pooled samples were shipped to the Earlham Institute for sequencing. Quality control and data demultiplexing was performed by the Earlham Institute Genomics Pipelines facilities. Samples were sequenced using Illumina HiSeq2500 50 base single-end rapid run sequencing for the *A. thaliana* wounding and *A. thaliana* flg22 infiltration datasets Illumina NextSeq500 75 base single-end for the *A. thaliana* untreated leaf dataset and Illumina HiSeq4000 150 base paired-end sequencing for the *A. thaliana* flg22 droplet spotting experiment.

Data quality control and mapping

The sequencing reads were quality controlled using FastQC-0.11.5 [156]. After quality control I used cutadapt-1.17 [157] to trim low-quality bases (-q 20) and remove Oligo-dT, template switching oligos, primer and Illumina Nextera library preparation sequences (-n 5 -e 0.05 –overlap 10). I also removed sequences with less than 40 bases (--minimum-length 40) and sequences containing N's (--max-n 0) from the dataset with cutadapt-1.17. After adapter and quality trimming, I re-assessed the reads a second time with FastQC-0.11.5. I mapped the reads to the *A. thaliana* TAIR10 release 37 genome assembly using STAR-2.5.1b [158] default settings and assessed mapping scores, duplication levels, GC-bias and gene-body coverage after mapping with RSeQC-2.6.4 [159]. Reads were counted with HTSeq-count-0.6.0 [160] default settings. To obtain a single quality control report as an overview for all samples, I aggregated the outputs of all used quality-control tools described above to a single report using MultiQC-1.7 [161].

Differential-expression analysis, GO-term enrichment

Differential expression analysis was performed using DESEq2-1.20.0 [162] in the statistical language R-3.5.1 using the workflow described by Love et al. [163] but by pre-filtering the dataset for rows with less than 10 rather than 1 raw read counts. DE genes were called with a q-value threshold < 0.05.

Across leaf DE-gene expression plots were prepared using R-3.5.1; in brief: I imported all samples with DESeq2-1.20.0 and calculated a table with normalised expression values as in the workflow described by Love et al. [163]. Next, I calculated the average expression value of each gene in each leaf square across all biological replicates. As a final step I normalised the expression values of the leaf squares. For this I divided the mean expression value of each leaf square of a gene with the mean expression value across all leaf squares of the same gene. The log2 transformed plots of the normalised data were generated using ggplot2-3.1.0 [164].

Affinity propagation clustering of the normalised expression tables was performed using the R-3.5.1 library apcluster-1.4.7 and a Pearson distance matrix of the normalised data for apclustK and apclust [137] default settings. The number of clusters was either empirically determined by continuously increasing the preferred cluster number in the apclustK function and visualising the expression profiles of the clusters using ggplot2-3.1.0 or determined without providing a cluster number preference value using the apclust function.

GO-term enrichment analysis on DE genes was performed using the R-3.5.1 Bioconductor library ClusterProfiler-3.8.1 [165] with the settings (Statistical test: Hypergeometric test, Multiple testing correction: Benjamini & Hochberg False Discovery Rate correction, False Discovery Rate cut-off: 0.01) and the Bioconductor library org.At.tair.db-3.6.0 as organism database [166].

ST-seq library preparation reagent cost calculations

Table 2 Reagent cost calculation for ST-seq: Reagent costs per ST-seq reaction add to approximately £ 6 per sequencing library. The consumable costs were determined according to list prices in September 2019. For a more detailed cost calculation see **Chapter2_additional_file2.xlsx: 'Costs'**.

Reagent	Supplier	Product code supplier	Cost per reaction [£]
Tris-HCl pH 7.5	Fisher Scientific	BP1757	0.046
LiCl	Sigma Aldrich	L7026	0.002
EDTA	Sigma Aldrich	E7889	0.0003
LiDS	Sigma Aldrich	L4632	0.002
Oligo-dT beads	New England Biolabs Ltd	E7490	0.128
Grinding beads	Simply Bearings Ltd	1 mm diameter grade 1000 hardened 1010 carbon steel ball bearings	0.003
Oligo-dT30VN	IDT	order with sequence	0.008
dNTPs	Thermo Fisher Scientific	10297018	0.074
Betaine	Sigma Aldrich	B0300	0.020
MgCl2	Thermo Fisher Scientific	AM9530G	0.000
RNAse inhibitor	2313A	Takara / Clontech	0.192
TSO exiqon	Exiqon	500100	0.092
Superscript	Thermo Fisher Scientific	18064014	1.928
HiFi Hotstart Readymix	Roche	7958935001	0.996
ISPCR primers	IDT	IDT custom order	0.0004
P5 primer	IDT	IDT custom order	0.003
P7 primer	IDT	IDT custom order	0.003
Nextera enzyme	Illumina	FC-121-1030	1.488
Kapa 2G robust	Sigma Aldrich	KK5024	0.236
Ampure XP beads	A63881	Beckman Coulter	0.784
		TOTAL:	6.00

Sequencing and sample pooling metrics

Table 3 Sequencing and sample pooling metrics for the wounding, flg22 (spotting and infiltration) and untreated control leaf experiments: The table lists the number of samples used for each experiment (separated by number of time-points, conditions and the number of used leaves), the number of extracted squares per leaf, the total number of squares processed to sequencing libraries, the number of combined samples for sequencing (per experiment) and the used Sequencing platform and chemistry (SE = single end, PE = paired end). Each leaf was taken from a separate plant.

Experiment	Samples per experiment	Extracted squares per leaf	Total number of leaf squares (i.e. sequencing libraries)	Number of samples combined for sequencing	Sequencing platform and chemistry
Wounding experiment	4 time-points, 2 conditions, 3 leaves each	1	24	All at once	Illumina HiSeq 2500 (50 SE)
Flg22/water droplet spotting	1 time-point, 2 conditions, 6 leaves each	8	96	All at once	Illumina HiSeq 4000 (150 PE)
Untreated leaves	1 time-point, 1 condition, 3 leaves each	8	24	All at once	Illumina NextSeq 500 (75 SE)
Flg22/water infiltration	1 time-point, 2 conditions, 3 leaves each	8	48	All at once	Illumina NextSeq 500 (75 SE)

Contributions

Walter Verweij (Earlham Institute, UK) established the ST-seq v1.0 workflow, prepared the wounding experiment and the ST-seq with Illumina TruSeq comparison. Walter Verweij and I conducted the ST-seq flg22 droplet spotting experiment. Ashleigh Lister (Earlham Institute, UK) helped with flg22 infiltration ds-cDNA synthesis reactions.

I established the 96-well plate mRNA extraction, the ST-seq v1.1 method, modified the Nextera workflow described in [146, 155] to robustly work on full length ds-cDNA, planned and conducted all other experiments.

Sequencing was performed by the Earlham Institute Genomics Pipeline facilities. I analysed the data.

Pre-release of materials

The here described work has been submitted as preprint to the biorXiv server. The preprint [167] has been written by me and edited and uploaded to biorXiv by Matthew D. Clark (Earlham Institute, Natural History Museum, UK) as corresponding author. Based on the biorXiv copyright statement 'The copyright holder for this preprint is the author/funder, who has granted biorXiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license. ' Matthew D. Clark gives his consent for the work to be included in this thesis. The described work is published in Plant Methods under the reference [167].

Chapter-3 Spatially resolved transcriptomics reveals plant host responses to insect elicitors

Introduction

To date more than one million insect species have been characterised [168]. Half of the described insect species feed on plants [46]. Plants employ a series of protective strategies such physical barriers (e.g. hairs, trichomes and waxes) and metabolic as well as chemical cues (i.e. volatile signals and secondary metabolites that affect host choices, act as repellents, function as insect poisoning toxins or attract insect predatory enemies) against insect herbivores [46–50].

One particularly damaging clade for agricultural systems are insects of the order Hemiptera [81]. This order includes aphids, leafhoppers, planthoppers, whiteflies and true bugs and is composed of approximately 100,000 insect species [168]. Among the most destructive hemipteran pests in agriculture are aphids [80]. More than 4,000 aphid species have been described [80]. Although many aphid species are specialists that colonise one or a few closely related plant species [169], the green peach aphid *M. persicae* is a generalist with the potential to colonise hundreds of plant species in over 40 plant families [79, 80]. Detailed knowledge about the defence mechanisms to this generalist could provide important information about strategies to protect plants against many insect herbivores [46, 170].

Plants perceive information about an attacking herbivore with plant immune receptors that recognise HAMPs [53, 58, 59, 171]. Upon HAMP perception these receptors trigger a series of PTI responses such as the production of ROS, calcium bursts, kinase cascades, gene expression changes and physical or structural changes of plants [46, 49, 57, 172]. Together with PTI, plant resistance-gene (R-gene) mediated ETI, which often leads to a rapid HR and local cell death [53, 57,

173], plays a role in mediating insect resistance (e.g. Mi-1 [174], Mi-1.2 [175] and Vat [176]).

Of special importance in mediating herbivore resistance are the JA, ET and SA signalling pathways [46, 62], as well as camalexin [66] and glucosinolate [98] accumulating secondary metabolite synthesis pathways. However, especially biosynthetic defence pathways consume energy and nutrients and plants therefore carefully balance energy levels during an insect attack [62, 177]. The immune response to an attacking herbivore is therefore likely temporally and spatially controlled to use most resources for development, growth and reproduction [177]. This is supported by recent studies that observed strongly localised consecutive layers of defence responses such as calcium bursts [61], hypersensitive responses [170] and gene expression changes [66, 123] at and near sites of insect attack.

For insect pests little is known about the specific genes that regulate and fine-tune defence responses to herbivory, which checkpoints are important to determine the level of attack and which layers of gene expression changes are utilised to progress an immune response over time [99]. This motivated me to identify potential regulators in sensing and processing spatiotemporal plant to herbivore responses using green peach aphid extract infiltration.

Results

Sample preparation and sequencing

To prepare a spatiotemporal *A. thaliana* Col-0 response series to green peach aphid extract I infiltrated a droplet of crude extract and as a control a droplet of potassium phosphate buffer prepared according to Prince et al. [59] on the left-hand, abaxial side of a leaf. Starting with the infiltration spot I dissected four ~ 1 x 1 mm squares laterally moving towards the midrib from the leaf at the time-points of 0.5, 1, 2, 3, 5, 7 and 24 hours after infiltration: e+0mm describes the extract or buffer control infiltration spot, e+1mm the laminal spot adjacent to the infiltration

spot, e+2mm the dissected midvein and e+3mm the laminal tissue next to the midvein (**Figure 13A**). I prepared 4 biological replicates per time-point and crude extract or buffer infiltration (**Table 6**). I sequenced each library (448 libraries in total) using the Illumina NextSeq500 75 bp single-end chemistry to a depth of 923,344 ± 320,515 reads (898,593 ± 310,621 reads after adapter trimming and quality filtering) and detected an average number of 14,309 ± 1,826 genes with ≥ 1 mapping reads per library (89.3 ± 2.1 % of reads were assigned to *Arabidopsis thaliana* gene features (TAIR10; www.arabidopsis.org).

Spatiotemporal progression of DE gene expression

To characterise the spatiotemporal progression of the DE gene response I compared the extract with the buffer infiltrated squares and detected a total number of 7,536 statistically significant (q-value < 0.05) DE genes over the entire sampled time-course (**Chapter3_additional-file1.xlsx: all-DE-genes**). I found most DE genes (6,105) at the infiltration area (e+0mm). 1,601 DE genes I detected in the adjacent e+1mm to e+3mm squares. The number of DE genes gradually decreased from the infiltration area to the e+3mm area and over time (**Figure 13B**). At the 24 hour time-point I could not detect differential gene expression in the e+2mm and e+3mm areas, which indicates the potential return to the non-stimulated state in these sections between 7 and 24 hours (**Figure 13B**).

Table 4 Numbers of detected DE genes over the time course of 0.5 24 hours after infiltration: At the area of infiltration (sq1) the number of DE genes increases from 0.5 to 2 hours after infiltration, shows a decrease at 2 hours after infiltration and then increases again over time (up to the 7 hour time-point). In the spatial areas (sq2, sq3 and sq4) a similar pattern with a high number of spatially higher expressed genes at the beginning and a decrease towards the middle time segments with a new burst at 5 hours and 7 hours after infiltration is visible. At 24 hours after infiltration I could not detect any DE genes in sq3 and sq4. Data is presented as total DE genes, with 'h:' describing the number of higher expressed DE genes and 'l:' the number of lower expressed DE genes.

Time-point (hours after infiltration)	**Square 1 (sq1)**	**Square 2 (sq2)**	**Square 3 (sq3)**	**Square 4 (sq4)**
0.5 hours	646 (h:366, l:280)	490 (h:218, l:272)	79 (h:48, l:31)	149 (h:41, l:108)
1 hour	1,421 (h:831, l:590)	107 (h:54, l:53)	30 (h:12, l:18)	377 (h:142, l:235)
2 hours	2,274 (h:1477, l:797)	106 (h:84, l:22)	11 (h:6, l:5)	11 (h:0, l:11)
3 hours	1,333 (h:862, l:471)	50 (h:38, l:12)	38 (h:7, l:31)	198 (h:48, l:150)
5 hours	3,151 (h:1545, l:1606)	55 (h:30, l:25)	40 (h:22, l:18)	106 (h:26, l:80)
7 hours	4,535 (h:2218, l:2317)	188 (h:173, l:15)	278 (h:121, l:157)	56 (h:22, l:34)
24 hours	1,530 (h:799, l:731)	260 (h:236, l:24)	0 (h:0, l:0)	0 (h:0, l:0)

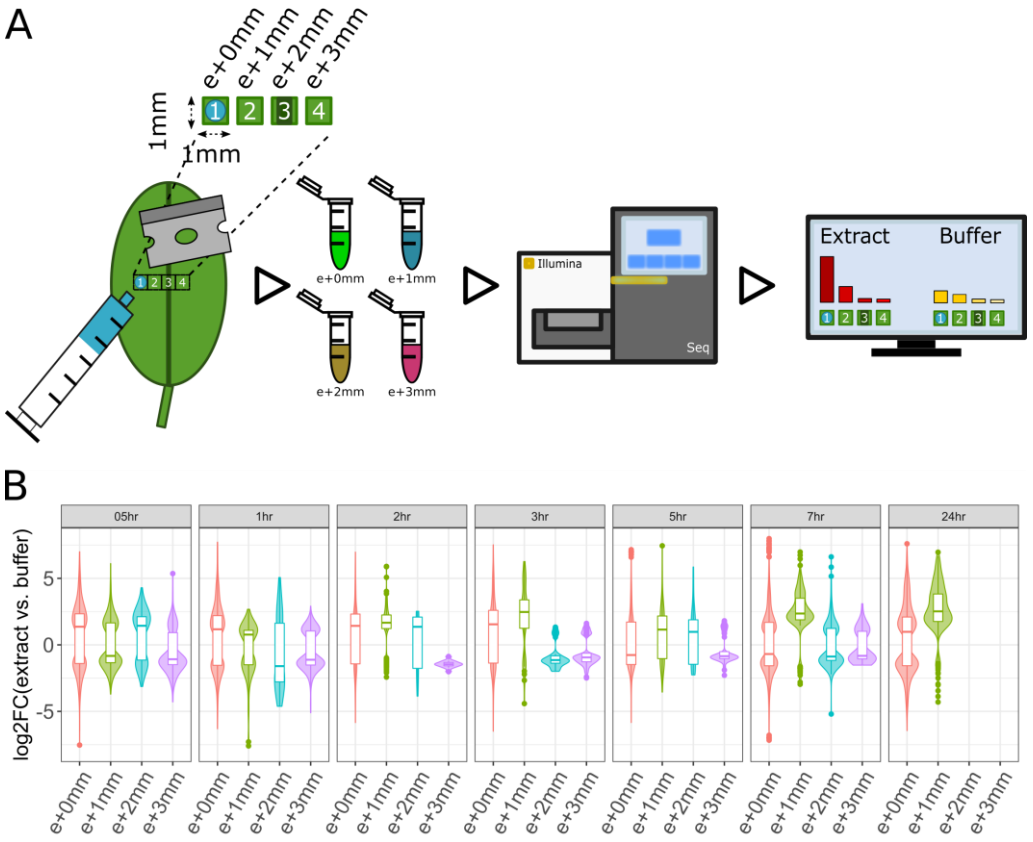

Figure 13 Experimental overview and spatiotemporal DE gene response upon crude *M. persicae* extract infiltration: (A) explains the experimental design: I infiltrated a 5 µ droplet of crude aphid extract or buffer control at the area e+0mm. From the boundary of the e+0mm area I sampled a lateral gradient moving towards the midvein of the leaf with e+1mm as laminal area, e+2mm as midvein and e+3mm as laminal area next to the midvein. Of each sampled area I prepared an Illumina sequencing library. (B) shows the spatiotemporal progression of the DE gene response. Immediately after infiltration I observe spatial DE gene expression over the entire sampled leaf section. With ongoing time, I detect lower magnitudes of log2 fold-changes at distant areas in comparison to the infiltration spot and at 24 hours after infiltration no DE genes in the areas e+2mm and e+3mm were detected.

Spatiotemporal characterisation of GO-terms

Based on the number of spatially DE genes I assumed that the response to a local herbivore attack is maximised at the area of attack and fades with increasing distance from the attacked spot. I assumed that this leads to triggering of different biological processes because of the stimulus intensity at and near the attacked area. To test this in the sampled cross-section, I analysed the DE genes of the infiltration area (e+0mm) and the DE genes in the proximity of the infiltration area (e+1mm to e+3mm) for enriched GO-terms (q-value < 0.01). I searched for statistically enriched biological processes (BP) and molecular functions (MF) and found that the stimulus site and proximal DE gene sets enriched a high number of shared GO-terms (**Figure 14A**). As assumed, I detected stimulus site and proximity unique GO-terms. Many of the unique GO-terms enriched in the time-period of 0.5 and 1 hour post stimulus (**Figure 14A, Figure 14B** and **Chapter3_additional-file1.xlsx: 'GO-BP_stimulus-proximal'**). At later time-points (> 1 hour) I observed less unique biological processes in the proximity of the infiltration spot. The spatial extension of the immune response is therefore likely defined early in plant herbivore perception and at later time-points strongly locally restricted.

Spatiotemporal characterisation of protein families

In a spatiotemporal experiment of *A. thaliana* exposure to *B. cinerea* Mulema et al. [38] detected overrepresented motifs for specific transcription factor family binding sites at and near areas of fungal inoculation. This suggests that the spatial DE gene expression of the *A. thaliana* defence response to *B. cinerea* is mediated by transcription factor families with potentially different downstream targets. This could explain the statistical enrichment of unique biological processes at the attacked spot or its proximity. In a similar analysis I therefore tested the DE-gene response of my experiment for overrepresented protein families (PFAMs) [178, 179]. I used the same hypergeometric test as for the GO-term enrichment and searched for statistically enriched PFAMs (q-value < 0.05) among the DE genes of

the stimulus and proximal areas. My analysis revealed a set of 86 enriched PFAMs. I detected shared but also stimulus and proximal site uniquely enriched families (**Figure 14A**, **Chapter3_additional-file1.xlsx: 'PFAM-enrich_stimulus-proximal'**). This indicates selective spatiotemporal recruitment of PFAMs during the progression of the defence response. To identify gene families with roles in sensing pest induced stresses and transcriptional regulation, I studied the 86 PFAMs and found a set of 9, hereafter referred to as regulatory PFAMs, families (**Figure 14C**, **Chapter3_additional-file1.xlsx: 'PFAM-enrich_stimulus-proximal'**): (1) AP2 domain genes (PF00847) [140], (2) C2H2-type zinc fingers (PF13912) [180], (3) WRKY DNA-binding genes (PF03106) [139], (4) tify domain genes (PF06200) [181], (5) AN1-like zinc fingers (PF01428) [182], (6) leucine rich repeat (LRR) genes (PF08263, PF13855, PF00560) [53, 183], (7) protein tyrosine kinase genes (PF07714) [183], (8) TIR domain genes (PF01582) [184] and (9) Arabidopsis broad-spectrum mildew resistance RPW8 (PF05659) PFAM members [185, 186].

