Title

The ecology and evolution of polyploid niches: investigating the interaction of ploidy, microbiomes, and pathogens

Author

Song, Michael J

Publication Date

2021

Peer reviewed|Thesis/dissertation

The ecology and evolution of polyploid niches: investigating the interaction of ploidy, microbiomes, and pathogens

by

Michael J. Song

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Carl J. Rothfels, Chair
Associate Professor Britt Koskella
Associate Professor Benjamin Blackman

Summer 2021

The ecology and evolution of polyploid niches: investigating the interaction of ploidy, microbiomes, and pathogens

Abstract

The ecology and evolution of polyploid niches: investigating the interaction of ploidy, microbiomes, and pathogens

by

Michael J. Song

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Assistant Professor Carl J. Rothfels, Chair

Polyploidy or whole-genome duplication (WGD) is a ubiquitous phenomenon in the evolutionary history of land-plants. As WGD often induces many new novel, adaptive, or transgressive phenotypes, it can have cascading ecological effects on biotic interactions and potentially the microbiome. There is a growing body of research on the impact that the microbiome plays in plant ecology, but few studies have looked at the potential interactions between ploidy, microbiomes, and pathogens in shaping the ecology of newly established polyploids. This study uses synthetic auto-tetraploid *Arabidopsis* accessions and a synthetic microbiome representative of natural commensal bacteria in order to assay how these interactions impact host phenotype with respect to pathogen response. In Chapter 1, I describe how the induction of polyploidy does not change the beta diversity of the phyllosphere but does alter the selection of various taxa of the synthetic community. In Chapter 2, I describe a phenomenon whereby polyploids fare better than diploids when treated with a pathogen regardless of inoculation with a protective microbiome, but where diploids treated with a microbiome better arrest the growth of pathogens than the non-treated diploids. In Chapter 3, I perform an RNA-Seq experiment and find a pattern where defense-associated genes are expressed less in diploid accessions than in polyploids when treated with a microbiome. Together, these chapters for the first time demonstrate that a potential consequence of whole genome duplication may be a loss of control over the composition of the microbiome. Finally in chapter 4, I review and synthesize the literature on somatic polyploidy to assess whether endopolyploidy and whole-genome duplication have shared underlying evolutionary rules.

To my parents Joon and Veronica and to my grandparents

Contents

List of Figures

List of Tables

Acknowledgments

I would like to acknowledge all the people who made this possible. Mi gata gemela, Dr. Marianne Kaletzky, my life partner and brilliant interlocutor. My incredible P.I. Dr. Carl Rothfels who inspired me to become a botanist and introduced me to a life of ferns, one day I will name a fern after him. Dr. Sarah Schaack and Dr. Jeremy Coate who set me down this path of science and opened up all the doors. My colleagues Eli Mehlferber, Johan Jaenisch, and Julianne Pelaez who provided invaluable expertise and support on this project, truly I could not have done this without them. My wonderful committee members, Dr. Britt Koskella and Dr. Ben Blackman, who contributed so much time, thought, and resources and who believed in me. The Rothfels and Koskella labs for their invaluable feedback and contributions. Also colleagues Jesus and Carrie, one day we'll have our project done too... The union UAW 2865, except for the reactionary caucus that took over, settled a concessionary contract and sold us out. I would also like to anti-acknowledge the administration that continues to pay poverty-wages to the workers that make the university run. Back to acknowledgements: Comrades Maggie, Shannon, and Marianne (again) for signing me up to the union at the new ASE orientation and instilling in me good radical leftist politics. Union comrades Beezer, Julia, Juliette, Connor, Emily, Alborz, Paul, Blanca, Ronan, Spencer, Thornton, Fiona, Abby, and Audrey. My colleagues in Comparative Literature including but not limited to: my first friend in graduate school Mary Mussman, namesake of the famous Mussman appeal, and comrade Tara Phillips. My colleagues in Slavic, Dom and Kit, and Caroline for forcing me to give a conference paper about a language I hardly know. My colleagues in Classics Emily and Esther, I had a good time in the Medea reading group! My brother Joey who reminded me that "the canon is hegemonic." Jacob and Jules for helping me pass my classes in college. And all the people who have been playing chess with me on chess.com including Jimmy, Galen, Isaac, Ixchel, Keir, and Helen. I finished this dissertation not in spite of but because of all these distractions that kept me in graduate school...

