

Investigating the transcriptional response to bacterial induced growth suppression: pilot study

Sibbe Bakker*

December 23, 2022

Abstract

Interactions between plants and microbes are important for plant health. The collection of microbes that colonise a plant is called the microbiome. It consists of mutualist (members that benefit themselves and the plant), commensals (members that benefit from the plant without hurting the plant) and pathogens (which hurt the plant to their own benefit). Differences between commensals and pathogens are often blurry: Commensal bacteria and pathogenic ones are known to trigger the same immune response in plants. It is this immune response that is under investigation in the current work. Another research group investigated *Arabidopsis thaliana* ecotype Columbia-0 (Col-0) plants that were infected with commensal bacteria that behaved more like pathogens. This group collected RNA-seq data that was re-analysed in this work. A clear immune response was observed in the infected plants. No reason for the altered behaviour of the commensal bacteria could be found within the dataset. An alternative methodology has been proposed to deal with this shortcoming.

Contributions: The data analysis for this report was completed with the help of Suraj Hassan Muralidhar, Laurens Klinkien, Marjolein Relouw, and Erika Castañeda (group desesperados). My contributions involved: (i) DESeq2 R scripts, (ii) Helped design the pipeline and pick the programmes, (iii) FASTQC analysis scripts, (iv) Refactoring R scripts of my team members to include command line arguments, (v) Explaining the use of git, command line and R to my team mates, (vi) making the first version of KEGG and GO scripts (later improved on by Laurens), (vii) A script to determine which genes are shared between samples, (viii) Tying together (some) loose scripts in a snakefile, (ix) Final presentation.

Repository: Supplementary information, such as data, graphics and source codes for the bioinformatic analyses and this L^AT_EX document can be found on Git-Lab: <https://git.wur.nl/sibbe.bakker/BIF30806-project-report.git>.

Contents

1	Introduction	2
1.1	Importance and theory	2
1.2	Aim and research question	2

*Email: sibbe.bakker@wur.nl; student number: 1069349.



2	Methods	3
2.1	Experimental design & methods	3
2.2	Data analysis	3
3	Results	4
4	Discussion	6
4.1	Immune responses elicited by commensals tested in this work. . .	6
4.2	Recommendations for future work	8

1 Introduction

1.1 Importance and theory

Plants are colonised by a diverse array of microbes, these are part of the microbiome. The microbiome of a plant plays an important role in plant health: it can promote disease resistance & abiotic stress resistance [6, 11]. The microbiome may also include pathogenic members [9]. These can cause a lot of damage to the agricultural sector: if plant diseases are not managed, yield loss can easily reach up to 30 % [8]. Contrary to expectation, the differences between pythopathogens and commensal organisms in the microbiome is not clear cut: both are able to manipulate or outmanoeuvre the the immune system of the host plant [31]. For example: the beneficial bacteria *Pseudomonas simiae* WCS417 triggers a similar transcriptional response as its pathogenic ‘twin’ *P. aeruginosa* in *Arabidopsis thaliana* roots [28]. *P. simiae* WCS417 is able to successfully suppress 50 % of the *A. thaliana* immune responses. Whether a microbe is a commensal or a pathogenic member of the microbiome is dependent on context [27]. With this fussiness it is important to have more methods to research the differences in plant response pathogens compared to the response against commensal organisms.

1.2 Aim and research question

Investigating the transcriptomes via RNA sequencing is a good way to understand plant microbe interactions better [26]. A good model organism to investigate plant microbe interactions is *A. thaliana*, a diploid plant for which a lot of genetic information is known [12]. Before the transcriptomic response of *A. thaliana* ecotype Columbia-0 (Col-0) to commensal microbes can be compared to the response to pathogens, it is important to set up a pilot study. Such a pilot study can be used to determine what results can be expected. Additionally, issues with the method may become apparent even on a small dataset. This makes it possible to improve on the methods before the actual RNA-seq experiment starts. In this report such a pilot study will be conducted with a dataset from Maier et al. [21]. This study used 39 isolates commensal bacteria from the Col-0 phyllosphere. Five of these isolates were able to consistently induce a sickly looking ‘stunted growth’ phenotype in gnotobiotic Col-0. The rest of the strains did not cause a notable phenotypic difference. Investigating the underlying reasons for the aforementioned phenotypic difference will be the main aim of this work. Any information gained from this investigation will be used to inform future research.