For some of the regulatory PFAMs I observed distinct temporal expression patterns: Most AP2 domain family members were DE at the earliest time-point 30 minutes after infiltration. In contrast, RPW8 family members were most abundantly expressed after 3 hours, potentially highlighting the importance of an early ET signalling response and R-gene related defence at later time-points [185, 186] (**Figure 14C**). To study which PFAMs show the most extended spatiotemporal differential gene expression, I visualised spatiotemporal DE counts per family (**Figure 15A**). I also visualised the expression patterns of PFAM genes across the tested area (**Figure 15B**). This showed early spatially elevated or lowered DE patterns of AP2 domain and WRKY domain genes at the time-points of 30 minutes and 1 hour after stimulus. From 3 hours onwards I observed higher spatial loads of LRR members. At the 7-hour time-point I observed spatially extended tify domain DE patterns. The strongest localised PFAMs were protein tyrosine kinases, RPW8 proteins and AN1-like zinc fingers (**Figure 15B**).

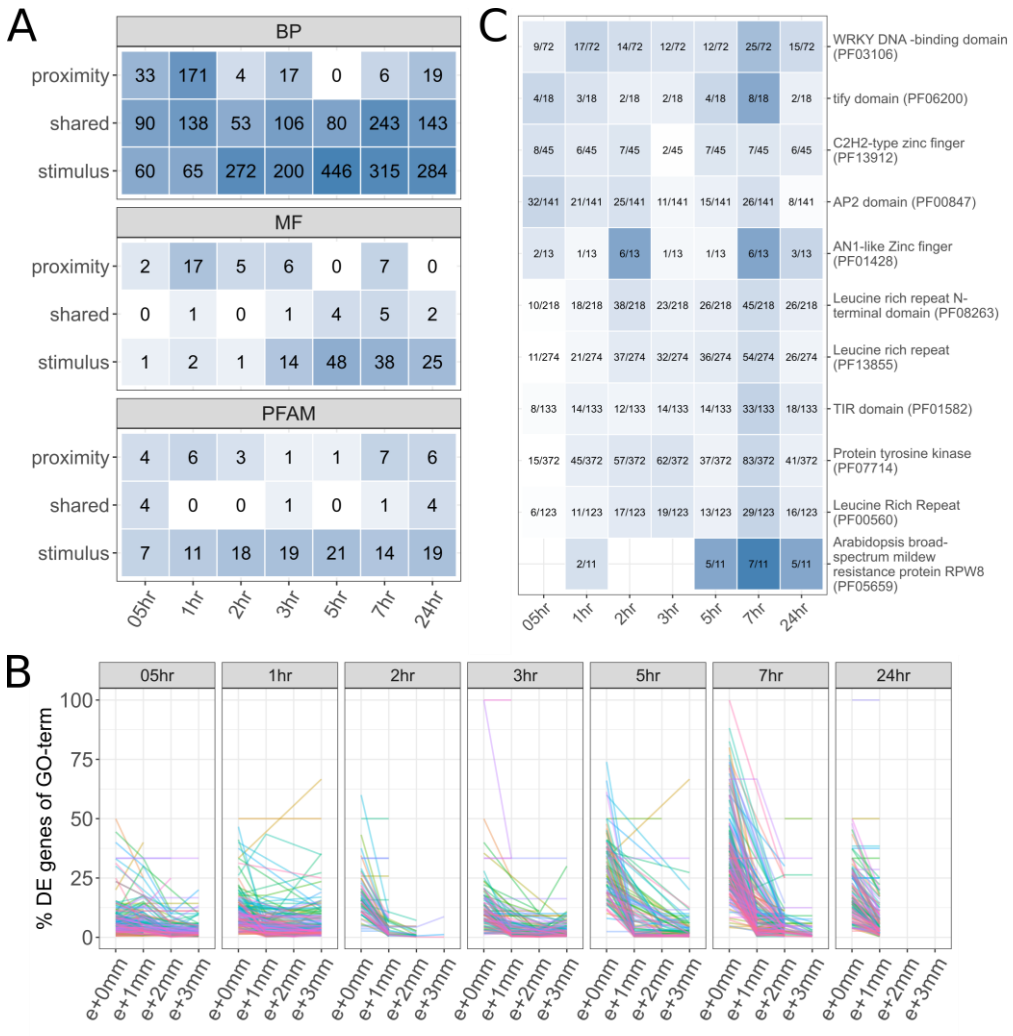

Figure 14 Stimulus site and proximal GO-term and PFAM enrichments: (A) shows the number of unique and shared statistically enriched biological processes (BP), molecular functions (MF) and protein families (PFAM) over time between the stimulus site (e+0mm) and the proximal tissue (e+1mm to e+3mm). (B) shows the percentage of enriched (biological process) GO-term genes at each time-point and area. Each biological process is represented as a line across the sampled leaf tissue. Whereas at early time-points many GO-terms show differential expression of their genes across the sampled leaf tissue, from 2 hours onwards most processes locate with higher DE gene numbers to the infiltration area. (C) Shows the detected enriched regulatory PFAMs and the number of detected DE gene members at each time-point in relation to all described PFAM members in the biomaRt [179, 187] database.

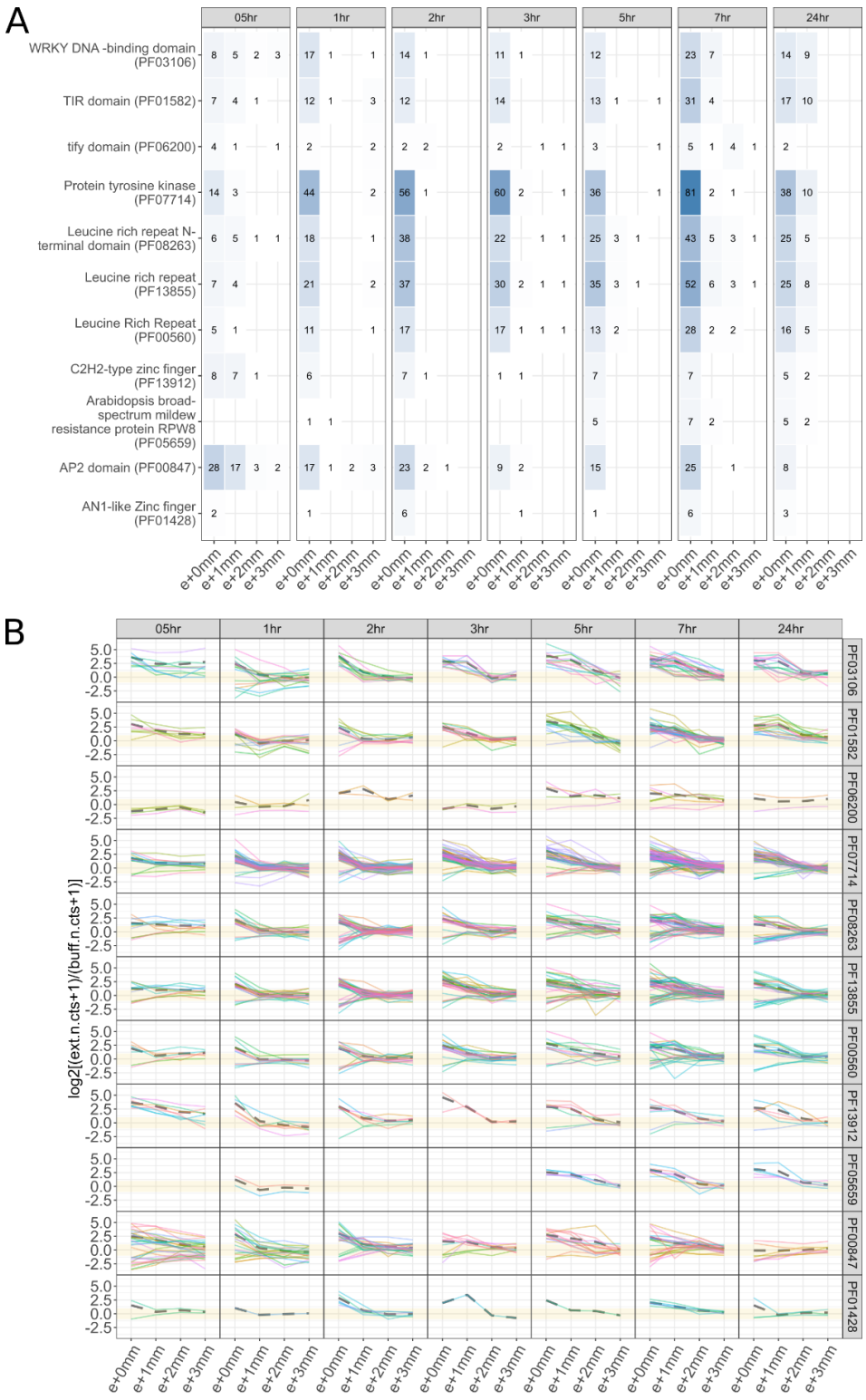

Figure 15 Spatiotemporal expression patterns of enriched PFAMs: (A) shows the spatiotemporal DE gene numbers of each regulatory PFAM over time. (B) displays the spatial expression patterns of DE PFAM members (in the same order from top to bottom as in (A)). Each line represents the spatial expression pattern of a single DE gene (detected as DE in at least one region). The expression of a DE gene is only shown at the time-point at which the gene was called as statistically significant DE.

Influence of PFAMs on the spatiotemporal plant to herbivore response

Although all of the identified regulator PFAM genes may play a crucial role in defence responses to herbivore attack, some regulators likely play a more significant role on the defence response than others [188]. The experimental identification of such important genes is often difficult as genetic studies of knocked-out or overexpressed key-regulators can lead to disrupted or non-viable phenotypes [189]. Here *in silico* network analyses have proven as useful tool to disseminate the position of genes in a network and determine the most influential nodes [190–192].

To identify important regulators in plant to herbivore responses I built a spatiotemporal protein-protein association network by parsing the DE genes of each time-point and square separately to the STRING [193] protein interaction database. I filtered the obtained network for DE-DE gene interactions and gene clusters with a size of ≥ 5 DE genes (**Chapter3_additional-file1.xlsx: 'Network-clusters'**). This revealed a large network with 62 clusters and 5815 DE genes. Of the 5815 detected DE genes 324 belonged to the regulatory PFAMs (**Chapter3_additional-file1.xlsx: 'Network-clusters'**). I found the largest clusters at the area of infiltration with decreasing cluster sizes towards the proximal areas (i.e. e+0mm 38 clusters with an average number of 265 ± 311 DE genes, e+1mm 15 clusters with an average number of 35 ± 36 DE genes, e+2mm 3 clusters with an average number of 26 ± 16 DE genes and e+3mm 6 clusters with an average number of 18 ± 11 DE genes) (**Chapter3_additional-file1.xlsx: 'Network-clusters'**).

The influence of a gene on a network can be estimated by its eigencentrality value (a relative measure for each gene with values ranging from 0 to 1; the higher the score, the higher the influence of a gene on a network) [194]. I therefore calculated the eigencentrality values for each gene in each cluster (**Chapter3_additional-file1.xlsx: 'Network-clusters'**). To test which regulatory PFAMs take most influence on networks over time I visualised the eigencentrality

values for each PFAM in a temporal boxplot (**Figure 16A**). This analysis showed temporally dynamic eigencentrality patterns for the detected regulatory PFAMs: I observed that WRKY domain, AP2 domain, C2H2-type zinc finger and TIR domain families showed highest eigencentrality values at early time-points starting with 0.5 hours after infiltration. Whereas WRKY domain, C2H2-type zinc finger and TIR domain genes possessed some influence in network clusters over the entire sampled time-course with a maximum at early time-points and some influential WRKY members at 24 hours after infiltration, AP2 domain eigencentrality loads were highest at the beginning and decreased over time to the time-point of 7 hours. For protein tyrosine kinase and LRR domain PFAMs I detected lower eigencentrality values as for the transcription factor families and TIR domain genes. RPW8 domain associated genes increasingly gained influence at late stages starting with 5 hours after infiltration. AN1 zinc finger genes had little effect on network clusters (**Figure 16A**).

Next, I extended my analysis to identify PFAMs that contribute to spatial network formation. I found that WRKY domain, C2H2-type zinc finger, TIR domain and AP2 domain PFAMs showed the most expanded spatial DE patterns within the first hour after infiltration. At later time-points LRR and tify domain PFAMs joined with spatial DE gene expression. Protein tyrosine kinase, RPW8 and AN1 zinc finger families were strongly localised with a maximal extension to the e+1mm area (**Figure 16B**).

This indicates that the early spatiotemporal plant to herbivore response could be mediated by rapidly responding WRKY domain, C2H2-type zinc finger, TIR domain and AP2 domain genes. Besides of AP2 domain transcription factors the detected PFAMs influence a plant's herbivore response over the entire sampled time-period and AP2 domain genes show regulatory effects up to 7 hours after attack but reduce influence from 0.5 hours on. Over time the response is converted from a spatial to a local one, with RPW8 genes gaining influence on late time-points. Protein tyrosine kinase and LRR PFAM eigencentrality patterns were less conclusive (**Figure 16B**).

Spatiotemporal network analysis shows key-regulators of plant to herbivore responses

I assumed that the identification of DE genes with the highest eigencentrality scores could reveal potentially important regulators in plant to herbivore responses. To identify DE genes with strong influence on network clusters I extracted all nodes with an eigencentrality score ≥ 0.75 from the spatiotemporal network. This filtering yielded 435 of 5,813 DE genes (7.5 %) of which 29 genes belonged to the regulatory PFAMs (**Table 5**, **Chapter3_additional-file1.xlsx: '29-PFAM-genes'** and **'29-PFAM-gene-clusters'**).

Table 5 Regulatory PFAM DE genes determined using association network analysis: The table shows the discovered regulatory PFAM DE genes with eigencentrality values > 0.75 in the constructed spatiotemporal network. TAIR identifiers are associated with known gene names.

PFAM domain association	Detected DE genes
LRR domain	AT1G35710, AT1G68780, AT1G74360 (*NILR1*), AT5G25930, AT1G07650, AT2G32680 (*RLP23*), AT5G48380 (*BIR1*)
AP2 domain	AT5G47230 (*ERF5*), AT4G17490 (*ERF6*), AT5G61590 (ERF107), AT5G61600 (*ERF104*), AT4G17500 (*ERF1A*)
WKRY domain	AT4G23810 (*WRKY53*), AT2G46400 (*WRKY46*), AT4G31800 (*WRKY18*), AT5G64810 (*WRKY51*), AT5G26170 (*WRKY50*)
TIR domain	AT2G20142, AT1G72940, AT1G72900
The tify domain	AT1G19180 (*TIFY10A* / *JAZ1*)
C2H2-type zinc fingers	AT5G59820 (*ZAT12*), AT5G04340 (*ZAT6*)
Protein tyrosine kinases	AT5G25440 (*SZE1*), AT1G21250 (*WAK1*), AT4G23190 (*CRK11*), AT1G18390, AT3G09830 (*PCRK1*), AT1G07650

To infer the functionality of the 29 regulatory PFAM genes I analysed the genes using STRING association and biological processes enrichment. I found statistically enriched processes for chitin responses, responses to bacteria and oxidative stress responses (**Chapter3_additional-file1.xlsx: 'GO-BP_stimulus-proximal'**). I also discovered the terms 'response to ET' at the time-point of 30 minutes after infiltration, 'response to SA' at the 7 hour after infiltration time-point and 'JA mediated signalling' at the 24 hour time-point indicating contribution of the

regulators to three highly important pathways in green peach aphid defence [46, 65] (**Chapter3_additional-file1.xlsx: 'GO-BP_stimulus-proximal'**). Some of the 29 regulatory PFAM genes were consecutively DE over multiple time-points: I discovered *WRKY53* as DE at each time-point of the sampled time-series, *ZAT12* as DE from the time-point of 30 minutes after infiltration to 5 hours after infiltration and *CRK11* as DE from 1 hour to 7 hours after infiltration (**Figure 16C**, **Figure 16D**, **Chapter3_additional-file3.cys**). Due to their abundant differential expression, these DE genes could be of importance in mediating plant to herbivore attack responses over multiple time-points. I analysed the 29 genes for their spatial loads and found *WRKY53* as DE in all four sampled areas at 30 minutes after infiltration (**Figure 16C**, Figure 17) and two TIR domain proteins as DE in the areas e+1mm to e+3mm: AT1G72900 and AT2G20142 over various time-points (**Figure 16C**, **Figure 16D**, **Chapter3_additional-file2.cys**). This indicates that especially *WRKY53* could be a spatially early and temporally important regulator of plant responses to an insect attack.

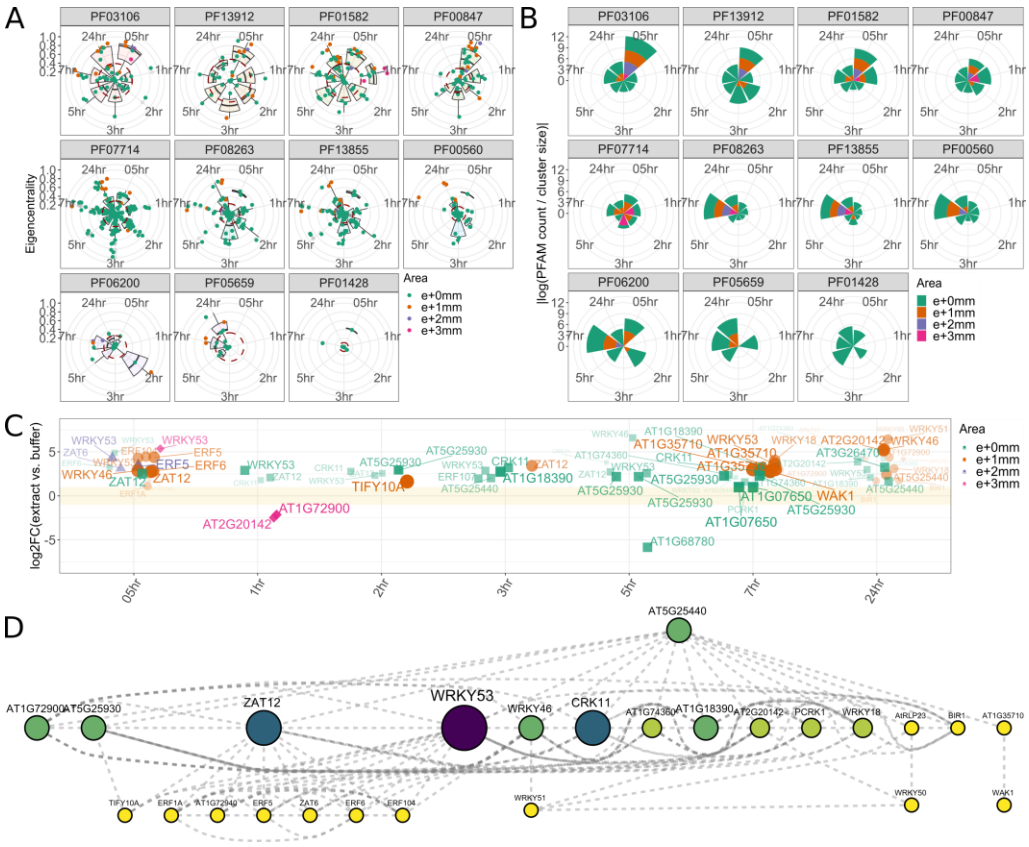

Figure 16 Spatiotemporal contribution of regulatory PFAMs and spatiotemporal expression of network-analysis filtered DE genes: (A) shows the eigencentrality values of all detected DE genes at a specific time-point and area and a summarising boxplot for all data-points split up for each PFAM. The red dashed line in each plot shows the average eigencentrality of all PFAM genes over all sampled squares and time-points. (B) shows the spatiotemporal network load of each PFAM calculated as the log(number of DE genes / average cluster size per time-point). (C) displays the expression of filtered DE genes with eigencentrality values > 0.75 over time. (D) shows the STRING database associations between potentially regulatory PFAM members. The size and colour (from yellow to purple) of each node indicates how often a gene was detected as DE over time. (PFAM IDs: PF03106 WRKY domain, PF13912 C2H2 zinc-finger, PF01582 TIR domain, PF00847 AP2 domain, PF07714 protein tyrosine kinase, PF08263 LRR domain N-terminal, PF13855 LRR domain, PF00560 LRR domain, PF06200 tify domain, PF05659 Arabidopsis broad spectrum mildew resistance protein RPW8, PF01428 AN1-like zinc finger).