...and, of course, Lyra cat.

Chapter 1

Neo-autopolyploidy induces changes to microbial community recruitment

Abstract

Polyploidy often induces novel, adaptive, or transgressive phenotypes and can have cascading ecological effects on biotic interactions such as the microbiome. I assayed diploid and synthetic autotetraploid accessions of Arabidopsis *that were inoculated with a synthetic microbial community comprising commensal bacteria representative of a broad swath of common plant microbial taxa to discover associations between genotype, ploidy, and the composition of the microbiome and to test whether induced autotetraploidy has an impact on microbiome community structure. I find that while polyploids do not significantly differ from diploids in the overall composition of their microbiome, they do select for different taxa. I also found that while biomass increases with ploidy which could potentially increase the carrying capacity for bacteria, the absolute abundance of bacteria is not statistically different. These findings together are consistent with the theory that the same mechanisms that increase defense in polyploids may alter the recruitment of bacteria associated with diploids.*

1.1 Introduction

Whole-genome duplication (WGD, polyploidy) is a dramatic mutation that can lead to the production of new species in sympatry and the extension or divergence of ecological niches from the parent range (Hijmans et al., 2007; Theodoridis et al., 2013; Molina-Henao and Hopkins, 2019) and it is often considered to be a major driver of evolution (reviewed in Van de Peer et al., 2017). However, whatever the evolutionary consequences of polyploidy are, they are first mediated by the proximate changes to the plant's ecology and the differences between diploids and polyploids with respect to the interactions of these plants and their environment. Polyploidy is associated with many novel and potentially adaptive phenotypes including changes to biomass, photosynthesis, water and nitrogen use efficiency, and

secondary metabolism (Ni et al., 2009; Coate et al., 2012; Huang et al., 2007; Levin, 1983). Polyploidy is known to both be associated with larger cell and organ size (Coate et al., 2012) and with having less endopolyploidy than diploids (Pacey et al., 2020), although with little difference in biomass due to a trade-off between defense and growth (Chen, 2010; Ng et al., 2012), and therefore I hypothesized that there would be differences in the bacterial carrying capacity of diploids and polyploids, even across genotypic differences of the accessions. Although polyploidy is often considered to be a short-term adaptive mechanism used to deal with changes to the environment or stress (reviewed in Van de Peer et al., 2017), we do not know whether whole-genome duplications change the microbiome in a way that influences host traits.

However, plants do not grow in axenic environments nor have they evolved to promote sterility in the root or the shoot system, but rather, plants are colonized by microorganisms that play essential roles in both form and function—notable examples being the mycorrhizal symbiosis, nitrogen fixation in root nodules, and the induction of plant immunity responses triggered by microbe-associated molecular patterns (MAMPs, reviewed in Turner et al. (2013); including both PTI (pathogen-triggered immunity) and ETI (effector-triggered immunity), which not only share many of the same genes and signaling networks while using distinct receptors (Ngou et al., 2021), but also exhibit many positive interactions between the two systems (Zhang et al., 2010)). The microbiome is the microbial community associated with an organism and is composed of bacteria, fungi, protozoa, and viruses–some of whom are facultative or obligate symbionts–that perform various functions for the host. The role of the microbiome in the evolution of the host has been widely theorized about over the last decade, and it is thought to have played a significant role in the evolution of humans (reviewed in Schnorr et al., 2016) and has led to the development of the idea of the holobiont as a unit of selection above the level of the organism *sensu stricto* (reviewed in Rosenberg and Zilber-Rosenberg, 2018). In as far as the function of the microbiome evolves, it can be considered through the interplay between competition between symbionts within the microbiome and control of the microbes themselves by the host (Foster et al., 2017). As the microbiome plays essential roles in immunity, development, nutritional uptake, and drought tolerance in plants (Fitzpatrick et al., 2018), it is apparent that exploring the factors that mediate the interaction between the microbiome and the host plant will be important to understand how these complex systems evolve. In particular, this study examines whether polyploidy influences the microbiome in a way that consistently changes community structure with potential implications for pathogen response, a topic treated in Chapter 2.