2 Methods

2.1 Experimental design & methods

Sample selection and treatment To determine the tentative transcriptional differences between healthy plants and plants with a infection related ‘stunted growth’ phenotype, Col-0 data from Maier et al. [21] was re-analysed. Please see the the original work for exact details on the methodology. Maier et al. [21] used RNA-sequencing technology to investigate the transcriptional response to bacterial colonisation. Treatment consisted of growing 10 d old Col-0 plants with only a single bacteria picked from the *At*-SPHERE collection [5]. Sterile Col-0 plants were used as a negative control condition. All plants were harvested after 9 d of growing. Plants were grown in batches of 24 plants per biological replicate. Per experiment, 5 biological replicates were included. Maier et al. [21] observed that the following strains decreased the ability of Col-0 to grow : (i) *Serratia* Leaf51, (ii) *Erwina* Leaf53, (iii) *Xanthomonas* Leaf131, (iv) *Burkholderia* Leaf177.

RNA sequencing All inoculated plants investigated by Maier et al. [21] were subject to RNA-sequencing: The RNA was extracted using a Zymo Quick-RNA plant kit and enriched for poly A tails using the Illumina TruSeq RNA sample prep kit. All samples were single-end sequenced on an Illumina NovaSeq (100 cycles) with a TruSeq SBS v.3 kit. Basecalls were made using the Illumina data processing pipeline. The FASTQ file for each of the samples is hosted on Brightspace.¹ Only the samples: Anxenic, Leaf51, Leaf53, Leaf131 and Leaf177 were analysed further.

2.2 Data analysis

Overview of the analysis All members of the desesperados group contributed to the code base of the bioinformatic data processing and analysis. All programmes were executed on the altschul server of the WUR bioinformatics department, running Ubuntu 18.04.6 LTS (64 bit platform). Where applicable, **snakemake** [18] version 4.3.1 was used to automate the execution of loose scripts and programmes.

Cleanup of FASTQ files Quality checks of FASTQ files were made with **FastQC** v0.11.5 [2] and **MultiQC** v1.13 [14]. To remove the adapter sequences that were still present, **Trimmomatic** v0.33 was used [7]. After trimming, removal of the adapters was again checked using **FastQC** and **MultiQC**.

Mapping RNA sequences to produce read counts The trimmed FASTQ RNA sequences were aligned to the TAIR 10 reference genome [29] using **HISAT2** 2.2.1 [33]. The index for the HISAT2 step was provided by the staff of the BIF308006 course. Results from HISAT2 analysed with **stringtie** v1.3.2d [19]. Summary information from the HISAT2 run was compiled into a report with **MultiQC**. Using **prepDE.py** provided by authors of **stringtie**, the **gff** files produced by the **stringtie** programme were converted to a gene count matrix.

¹https://brightspace.wur.nl/content/enforced/192229-BIF30806_2022_2/RNA_seq_overview.xlsx?_&d2lSessionVal=9cANJDGYdX8bcH2uAHMiMJWox&ou=192229