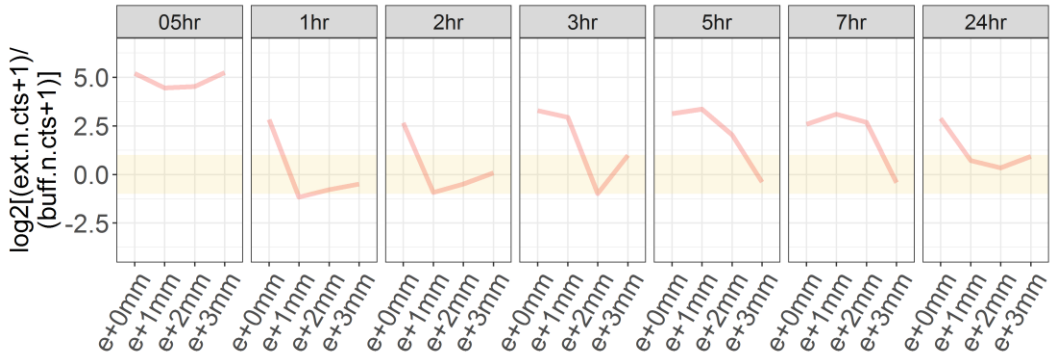

Figure 17 Spatial expression pattern of WRKY53 over the tested time-course: WRKY53 expression is strongly spatially elevated in comparison to the buffer control at 0.5 hours after infiltration. The WRKY53 expression levels remain elevated at the area of infiltration over the entire time-course. At e+1mm to e+3mm areas spatial WRKY53 expression levels decrease between 1 and 3 hours after infiltration and are elevated again at later time-points.

Discussion

To date little is known about the complex spatiotemporal networks that mediate the perception and progression of defence responses to a local herbivore attack. Especially the knowledge about plant immune responses to the generalist green peach aphid could help to protect plants against many insect herbivores, e.g. by producing transgenic plants that express insect resistant alleles [75, 76] or targeted alteration of susceptibility factors by genome editing [77].

Prince et al. recently described aphid extract infiltration as a successful mean to study and identify important plant innate immune components involved in responses to herbivore attack; i.e. BAK1 and PAD3 [59, 66].

I utilise this method in combination with my recently developed ST-seq workflow [167] to characterise DE gene association networks of a pest attack event. Motivated by recent studies that describe the localised immune response to phloem feeding insect attack [61, 66, 123, 170] I am interested in the genes that regulate and fine-tune the local defence response at and near herbivore attack sites and the protein families that orchestrate the progression of the response to herbivory in the first 24 hours after attack.

In my study I find overlap with the data from by Prince et al. [59], who showed that PTI and camalexin pathway components such as FRK1, BAK1, CYP81F2 and PAD3 are induced upon *M. persicae* extract stimulus.

In addition to DE genes involved in camalexin synthesis I also found differential gene expression activities in important plant hormone and secondary metabolite synthesis pathways mediating plant resistance; i.e. JA (21 / 36 DE genes), ET (10 / 24 DE genes), camalexin (11 / 36 DE genes), glucosinolate (47 / 97 DE genes) and SA synthesis processes (4 / 8 DE genes) [46, 65, 66, 195] - (**Chapter3_additional-file3.csv: 'GO-BP_stimulus-proximal'**, **Chapter3_additional-file4.pdf**).

In contrast to many studies that observe phloem feeding insect induced transcriptome changes after few hours of insect exposure [99], I detected the first DE genes, of a rapidly induced and spatially quickly spreading transcriptome response already 30 minutes after infiltration. I found that between the time-period of 7 to 24 hours, the spatially wider DE gene response is converted to a spatially restricted one with DE genes localised to the site of stimulus and the adjacent millimetre. During this time-period I also observed a shift from potentially early transcription factor mediated signalling towards R-gene and R-gene helper mediated resistance based on the expression of *ADR1*, *ADR1-LIKE1*, *ADR1-LIKE2* [185, 186].

I analysed the detected DE genes for statistically enriched PFAMs to identify potential mediators of plant to herbivore responses. My analysis statistically enriched 86 (of 4,165) PFAMs [178] of which 9 are associated with important roles in plant defence responses, immune signalling and transcriptional control of plant stresses: i.e. (1) WRKY domain genes [139], (2) C2H2 zinc-fingers [180], (3) TIR domain genes [184], (4) AP2 domain genes [140], (5) protein tyrosine kinase genes [183], (6) LRR domain genes [53, 183], (7) tify domain genes [181, 196], (8) Arabidopsis broad spectrum mildew resistance protein RPW8 [185] and (9) AN1-like zinc-finger genes [182].

To identify regulators of herbivory induced plant responses I studied the topology of a 5,815 DE gene large spatiotemporal association network [193]. This network

was composed of 62 clusters and contained 324 DE gene members of the identified regulatory PFAMs. Analysis of PFAM contributions to networks over time (based on the eigencentrality scores of their DE genes [194]), confirmed that especially WRKY domain proteins and AP2 domain proteins, accompanied by C2H2 type zinc fingers and TIR domain members are important in early spatiotemporal network clusters.

I analysed the highest 5 % of influential regulatory PFAM DE genes and found 29 of 324 potential key-regulators distributed over AP2 domain (5 DE genes), WRKY domain (5), C2H2-type zinc finger (2), TIR domain (3), tify domain (1), LRR domain (7) and protein tyrosine kinase (6) PFAMs. Among the WRKY, tify and C2H2-type zinc fingers I discovered higher expressed regulators of JA and SA signalling that could act on the antagonistic interplay of both pathways: i.e. ZAT6 has been described to act as SA activator [197], WRKY46 and WRKY53 as potential SA enhancers [198], WRKY50 and WRKY51 act as repressors of JA signalling [199] and also JAZ1 is described as a repressor of JA responses [200].

Literature search for AP2 domain PFAMs indicated a broader activity spectrum [107]; i.e. ERF107 is described to play a role in glucosinolate metabolic processes [201] and so could be an important regulator of glucosinolate signalling in aphid extract responses. ERF104 is described to function as MAPK interactor [202], ERF5 is known to be involved in the chitin immune response [203] and acts together with ERF6 as positive regulator of JA/ET mediated defences [204].

Some regulatory DE genes I found at multiple time-points. Especially *WRKY53* was DE over the entire sampled time-period and in all tested areas 30 minutes after stimulus, which could point to the importance of WRKY53 in guiding and modulating transcriptome responses over multiple time-points. WRKY53 mutant plants have already been characterised to be more susceptible to *Pseudomonas syringae* and Hu et al. hypothesise that together with other WRKY transcription factors (i.e. WRKY46 and WRKY70) WRKY53 could be involved in orchestrating basal defence against bacterial pathogens [198]. Miao and Zentgraf report of WRKY53 as upstream regulator of other WRKY transcription factors and further propose a model of WRKY53 senescence gene induction which is antagonistically

regulated by JA and SA pathways – hence altogether major pathways in plant to aphid resistance [95, 112, 205]. WRKY53 has recently also been associated with a potential role in promoting resistance to the Russian wheat aphid [206] (e.g. WRKY53 silenced plants showed a higher number of aphid feeding sites) and could therefore be a key-mediator of spatiotemporal plant to herbivore responses.

Material and Methods

Plant growing conditions

For my experiments I used 5-week-old *A. thaliana* Col-0 plants that I grew in a controlled environment room with an 8 hours light, 16-hour dark cycle at a constant temperature of 22 °C and 70 % humidity.

Aphid extract preparation and infiltration

I prepared aphid extract at the day of the experiment as described by Prince et al. [59]. In brief: 0.16 mg 6-day old aphids were collected from *A. thaliana* Col-0 plants, placed in a 1.5 ml conical tube and snap frozen in liquid nitrogen. The insects were ground to a fine powder using a prechilled polypropylene pestle (Z359947, Sigma Aldrich, St. Louis, USA) and dissolved in 0.8 ml 0.025 M KH_2PO_4 (pH 6.8) to a wet weight concentration of 20 mg / ml. The crude extract was immediately centrifuged for 15 minutes at 13,200 rcf at 4 ^{o}C, the supernatant transferred to a new 1.5 ml conical tube and stored on ice until use. Alongside with the aphid extract I stored a conical tube containing 0.025 M KH_2PO_4 (pH 6.8) as infiltration control on ice.

Approximately 15 minutes before the infiltration experiment, I transferred 5-week-old *A. thaliana* Col-0 plants from the controlled environment room to a laboratory working bench (a room with constant light exposure and temperature). To elicit plant responses, I syringe infiltrated a 5 µl droplet of the crude aphid extract or the buffer control at the same time-point (starting with the infiltration at noon) and collected the samples 0.5, 1, 2, 3, 5, 7, 24 hours after infiltration.

During the time of extract exposure, the plants were kept on the laboratory working bench. I prepared 8 plants for each time-point (4 for the crude extract infiltration and 4 for the buffer control). I used one leaf per plant for infiltration. By application of mild pressure on the plunger of the syringe when infiltrating I produced an approximately 2 - 3 mm diameter infiltration spot on the left-hand side of a leaf leaving approximately 1 mm to the midvein. I marked the boundary of the infiltration spot with a small spot of pencil dust, which I transferred on the leaf as suspension in 1 µl of 0.025 M KH_2PO_4 (pH 6.8) buffer. Plants which were wounded by the infiltration, or which showed large spread of the infiltrated liquid instead of a small concise spot were discarded and not processed for sequencing. The entire experimental time-series was prepared with one batch of aphid extract on the same day.

Leaf sectioning and sample harvesting

I used single margin razor blades (T586, Agar Scientific Ltd., Stansted, UK) to cut leaves into approximately 1 mm^2 small leaf squares on the pealed and clean, non-sticky paper cover of a 96-well plate seal (AB0580, Thermo Fisher Scientific, Waltham, USA). I sampled the infiltration spot on the abaxial left hand side of the leaf as e+0mm and from the infiltration spot towards the midvein e+1mm as the area between the infiltration spot and the midvein, e+2mm as the midvein and e+3mm as the laminal area next to the midvein. After cutting I transferred each leaf section immediately into a well of a dry ice cold (pre-cooled with a metal block stored on dry ice) well of a 96-well plate (E1403-0100-C, Starlab, Milton Keynes, UK) using a RNaseZAP (AM9780, Thermo Fisher Scientific, Waltham, USA) washed and air-dried forceps (T083, TAAB Laboratories Equipment Ltd, Berks, UK). I sealed the wells of the 96-well plate with domed PCR cap strips (AB0602, Thermo Fisher Scientific, Waltham, USA) and stored the samples at -80 ^{o}C until use.

Leaf sample lysis and mRNA extraction

I extracted the mRNA of each leaf square according to the 96-well plate format workflow as described in **Chapter 2 Leaf mRNA purification – page 49**.

Illumina sequencing library construction

I constructed the sequencing libraries according to the ST-seq v1.1 as described in **Chapter 2 Double-stranded cDNA synthesis reaction - page 50** and **Chapter2 Illumina library preparation from ds-cDNA – page 54**.

Sample pooling and sequencing

I pooled libraries at equal mass (1 ng of each cleaned library) based on the Qubit 2.0 Fluorometer dsDNA HS Assay Kit yields to balance the amount of each library in the final pool and obtain an approximately equal number of reads for each library in sequencing. As AMPure XP clean-ups in 96-well plates are prone to shorter fragment carryover, which in sequencing could lead to an increased amount of adapter dimer reads, I concentrated (i.e. removed potential short fragments) the library pool with two 0.60x AMPure XP bead clean-ups and eluted the final library in a volume of 30 µl 1x EB buffer. I assessed the profile of the final library pool on the Agilent Bioanalyser using Agilent High Sensitivity DNA Kit Sequencing reagents. I determined the concentration of the library pool on a 1 / 1,000 and 1 / 10,000 dilution (samples have to be highly diluted otherwise the qPCR fluorescence signal saturates early and confident Cp calls in comparison to quantification standards are not possible)of the final library using the KAPA SYBR FAST qPCR Library Quantification Kit (KK4824, KAPA BioSystems) on a LightCycler 480 Instrument II (Roche) in 384-well plates using scaled down 5 µl reactions and the programme: (1) 3 minutes at 95 ^{o}C, (2) 35 cycles with 30 seconds at 95 ^{o}C (4.4 ^{o}C/s ramp rate) and 90 seconds at 60 ^{o}C (4.4 ^{o}C/s ramp rate) with an imaging step at the end and (3) after the cycling phase the default LightCycler 480 melting curve step as final analysis.

I used the NextSeq 500/550 High Output v2 kit 75 bp cycle reagents (FC-404-2005, Illumina) for 75 bp single-end sequencing. To avoid overclustering and anneal library molecules to the grafted flow-cell primers I denatured and diluted the final library pool according to the 'NextSeq System Denature and Dilute Libraries Guide' (Illumina Document # 15048776v02) guidelines and sequenced the library with a 1 % PhiX spike-in ratio as a potential sequencing control and to increase the diversity of the sequencing pool. I prepared and loaded the NextSeq500 system according to the NextSeq500 system guide (Document # 15046563v02, Illumina). I specified a length of 9 nt for the Index 1 (i7) primer and 6 nt for the Index 2 (i5) primers.

Data demultiplexing, quality control and mapping

To convert binary base call files to FASTQ data and produce single FASTQ files per sample I demultiplexed the sequencing data using bcl2fastq-2.20.0 standard settings. Quality control and mapping was performed as described in **Chapter 2 Data quality control and mapping – page 56**.

Differential-expression analysis, GO-term and PFAM enrichment, pathway analysis

I performed the differential gene expression and GO-term analysis as described in **Chapter 2 Differential-expression analysis, GO-term enrichment – page 56**. I statistically enriched PFAMs using the same settings as for the GO-term enrichment analysis in the ClusterProfiler-3.8.1 [165] package (q-value < 0.05). I retrieved the PFAM domains for enrichment from biomaRt [179] using the boimaRt-2.40.0 package [187] in R-3.5.1.

Plant metabolic pathway analyses were performed on the AraCyc plant metabolic network database version 20180702 [207].

Across leaf DE-gene expression plots were prepared in R-3.5.1 on a data frame containing all DESeq2-1.20.0 normalised samples. Log2 fold-changes of buffer

infiltrated samples vs. controls were calculated as quotient of extract infiltrated normalised counts / buffer infiltrated normalised counts. All plots were generated using ggplot2-3.1.0 [164] in R-3.5.1.

StringDB network analysis

Association networks (of DE genes detected as DE at the same time-point and area) were built using the R-3.5.1 StringDB-1.22.0 [193] package with standard settings (species=3702, version=10, score_threshold=400). I filtered each network for connections between DE genes by removing all connections between non-DE x DE genes and non-DE genes x non-DE genes. I calculated network measures (eigencentrality, closeness centrality, degree centrality and betweenness centrality) on the network clusters and visualised the filtered networks both by using the R-3.5.1 igraph-1.2.4.1 [208] library. Networks were also visualised using Cytoscape-3.7.1 [209].

Sequencing and sample pooling metrics

Table 6 Sequencing and sample pooling metrics for the crude aphid extract infiltration experiment: The table lists the number of samples used for each experiment (separated by number of time-points, conditions and the number of used leaves), the number of extracted squares per leaf, the total number of squares processed to sequencing libraries, the number of combined samples for sequencing (per experiment) and the used Sequencing platform and chemistry (SE = single end, PE = paired end). Each leaf was taken from a separate plant.

Experiment	Samples per experiment	Extracted squares per leaf	Total number of leaf squares (i.e. sequencing libraries)	Number of samples combined for sequencing	Sequencing platform and chemistry
Aphid extract /control buffer infiltration	7 time-points, 2 conditions, 4 leaves each	8	448	All at once	Illumina NextSeq 500 (75 SE)

Contributions

I planned and conducted all experiments, constructed the sequencing libraries, performed the quality control, sequenced the libraries and analysed the data.

Chapter-4 Spatially resolved transcriptomics reveals plant host responses to *Myzus persicae*

Introduction

Plants are in constant interaction with their environment and in regular exposure to pests and pathogens [147]. The prevalent plant pests, with more than one million of species, are insects [50]. These are already major economic pests but climate change creates new opportunities for insects to invade and threaten new agrosystems [210]. Amongst these the green peach aphid *M. persicae* is one of the most destructive pests worldwide [80]. While many pests are specialised to specific hosts or host families, the generalist aphid pest *M. persicae*, can infest a broad range of up to 40 plant families and so cause devastating agricultural losses [80]. Infestations with *M. persicae* develop quickly due to the asexual life cycle of the insect and the rapid increase of aphid population can have detrimental effect on a host plants health (i.e. reduced plant growth, water and nutrient stress [95]). As the insect can also act as virus transmitter [81], the combination of the rapid asexual life cycle with viral disease transmission can accelerate viral disease dispersal [211].

One efficient means for controlling insect pests is the use of pesticides. However, due to environmental and health impacts new legislation increasingly restricts pesticide use in many countries [68] with the development of pesticide resistances by insects [82], other ways of controlling insect pests are attractive [69].

Over the last decades many studies of plant insect interactions provided valuable insights in plant defence mechanisms against insects [51, 212–215]. It has been shown that insect pests use effector proteins to interact and manipulate a host plants defence response [87, 147]. This complex molecular interaction defines the cycle between a pests attempts to overcome a plant's immune system versus the successful triggering of plant defence responses against the pest [57].

For *M. persicae* this interaction is well characterised [61, 66, 86, 92, 216]. *M. persicae*, actively secretes effector protein rich saliva [86], which acts on plant defence responses [87]. This is mediated with the use of specialised mouth parts called stylets, which are not only used to secrete effectors, but also to test a plant for its suitability as a host – a behaviour called probing, followed by feeding (i.e. ingestion of nutrient rich phloem sap) and drinking (i.e. uptake of water from xylem sieve element bundles) [96, 217, 217].

This probing behaviour is not fully understood yet. Aphids probe plants regardless of their host compatibility [170] and it has been observed that aphids show increased mobility and probing rates on non-host plant species or resistant varieties [44, 97] – likely driven by unsuccessful probing or inherent plant defence responses [218]. This behaviour resembles the search for a suitable site to settle [44] and suggests a complex interaction of the insect with the plant during this time-period. Disrupting this interaction could be an effective control, but to date little is known about this interaction. Recent studies have indicated consecutive layers of defence responses i.e. calcium bursts [61], reactive oxygen bursts [122], gene expression patterns [66, 123] and hypersensitive responses [170] closely in and around the site of aphid activity.