Our study focuses on the phyllosphere, which is the above-ground habitat for microorganisms. Across all living plants, the phyllosphere collectively harbors enough bacteria to influence carbon and nitrogen cycles, and many metabolic functions and commensal bacteria have been found to be both shared and abundant across plant taxa (Delmotte et al., 2009). Nonetheless, the same metabolic function and ecological niche can be filled by several different taxa, and the complex interplay between host and microbiome is mediated by and can be modulated significantly by even small changes to the host genome (Turner et al., 2013).

Previous studies have looked at the potential role that commensal bacteria found in

the phyllosphere could play in host pathogen response. In particular, a controlled study assessing two common genera, *Methylobacterium* (Rhizobiales) and *Sphingomonas* (Sphingomonadales), found that *Sphingomonas* both suppressed disease symptoms and diminished the growth of the plant pathogen *Pseudomonas syringae* pv. tomato DC3000 (Innerebner et al., 2011). However, this study was limited in the scope of taxa examined, and therefore I utilized a synthetic community comprised of taxa that are common commensals in tomato (courtesy of the Koskella Lab) to examine the interplay of these commensals as potentially influenced by ploidy level.

Likewise, if microbiomes evolve with their hosts, we may expect the compositional similarity of microbiomes to parallel the phylogenetic relationships of the host–a pattern called phylosymbiosis (Brooks et al., 2016). Although phylosymbiosis has been found to be a common phenomenon, it is not ubiquitous and therefore provides a framework for testable hypotheses about host-microbial associated ecology and evolution (Lim and Bordenstein, 2020). One common pattern that has emerged is that there is less intraspecific variation of the microbiome than interspecific variation (Brooks et al., 2016). However, one major form of speciation in plants is WGD, whereby a new species could be formed, and it is unknown how this process impacts the microbiome.

It has already been shown that modest levels of host-genotype-dependent interactions characterize the metagenome of *Arabidopsis thaliana* (Lundberg et al., 2012) and that polyploidy is easy to induce using colchicine (reviewed in Dermen, 1940). By using synthetic tetra-autoploid accessions of Arabidopsis in conjunction with a synthetic microbial community comprising commensal bacteria representative of a broad swath of common plant microbial taxa, we are able to have an *a priori* experimental design to assay whether there are associations between genotype, ploidy, microbiome, and phenotype. I test the hypothesis that the induction of polyploidy will consistently change the recruitment of host-associated microbiota.

1.2 Methods

Arabidopsis accessions

I received from Luca Comai's lab seeds for 14 total lines from seven Arabidopsis diploid accessions from natural populations and their colchicine-induced autotetraploids: Columbia (Col-0), Warschau (Wa-1), Wassilewskija (Ws-2), Gudow (Gd-1), HR (HR-5), Sorbo (Sorbo), St. Maria d. Feiria (Fei-0). Euploidy was confirmed using RNA-seq (Chapter 3).

Plant Growth Conditions

Seeds were surface sterilized by treatment with 70% ethanol for 2 min and then sodium hypochlorite solution (7% available chlorine) containing 0.2% Triton X-100 for 8 min. Samples were then washed seven times with sterile double-distilled H2O (Bhardwaj et al., 2011).

Seeds were then placed on MS media with .8% agar and cold stratified for two to three days at 4°C in the dark (Bhardwaj et al., 2011). After germination, seedlings were transferred to a controlled environment with a long-day photoperiod (16-h photoperiod) at 22°C and 55% relative humidity with cool white fluorescent light (Bhardwaj et al., 2011). After seven days the seedlings were transferred to sterile peat and the lighting was changed to short-day conditions (9-h photoperiod) (Innerebner et al., 2011).

Innoculation with synthetic community

The synthetic community is composed of 25 taxa that span the diversity of microbial variation in tomato (Elijah Mehlferber, personal correspondence, Table 1.1).