Analysis of differentially expressed genes Using the R: a language environment for statistical computing (R) v4.2.1 [24] programme DESeq2 v1.34.0 [20]. DESeq2 was used according to the official documentation to calculate \log_2 fold change (FC) (\log_2 FC) values and false detection rates (FDRs) for all genes per experimental conditions, as estimated using the Benjamini-Hochberg method. Because of a lack of information supplied by Maier et al. [21], batch effects could not be accounted for. Before starting the calculation of differentially expressed genes (DEGs) total gene counts over the whole experiment were determined. Any genes that had a total gene count less than 10 were omitted to reduce noise in the analysis. Each experimental treatment was compared to the control value using the Wald test. Additionally, the normalised counts for all genes per experimental condition were calculated. The list with the false discovery rates and \log_2 FC was filtered for an false discovery rate < 0.01 and a FC > 2 .² These thresholds were chosen because a drastic transcriptional shift was expected and only genes that differ in a major way from the negative control are expected to be biologically relevant.

Analysis of differentially expressed genes The list of differentially expressed genes was analysed with R to produce heatmaps and volcano plots to visualise differences in gene expression between the different experimental conditions. Using R, the list of differentially expressed genes was filtered for genes that were present in *all* of the samples. This core set was cross referenced with gene ontology (GO) [3] and Kyoto database of genes and genomes (KEGG) [16], using the R packages topgo [1] and KEGGREST [32] respectively. For significance testing with KEGGREST and topGO, an α value of 0.05 was used.

Clustering analysis Clustering analysis was done using the build-in R function heatmap2.

3 Results

Data quality Quality of the RNA sequencing was high, with Fastqc and Multiqc reporting that all of the samples have FASTQ files with pred scores above 30. Some FASTQ had an adapter content above 1%. After adapter trimming with Trimmomatic, all of the FASTQ files had an adapter content below 1%. Rate of sequence duplication was around 70% for most samples. Leaf51 replicate 1 had the highest percentage of duplicated sequences, with 92.8%. Additionally, about 2% of the sequences in Leaf51 replicate 1 are over represented. Finally, the sequencing depth was variable averaging around 28 million reads per sample, with a standard deviation around 7.

Alignment to TAIR 10 reference genome Using HISAT2, the trimmed FASTQ files were aligned to the TAIR 10 reference genome. Alignment rate of samples was consistently above 99% for all samples except for Leaf51 replicate 1 (Figure 2). Leaf51 replicate 1 aligned differently than the other samples: it has a lower percentage of uniquely mapped reads (around 80.9%) while having a higher percentage multimapped reads (12.0%) and non aligned reads (7.2%).

²Meaning that a $\log_2(4) = 2$ was used.

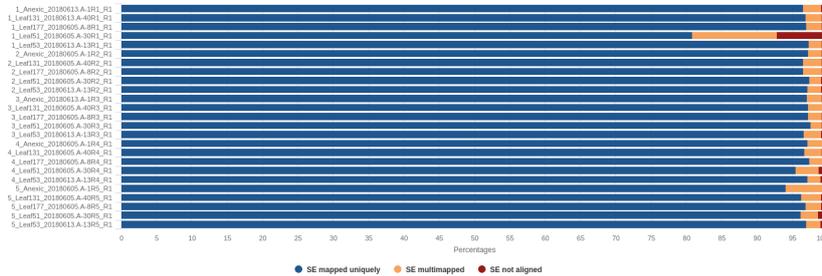


Figure 1: Alignment statistics of the HISAT run.

While omitting the Leaf51 replicate 1 was considered, it was decided against because of the low sample size of the whole experiment.

Determination of degs To find the genes in the stunted Col-0 differed significantly in expression from the axenic Col-0 plants, the DESeq2 programme was used. DESeq2 was able to estimate the dispersion of each gene based on the mean of the counts. Using a Wald test, the p value of each detected gene was calculated. The Benjamini-Hochberg method was used to estimate the FDR for all the genes. Since the phenotypic difference between the control Col-0 and stunted Col-0 were quite severe, the list of detected genes was filtered on a $FC > 2$ and $FDR < 0.01$ to only select for the genes with the most dramatic difference. Using this method, 158 DEGs were detected for Leaf131, 260 for Leaf171, 486 for Leaf51 and 308 for Leaf53.