These localised plant responses to pests probing and manipulation motivated me to question the differences in host responses between unsuccessful probing, and successful colonisation e.g. prolonged phloem sap uptake. One means to achieve this is by using transcriptome profiling techniques. However, to date many transcriptomic experiments studying plant insect interactions have been performed at a gross scale (e.g. whole plant or leaf level) and by exposing plant tissues to bulks of insects [99] (likely due to the RNA amounts required for many RNA assays). To fully dissect plant defence responses at small sites of biotic attack, a higher spatial resolution for the detection of gene expression patterns is necessary [24]. I therefore applied my newly developed low-input spatial transcriptomic technique with millimetre scale resolution to profile gene expression levels of leaf areas during aphid attack. Such low-input RNA-seq technologies have proven as valuable tools to finely characterise an organism's

transcriptome [24], especially to detect the most lowly expressed genes which are only responding in a small number of an organisms cells and which assign valuable spatial and temporal information to the entire transcriptome of an organism [20, 24]. Here I present insights in the gene expression differences between a host plant's transcriptional responses at abandoned probing spots and from successful probing (i.e. feeding) areas.

Results

EPG sampling and sequencing

To establish abandoned aphid probing spots, I exposed a ~1 x 1 mm small leaf area immediately next to the midrib (**Figure 18A**) to a single *M. persicae* aphid with recording electrodes on the aphid and host plant over a 30-minute EPG experiment. I only included plants with at least 15 minutes of activity (i.e. detected aphid stylet activity in the plant) over the 30-minute EPG experiment with at least 20 observed probes. The aphid was subsequently removed and I sampled a time-series of 0, 0.5, 1, 2, 3, 5, 7, 24 hours after the first detected probing event (hap). For every time-point (measured in hap) I collected 4 biological replicates with corresponding non-aphid exposed controls. The control samples were collected in an identical manner to those of the aphid exposed plants (**Table 9**).

To sample areas of sustained phloem sap uptake (hereafter referred as feeding), I collected a ~1 x 1 mm large leaf area where a single *M. persicae* insect successfully fed for sustained EPG E2 phase (sE2) > 10 minutes [219]. In my biological replicates the sE2 phases were reached: 30, 31, 52 and 77 minutes after the first detected probe at the sites of feeding (**Figure 18B**). For every feeding time-point I collected a corresponding non-aphid exposed control (**Table 9**).

In addition to the aphid exposed areas I also harvested the adjacent leaf squares to the aphid exposed areas in a lateral cross section composed of eight ~1 x 1 mm squares. This allows me to dissect spatiotemporal gene expression patterns (**Figure 18C**).

To account for potential biological variation of spatiotemporal gene expression patterns, I also sampled a 1D cross section of 8 ~1 x 1 mm non-aphid exposed, untreated leaves as additional control on three consecutive days (4 biological replicates per day), (**Table 9**).

All samples were processed using my ST-seq protocol and sequenced on an Illumina NextSeq500 system using 75 bp single-end reads to an average of 1.03 ± 0.30 M reads per leaf square. I observed 89.7 ± 4.9 % of reads assigned to TAIR10 gene features. On average I detected 14,426 ± 1,745 genes with ≥ 1 mapped reads.

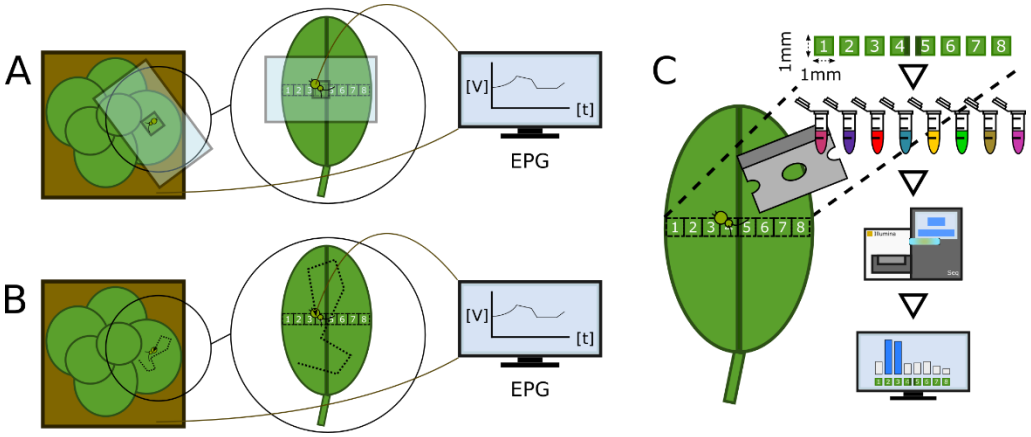

Figure 18 EPG sampling strategy of probing and feeding samples and ST-seq library preparation: (A) Shows the sampling strategy for the probing time-series. To obtain brief periods of probing at the same area I restricted aphids to the same spot of a leaf a 30-minute EPG experiment. (B) Shows the sampling strategy for feeding. To allow aphids to find a suitable feeding site I tracked insects on a leaf. At the time-point when insects reached the phase of sustained phloem feeding, I collected the sample. (C) shows the library preparation strategy. Aphid are imaged during EPG and so the position of the insect is known. At the height of an insect a lateral leaf-cross strip is dissected and 8 ~1 x 1mm leaf squares are transformed to ST-seq sequencing libraries and analysed for spatiotemporal differential gene expression.

RNA velocity estimation of aphid probing areas

To better understand the host plant transcriptional response e.g. if the plant executes a series of consecutive transcriptome responses, I measured the RNA velocity of the transcriptional changes using the Velocyto software [220]. RNA velocity measures the ratio of final fully spliced mRNAs to early incompletely spliced mRNA transcripts [220]. With multiple time points this provides directionality of transcriptional change as well as speed [220]. On my RNA velocity estimation plot the abandoned probing data-points of the same time grouped closely together and velocity arrows indicated progression along the sampled time-line (**Figure 19**). I also observed that the aphid data for the different time-points arranged into a circle, or swirl. La Manno et al. [220] associate circular RNA velocity estimations with circadian changes. My samples were collected during the afternoon and evening hours (**Chapter4_additional_file1.pdf**), and even though the control samples at the same times do not show the same strong circular progression I wanted to see if there was influence of circadian rhythm. For this I conducted velocyto analysis on subsets of my transcriptome data: (1) all detected genes (17,285), (2) all DE genes (4,649 DE genes, **Chapter4_additional_file2.xlsx: 'Detected DE genes'**), (3) all detected 131 circadian / clock associated genes (retrieved using biomaRt [187], **Chapter4_additional_file2.xlsx: 'Circadian DE gene associations'**), (4) all detected genes without DE genes (12,636 genes), (5) all detected genes without circadian genes (17,154), (6) all DE genes without circadian genes (4,597 genes). I observed that the circular pattern seen with aphid DE genes in the velocyto plot was lost when exclusively analysing circadian genes. A similar circular pattern is seen with the velocyto plot of all genes, and all non-DE genes but not in the non-aphid controls. This suggests there is little contribution of circadian genes to the circular pattern and points to the influence of the aphid induced DE genes on the circular RNA velocity estimation (**Figure 19**). As also non-DE genes contributed to the circular pattern this suggests some influence of genes with a q-value > 0.05 on the circular pattern.

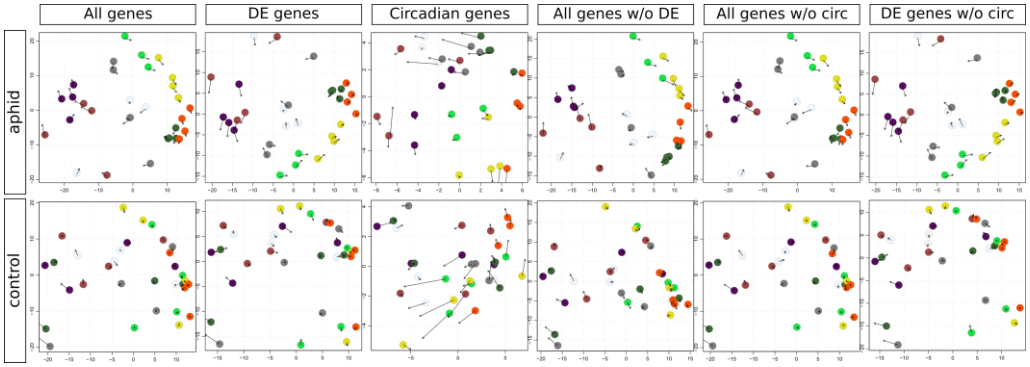

Figure 19 RNA velocity estimation plot of the aphid probing dataset: RNA velocity measures the ratio of spliced to incompletely spliced mRNA transcripts and so provides directionality of transcriptional changes and speed **[220]**. The upper panel (aphid) shows the velocyto analysis PCA plot of aphid exposed areas for the aphid probing experiment time-series. The lower panel (control) shows the corresponding non–aphid control areas. Colours of data points change with time: 0h white, 0.5 h ochre, 1 h purple, 2 h grey, 3 h light green, 5 h yellow, 7 h orange, 24 h, dark green. I observed that probing data-points of the same time-point grouped closely together and are arranged in a circle. To control my data for circadian changes (as the samples were collected during the afternoon and evening hours) I tested the data for influences of circadian rhythm using (1) all detected genes (17,285), (2) all DE genes (4,649 DE genes), (3) all detected 131 circadian / clock associated genes, (4) all detected genes without DE genes (12,636 genes), (5) all detected genes without circadian genes (17,154) and (6) all DE genes without circadian genes (4,597 genes). I observed that the circular pattern for the aphid DE genes was lost when I exclusively analysed circadian genes. I observed a similar circular pattern with the velocyto plot of all genes, and all non-DE genes but not in the non-aphid controls. This suggests that there is little contribution of circadian genes to the circular pattern and indicates that the aphid induced DE genes influence the circular RNA velocity estimation. As also the non-DE genes show a circular pattern this indicates that genes with a q-value > 0.05 influence the circular pattern as well. The x-axis of each plot represents the PC1 and the y-axis the PC2 (PC1/PC2 labels have been removed to condense the plot).

Characterisation of enriched biological processes over time

I observed two waves of transcriptome responses with a number of increasing DE genes post aphid probing: firstly, the numbers of observed DE genes increased early between the 0-hap to 1-hap time-point from 33 DE genes (0-hap) to 425 DE genes (1-hap). Then I detected a minimal response of 2 DE genes at the 2-hap time-point. This was followed by a new increase from the 3-hap time-point onwards with 74 DE genes through to 3,721 DE genes at the 24-hap time-point. In detail: 0-hap: 33 DE genes, 0.5-hap: 109 DE genes, 1-hap: 425 DE genes, 2-hap: 2 DE genes, 3-hap: 74 DE genes, 5-hap: 398 DE genes, 7-hap: 1,354 DE genes, 24-hap: 3,721 DE genes (**Table 7**, **Figure 20**). I compared the detected DE genes to the DE genes of the aphid extract infiltration time-series. For the extract time-series I found a higher (total) number of 7,706 DE genes, for the probing dataset a total number of 4,649 DE genes. Both datasets shared 2,886 DE genes (30.5 %) with 4,820 unique DE genes for the extract and 1,763 unique DE genes for the probing dataset, possibly suggesting that other processes are triggered by the infiltration of whole aphid extract in comparison to aphid stylet probing.

Table 7 Numbers of detected DE genes over the time course of 0-hap to 24-hap: I observed two waves of gene expression from 0-hap to 1-hap and from 3-hap to 24-hap, with a minimal number of DE genes at 2-hap.

Time-point (hap)	DE genes total	DE genes higher expressed	DE genes lower expressed
0-hap	33	16	17
0.5-hap	109	49	60
1-hap	425	244	181
2-hap	2	2	0
3-hap	74	28	46
5-hap	398	164	234
7-hap	1,354	632	722
24-hap	3,721	1,972	1,749

To determine the time-point at which the earliest plant stress responses can be tested for statistical enrichment at the abandoned aphid probing spots, I analysed the DE genes of each time-point using GO-term enrichment (**Chapter4_additional_file2.xlsx: 'GO-enrichment-probing'**) (q-value < 0.01). I comprehended the obtained GO-term lists for each time-point with REVIGO [136] and summarised the results using the REVIGO representative GO-terms in **Table 8**. At the 0-hap time-point I found the single biological process 'regulation of flavonoid biosynthesis' as enriched. However, soon after this at 0.5-hap I observed the first series of stress related biological processes.

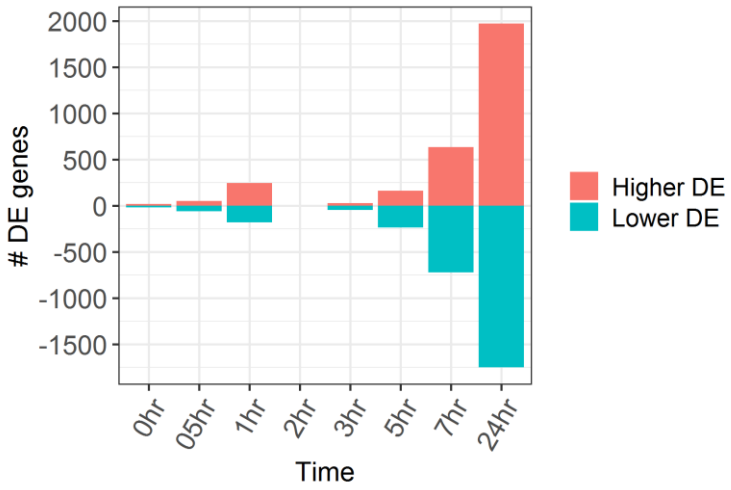

Figure 20 Higher or lower DE genes over the samples time course of 0-hap to 24-hap: I observed two waves of transcriptome responses with a smaller waves between 0-hap to 1-hap and a second, more DE gene numerous response between 3-hap to 24-hap.

Table 8 Enriched biological processes of aphid probing DE genes: With ongoing time I detected an increasing amount of enriched biological processes with first enriched stress response related GO-terms at 0.5 hours after the first observed aphid probing event. The table shows the REVIGO **[136]** summarised enriched biological processes at each time-point.

Time after probing [h]	REVIGO comprehended, most representative GO-terms
0	regulation of flavonoid biosynthesis
0.5	response to organonitrogen compound, establishment of protein localization to membrane, olefin metabolism, reactive oxygen species metabolism, ethylene biosynthesis, antibiotic metabolism, phenol-containing compound metabolism, respiratory burst
1	vacuolar transport, response to jasmonic acid, ribosome biogenesis, isopentenyl diphosphate metabolism, photosynthesis, cuticle development, tetrapyrrole metabolism, cellular glucan metabolism
2	Any GO-terms enriched
3	response to wounding, jasmonic acid metabolism
5	flavonoid metabolism, programmed cell death, regulation of cellular response to stress, reactive oxygen species metabolism, antibiotic metabolism, photosynthesis, benzene-containing compound metabolism, plant ovule development, glycosyl compound biosynthesis
7	RNA methylation, circadian rhythm, regulation of proton transport, ribosome biogenesis, response to cadmium ion, rhythmic process, wax biosynthesis, wax metabolism, flavonoid metabolism, protein folding, mucilage metabolism, multidimensional cell growth, leaf morphogenesis, carbon fixation, photosynthesis, reactive oxygen species metabolism, carbohydrate catabolism, oxidoreduction coenzyme metabolism, dephosphorylation, cutin biosynthesis, photorespiration, fatty acid derivative metabolism
24	RNA methylation, shoot system morphogenesis, ribosome biogenesis, positive regulation of catalytic activity, response to cadmium ion, establishment of protein localization to organelle, autophagy, protein folding, photosynthesis, reactive oxygen species metabolism, cellular aldehyde metabolism, oxidoreduction coenzyme metabolism, sulfur compound biosynthesis, starch metabolism

Although many *M. persicae* plant interaction studies highlight transcriptome changes upon plant to insect stress hours to days after prolonged insect infestation [99], I observed the first set of stress responses enriching for the biological processes 'responses to organonitrogen compound', 'respiratory burst' and 'reactive oxygen metabolism' at the 0.5-hap time-point (**Figure 21**). I also found few DE genes (4 of 33) enriching for 'flavonoid synthesis' at 0-hap (flavonoids play an important role in modulation of insect defences [221]). At later time points (> 1-hap) the enriched biological processes indicated 'response to JA', 'JA metabolism', 'cell death', 'glycosyl compound synthesis', 'starch metabolism', altogether listing many of the known processes described in plant to insect responses [49, 95].

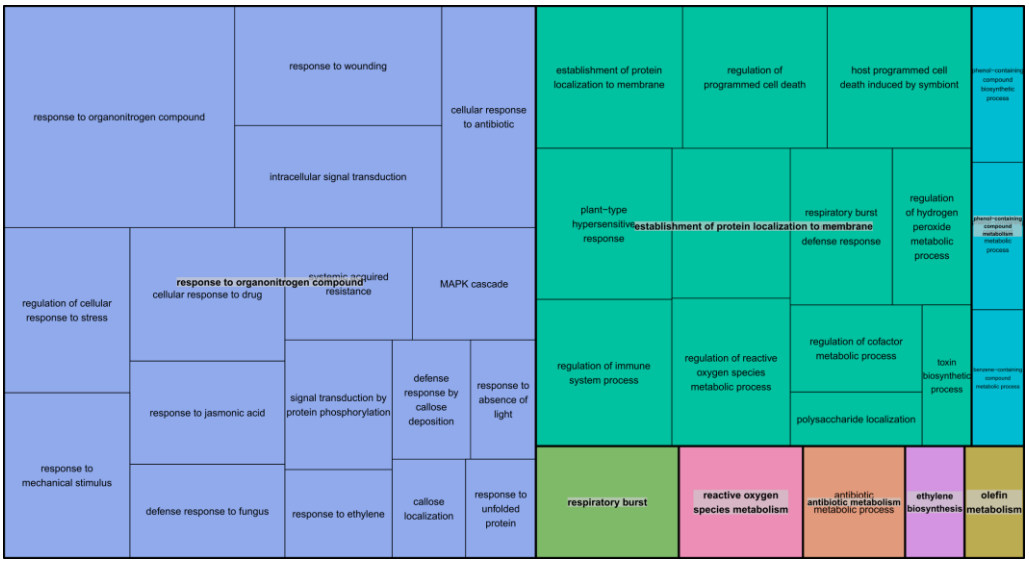

Figure 21 REVIGO GO-term enrichment summary of earliest detected plant to aphid responses: At the 0.5 hap time-point I detected biological processes related to aphid perception and early stress responses as well as enriched biological processes associated with reactive oxygen species: 'response to organonitrogen compound', 'establishment of protein localisation', 'respiratory burst', 'reactive oxygen species metabolism', 'phenol containing compound metabolism', 'antibiotic metabolism', 'ethylene biosynthesis' and 'oleofin metabolism'. The size of each rectangle represents the absolute log10(q-value) of each GO-term. The bold identifiers for each colour are the REVIGO determined representative GO-terms.

Transcriptome differences between probing and feeding sites

As *M. persicae* actively manipulates plant defences [87], I hypothesised that successful overcoming of a plants defence response by probing leads to feeding whereas unsuccessful probing leads to the search of a new feeding spot [44]. I therefore assumed that the difference in the transcriptome response between probing and feeding could point to important defence components that *M. persicae* manipulates and uses as molecular cues to establish feeding sites. To test this, I sampled 4 biological replicates where insects reached sustained phloem uptake at 30, 31, 52 and 77 minutes after the first probe (as for the probing time-series in a 1D leaf cross strip) and analysed the feeding samples together with the most similar 0.5-hap and 1-hap time-points where transcriptome responses are tested 30 and 60 minutes after the first probe (**Table 9**).