In particular, our synthetic community includes *Pantoea agglomerans* (synonym *Erwinia herbicola*), which is known to be an antogonist to pathogens such as other *Pantoea* and *Pseudomonas syringae* pv. syringae (Beer et al., 1983; Braun-Kiewnick et al., 2000), as well as commensal members of *Pseudomonas*, which is a genus that includes beneficial species that suppress pathogens and promote growth (Mercado-Blanco and Bakker, 2007; Yao et al., 2010).

Two weeks after germination, each plant was inoculated with either the synthetic community suspended in 10Mm MgCl buffer or just the 10Mm MgCl buffer as a control. The plants were inoculated by spraying the plant until saturation.

Amplification and Sequencing of Microbial 16S rDNA

The synthetic community was then assayed three weeks following germination where the plants were all approximately at the same stage of development (Johan Jaenisch, personal correspondence): Stage 1.10, ten rosette leaves >1 mm in length (Boyes et al., 2001). Leaf samples were frozen and kept at -4°C and sent out to Microbiome Insights for 16S V4 sequencing and qPCR analysis within one month of freezing (quality statistics and metadata summarized in Supplemental Table 1).

Data Analysis

Forward and reverse reads were filtered and trimmed to 230 and 160 base pairs (bp), respectively using the `DADA2` pipeline with default parameters (Callahan et al., 2016). Following denoising and merging reads and removing chimeras (Table 2), I used `DADA2` to infer amplicon sequence variants (ASVs), which are analogous to operational taxonomic units (OTUs) and assigned taxonomy to these ASVs using the `DADA2`-trained SILVA database (Version 132, https://benjjneb.github.io/dada2/training.html). A `GTR+G+I` maximum likelihood (ML) tree was then inferred using the `phangorn` package in `R` (Schliep, 2011) using a neighbor joining tree as the starting point. The ML tree, assigned ASVs, read count data, and sample metadata were combined in a `phyloseq` object (McMurdie and Holmes, 2013) for downstream analyses. Differential microbial changes were calculated using `DESeq2` (Love et al., 2014),

Table 1.1: 16S V4 region of synthetic community members

synthetic community member	16S sequence
Genome1_Frigoribacterium_endophyticum	TACGTAGGGTGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAATCTGGGGGCTCAACCCCCAGCCTGCAGTGGGTACGGGCAGACTAGAGTGCGGTAGGGGAGATTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGATCTCTGGGCCGTAACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGG
Genome2_Pantoea_agglomerans	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTTAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTGAAACTGGCAGGCTTGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome3_Curtobacterium_herbarum	TACGTAGGGTGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAATCCCGAGGCTCAACCTCGGGCTTGCAGTGGGTACGGGCAGACTAGAGTGCGGTAGGGGAGATTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGATCTCTGGGCCGTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGG
Genome4_Erwinia_tasmaniensis	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTCAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome6_[Brevibacterium]_frigoritolerans	TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTCCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGGAACTTGAGTGCAGAAGAGGAAAGTGGAATTCCAAGTGTAGCGGTGAAATGCGTAGAGATTTGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGCAAACAGG
Genome7_[Brevibacterium]_frigoritolerans	TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTCCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGGAACTTGAGTGCAGAAGAGGAAAGTGGAATTCCAAGTGTAGCGGTGAAATGCGTAGAGATTTGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGCAAACAGG
Genome9_Massilia_aurea	TACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGTTTTGTAAGTCTGACGTGAAATCCCCGGGCTCAACCTGGGAATTGCGTTGGAGACTGCAAGGCTAGAATCTGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGAGGAACACCGATGGCGAAGGCAGCCCCCTGGGTCAAGATTGACGCTCATGCACGAAAGCGTGGGGAGCAAACAGG
Genome10_Pseudomonas_asturiensis	TACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGAATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTGGCAAGCTAGAGTAGGGCAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGGCTCATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome13__Exiguobacterium_sibiricum	TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGCCTTTTAAGTCTGATGTGAAAGCCCCCGGCTCAACCGGGGAGGGTCATTGGAAACTGGAAGGCTTGAGTACAGAAGAGAAGAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTTTGGTCTGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGG
Genome14_Pseudomonas_rhizosphaerae	TACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGAATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTGGCAAGCTAGAGTAGGGCAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGGCTCATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome15_Leclercia_adecarboxylata	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome16_Pseudomonas_moraviensis	TACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTGGCAAGCTAGAGTATGGTAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome17_Pseudomonas_moraviensis	TACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTGGCAAGCTAGAGTATGGTAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome19_Pantoea_agglomerans	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTTAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTGAAACTGGCAGGCTTGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome20_Pseudomonas_rhizosphaerae	TACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGAATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTGGCAAGCTAGAGTAGGGCAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGGCTCATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome21_Rathayibacter_festucae	TACGTAGGGTGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAACCCGAGGCTCAACCTCGGGCCTGCAGTGGGTACGGGCAGACTAGAGTGCGGTAGGGGAGAATGGAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGTTCTCTGGGCCGTTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGG
Genome22_Pantoea_agglomerans	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTTAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTGAAACTGGCAGGCTTGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome23_Curtobacterium_pusillum	TACGTAGGGTGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAATCCCGAGGCTCAACCTCGGGCTTGCAGTGGGTACGGGCAGACTAGAGTGCGGTAGGGGAGATTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGATCTCTGGGCCGTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGG
Genome24_Pantoea_allii	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTTAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTGAAACTGGCAGGCTTGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome25_Pseudomonas_rhodesiae	TACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTCAAAACTGACTGACTAGAGTATGGTAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome26_Pantoea_allii	TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTTAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTTGAAACTGGCAGGCTTGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGG
Genome33_Microbacterium_oleivorans	TACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAATCCCGAGGCTCAACCTCGGGTCTGCAGTGGGTACGGGCAGACTAGAGTGCGGTAGGGGAGATTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGATCTCTGGGCCGTTACTGACGCTGAGGAGCGAAAGGGTGGGGAGCAAACAGG
Genome35_Bacillus_wiedmannii	TACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGCAAACAGG
Genome38_Pseudomonas_moraviensis	TACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTGGCAAGCTAGAGTATGGTAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGG

Table 1.2: Sample metadata and counts of reads through each processing step. The *input* reads were the number of initial paired fastq reads. These reads were then trimmed and *filtered* by quality score. *F*orward and *R*everse reads were *denoised* using DADA2, then *merged* to their respective pair. Finally *chimeras* were removed

sample	accession	input	filtered	denoisedF	denoisedR	merged	nonchim
A2	Wa-1 2x	30042	23799	23669	23679	23027	22900
A4	Wa-1 4x	67447	58565	58294	58400	57655	54656
B2	Col-0 2x	94338	75710	75489	75538	74094	70134
B2C	Col-0 2x control	43145	35701	35404	35474	34432	33301
B4	Col-0 4x	68575	59688	59258	59439	58128	55061
B4C	Col-0 4x control	63554	54541	54032	54227	52776	50394
C2	HR5 2x	32021	24622	24418	24455	23791	23763
C4	HR5 4x	68157	57353	56976	57088	56018	54602
D2	Sorbo 2x	48354	41491	41182	41323	40470	39027
D4	Sorbo 4x	84466	73463	72996	73197	71900	69894
E2	Ws-2 2x	74075	64843	64353	64537	63297	60643
E4	Ws-2 4x	62265	53236	52928	53050	52109	50580
F2	Fei-0 2x	12918	9642	9468	9475	9109	9101
F4	Fei-0 4x	37023	31316	31031	31099	30587	30394
G2	Gd-l 2x	56929	49983	49814	49829	49566	49497
G4	Gd-l 4x	52392	45311	44927	45050	43936	42832

and the `phyloseq` and `microbiomeseq` (Ssekagiri et al., 2018) packages were implemented in `R` to calculate changes in alpha and beta diversity. For a permutational analysis of variance (PERMANOVA), data was rarified to 90% of the reads of the least abundant sample and the test was performed using the adonis function in the `vegan` package (v2.5-2, Oksanen et al. (2007)) in `R` with 999 permutations to test whether ploidy or genotype had an effect on beta diversity measures. The `betadisper` function was also used as implemented in the `vegan` package in `R` for the analysis of multivariate homogeneity of group dispersions (variances).