Replicates of the same condition mostly cluster together, with exception of Leaf51 and Leaf53 (Figure 2). Replicate 4 of Leaf53 has exactly the same gene expression as the Leaf51 samples. Leaf51 replicate 1 is an outlier, as it is not similar to any sample analysed.

With the exception of AT5G01542 (annotated in TAIR as a long non coding RNA), all DEGs detected were upregulated. Of all DEG genes, 103 DEGs are shared between all treatments and are therefore considered to be part of the core set.

The common transcriptional response to bacterial induced stunted growth The whole core gene set is up regulated with all genes having an expression that does not differ much between treatments (Figure 3). Gene AT1G19550, AT5G40990 and AT4G31970 stand out as being consistently more up regulated than the others. AT1G19550 is annotated by KEGG as Glutathione S-transferase family protein. AT5G40990 is annotated by GO as a GDSL lipase 1 involved in immunity against fungi, salicylic acid response and systemic acquired resistance. AT4G31970 is annotated by GO as a cytochrome P450, which is involved in response to hypoxia and defence responses. A go enrichment analysis on the whole core gene set reveals that most of the genes that are differentially expressed in the stunted plants play roles in stress response (Figure 4). There are 2 types of GO terms identified using the Kolmogorov-Smirnov test offered by topGO at a significance threshold of 0.05: (*i*) defence response, (*ii*) decreased oxygen/hypoxia. Using KEGGREST, KEGG pathway ath00380 was found to be significantly up-regulated ($p = 0.04803$). This tryptophan metabolism path-

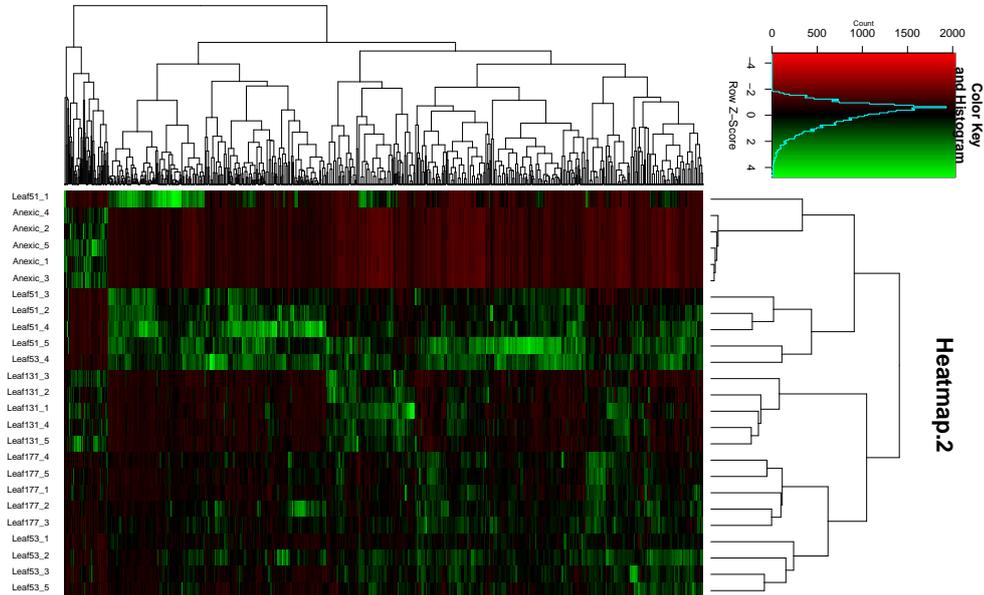


Figure 2: Clustering of scaled normalised counts for all genes with a $FC > 2$. Complete hierarchical clustering using non-parametric spearman correlation as a distance metric.

way is annotated in KEGG as responsible for the conversion of tryptophan to glucobrassicin.