Although the feeding spots were not established at the same time, I expected that genes affected by aphid manipulation could share time independent spatial expression patterns in all feeding replicates. To identify the genes that contribute to the difference between probing and feeding spots I used principal component analysis (PCA). I assumed that especially the DE gene response at areas where aphids aborted probing (i.e. my previously sampled hap time-series) could help to identify components that plants use in defence responses to *M. persicae*.

PCA using log2 fold-changes of aphid vs. non-aphid exposed controls separated probing and feeding areas on PC1 explained 35.5 % of the variance. In contrast to PC2 – PC10 only PC1 separated the areas well (**Chapter4_additional_file3.pdf**). To affirm that the high load on PC1 is due to the 0.5 and 1-hap DE genes and not a random effect, I also PCA tested the probing and feeding spots using log2 fold-changes of all detected genes and 1 million random 516 gene large subsets of the 4,649 hap DE genes. Both analyses showed lower contributions to PC1: I observed 12 % of explained variance for the entire transcriptome data and 18.9 ± 1.9 % explained variance on PC1 for the random subsets (**Chapter4_additional_file4.pdf**).

Among the 516 DE genes of the 0.5 and 1-hap time-points the PCA revealed a set of 71 well represented genes on PC1 (cos2 ≥ 0.7) with statistically significantly different log2 fold-changes (two sided t.test, q-value < 0.05, Benjamini-Hochberg p-value correction) between probing and feeding areas (**Figure 22A, Figure 22B, Chapter4_additional_file2.xlsx: 'Feeding_gene-list'** and **'Feeding-list-expression-values', Chapter4_additional_file4.pdf**).

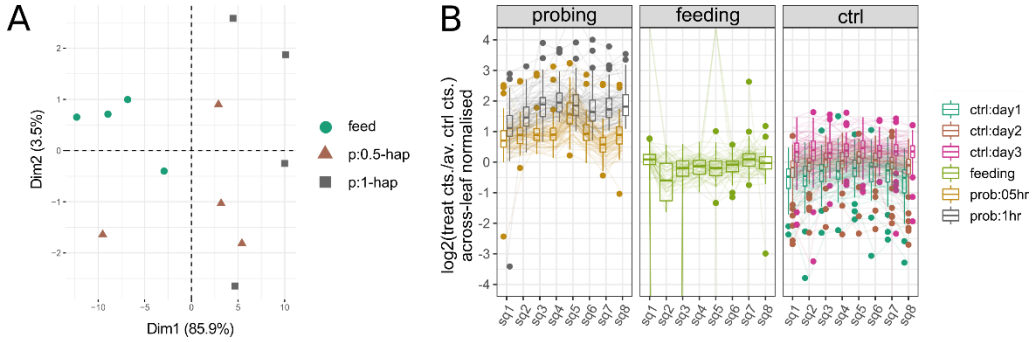

Figure 22 Spatiotemporal expression differences of a 71 probing DE gene set that discriminates probing from feeding areas: (A) Shows a PCA plot (PC1 and PC2) visualising feeding and 0.5 as well as 1-hap samples separated by the log2 fold-changes of the selected 71 genes. (B) Shows the spatiotemporal expression plot of the 71 genes for the probing (0.5-hap yellow and 1-hap grey) samples, the feeding replicates (green) and aphid untreated control replicates (turquoise, brown and purple). The aphid untreated control fold-changes were calculated to the average normalised counts of all control samples collected at three consecutive days.

I tested the identified genes for statistically enriched GO-terms and found a set of protein catabolism and protein vacuole targeting biological processes (**Chapter4_additional_file2.xlsx: 'Feeding_genes_GO-enrichment', Chapter4_additional_file4.pdf**). I assumed that the 71 genes potentially belonged to a shared pathway which could not be induced in an unsuccessful defence response and therefore tested the gene set using association network analysis. This analysis revealed a network with 8 clusters of which I filtered all clusters with ≥ 5 genes. This left me with 3 clusters built by 12, 9 and 8 genes (**Figure 23, Chapter4_additional_file2.xlsx: 'String clusters feeding genes', Chapter4_additional_file5.cys**). To determine the most influential components of each cluster I weighed the nodes based on their eigencentrality values [194] and

found that the most influential genes in the small network clusters are involved in protein degradation (UBQ3 [222]), energy sensing (KING1 and AKINbeta1 [223, 224]) and JA pathway modulation (BBD2 [225]) - **Figure 23**. This indicates that *A. thaliana* could be manipulated by *M. persicae* probing to not express a set of components involved in important resistance pathways to the insect [95].

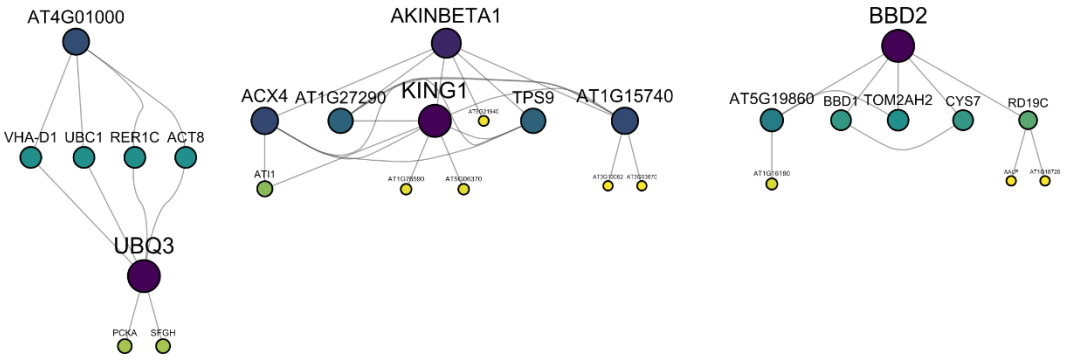

Figure 23 STRING network clusters with ≥ 5 genes detected of the 71 genes distinguishing probing from feeding sites: The figure shows 3 of 8 small gene association network clusters. The size (the larger) and the color (the darker) of each node correlate with the eigencentrality value of the gene in the cluster.

Discussion

An important but not well understood behaviour of *M. persicae* is the active probing of plant cells [96]. With this behaviour a phloem feeding insect tests host cells by ingesting small amounts of cell sap [96] and actively secretes effector protein rich saliva [86] to act on plant defences [87]. Published data suggests that aphids may show higher probing frequencies on non-host plants or resistant varieties [44, 97]. This could be driven by the cycle of a plants defence response and the insects attempt to overcome the plant immune system [218]. Knowledge about plant responses involved in this process could help us to understand the defence mechanisms that phloem feeding insects manipulate to colonise new hosts [89, 93].

To dissect plant defence responses at small sites of *M. persicae* attack, a high spatial resolution for the detection of gene expression patterns is necessary [24].

I therefore applied my newly developed low-input spatial transcriptomic technique with millimetre scale resolution in combination with the established EPG method [124] to profile gene expression levels of leaf areas during aphid attack (i.e. probing and feeding).

Analysis of the probing time-series from 0 to 24-hap showed statistically significant DE gene responses (of 33 genes) to phloem feeding already at the time-point of the first cell puncture event. I detected the first statistically enriched set of insect defence associated processes 30 minutes after the first probe and so much earlier than many published experiments suggest [99]. RNA velocity analysis [220] of aphid probing areas showed highly dynamic transcription rates immediately after the first probing event up to the 1-hap time-point and circular progression of the immune response suggesting the return to a non-stimulated state sometime after 24 hours. This indicates that many processes involved in the defence response to phloem feeding insects are triggered immediately after insect perception but substantially between 30 minutes and 1 hour after the first cell puncture event.

Recently studied calcium responses to *M. persicae* indicate that especially the first cell puncturing events induce strong calcium bursts [61]. Plants utilise calcium as secondary messenger at the onset of defence signalling cascades [49], therefore rapidly probing elicited genes could help to better understand the complex interaction of phloem feeding insects with a plants defence mechanism. Although I also found many genes with a described role in plant colonisation by *M. persicae* (i.e. *DCL1*, *AGO1*, *SSI2*, *FAD7*, *PP2-A1*, *MYB73*, *CYP79B2*, *EIN2*, *RBOHD*, *MYB34*, *MYBR1*) [95] at time-points > 5-hap (**Chapter4_additional_file2.xlsx: 'Louis_Shah_DE-gene-overlap'**), I detected the two WRKY transcription factors *WRKY33* and *WRKY46* (both with roles in insect defences [226, 227]) as lower DE at 0.5-hap and higher DE of the *TPS11* gene with an important role in *A. thaliana* to *M. persicae* defence [106] at 1-hap. This suggest potential to detect novel and important genes involved in the onset of plant to aphid responses using EPG coupled ST-seq.

So far little is known about the molecular defence mechanisms that phloem feeding insects overcome in establishing a feeding site. During probing *M. persicae*, actively secretes effector protein rich saliva [86] to manipulate plant defences [87]; e.g. calcium binding proteins in the saliva prevent calcium-dependent signalling cascades [228] and the *M. persicae* effector Mp10 suppresses ROS bursts [89, 92]. Recently Wang et al. [226] identified the camalexin synthesis promoting transcription factor WRKY33 as direct target of the whitefly Bsp9 effector [66, 229]. This suggests that phloem feeding insect effectors can act on early defence signalling cascades and also on the transcription of defence genes [49, 87, 228]. I therefore assumed that differences in the transcriptome response between probing and feeding could unravel important immune system components that *M. persicae* manipulates and uses as molecular cue to establish feeding sites. I tested this by comparing probing and feeding sites for discriminative genes. Using PCA analysis I found 71 gene candidates which contribute to the difference between probing and feeding sites. The 71 genes were spatially higher expressed in the probing replicates but not elicited in the feeding samples. The 71 genes enriched for protein catabolic and vacuole transport processes. As genes of protein degrading systems are linked with defence responses [230] and proteasomal systems are targets for plant pathogen effectors [231, 232], perturbation of proteasomal component expression could play an important role in establishing a feeding site for *M. persicae* [233]. Using STRING [193] network association analysis I detected 3 small clusters composed of 8, 9 and 12 genes with *UBQ3*, *AKINbeta1*, *KING1* and *BBD2* as most influential nodes.

Literature search revealed that *UBQ3* is described to be involved in protein degradation [222]. *AKINbeta1* and *KING1* encode subunits for the SnRK1 kinase - an important regulator of the sugar and energy metabolism [223, 224]. SnRK1 is inhibited by trehalose 6-phosphate [234] and involved in delaying senescence [235] - both mechanisms that are associated with defences to *M. persicae* [95]. *BBD2* is involved in negative regulation of the JA signalling components JMT and a potential positive regulator of JR2 [236] and so also involved in an effective

pathway of insect resistance [95]. This indicates that for the establishment of a feeding site *M. persicae* potentially modulates a plants energy sensing system, JA signalling pathway and protein degradation system during early probing.

Material and Methods

Plant growing conditions

Plants were grown as described in **Chapter 3 Plant growing conditions – page 78**.

Preparation staged of *M. persicae A. thaliana* colonies

All experiments were performed using *M. persicae* aphids at an age of 7 – 11 days. To obtain staged aphid colonies, I placed 20 - 25 aphids from an existing clonal Chinese cabbage population on a 4 - 5 week old *A. thaliana* plant and incubated the plant for 24 hours at 22 °C day and night temperature, 48 % humidity, 16 hours photoperiod (2 am – 6 pm) to induce nymph production. After 24 hours the adult aphids were removed and the plants with the nymphs were transferred to a controlled environment room with the following conditions: 18 °C day and 16 °C night temperature, 48 % humidity, 8 hours photoperiod (10 am – 6 pm). Over the next 11 days I incubated the plants in this room and used the aphids at an age of 7 – 11 days for experiments.

EPG setup

I performed EPG experiments using a GIGA-8 DC EPG Amplifier (EPG Systems EU) set up in a faraday cage. To build insect electrodes, I connected the aphids with a 12.5 µm strong gold fibre to a 0.2 mm thick copper wire using a water-based silver glue (SCP03B, silver conductive paint, Electrolube).

EPG sampling: probing phase

To collect samples for the aphid probing dataset, I restricted the position of the insect to a specific area on a leaf (i.e. square-4 next to the midrib on the abaxial left half of the leaf). For this I used the following set-up: Leaves of 4 – 5 week old *A. thaliana* plants were selected for similar size and age. Before each EPG experiment selected leaves of a fresh batch of plants were very carefully rotated by 180 degrees around the stem axis while wearing clean nitrile gloves (67-6233, Slaughter Ltd R&L, U.K.). The abaxial leaf side was then carefully covered with a plastic strip (B8598, Guest Medical Ltd) previously prepared with razor blades excising a ~1 x 1 mm square to allow the insect access to the leaf. I carefully fixated the plastic strip in the soil using syringe needles. At the beginning of each EPG experiment I placed an insect electrode on the open area. I performed a 30-minute EPG experiment. To keep aphids on the open leaf spot, I continuously moved aphids wandering off back using a fine No2 brush (13609, Eastern Shires Purch Org. (ESPO), UK). During the experiment aphid pathway activities were ongoingly observed using the EPG Stylet+d v01.28 software (EPG Systems EU).

Samples with a stylet pathway activity > 15 minutes and with a minimal number of 10 probes were harvested after 30 minutes (the end of the EPG experiment). I carefully removed the plastic strip and pipetted a 1 µl droplet of DNase/RNase free water containing a small amount of inert pencil dust on the aphid exposed area. I then incubated the plants in a controlled environment room with 22 °C day and night temperature, 48% humidity and 10-hour photoperiod (8 am - 6pm). The plants were sampled 0, 0.5, 1, 2, 3, 5, 7 and 24 hours after the first observed insect probing event. At the time-point of sampling I dissected a single lateral cross strip from the leaf composed of 8 ~ 1 x 1 mm squares containing the probing area at square-4.

Each plant was used once; unsuccessful samples with less than 15 minutes pathway activity and less than 10 probes were discarded. I also prepared a control plant for each aphid exposed plant. Control plants were prepared in the same way as the aphid exposed plants but not connected to the EPG device and not loaded

with an insect. I incubated the control plants and successful probing replicates as pairs in the same controlled environment room both labelled with a pencil dust droplet. I sampled corresponding treatment / control pairs at the same time.

EPG sampling: feeding phase

To collect samples for the feeding experiment I used plants (and leaves) of the same age as for the probing experiment. In comparison to the probing experiment I found it necessary to adjust the sampling strategy: in probing experiment aphids frequently wandered away from the presented open leaf area. I also observed that the insects only reached sustained E2 phases when roaming freely on leaves. To associate the feeding spot therefore with the position of the insect, I imaged the insects continuously during the EPG experiment using USB microscope cameras (MS100, 10-200x magnification, 1280x720 resolution, Shekar Android USB Microscope). I kept the plant leaves in camera focus during the time of the experiment by carefully leaning them to a 10/20 µl graduated pipette tip (S1120-3810-C, Starlab UK Ltd.).

During the EPG experiment I imaged the insects with a snapshot every 30 seconds. After an E2 phase > 10 minutes I marked the spot next to the aphid with a black marker pen (183.171, Lyreco UK Ltd.), immediately removed the leaf from the plant and instantly dissected a lateral cross strip at the height of the feeding area in ~ 1 x 1 mm squares. Control plants were not imaged and not exposed to an insect during the experiment but prepared and collected in the same way as the aphid exposed plants.

Leaf dissection

Leaves were prepared as in **Chapter 3 Leaf sectioning and sample harvesting – page 47**. Instead of 4 areas I dissected a lateral cross-section of the leaf in 8 ~1 x 1 mm squares.

Leaf sample lysis and mRNA extraction

I extracted the mRNA of each leaf square following the 96-well plate format workflow as described in **Chapter 2 Leaf mRNA purification – page 49**.

Illumina sequencing library construction

I constructed the sequencing libraries according to the v1.1 of my spatial transcriptomics workflow as described in **Chapter 2 Double-stranded cDNA synthesis reaction – page 50** and **Chapter2 Illumina library preparation from ds-cDNA – page 50**.

Sample pooling and sequencing

I pooled and sequenced libraries as described in **Chapter 3 Sample pooling and sequencing – page 80**.

Data demultiplexing, quality control and mapping

I demultiplexed the sequencing data using bcl2fastq-2.20.0 standard settings. Quality control and mapping was performed as specified in **Chapter 2 Data quality control and mapping – page 56**.

Differential-expression analysis and GO-term analysis

I performed the differential gene expression and GO-term analysis as described in **Chapter 2 Differential-expression analysis, GO-term enrichment – page 56**.

Across leaf DE-gene expression plots were prepared in R-3.5.1 by importing all samples (of all time-points) with DESeq2-1.20.0 and calculating the normalised expression values for all samples. Log2 fold-changes of aphid exposed probing samples vs. controls were calculated as quotient of aphid exposed normalised counts / non-aphid exposed normalised counts. All plots were generated using ggplot2-3.1.0 [164] in R-3.5.1.

Velocyto analysis

I estimated the RNA velocity of leaf squares using the R-3.5.1 package velocyto.R-0.6 [220]. To identify and filter genes associated with circadian rhythms I used the biomaRt-2.40.0 [179, 187] package.

StringDB network analysis

I associated genes to networks using the R-3.5.1 StringDB-1.22.0 [193] package and standard settings (species=3702, version=10, score_threshold=400). I filtered each network for connections between DE genes by removing all connections between non-DE x DE genes and non-DE genes x non-DE genes. I calculated network measures (eigencentrality, closeness centrality, degree centrality and betweenness centrality) on the network clusters using the R-3.5.1 igraph-1.2.4.1 [208] library. Networks were visualised using Cytoscape-3.7.1 [209].

Principal component analysis

Principal component analysis, PCA plots and identification of genes contributing to specific principal components was performed using the R-3.5.1 factoextra-1.0.5.999 [237] package. To test for statistically significant log2 fold-changes between probing and feeding I used the R-3.5.1 built in t.test() and p.adjust(x, method='BH') functions.

Sequencing and sample pooling metrics

Table 9 Sequencing and sample pooling metrics for the aphid probing and aphid feeding experiments: The table lists the number of samples used for each experiment (separated by number of time-points, conditions and the number of used leaves), the number of extracted squares per leaf, the total number of squares processed to sequencing libraries, the number of combined samples for sequencing (per experiment) and the used Sequencing platform and chemistry (SE = single end, PE = paired end). Each leaf was taken from a separate plant.

Experiment	Samples per experiment	Extracted squares per leaf	Total number of leaf squares (i.e. sequencing libraries)	Number of samples combined for sequencing	Sequencing platform and chemistry
Aphid probing	8 time-points, 1 condition, 4 leaves each	8	256	All at once	Illumina NextSeq 500 (75 SE)
Aphid probing controls	8 time-points, 1 condition, 4 leaves each	1	32	All at once	Illumina NextSeq 500 (75 SE)
Aphid feeding	1 'state' (i.e. feeding), 2 conditions, 4 leaves each	8	64	All at once	Illumina NextSeq 500 (75 SE)
Untreated control leaves	3 time-points, 1 condition, 3 leaves each	8	72	All at once	Illumina NextSeq 500 (75 SE)

Contributions

I planned and conducted all experiments. Ashleigh Lister (Earlham Institute, UK) prepared the ds-cDNA synthesis reactions. I constructed the sequencing libraries, performed the quality control, sequenced the libraries and analysed the data.

Chapter-5 Summarising discussion and outlook

Profiling of gene expression patterns in small areas without bulk effects from sample pooling or heterogeneous cell populations allows deeper insights in how a complex organism functions in comparison to gross-scale RNA-seq methods [3, 15–17]. By establishing a robust workflow that combines rapid sample dissection and easy sequencing library preparation, I was able to reconstruct spatial transcriptome patterns in thin tissue slices with a great level of resolution across plant organs and study localised defence responses to pests and pathogens from minute amounts of tissue.