4 Discussion

4.1 Immune responses elicited by commensals tested in this work.

Immune responses in Col-0 plants infected with commensal bacteria

In this work, the RNA-seq data from of Col-0 plants infected with commensal leaf bacteria collected by Maier et al. [21] was analysed. All the 25 plants that were analysed had a stunted growth phenotype. From the GO enrichment, a variety of stress response processes related to plant immune responses were found to be enriched. Terms for hypoxia were identified, these may also be related to plant immune response [30]: when plants are hypoxic, they respond by upregulating ERF-VII transcription factors, these same genes may also have a double role in the response to pathogens [30]. Go term enrichment for cell wall biosynthesis & oxidative stress also indicates some immune response: during infection, plants may thicken their cell wall as a defence [22] and release reactive oxygen species that harm infecting microbes [13]. A KEGG pathway enrichment corroborates the findings of an immune response: The single enriched pathway was for the conversion of tryptophan to glucobrassicin. Glucobrassicin is a phytoanticipin, a molecule that always made by plants to promote innate immunity [23].

Relation to other works When *A. thaliana* is infected with *P. syringae* pv. tomato (*Pst*) DC3000, production of the immunogenic phytohormone salicylic

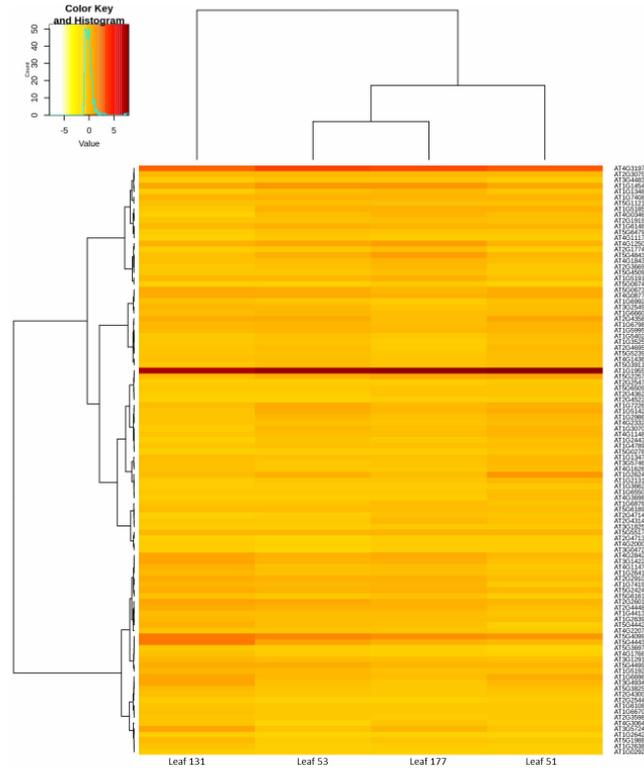


Figure 3: Complete clustering of core genes using Pearson distance metric.

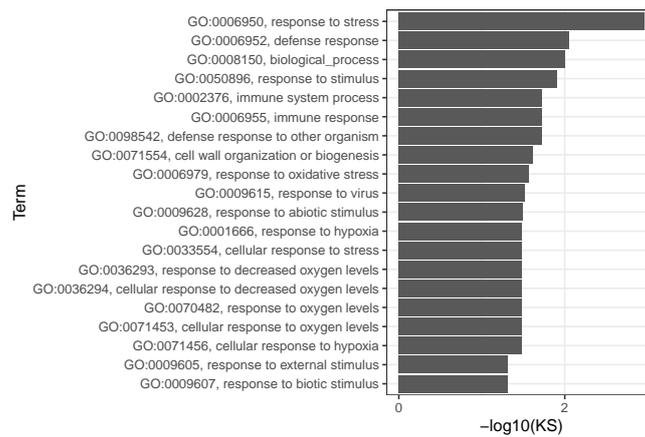


Figure 4: GO enrichment of the core gene set using topGO with the Kolmogorov-Smirnov test. Top 20 values ordered by Kolmogorov-Smirnov statistic. All $p < 0.05$.