In contrast to specialised plant spatial transcriptome profiling protocols such as LCM [24, 25], FACS of protoplasts [37] or array based technologies [15–18] my workflow (for the first time in plants) allows access to spatial transcriptome data in an easy manner based on easily accessible methods. Similar methods have already been successfully applied to model the 3D transcriptome landscape of a developing zebrafish embryo (i.e. Tomo-seq) [20].

I evolved my spatial transcriptomics workflow by combining elements from existing single-cell RNA-seq [154] and genome (re-)sequencing [155, 238] methods and refined these for stable, low-cost generation (£ 6.00 per library) of sequencing libraries. I tested and optimised several features to efficiently generate ds-cDNA with reduced PCR amplification in ds-cDNA synthesis and Illumina library amplification after Nextera tagmentation. I introduced sample specific dual-indexing barcodes in the Nextera amplification step that enable pooling of 2,304 samples per sequencing run [155].

I compared my workflow with the Illumina TruSeq library preparation protocol and showed that my method compares well with this widely used commercial RNA-seq method. I showed that my workflow enables detection of transcript level differences across 1D leaf sections in leaf elements such as vascular tissues and

margins and therefore the visualisation of transcriptional expression profiles across tissues is possible.

To test my methods performance in disseminating biotic interactions, I challenged *A. thaliana* leaves locally with the bacterial peptide flg22, simulating an encounter with potentially plant pathogenic bacteria [151]. In my experiments I showed the triggering of a local defence response of known flg22 induced biological processes and I found overlap with described FLARE genes from a gross-scale flg22 exposure experiment [134].

Cluster analysis of spatial expression patterns allowed me to describe genes with similar expression patterns as FLARE genes and so potential novel FLAREs. Detailed spatial analysis of gene expression profiles highlighted plant regulatory elements that potentially act as short distance signalling components to a local flg22 stimulus.

I applied my method to detect regulators involved in the defence response to a herbivore attack by challenging plants locally with crude *M. persicae* extract [59]. I showed DE gene overlap with a published experiment [59] and found several DE genes in plant hormone and secondary metabolite synthesis pathways that are involved in mediating plant resistance to herbivores (i.e. JA, SA, ET, glucosinolate and camalexin synthesis) [46, 65, 66, 195]. I found a rapidly induced and spatially quickly spreading transcriptome response to 30 minutes after stimulus and conversion of the spatial to a local immune response in the time-period between 7 to 24 hours.

My analysis showed that the spatiotemporal transcriptome response to aphid extract enriched 86 different PFAMs [178]. I identified 9 regulatory PFAMs with described roles in pathogen perception and immune signalling: (1) WRKY domain genes [139], (2) C2H2 zinc-fingers [180], (3) TIR domain genes [184], (4) AP2 domain genes [140], (5) protein tyrosine kinase genes [183], (6) LRR domain genes [53, 183], (7) tify domain genes [181, 196], (8) Arabidopsis broad spectrum mildew resistance protein RPW8 [185] and (9) AN1-like zinc-finger genes [182].

I found that the early response to herbivore attack involves WRKY domain proteins and AP2 domain proteins, accompanied by C2H2 type zinc fingers and TIR domain members, whereas late responses may be mediated by Arabidopsis broad spectrum mildew resistance protein RPW8 and LRR domain genes. To find the most influential genes among the identified PFAMs I applied network analysis. Network analysis strategies have proven as efficient means in identifying important genes especially when candidates exhibited pleiotropic effects (i.e. influence multiple traits), acted redundantly together with other genes or were lethal when mutated [191, 192, 239, 240]. I detected 324 genes with a high influence in a spatiotemporal DE gene association network. 29 genes I could associate to the previously identified regulators.

Among the most influential genes I found the SA activator ZAT6 [197], the potential SA enhancers WRKY46 and WRKY53 [198], the JA repressors WRKY50, WRKY51 [199] and JAZ1 [200]. Walling et al. hypothesised that *M. persicae* acts on the elicitation of SA signalling to negatively feed-back on the resistance promoting JA signalling pathway [241]. The found DE genes could therefore be important susceptibility factors for *M. persicae* in this negative cross talk.

For the *WRKY53* gene I detected differential expression over multiple time-points from 30 minutes to 24 hours and spatial differential expression at 30 minutes after stimulus, possibly indicating a role as important spatiotemporal regulator of plant to herbivore responses. WRKY53 has recently been associated with a potential role in promoting plant resistance to the Russian wheat aphid [206]. Among the AP2 domain genes I detected the MAPK interactor ERF104 [202], the positive JA/ET mediators ERF5 and ERF6 [204] and ERF107 with a role in regulating glucosinolate synthesis [201] and plant defensin expression [225]. Especially *ERF5* could be of interest for genetic follow up studies as the transcription factor was characterised to be an important regulator in an experimentally tested chitin sensing transcription factor network [203].

In a final experiment I combined the EPG technique and ST-seq to study plant responses to probing and feeding. In contrast to many gross-scale studies that expose leaves to multiple insects for hour-long time-periods [99], I observed a

quickly elicited DE gene response with highly dynamic transcription rates and multiple defence associated biological processes to insects already 30 minutes after probing. In a comparative analysis of early probing time-points (0.5 and 1-hap) with areas where *M. persicae* rapidly established feeding sites (from 30 – 77 minutes), I discovered 71 genes that were spatially higher expressed in probing but not in feeding samples. The 71 genes enriched for protein catabolic and vacuole transport processes. Plant protein degradation systems are linked with defence responses [230] and known targets of pathogen effectors [231, 232]; therefore perturbed proteasomal expression levels could be important cues for *M. persicae* to establish feeding sites. Using association network analysis, I detected UBQ3, AKINbeta1, KING1 and BBD2 as most influential nodes in three small network clusters. UBQ3 is involved in the protein degradation system [222]. AKINbeta1 and KING1 encode subunits for the SnRK1 kinase which over negative trehalose 6-phospate feed-back signalling is connected with the trehalose metabolism [234] and involved in delaying senescence [235] - both pathways are associated with PAD4 mediated defences to *M. persicae* [95]. *BBD2* plays a role in the regulation of JA signalling components [236] and so in an effective pathway that contributes to *M. persicae* resistance [95]. To successfully establish feeding sites, *M. persicae* could therefore actively manipulate a plants energy sensing system, JA signalling pathway and protein degradation components.

To investigate on this finding, I would suggest three experiments:

(1) Genetic analysis of plant single gene knock-out [242] or overexpression lines with high-throughput insect phenotyping to assess feeding performance [44] and fecundity rates [216] on mutant plants. Methods for this analysis are well established [44, 216] and should reveal important resistance or susceptibility factors to *M. persicae* for in depth follow-up studies.

(2) Identification of plant components interacting with already described *M. persicae* effectors [87] in protein-protein interaction pulldown experiments [243]. Identified effector targets that locate closely to the above described 71 genes in an association network could be susceptibility factors that *M. persicae* manipulates to silence defence responses. This analysis could enable resistance

breeding strategies e.g. by exchanging susceptible alleles with potentially resistant alleles from relatives [76] or alteration of susceptibility factors using knock-out/-in technologies [77].

(3) Adaptation of EPG coupled ST-seq to perform RNA tomography sequencing [244] on a time-series of insects attacking a plant (i.e. from landing on a plant to sustained feeding). I assume that during the tested time-period transcriptome changes in the salivary gland and the insect gut could identify receptors and effectors *M. persicae* uses to interpret or manipulate host defence responses. Such genes could be identified by comparing detected transcripts with available genome or transcriptome data from blood [245] and phloem feeding insects (or for effectors functional *in planta* assays [92]) and lead to novel pest control strategies using transgenic methods such as RNAi [216].

I demonstrated that ST-seq is a rapid and robust protocol for spatial whole transcriptome sequencing. However, the method itself still poses two bottlenecks: (1) A lower resolution (i.e. millimetre vs. micrometre scale) in comparison to other NGS spatial transcriptome profiling techniques such as Tomo-seq [244] and the array based spatial transcriptomics methods [16, 17, 246]. (2) A limited throughput of samples due to the mechanical dissection of leaf squares. Whereas array based methods blot entire transcriptomes of small, thinly sectioned tissues [16, 17, 246], mechanical sample dissection limits ST-seq's throughput (especially when studying larger 2D grids). In the future I would therefore improve the methods to test 2D areas at a higher resolution using a rapid mechanical sample dissection. I would improve the protocol as follows:

(1) Establishment of a 384-well / 1,536-well compatible biopsy puncher or microneedle patch (e.g. similar to the patch described by Paul et al. [247]) for rapid 'patch and peel ' plant tissue sampling.

(2) Development of a G&T-seq [150] variant that enables lower volume ds-cDNA synthesis and preparation of Nextera libraries in a one-tube reaction (e.g. by consecutively adding reagents after each protocol step).

(3) Integration of Illumina and full length isoform sequencing compatible unique molecular identifiers (UMIs) to enable accurate quantification of mRNA levels (by counting of the UMI numbers) [248].

■

Appendices

High resolution figures in scalable vector graphics format and all tables necessary to recapitulate the described analysis steps and results are supplied on the accompanying CD. The contents of the CD are:

Appendix_figure2.pdf: High resolution image of thesis figure 2.

Appendix_figure3.pdf: High resolution image of thesis figure 3.

Appendix_figure4.pdf: High resolution image of thesis figure 4.

Appendix_figure5.pdf: High resolution image of thesis figure 5.

Appendix_figure6.pdf: High resolution image of thesis figure 6.

Appendix_figure7.pdf: High resolution image of thesis figure 7.

Appendix_figure8.pdf: High resolution image of thesis figure 8.

Appendix_figure9.pdf: High resolution image of thesis figure 9.

Appendix_figure13.pdf: High resolution image of thesis figure 13.

Appendix_figure14.pdf: High resolution image of thesis figure 14.

Appendix_figure15.pdf: High resolution image of thesis figure 15.

Appendix_figure16.pdf: High resolution image of thesis figure 16.

Appendix_figure17.pdf: High resolution image of thesis figure 17.

Appendix_figure18.pdf: High resolution image of thesis figure 18.

Appendix_figure19.pdf: High resolution image of thesis figure 19.

Appendix_figure20.pdf: High resolution image of thesis figure 20.

Appendix_figure21.pdf: High resolution image of thesis figure 21.

Appendix_figure22.pdf: High resolution image of thesis figure 22.

Appendix_figure23.pdf: High resolution image of thesis figure 23.

Chapter2_additional_file1.pdf: Comparison the described spatial transcriptomics workflow with Illumina TruSeq, qRT-PCR validation of DE gene expression, wounding time-series GO enrichment, wet-lab protocols of the described workflow.

Chapter2_additional_file2.xlsx: Supplementing information to the analysis of chapter 2: costs of the protocol, modified Illumina adapter sequences, wounding DE genes and enriched GO-terms, untreated leaf spatial gene expression data, flg22 infiltration experiment DE genes and enriched GO-terms, flg22 droplet spotting experiment DE genes and enriched GO-terms, detected FLARE overlap, clustering analysis numbers, transcription factor identification, read numbers, protocol use and sequencing information.

Chapter2_additional_file3.xlsx: GO-term enrichments of clustering analysis.

Chapter3_additional_file1.xlsx: Supplementing information to the analysis of chapter 3: detected DE genes, GO-term enrichments, PFAM enrichments, association network clusters with calculated measures (i.e. degree centrality, betweenness centrality, eigencentrality, closeness centrality, cluster size), the identified 29 PFAM, STRING enrichments on the 29 PFAM DE genes.

Chapter3_additional_file2.cys: The STRING association of the detected 29 PFAM genes - Cytoscape-3.7.1 file.

Chapter3_additional_file3.xlsx: AraCyc [207] metabolic pathway associations for all detected DE genes.

Chapter3_additional_file4.pdf: Temporal plots of AraCyc [207] metabolic pathway associations for all detected DE genes. The yellow shaded line in each diagram marks the log2 fold-change range between -1 to 1.

Chapter4_additional_file1.pdf: Sampling information of the harvested EPG traces.

Chapter4_additional_file2.xlsx: Supplementing information to the analysis of chapter 4: DE genes of the probing time-series, detected circadian genes, GO-enrichments of the probing dataset, probing DE gene overlap with published

experiments, the discovered feeding genes, GO-term enrichment and association network analysis of the feeding genes.

Chapter4_additional_file3.pdf: PCA analysis, all vs. all principal components.

Chapter4_additional_file4.pdf: Additional information to the PCA analysis on probing vs. feeding data: PC1 contribution of random subsets or the entire detected transcriptome, log2 fold-changes of feeding vs. probing genes, GO-term enrichment on genes.

Definitions

Base pair (bp)

Damage associated molecular patterns (DAMPs)

Differentially expressed (DE)

Double stranded cDNA (ds-cDNA)

Effector triggered immunity (ETI)

Electrical penetration graph (EPG)

Ethylene (ET)

Fluorescent activated cell sorting (FACS)

Flagellin-22 (flg22)

Flagellin-22 rapidly elicited (FLARE)

Green peach aphid (GPA)

Herbivore associated molecular patterns (HAMPs)

High Density Spatial Transcriptomics (HDST)

Hours after probing (hap)

Hypersensitive response (HR)

Jasmonic acid (JA)

Laser capture microdissection (LCM)

Leucine-rich repeat receptor like kinase (LRR-RLK)

Messenger RNA (mRNA)

Million (M)

Mitogen-activated protein kinase (MAPK)

Next-generation sequencing (NGS)

Pattern triggered immunity (PTI)

Microbe-associated molecular patterns (MAMPs)

Pattern recognition receptors (PRRs)

Principal component analysis (PCA)

Reactive oxygen species (ROS)

Resistance gene (R-gene)

Salicylic acid (SA)

Spatial-transcriptome sequencing (ST-seq)

Unique molecular identifiers (UMIs)

9-Lipoxygenase (9-LOX)

13-Lipoxygenase (13-LOX)

Bibliography

1. Davidson EH. Gene Regulatory Networks and the Evolution of Animal Body Plans. Science. 2006;311:796–800.

2. Peter IS, Davidson EH. Evolution of Gene Regulatory Networks Controlling Body Plan Development. Cell. 2011;144:970–85.

3. Crosetto N, Bienko M, van Oudenaarden A. Spatially resolved transcriptomics and beyond. Nat Rev Genet. 2015;16:57–66.

4. Bridger JM, Volpi EV, editors. Fluorescence in situ Hybridization (FISH). Totowa, NJ: Humana Press; 2010. http://link.springer.com/10.1007/978-1-60761-789-1. Accessed 7 Mar 2015.

5. Ramos-Vara JA. Technical Aspects of Immunohistochemistry. Vet Pathol. 2005;42:405–26.

6. Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature. 2019;568:235–9.

7. Mincarelli L, Lister A, Lipscombe J, Macaulay IC. Defining Cell Identity with Single-Cell Omics. PROTEOMICS. 2018;:1700312.

8. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nat Rev Genet. 2016;17:175–88.

9. Kelsey G, Stegle O, Reik W. Single-cell epigenomics: Recording the past and predicting the future. Science. 2017;358:69–75.

10. Schwartzman O, Tanay A. Single-cell epigenomics: techniques and emerging applications. Nat Rev Genet. 2015;16:716–26.

11. Lovatt D, Ruble BK, Lee J, Dueck H, Kim TK, Fisher S, et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. Nat Methods. 2014;11:190–6.

12. Medaglia C, Giladi A, Stoler-Barak L, De Giovanni M, Salame TM, Biram A, et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. Science. 2017;358:1622–6.

13. Nichterwitz S, Chen G, Aguila Benitez J, Yilmaz M, Storvall H, Cao M, et al. Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. Nat Commun. 2016;7:12139.

14. Chen J, Suo S, Tam PP, Han J-DJ, Peng G, Jing N. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. Nat Protoc. 2017;12:566–80.

15. Giacomello S, Salmén F, Terebieniec BK, Vickovic S, Navarro JF, Alexeyenko A, et al. Spatially resolved transcriptome profiling in model plant species. Nat Plants. 2017;3:17061.

16. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science. 2016;353:78–82.

17. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science. 2019;363:1463–7.

18. Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Äijö T, et al. High-density spatial transcriptomics arrays for *in situ* tissue profiling. preprint. Genomics; 2019. doi:10.1101/563338.

19. 10x Genomics Acquires Spatial Transcriptomics - 10x Genomics. https://www.10xgenomics.com/news/10x-genomics-acquires-spatial-transcriptomics/. Accessed 8 Sep 2019.

20. Junker JP, Noël ES, Guryev V, Peterson KA, Shah G, Huisken J, et al. Genome-wide RNA Tomography in the Zebrafish Embryo. Cell. 2014;159:662–75.

21. Lieben L. Plant genetics: Spatial transcriptomics in plants. Nat Rev Genet. 2017;18:394.

22. Walker RP, Chen Z-H, Johnson KE, Famiani F, Tecsi L, Leegood RC. Using immunohistochemistry to study plant metabolism: the examples of its use in the localization of amino acids in plant tissues, and of phosphoenolpyruvate carboxykinase and its possible role in pH regulation. J Exp Bot. 2001;52:565–76.

23. Duncan S, Olsson T, Hartley M, Dean C, Rosa S. Single Molecule RNA FISH in Arabidopsis Root Cells. BIO-Protoc. 2017;7. doi:10.21769/BioProtoc.2240.

24. Chandran D, Inada N, Hather G, Kleindt CK, Wildermuth MC. Laser microdissection of Arabidopsis cells at the powdery mildew infection site reveals site-specific processes and regulators. Proc Natl Acad Sci. 2010;107:460–5.

25. Kerk NM. Laser Capture Microdissection of Cells from Plant Tissues. PLANT Physiol. 2003;132:27–35.

26. Efroni I, Birnbaum KD. The potential of single-cell profiling in plants. Genome Biol. 2016;17:65.

27. Yuan Y, Lee H, Hu H, Scheben A, Edwards D. Single-Cell Genomic Analysis in Plants. Genes. 2018;9:50.

28. Libault M, Pingault L, Zogli P, Schiefelbein J. Plant Systems Biology at the Single-Cell Level. Trends Plant Sci. 2017;22:949–60.

29. Wai CM, VanBuren R, Zhang J, Huang L, Miao W, Edger PP, et al. Temporal and spatial transcriptomic and micro RNA dynamics of CAM photosynthesis in pineapple. Plant J. 2017;92:19–30.

30. Wang K, Jiang S, Sun C, Lin Y, Yin R, Wang Y, et al. The Spatial and Temporal Transcriptomic Landscapes of Ginseng, Panax ginseng C. A. Meyer. Sci Rep. 2016;5:18283.

31. Cosgrove DJ. Growth of the plant cell wall. Nat Rev Mol Cell Biol. 2005;6:850–61.

32. Bourgaud F, Gravot A, Milesi S, Gontier E. Production of plant secondary metabolites: a historical perspective. Plant Sci. 2001;161:839–51.

33. Damiani I, Baldacci-Cresp F, Hopkins J, Andrio E, Balzergue S, Lecomte P, et al. Plant genes involved in harbouring symbiotic rhizobia or pathogenic nematodes. New Phytol. 2012;194:511–22.

34. Tremblay A, Li S, Scheffler BE, Matthews BF. Laser capture microdissection and expressed sequence tag analysis of uredinia formed by Phakopsora pachyrhizi, the causal agent of Asian soybean rust. Physiol Mol Plant Pathol. 2008;73:163–74.

35. Lenzi L, Caruso C, Bianchedi PL, Pertot I, Perazzolli M. Laser Microdissection of Grapevine Leaves Reveals Site-Specific Regulation of Transcriptional Response to *Plasmopara viticola*. Plant Cell Physiol. 2016;57:69–81.

36. Hacquard S, Delaruelle C, Legué V, Tisserant E, Kohler A, Frey P, et al. Laser Capture Microdissection of Uredinia Formed by *Melampsora larici-populina* Revealed a Transcriptional Switch Between Biotrophy and Sporulation. Mol Plant Microbe Interact. 2010;23:1275–86.

37. Coker TLR, Cevik V, Beynon JL, Gifford ML. Spatial dissection of the Arabidopsis thaliana transcriptional response to downy mildew using Fluorescence Activated Cell Sorting. Front Plant Sci. 2015;6:527.

38. Mulema JMK, Denby KJ. Spatial and temporal transcriptomic analysis of the Arabidopsis thaliana-Botrytis cinerea interaction. Mol Biol Rep. 2012;39:4039–49.

39. Huibers RP, de Jong M, Dekter RW, Van den Ackerveken G. Disease-specific expression of host genes during downy mildew infection of Arabidopsis. Mol Plant-Microbe Interact MPMI. 2009;22:1104–15.

40. Hok S, Danchin EGJ, Allasia V, Panabières F, Attard A, Keller H. An Arabidopsis (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease. Plant Cell Environ. 2011;34:1944–57.

41. Wang W, Barnaby JY, Tada Y, Li H, Tör M, Caldelari D, et al. Timing of plant immune responses by a central circadian regulator. Nature. 2011;470:110–4.

42. Zierold U, Scholz U, Schweizer P. Transcriptome analysis of mlo-mediated resistance in the epidermis of barley. Mol Plant Pathol. 2005;6:139–51.

43. Bruggmann R, Abderhalden O, Reymond P, Dudler R. Analysis of epidermis- and mesophyll-specific transcript accumulation in powdery mildew-inoculated wheat leaves. Plant Mol Biol. 2005;58:247–67.

44. Kloth KJ, ten Broeke CJ, Thoen MP, Hanhart-van den Brink M, Wiegers GL, Krips OE, et al. High-throughput phenotyping of plant resistance to aphids by automated video tracking. Plant Methods. 2015;11:4.

45. Gatehouse JA. Plant resistance towards insect herbivores: a dynamic interaction. New Phytol. 2002;156:145–69.

46. Wu J, Baldwin IT. New Insights into Plant Responses to the Attack from Insect Herbivores. Annu Rev Genet. 2010;44:1–24.

47. Clavijo McCormick A, Unsicker SB, Gershenzon J. The specificity of herbivore-induced plant volatiles in attracting herbivore enemies. Trends Plant Sci. 2012;17:303–10.

48. Erb M, Reymond P. Molecular Interactions Between Plants and Insect Herbivores. Annu Rev Plant Biol. 2019;70:527–57.

49. Fürstenberg-Hägg J, Zagrobelny M, Bak S. Plant Defense against Insect Herbivores. Int J Mol Sci. 2013;14:10242–97.

50. Howe GA, Jander G. Plant Immunity to Insect Herbivores. Annu Rev Plant Biol. 2008;59:41–66.

51. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant–pathogen interactions. Nat Rev Genet. 2010;11:539–48.

52. Boehm T, Swann JB. Origin and Evolution of Adaptive Immunity. Annu Rev Anim Biosci. 2014;2:259–83.

53. Caplan J, Padmanabhan M, Dinesh-Kumar SP. Plant NB-LRR Immune Receptors: From Recognition to Transcriptional Reprogramming. Cell Host Microbe. 2008;3:126–35.

54. Zipfel C. Pattern-recognition receptors in plant innate immunity. Curr Opin Immunol. 2008;20:10–6.

55. Jones JDG, Dangl JL. The plant immune system. Nature. 2006;444:323–9.

56. Thomma BPHJ, Nürnberger T, Joosten MHAJ. Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. Plant Cell. 2011;23:4–15.

57. Jones JDG, Dangl JL. The plant immune system. Nature. 2006;444:323–9.

58. Steinbrenner AD, Muñoz-Amatriaín M, Aguilar Venegas JM, Lo S, Shi D, Holton N, et al. A receptor for herbivore-associated molecular patterns mediates plant immunity. preprint. Plant Biology; 2019. doi:10.1101/679803.

59. Prince DC, Drurey C, Zipfel C, Hogenhout SA. The Leucine-Rich Repeat Receptor-Like Kinase BRASSINOSTEROID INSENSITIVE1-ASSOCIATED KINASE1 and the Cytochrome P450 PHYTOALEXIN DEFICIENT3 Contribute to Innate Immunity to Aphids in Arabidopsis. PLANT Physiol. 2014;164:2207–19.

60. Zebelo SA, Matsui K, Ozawa R, Maffei ME. Plasma membrane potential depolarization and cytosolic calcium flux are early events involved in tomato (Solanum lycopersicon) plant-to-plant communication. Plant Sci. 2012;196:93–100.

61. Vincent TR, Avramova M, Canham J, Higgins P, Bilkey N, Mugford ST, et al. Interplay of Plasma Membrane and Vacuolar Ion Channels, Together with BAK1, Elicits Rapid Cytosolic Calcium Elevations in Arabidopsis during Aphid Feeding. Plant Cell. 2017;:tpc.00136.2017.

62. Mithöfer A, Boland W. Plant Defense Against Herbivores: Chemical Aspects. Annu Rev Plant Biol. 2012;63:431–50.

63. Arimura G -i., Matsui K, Takabayashi J. Chemical and Molecular Ecology of Herbivore-Induced Plant Volatiles: Proximate Factors and Their Ultimate Functions. Plant Cell Physiol. 2009;50:911–23.

64. von Dahl CC, Baldwin IT. Deciphering the Role of Ethylene in Plant–Herbivore Interactions. J Plant Growth Regul. 2007;26:201–9.

65. Morkunas I, Mai VC, Gabryś B. Phytohormonal signaling in plant responses to aphid feeding. Acta Physiol Plant. 2011;33:2057–73.

66. Kettles GJ, Drurey C, Schoonbeek H, Maule AJ, Hogenhout SA. Resistance of *Arabidopsis thaliana* to the green peach aphid, *Myzus persicae* , involves camalexin and is regulated by microRNAs. New Phytol. 2013;198:1178–90.

67. War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, et al. Mechanisms of plant defense against insect herbivores. Plant Signal Behav. 2012;7:1306–20.

68. Donley N. The USA lags behind other agricultural nations in banning harmful pesticides. Environ Health. 2019;18:44.

69. Mallet J. The evolution of insecticide resistance: Have the insects won? Trends Ecol Evol. 1989;4:336–40.

70. Hawkins NJ, Bass C, Dixon A, Neve P. The evolutionary origins of pesticide resistance. Biol Rev. 2019;94:135–55.

71. Ghosh S, Watson A, Gonzalez-Navarro OE, Ramirez-Gonzalez RH, Yanes L, Mendoza-Suárez M, et al. Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. Nat Protoc. 2018;13:2944–63.

72. Giolai M, Paajanen P, Verweij W, Percival-Alwyn L, Baker D, Witek K, et al. Targeted capture and sequencing of gene-sized DNA molecules. BioTechniques. 2016;61. doi:10.2144/000114484.

73. Giolai M, Paajanen P, Verweij W, Witek K, Jones JDG, Clark MD. Comparative analysis of targeted long read sequencing approaches for characterization of a plant's immune receptor repertoire. BMC Genomics. 2017;18. doi:10.1186/s12864-017-3936-7.

74. Steuernagel B, Periyannan SK, Hernández-Pinzón I, Witek K, Rouse MN, Yu G, et al. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. Nat Biotechnol. 2016;34:652–5.

75. Wulff BBH, Moscou MJ. Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. Front Plant Sci. 2014;5. doi:10.3389/fpls.2014.00692.

76. Bent A. Resistance from relatives. Nat Biotechnol. 2016;34:620–621.

77. Borrelli VMG, Brambilla V, Rogowsky P, Marocco A, Lanubile A. The Enhancement of Plant Disease Resistance Using CRISPR/Cas9 Technology. Front Plant Sci. 2018;9:1245.

78. Zaidi SS-A, Mukhtar MS, Mansoor S. Genome Editing: Targeting Susceptibility Genes for Plant Disease Resistance. Trends Biotechnol. 2018;36:898–906.

79. Mathers TC, Chen Y, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, et al. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. Genome Biol. 2017;18:27.

80. Blackman RL, Eastop VF. Aphids on the world's crops: an identification and information guide. 2nd ed. Chichester, West Sussex, England ; New York: Wiley; 2000.

81. Hogenhout SA, Ammar E-D, Whitfield AE, Redinbaugh MG. Insect Vector Interactions with Persistently Transmitted Viruses*. Annu Rev Phytopathol. 2008;46:327–59.

82. Bass C, Puinean AM, Zimmer CT, Denholm I, Field LM, Foster SP, et al. The evolution of insecticide resistance in the peach potato aphid, Myzus persicae. Insect Biochem Mol Biol. 2014;51:41–51.

83. Van Emden HF, Harrington R, editors. Aphids as crop pests / edited by Helmut F. van Emden and Richard Harrington. Wallingford, UK ; Cambridge, MA: CABI; 2007.

84. Chougule NP, Bonning BC. Toxins for transgenic resistance to hemipteran pests. Toxins. 2012;4:405–29.

85. Cook PA, Davies WJ. The use of epifluorescence microscopy as a method for observing aphid stylet tracks. Entomol Exp Appl. 1994;72:91–95.

86. Mugford ST, Barclay E, Drurey C, Findlay KC, Hogenhout SA. An Immuno-Suppressive Aphid Saliva Protein Is Delivered into the Cytosol of Plant Mesophyll Cells During Feeding. Mol Plant Microbe Interact. 2016;29:854–61.

87. Hogenhout SA, Bos JI. Effector proteins that modulate plant–insect interactions. Curr Opin Plant Biol. 2011;14:422–8.

88. Perilla-Henao LM, Casteel CL. Vector-Borne Bacterial Plant Pathogens: Interactions with Hemipteran Insects and Plants. Front Plant Sci. 2016;7:1163.

89. Drurey C, Mathers TC, Prince DC, Wilson C, Caceres-Moreno C, Mugford ST, et al. Chemosensory proteins in the CSP4 clade evolved as plant immunity suppressors before two suborders of plant-feeding hemipteran insects diverged. bioRxiv. 2017;:173278.

90. Pitino M, Coleman AD, Maffei ME, Ridout CJ, Hogenhout SA. Silencing of Aphid Genes by dsRNA Feeding from Plants. PLoS ONE. 2011;6:e25709.

91. Rodriguez PA, Stam R, Warbroek T, Bos JIB. Mp10 and Mp42 from the Aphid Species *Myzus persicae* Trigger Plant Defenses in *Nicotiana benthamiana* Through Different Activities. Mol Plant Microbe Interact. 2014;27:30–9.

92. Bos JIB, Prince D, Pitino M, Maffei ME, Win J, Hogenhout SA. A Functional Genomics Approach Identifies Candidate Effectors from the Aphid Species Myzus persicae (Green Peach Aphid). PLoS Genet. 2010;6:e1001216.

93. Pitino M, Hogenhout SA. Aphid protein effectors promote aphid colonization in a plant species-specific manner. Mol Plant Microbe Interact. 2013;26:130–139.

94. Schulz A. Phloem. Structure Related to Function. In: Behnke H-D, Esser K, Kadereit JW, Lüttge U, Runge M, editors. Progress in Botany. Berlin, Heidelberg: Springer Berlin Heidelberg; 1998. p. 429–75. http://www.springerlink.com/index/10.1007/978-3-642-80446-5_16. Accessed 17 Apr 2016.

95. Louis J, Shah J. Arabidopsis thaliana—Myzus persicae interaction: shaping the understanding of plant defense against phloem-feeding aphids. Front Plant Sci. 2013;4. doi:10.3389/fpls.2013.00213.

96. Tjallingii WF, Esch ThH. Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. Physiol Entomol. 1993;18:317–28.

97. Sun M, Voorrips RE, Steenhuis-Broers G, van't Westende W, Vosman B. Reduced phloem uptake of Myzus persicae on an aphid resistant pepper accession. BMC Plant Biol. 2018;18:138.

98. Badenes-Perez FR, Gershenzon J, Heckel DG. Insect Attraction versus Plant Defense: Young Leaves High in Glucosinolates Stimulate Oviposition by a Specialist Herbivore despite Poor Larval Survival due to High Saponin Content. PLoS ONE. 2014;9:e95766.

99. Foyer CH, Verrall SR, Hancock RD. Systematic analysis of phloem-feeding insect-induced transcriptional reprogramming in Arabidopsis highlights common features and reveals distinct responses to specialist and generalist insects. J Exp Bot. 2015;66:495–512.

100. Kim JH, Jander G. Myzus persicae (green peach aphid) feeding on Arabidopsis induces the formation of a deterrent indole glucosinolate. Plant J Cell Mol Biol. 2007;49:1008–19.

101. Kim JH, Lee BW, Schroeder FC, Jander G. Identification of indole glucosinolate breakdown products with antifeedant effects on Myzus persicae (green peach aphid). Plant J Cell Mol Biol. 2008;54:1015–26.

102. Levy M, Wang Q, Kaspi R, Parrella MP, Abel S. Arabidopsis IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense: IQD1, a positive regulator of glucosinolate accumulation. Plant J. 2005;43:79–96.

103. Beneteau J, Renard D, Marché L, Douville E, Lavenant L, Rahbé Y, et al. Binding Properties of the N-Acetylglucosamine and High-Mannose N-Glycan PP2-A1 Phloem Lectin in Arabidopsis. Plant Physiol. 2010;153:1345–61.

104. Vandenborre G, Smagghe G, Van Damme EJM. Plant lectins as defense proteins against phytophagous insects. Phytochemistry. 2011;72:1538–50.

105. Singh V, Shah J. Tomato responds to green peach aphid infestation with the activation of trehalose metabolism and starch accumulation. Plant Signal Behav. 2012;7:605–7.

106. Singh V, Louis J, Ayre BG, Reese JC, Shah J. TREHALOSE PHOSPHATE SYNTHASE11-dependent trehalose metabolism promotes Arabidopsis thaliana defense against the phloem-feeding insect Myzus persicae: Trehalose signaling in plant defense. Plant J. 2011;67:94–104.

107. Louis J, Shah J. Plant defence against aphids: the PAD4 signalling nexus. J Exp Bot. 2015;66:449–54.

108. Wasternack C, Feussner I. The Oxylipin Pathways: Biochemistry and Function. Annu Rev Plant Biol. 2018;69:363–86.

109. Nalam VJ, Keereetaweep J, Shah J. The green peach aphid, Myzus persicae, acquires a LIPOXYGENASE5-derived oxylipin from Arabidopsis thaliana, which promotes colonization of the host plant. Plant Signal Behav. 2013;8. doi:10.4161/psb.22735.

110. Ellis C, Karafyllidis I, Turner JG. Constitutive activation of jasmonate signaling in an Arabidopsis mutant correlates with enhanced resistance to Erysiphe cichoracearum, Pseudomonas syringae, and Myzus persicae. Mol Plant-Microbe Interact MPMI. 2002;15:1025–30.

111. Nalam VJ, Keereetaweep J, Shah J. The green peach aphid, *Myzus persicae* , acquires a LIPOXYGENASE5-derived oxylipin from *Arabidopsis thaliana* , which promotes colonization of the host plant. Plant Signal Behav. 2013;8:e22735.

112. Pegadaraju V, Knepper C, Reese J, Shah J. Premature Leaf Senescence Modulated by the Arabidopsis PHYTOALEXIN DEFICIENT4 Gene Is Associated with Defense against the Phloem-Feeding Green Peach Aphid. Plant Physiol. 2005;139:1927–34.

113. Pegadaraju V, Louis J, Singh V, Reese JC, Bautor J, Feys BJ, et al. Phloem-based resistance to green peach aphid is controlled by Arabidopsis PHYTOALEXIN DEFICIENT4 without its signaling partner ENHANCED DISEASE SUSCEPTIBILITY1. Plant J. 2007;52:332–41.

114. Machado-Assefh CR, Lucatti AF, Alvarez AE. Induced Senescence Promotes the Feeding Activities and Nymph Development of Myzus persicae (Hemiptera: Aphididae) on Potato Plants. J Insect Sci. 2014;14. doi:10.1093/jisesa/ieu017.

115. Aldamen H, Gerowitt B. Influence of selected potato cultivars on the reproduction rate of the aphid species Myzus persicae (Sulzer) and Macrosiphum euphorbiae (Thomas). J Plant Dis Prot. 2009;116:278–82.

116. Moran PJ, Thompson GA. Molecular Responses to Aphid Feeding in Arabidopsis in Relation to Plant Defense Pathways. Plant Physiol. 2001;125:1074–85.

117. Wu Y, Zhang D, Chu JY, Boyle P, Wang Y, Brindle ID, et al. The Arabidopsis NPR1 Protein Is a Receptor for the Plant Defense Hormone Salicylic Acid. Cell Rep. 2012;1:639–47.

118. Mewis I, Appel HM, Hom A, Raina R, Schultz JC. Major signaling pathways modulate Arabidopsis glucosinolate accumulation and response to both phloem-feeding and chewing insects. Plant Physiol. 2005;138:1149–62.

119. Walling LL. Avoiding Effective Defenses: Strategies Employed by Phloem-Feeding Insects. Plant Physiol. 2008;146:859–66.

120. Zhu L, Guo J, Ma Z, Wang J, Zhou C. Arabidopsis Transcription Factor MYB102 Increases Plant Susceptibility to Aphids by Substantial Activation of Ethylene Biosynthesis. Biomolecules. 2018;8.

121. De Vos M, Jander G. *Myzus persicae* (green peach aphid) salivary components induce defence responses in *Arabidopsis thaliana*. Plant Cell Environ. 2009;32:1548–60.

122. Sun M, Voorrips RE, Vosman B. Aphid populations showing differential levels of virulence on *Capsicum* accessions: Resistance to aphids in Capsicum. Insect Sci. 2018. doi:10.1111/1744-7917.12648.

123. De Vos M, Van Oosten VR, Van Poecke RMP, Van Pelt JA, Pozo MJ, Mueller MJ, et al. Signal Signature and Transcriptome Changes of *Arabidopsis* During Pathogen and Insect Attack. Mol Plant Microbe Interact. 2005;18:923–37.

124. Tjallingii WF. Electrical nature of recorded signals during stylet penetration by aphids. Entomol Exp Appl. 1985;38:177–86.

125. Garzo E, Bonani JP, Lopes JRS, Fereres A. Morphological description of the mouthparts of the Asian citrus psyllid, Diaphorina citri Kuwayama (Hemiptera: Psyllidae). Arthropod Struct Dev. 2012;41:79–86.

126. Müller B, Grossniklaus U. Model organisms — A historical perspective. J Proteomics. 2010;73:2054–63.

127. Rusk N. Genomics: Spatial transcriptomics. Nat Methods. 2016;13:710–1.

128. Ghawana S, Paul A, Kumar H, Kumar A, Singh H, Bhardwaj PK, et al. An RNA isolation system for plant tissues rich in secondary metabolites. BMC Res Notes. 2011;4:85.

129. Borges F, Gardner R, Lopes T, Calarco JP, Boavida LC, Slotkin R, et al. FACS-based purification of Arabidopsis microspores, sperm cells and vegetative nuclei. Plant Methods. 2012;8:44.

130. Deal RB, Henikoff S. The INTACT method for cell type–specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat Protoc. 2010;6:56.

131. Gaiero P, Šimková H, Vrána J, Santiñaque FF, López-Carro B, Folle GA, et al. Intact DNA purified from flow-sorted nuclei unlocks the potential of next-generation genome mapping and assembly in Solanum species. MethodsX. 2018;5:328–36.

132. Chan AC, Khan D, Girard IJ, Becker MG, Millar JL, Sytnik D, et al. Tissue-specific laser microdissection of the *Brassica napus* funiculus improves gene discovery and spatial identification of biological processes. J Exp Bot. 2016;67:3561–71.