acid (SA) is increased [15]. Somewhat un-expectantly, such a response was not found to be common in all of the Maier et al. [21] ‘stunted growth’ plants, even though SA is important in the defence against biotrophic pathogens [15]. This may be because the bacteria used by Maier et al. [21] are not pathogens, but commensals that induce a more drastic response in Col-0. Another option is that some of the bacteria used by Maier et al. [21] are able to suppress the Col-0 immune system in a manner that is similar the suppression of immune responses by *P. simiae* WCS417 [28]. These interactions can be investigated by isolating the flagella from the bacteria in question, and testing which reaction the flagella elicit in the plant. If the flagella elicit a more intense transcriptomic signal than the live bacteria, then there is more evidence for bacterial suppression of plant immunity.

4.2 Recommendations for future work

Improvements to current work. No investigation was made as to how the results of the pipeline differ when different tools would have been used. In a comparison of mappability statistics of Col-0 transcriptomes among the the popular alignment programmes (bwa, HISAT, Kallisto, RSEM, and STAR) conducted by Schaarschmidt et al. [25] was found to have the HISAT2 highest number of reads mapped. Still it would be a good idea to determine whether this holds for the data of Maier et al. [21].

More time should have been taken to investigate the effects of changing parameters of the pipeline. Additionally, distance metrics of the clustering should have been considered with more thought. For example, the core genes were clustered with the pearson distance metric, while the data has a non-normal distribution. In this situation the non-parametric spearman distance metric would have been more fitting.

Lastly, the use of newer reference genomes should have been investigated: The RNA sequencing data was aligned to the TAIR 10 reference genome. This reference genome is a older annotation from 2010. Since then, 12973 more transcript isoforms have been discovered in *A. thaliana* which have been released in [10]. This some of these new isoforms might be have differentially expressed in the samples described in this work, but remained undetected because of the older annotation used here.

Follow up research Even though this work has found an immune reaction in Col-0 plants in response to infection with commensal bacteria, no knowledge about the role of the bacteria in this response was gained, despite the fact that this information is crucial to understand the interactions. By sampling bacterial RNA separately from the plant RNA, it would be possible to learn more about what role the commensal bacteria play in the stunted growth phenotype [17]. The best way to apply this technique would be to repeat the experimental set up of Maier et al. [21] with all the of isolates that cause a stunted growth phenotype as one treatment type, and a subset of the isolates that do not cause a change in phenotype as another treatment type. By increasing the number of replicates to 10 per treatment, it is possible to use 5 samples to do plant RNA-seq, and 5-samples to do RNA seq of the bacteria in the plants. Extraction protocols that enrich for bacterial RNA are available [4].



Concluding remarks Using RNA sequencing, various immune responses were found to be up-regulated in Col-0 plants that were infected with commensal bacteria from the phyllosphere. These plants showed a ‘stunted growth’ phenotype. While this work was not successful in determining the underlying cause for the ‘stunted growth’ phenotype, some methodology was proposed that might be useful for investigating the reasons behind the ‘stunted growth’ phenotype.