133. Martin LBB, Nicolas P, Matas AJ, Shinozaki Y, Catalá C, Rose JKC. Laser microdissection of tomato fruit cell and tissue types for transcriptome profiling. Nat Protoc. 2016;11:2376–88.

134. Navarro L. The Transcriptional Innate Immune Response to flg22. Interplay and Overlap with Avr Gene-Dependent Defense Responses and Bacterial Pathogenesis. PLANT Physiol. 2004;135:1113–28.

135. Farmer E, Farmer E, Mousavi S, Lenglet A. Leaf numbering for experiments on long distance signalling in Arabidopsis. Protoc Exch. 2013. doi:10.1038/protex.2013.071.

136. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE. 2011;6:e21800.

137. Frey BJ, Dueck D. Clustering by Passing Messages Between Data Points. Science. 2007;315:972–6.

138. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome: Tair: Making and Mining the “Gold Standard” Plant Genome. genesis. 2015;53:474–85.

139. Pandey SP, Somssich IE. The Role of WRKY Transcription Factors in Plant Immunity. PLANT Physiol. 2009;150:1648–55.

140. Xie Z, Nolan TM, Jiang H, Yin Y. AP2/ERF Transcription Factor Regulatory Networks in Hormone and Abiotic Stress Responses in Arabidopsis. Front Plant Sci. 2019;10:228.

141. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010;15:573–81.

142. Kulkarni SR, Vaneechoutte D, Van de Velde J, Vandepoele K. TF2Network: predicting transcription factor regulators and gene regulatory networks in Arabidopsis using publicly available binding site information. Nucleic Acids Res. 2018;46:e31–e31.

143. Ryu KH, Huang L, Kang HM, Schiefelbein J. Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. Plant Physiol. 2019;179:1444–56.

144. Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MCP. Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. Dev Cell. 2019;48:840-852.e5.

145. Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, et al. Dynamics of gene expression in single root cells of A. thaliana. Plant Cell. 2019;:tpc.00785.2018.

146. Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, et al. Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host Microbe. 2018;24:168-179.e4.

147. Dangl JL, Horvath DM, Staskawicz BJ. Pivoting the Plant Immune System from Dissection to Deployment. Science. 2013;341:746–51.

148. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. BioTechniques. 2001;30:892–7.

149. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014;9:171–81.

150. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015;12:519–22.

151. Denoux C, Galletti R, Mammarella N, Gopalan S, Werck D, De Lorenzo G, et al. Activation of Defense Response Pathways by OGs and Flg22 Elicitors in Arabidopsis Seedlings. Mol Plant. 2008;1:423–45.

152. Underwood W, Somerville SC. Perception of conserved pathogen elicitors at the plasma membrane leads to relocalization of the Arabidopsis PEN3 transporter. Proc Natl Acad Sci. 2013;110:12492–7.

153. Zhang W, Fraiture M, Kolb D, Löffelhardt B, Desaki Y, Boutrot FFG, et al. *Arabidopsis* RECEPTOR-LIKE PROTEIN30 and Receptor-Like Kinase SUPPRESSOR OF BIR1-1/EVERSHED Mediate Innate Immunity to Necrotrophic Fungi. Plant Cell. 2013;25:4227–41.

154. Picelli S, Björklund \AAsa K., Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods. 2013. http://www.nature.com/nmeth/journal/vaop/ncurrent/full/nmeth.2639.html. Accessed 6 Jan 2015.

155. Beier S, Himmelbach A, Colmsee C, Zhang X-Q, Barrero RA, Zhang Q, et al. Construction of a map-based reference genome sequence for barley, Hordeum vulgare L. Sci Data. 2017;4:170044.

156. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 31 Jan 2017.

157. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:pp–10.

158. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR: Mapping RNA-seq Reads with STAR. In: Bateman A, Pearson WR, Stein LD, Stormo GD, Yates JR, editors. Current Protocols in Bioinformatics. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2015. p. 11.14.1-11.14.19. http://doi.wiley.com/10.1002/0471250953.bi1114s51. Accessed 9 Jan 2017.

159. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinforma Oxf Engl. 2012;28:2184–5.

160. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.

161. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinforma Oxf Engl. 2016;32:3047–8.

162. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15. doi:10.1186/s13059-014-0550-8.

163. Love MI, Anders S, Kim V, Huber W. RNA-Seq workflow: gene-level exploratory analysis and differential expression. F1000Research. 2016;4:1070.

164. Wickham H. Ggplot2: elegant graphics for data analysis. New York: Springer; 2009.

165. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J Integr Biol. 2012;16:284–7.

166. Carlson M. org.At.tair.db: Genome wide annotation for Arabidopsis. 2018.

167. Giolai M, Verweij W, Lister A, Heavens D, Macaulay I, Clark MD. Spatially resolved transcriptomics reveals plant host responses to pathogens. preprint. Genomics; 2019. doi:10.1101/720086.

168. Stork NE. How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? Annu Rev Entomol. 2018;63:31–45.

169. Lankau RA. Specialist and generalist herbivores exert opposing selection on a chemical defense. New Phytol. 2007;175:176–84.

170. Jaouannet M, Morris JA, Hedley PE, Bos JIB. Characterization of Arabidopsis Transcriptional Responses to Different Aphid Species Reveals Genes that Contribute to Host Susceptibility and Non-host Resistance. PLOS Pathog. 2015;11:e1004918.

171. Zipfel C. Pattern-recognition receptors in plant innate immunity. Curr Opin Immunol. 2008;20:10–6.

172. Bigeard J, Colcombet J, Hirt H. Signaling Mechanisms in Pattern-Triggered Immunity (PTI). Mol Plant. 2015;8:521–39.

173. Cui H, Tsuda K, Parker JE. Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. Annu Rev Plant Biol. 2015;66:487–511.

174. Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM. The nematode resistance gene Mi of tomato confers resistance against the potato aphid. Proc Natl Acad Sci U S A. 1998;95:9750–4.

175. Pallipparambil GR, Sayler RJ, Shapiro JP, Thomas JMG, Kring TJ, Goggin FL. Mi-1.2, an R gene for aphid resistance in tomato, has direct negative effects on a zoophytophagous biocontrol agent, Orius insidiosus. J Exp Bot. 2015;66:549–57.

176. Boissot N, Schoeny A, Vanlerberghe-Masutti F. Vat, an Amazing Gene Conferring Resistance to Aphids and Viruses They Carry: From Molecular Structure to Field Effects. Front Plant Sci. 2016;7. doi:10.3389/fpls.2016.01420.

177. Naseem M, Kaltdorf M, Dandekar T. The nexus between growth and defence signalling: auxin and cytokinin modulate plant immune response pathways. J Exp Bot. 2015;66:4885–96.

178. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, et al. The Pfam Protein Families Database. Nucleic Acids Res. 2002;30:276–80.

179. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database J Biol Databases Curation. 2011;2011:bar030.

180. Ciftci-Yilmaz S, Mittler R. The zinc finger network of plants. Cell Mol Life Sci. 2008;65:1150–60.

181. Vanholme B, Grunewald W, Bateman A, Kohchi T, Gheysen G. The tify family previously known as ZIM. Trends Plant Sci. 2007;12:239–44.

182. Vij S, Tyagi AK. A20/AN1 zinc-finger domain-containing proteins in plants and animals represent common elements in stress response. Funct Integr Genomics. 2008;8:301–7.

183. Antolín-Llovera M, Ried MK, Binder A, Parniske M. Receptor Kinase Signaling Pathways in Plant-Microbe Interactions. Annu Rev Phytopathol. 2012;50:451–73.

184. Nandety RS, Caplan JL, Cavanaugh K, Perroud B, Wroblewski T, Michelmore RW, et al. The Role of TIR-NBS and TIR-X Proteins in Plant Basal Defense Responses. PLANT Physiol. 2013;162:1459–72.

185. Barragan CA, Wu R, Kim S-T, Xi W, Habring A, Hagmann J, et al. RPW8/HR repeats control NLR activation in Arabidopsis thaliana. PLOS Genet. 2019;15:e1008313.

186. Castel B, Ngou P, Cevik V, Redkar A, Kim D, Yang Y, et al. Diverse NLR immune receptors activate defence via the RPW 8- NLR NRG 1. New Phytol. 2019;222:966–80.

187. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4:1184–91.

188. Chan SS-K, Kyba M. What is a Master Regulator? J Stem Cell Res Ther. 2013;3. doi:10.4172/2157-7633.1000e114.

189. Gluecksohn-Waelsch S. Lethal Genes and Analysis of Differentiation. Science. 1963;142:1269–76.

190. van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform. 2018;19:575–92.

191. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. Nat Protoc. 2012;7:670–85.

192. Ahmed H, Howton TC, Sun Y, Weinberger N, Belkhadir Y, Mukhtar MS. Network biology discovers pathogen contact points in host protein-protein interactomes. Nat Commun. 2018;9:1–13.

193. 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:D433-437.

194. Ma'ayan A. Introduction to Network Analysis in Systems Biology. Sci Signal. 2011;4:tr5–tr5.

195. Jaouannet M, Rodriguez PA, Thorpe P, Lenoir CJG, MacLeod R, Escudero-Martinez C, et al. Plant immunity in plant-aphid interactions. Front Plant Sci. 2014;5. doi:10.3389/fpls.2014.00663.

196. Turner JG, Ellis C, Devoto A. The jasmonate signal pathway. Plant Cell. 2002;14 suppl 1:S153–S164.

197. Shi H, Wang X, Ye T, Chen F, Deng J, Yang P, et al. The Cysteine2/Histidine2-Type Transcription Factor *ZINC FINGER OF ARABIDOPSIS THALIANA6* Modulates Biotic and Abiotic Stress Responses by Activating Salicylic Acid-Related Genes and

C-REPEAT-BINDING FACTOR Genes in Arabidopsis. Plant Physiol. 2014;165:1367–79.

198. Hu Y, Dong Q, Yu D. Arabidopsis WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen Pseudomonas syringae. Plant Sci Int J Exp Plant Biol. 2012;185–186:288–97.

199. Gao Q-M, Venugopal S, Navarre D, Kachroo A. Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the WRKY50 and WRKY51 proteins. Plant Physiol. 2011;155:464–76.

200. Grunewald W, Vanholme B, Pauwels L, Plovie E, Inzé D, Gheysen G, et al. Expression of the Arabidopsis jasmonate signalling repressor JAZ1/TIFY10A is stimulated by auxin. EMBO Rep. 2009;10:923–8.

201. Li B, Tang M, Nelson A, Caligagan H, Zhou X, Clark-Wiest C, et al. Network-Guided Discovery of Extensive Epistasis between Transcription Factors Involved in Aliphatic Glucosinolate Biosynthesis. Plant Cell. 2018;30:178–95.

202. Bethke G, Unthan T, Uhrig JF, Poschl Y, Gust AA, Scheel D, et al. Flg22 regulates the release of an ethylene response factor substrate from MAP kinase 6 in Arabidopsis thaliana via ethylene signaling. Proc Natl Acad Sci. 2009;106:8067–72.

203. Son GH, Wan J, Kim HJ, Nguyen XC, Chung WS, Hong JC, et al. Ethylene-Responsive Element-Binding Factor 5, ERF5, Is Involved in Chitin-Induced Innate Immunity Response. Mol Plant Microbe Interact. 2012;25:48–60.

204. Moffat CS, Ingle RA, Wathugala DL, Saunders NJ, Knight H, Knight MR. ERF5 and ERF6 Play Redundant Roles as Positive Regulators of JA/Et-Mediated Defense against Botrytis cinerea in Arabidopsis. PLoS ONE. 2012;7:e35995.

205. Miao Y, Zentgraf U. The Antagonist Function of Arabidopsis WRKY53 and ESR/ESP in Leaf Senescence Is Modulated by the Jasmonic and Salicylic Acid Equilibrium. Plant Cell. 2007;19:819–30.

206. Van Eck L, Schultz T, Leach JE, Scofield SR, Peairs FB, Botha A-M, et al. Virus-induced gene silencing of WRKY53 and an inducible phenylalanine ammonia-lyase in wheat reduces aphid resistance: VIGS in wheat reduces aphid resistance. Plant Biotechnol J. 2010;8:1023–32.

207. Mueller LA. AraCyc: A Biochemical Pathway Database for Arabidopsis. PLANT Physiol. 2003;132:453–60.

208. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal. 2006;Complex Systems:1695.

209. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

210. Deutsch CA, Tewksbury JJ, Tigchelaar M, Battisti DS, Merrill SC, Huey RB, et al. Increase in crop losses to insect pests in a warming climate. Science. 2018;361:916–9.

211. Ng JCK, Perry KL. Transmission of plant viruses by aphid vectors. Mol Plant Pathol. 2004;5:505–11.

212. Tsuda K, Somssich IE. Transcriptional networks in plant immunity. New Phytol. 2015;206:932–47.

213. Miller RNG, Costa Alves GS, Van Sluys M-A. Plant immunity: unravelling the complexity of plant responses to biotic stresses. Ann Bot. 2017;119:681–7.

214. Nimchuk Z, Eulgem T, Holt III BF, Dangl JL. Recognition and Response in the Plant Immune System. Annu Rev Genet. 2003;37:579–609.

215. Mitchell C, Brennan RM, Graham J, Karley AJ. Plant Defense against Herbivorous Pests: Exploiting Resistance and Tolerance Traits for Sustainable Crop Protection. Front Plant Sci. 2016;7. doi:10.3389/fpls.2016.01132.

216. Coleman AD, Pitino M, Hogenhout SA. Silencing of aphid genes by feeding on stable transgenic Arabidopsis thaliana. Methods Mol Biol Clifton NJ. 2014;1127:125–36.

217. Powell G. Intracellular salivation is the aphid activity associated with inoculation of non-persistently transmitted viruses. J Gen Virol. 2005;86:469–72.

218. Giordanengo P, Brunissen L, Rusterucci C, Vincent C, van Bel A, Dinant S, et al. Compatible plant-aphid interactions: How aphids manipulate plant responses. C R Biol. 2010;333:516–23.

219. Tjallingii WF. Salivary secretions by aphids interacting with proteins of phloem wound responses. J Exp Bot. 2006;57:739–45.

220. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560:494–8.

221. Falcone Ferreyra ML, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front Plant Sci. 2012;3. doi:10.3389/fpls.2012.00222.

222. Marino D, Peeters N, Rivas S. Ubiquitination during plant immune signaling. Plant Physiol. 2012;160:15–27.

223. Margalha L, Confraria A, Baena-González E. SnRK1 and TOR: modulating growth–defense trade-offs in plant stress responses. J Exp Bot. 2019;70:2261–74.

224. Wurzinger B, Nukarinen E, Nägele T, Weckwerth W, Teige M. The SnRK1 Kinase as Central Mediator of Energy Signaling between Different Organelles1[OPEN]. Plant Physiol. 2018;176:1085–94.

225. Ju S, Go YS, Choi HJ, Park JM, Suh MC. DEWAX Transcription Factor Is Involved in Resistance to Botrytis cinerea in Arabidopsis thaliana and Camelina sativa. Front Plant Sci. 2017;8:1210.

226. Wang N, Zhao P, Ma Y, Yao X, Sun Y, Huang X, et al. A whitefly effector Bsp9 targets host immunity regulator WRKY33 to promote performance. Philos Trans R Soc B Biol Sci. 2019;374:20180313.

227. Hu L, Wu Y, Wu D, Rao W, Guo J, Ma Y, et al. The Coiled-Coil and Nucleotide Binding Domains of BROWN PLANTHOPPER RESISTANCE14 Function in Signaling and Resistance against Planthopper in Rice[OPEN]. Plant Cell. 2017;29:3157–85.

228. Will T, Tjallingii WF, Thönnessen A, Bel AJE van. Molecular sabotage of plant defense by aphid saliva. Proc Natl Acad Sci. 2007;104:10536–41.

229. Birkenbihl RP, Diezel C, Somssich IE. Arabidopsis WRKY33 Is a Key Transcriptional Regulator of Hormonal and Metabolic Responses toward Botrytis cinerea Infection1[W]. Plant Physiol. 2012;159:266–85.

230. Sharma B, Joshi D, Yadav PK, Gupta AK, Bhatt TK. Role of Ubiquitin-Mediated Degradation System in Plant Biology. Front Plant Sci. 2016;7:806.

231. Üstün S, Sheikh A, Gimenez-Ibanez S, Jones A, Ntoukakis V, Börnke F. The Proteasome Acts as a Hub for Plant Immunity and Is Targeted by Pseudomonas Type III Effectors. Plant Physiol. 2016;172:1941–58.

232. Üstün S, Bartetzko V, Börnke F. The Xanthomonas campestris Type III Effector XopJ Targets the Host Cell Proteasome to Suppress Salicylic-Acid Mediated Plant Defence. PLoS Pathog. 2013;9. doi:10.1371/journal.ppat.1003427.

233. Banfield MJ. Perturbation of host ubiquitin systems by plant pathogen/pest effector proteins. Cell Microbiol. 2015;17:18–25.

234. Wingler A, Delatte TL, O'Hara LE, Primavesi LF, Jhurreea D, Paul MJ, et al. Trehalose 6-Phosphate Is Required for the Onset of Leaf Senescence Associated with High Carbon Availability. Plant Physiol. 2012;158:1241–51.

235. Wang J, Guan H, Dong R, Liu C, Liu Q, Liu T, et al. Overexpression of maize sucrose non-fermenting-1-related protein kinase 1 genes, ZmSnRK1s, causes alteration in carbon metabolism and leaf senescence in Arabidopsis thaliana. Gene. 2019;691:34–44.

236. Seo JS, Koo YJ, Jung C, Yeu SY, Song JT, Kim J-K, et al. Identification of a Novel Jasmonate-Responsive Element in the AtJMT Promoter and Its Binding Protein for AtJMT Repression. PLOS ONE. 2013;8:e55482.

237. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. 2019. https://cran.r-project.org/web/packages/factoextra/index.html. Accessed 10 Jun 2019.

238. Karasov TL, Horton MW, Bergelson J. Genomic variability as a driver of plant–pathogen coevolution? Curr Opin Plant Biol. 2014;18:24–30.

239. Kafri R, Dahan O, Levy J, Pilpel Y. Preferential protection of protein interaction network hubs in yeast: Evolved functionality of genetic redundancy. Proc Natl Acad Sci. 2008;105:1243–8.

240. Nguyen T-P, Liu W, Jordán F. Inferring pleiotropy by network analysis: linked diseases in the human PPI network. BMC Syst Biol. 2011;5:179.

241. Walling LL. The myriad plant responses to herbivores. J Plant Growth Regul. 2000;19:195–216.

242. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, et al. Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. Science. 2003;301:653–7.

243. Louche A, Salcedo SP, Bigot S. Protein-Protein Interactions: Pull-Down Assays. Methods Mol Biol Clifton NJ. 2017;1615:247–55.

244. Holler K, Junker JP. RNA Tomography for Spatially Resolved Transcriptomics (Tomo-Seq). Methods Mol Biol Clifton NJ. 2019;1920:129–41.

245. Coutinho-Abreu IV, Guimarães-Costa AB, Valenzuela JG. Impact of insect salivary proteins in blood feeding, host immunity, disease, and in the development of biomarkers for vector exposure. Curr Opin Insect Sci. 2015;10:98–103.

246. Salmen F, Stahl PL, Mollbrink A, Navarro JF, Vickovic S, Frisen J, et al. Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. Nat Protoc. 2018;13:2501–34.

247. Extraction of Plant DNA by Microneedle Patch for Rapid Detection of Plant Diseases | ACS Nano. https://pubs.acs.org/doi/10.1021/acsnano.9b00193. Accessed 15 Sep 2019.

248. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods. 2013;11:163–6.