References

- [1] Jorg Rahnenfuhrer Adrian Alexa. *topGO*. Bioconductor, 2017. DOI: 10.18129/B9.BIOC.TOPGO.
- [2] Simon Andrews et al. *FastQC*. Babraham Institute. Babraham, UK, Jan. 2012.
- [3] Michael Ashburner et al. “Gene Ontology: Tool for the Unification of Biology”. In: *Nature Genetics* 25.1 (1 May 2000), pp. 25–29. DOI: 10.1038/75556.
- [4] *Bacterial RNA Isolation From Infected Eukaryotic Hosts - NL*. URL: <https://www.thermofisher.com/uk/en/home/references/ambion-tech-support/rna-isolation/tech-notes/bacterial-rna-isolation-from-infected-eukaryotic-hosts.html> (visited on 12/23/2022).
- [5] Yang Bai et al. “Functional Overlap of the Arabidopsis Leaf and Root Microbiota”. In: *Nature* 528.7582 (7582 Dec. 2015), pp. 364–369. DOI: 10.1038/nature16192.
- [6] Gabriele Berg et al. “The Plant Microbiome Explored: Implications for Experimental Botany”. In: *Journal of Experimental Botany* 67.4 (Feb. 1, 2016), pp. 995–1002. DOI: 10.1093/jxb/erv466.
- [7] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data”. In: *Bioinformatics* 30.15 (Aug. 1, 2014), pp. 2114–2120. DOI: 10.1093/bioinformatics/btu170.
- [8] Suresh G. Borkar and Rupert Anand Yumlembam. *Bacterial Diseases of Crop Plants*. CRC Press, Sept. 19, 2016. DOI: 10.1201/9781315367972.
- [9] Vasvi Chaudhry et al. “Shaping the Leaf Microbiota: Plant Microbe Microbe Interactions”. In: *Journal of Experimental Botany* 72.1 (Jan. 20, 2021), pp. 36–56. DOI: 10.1093/jxb/eraa417.
- [10] Chia-Yi Cheng et al. “Araport11: A Complete Reannotation of the Arabidopsis Thaliana Reference Genome”. In: *The Plant Journal* 89.4 (2017), pp. 789–804. DOI: 10.1111/tpj.13415.
- [11] Stéphane Compant et al. “A Review on the Plant Microbiome: Ecology, Functions, and Emerging Trends in Microbial Application”. In: *Journal of Advanced Research*. Special Issue on Plant Microbiome 19 (Sept. 1, 2019), pp. 29–37. DOI: 10.1016/j.jare.2019.03.004.
- [12] Keith R. Davis et al. “Arabidopsis Thaliana as a Model System for Studying Plant-Pathogen Interactions”. In: *Signal Molecules in Plants and Plant-Microbe Interactions*. Ed. by Ben J. J. Lugtenberg. NATO ASI Series. Berlin, Heidelberg: Springer, 1989, pp. 99–106. DOI: 10.1007/978-3-642-74158-6_10.
- [13] Jelena Dumanović et al. “The Significance of Reactive Oxygen Species and Antioxidant Defense System in Plants: A Concise Overview”. In: *Frontiers in Plant Science* 11 (2021).
- [14] Philip Ewels et al. “MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report”. In: *Bioinformatics* 32.19 (Oct. 1, 2016), pp. 3047–3048. DOI: 10.1093/bioinformatics/btw354.
- [15] Bethany Huot et al. “Dual Impact of Elevated Temperature on Plant Defence and Bacterial Virulence in Arabidopsis”. In: *Nature Communications* 8.1 (1 Nov. 27, 2017), p. 1808. DOI: 10.1038/s41467-017-01674-2.



- [16] Minoru Kanehisa et al. “KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs”. In: *Nucleic Acids Research* 45.D1 (Jan. 4, 2017), pp. D353–D361. DOI: 10.1093/nar/gkw1092.
- [17] Jeffrey A. Kimbrel et al. “RNA-Seq for Plant Pathogenic Bacteria”. In: *Genes* 2.4 (4 Dec. 2011), pp. 689–705. DOI: 10.3390/genes2040689.
- [18] Johannes Köster and Sven Rahmann. “Snakemake—a Scalable Bioinformatics Workflow Engine”. In: *Bioinformatics* 28.19 (Oct. 1, 2012), pp. 2520–2522. DOI: 10.1093/bioinformatics/bts480.
- [19] Sam Kovaka et al. “Transcriptome Assembly from Long-Read RNA-seq Alignments with StringTie2”. In: *Genome Biology* 20.1 (Dec. 16, 2019), p. 278. DOI: 10.1186/s13059-019-1910-1.
- [20] Michael I. Love, Wolfgang Huber, and Simon Anders. “Moderated Estimation of Fold Change and Dispersion for RNA-seq Data with DESeq2”. In: *Genome Biology* 15.12 (Dec. 5, 2014), p. 550. DOI: 10.1186/s13059-014-0550-8.
- [21] Benjamin A. Maier et al. “A General Non-Self Response as Part of Plant Immunity”. In: *Nature Plants* 7.5 (5 May 2021), pp. 696–705. DOI: 10.1038/s41477-021-00913-1.
- [22] Frederikke G. Malinovsky, Jonatan U. Fangel, and William G. T. Willats. “The Role of the Cell Wall in Plant Immunity”. In: *Frontiers in Plant Science* 5 (2014).
- [23] Anna Piasecka, Nicolas Jedrzejczak-Rey, and Paweł Bednarek. “Secondary Metabolites in Plant Innate Immunity: Conserved Function of Divergent Chemicals”. In: *New Phytologist* 206.3 (2015), pp. 948–964. DOI: 10.1111/nph.13325.
- [24] R Core Team. *R: A Language and Environment for Statistical Computing*. Version 4.2.1. Vienna, Austria: R Foundation for Statistical Computing, June 23, 2022.
- [25] Stephanie Schaarschmidt et al. “Evaluation of Seven Different RNA-Seq Alignment Tools Based on Experimental Data from the Model Plant Arabidopsis Thaliana”. In: *International Journal of Molecular Sciences* 21.5 (Mar. 3, 2020), p. 1720. DOI: 10.3390/ijms21051720.
- [26] Peer M. Schenk, Lilia C. Carvalhais, and Kemal Kazan. “Unraveling Plant–Microbe Interactions: Can Multi-Species Transcriptomics Help?” In: *Trends in Biotechnology* 30.3 (Mar. 1, 2012), pp. 177–184. DOI: 10.1016/j.tibtech.2011.11.002.
- [27] Ashley Stengel et al. “Rethinking the Roles of Pathogens and Mutualists: Exploring the Continuum of Symbiosis in the Context of Microbial Ecology and Evolution”. In: *Phytobiomes Journal* 6.2 (Jan. 2022), pp. 108–117. DOI: 10.1094/PBIOMES-05-21-0031-P.
- [28] Ioannis A. Stringlis et al. “Root Transcriptional Dynamics Induced by Beneficial Rhizobacteria and Microbial Immune Elicitors Reveal Signatures of Adaptation to Mutualists”. In: *The Plant Journal* 93.1 (2018), pp. 166–180. DOI: 10.1111/tpj.13741.
- [29] TAIR. *TAIR - Download - TAIR10 Genome Release*. URL: https://www.arabidopsis.org/download/index-auto.jsp%3Fdir%3D%252Fdownload_files%252FGenes%252FTAIR10_genome_release (visited on 12/19/2022).
- [30] Hu Tang and Huanhuan Liu. “Roles of Single Gene in Plant Hypoxia and Pathogen Responses”. In: *Plant Signaling & Behavior* 16.10 (2021), p. 1934295. DOI: 10.1080/15592324.2021.1934295.
- [31] Paulo José PL Teixeira et al. “Beyond Pathogens: Microbiota Interactions with the Plant Immune System”. In: *Current Opinion in Microbiology*. Environmental Microbiology 49 (June 1, 2019), pp. 7–17. DOI: 10.1016/j.mib.2019.08.003.
- [32] Dan Tenenbaum. *KEGGREST*. Bioconductor, 2017. DOI: 10.18129/B9.BI0C.KEGGREST.
- [33] Yun Zhang et al. “Rapid and Accurate Alignment of Nucleotide Conversion Sequencing Reads with HISAT-3N”. In: *Genome Research* 31.7 (July 2021), pp. 1290–1295. DOI: 10.1101/gr.275193.120.