Accession number

Links to DNA sequencing data are available in Supplemental Table A.1.

1.3 Results

qPCR was performed and the V4 16S region was sequenced for 16 total samples: seven diploid accessions of *Arabidopsis thaliana* and their colchicine-induced autotetraploids, and two controls. The sequencing generated an average of 56,000 reads per sample and 45,000 reads per

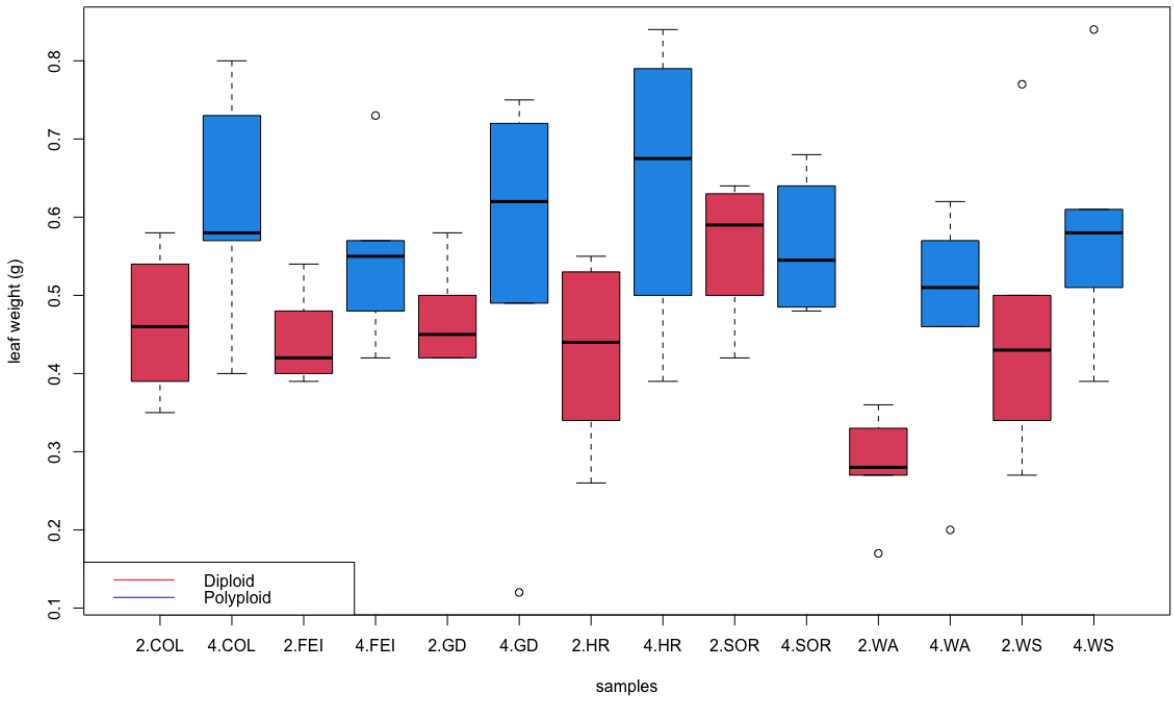


Figure 1.1: Box plot of the weights of the plants at collection

sample following the combined filtering, denoising, merging, and removal of chimeras steps (see Table 1.2 for metadata). Across all 16 samples 400 amplicon sequence variants across seven Linnean taxonomic ranks were identified.

Alpha diversity

Each treated plant was inoculated with a known synthetic microbial community of commensal bacteria known to associate with tomato and to be broadly representative of bacteria naturally associated with the phylosphere. Therefore, we expect there not to be any significant difference in alpha diversity between diploids and polyploids, especially species richness which is simply counts of species, but also Shannon and Simpson indices, which takes into account relative species abundance. A pair-wise ANOVA of diversity measures between groups for Fisher's Alpha, Pielou's evenness, species richness, Shannon, and Simpson indices was performed and we cannot reject the null hypotheses that there isn't a significant differences between diversity measure for diploids and polyploids (Figure 1.2).

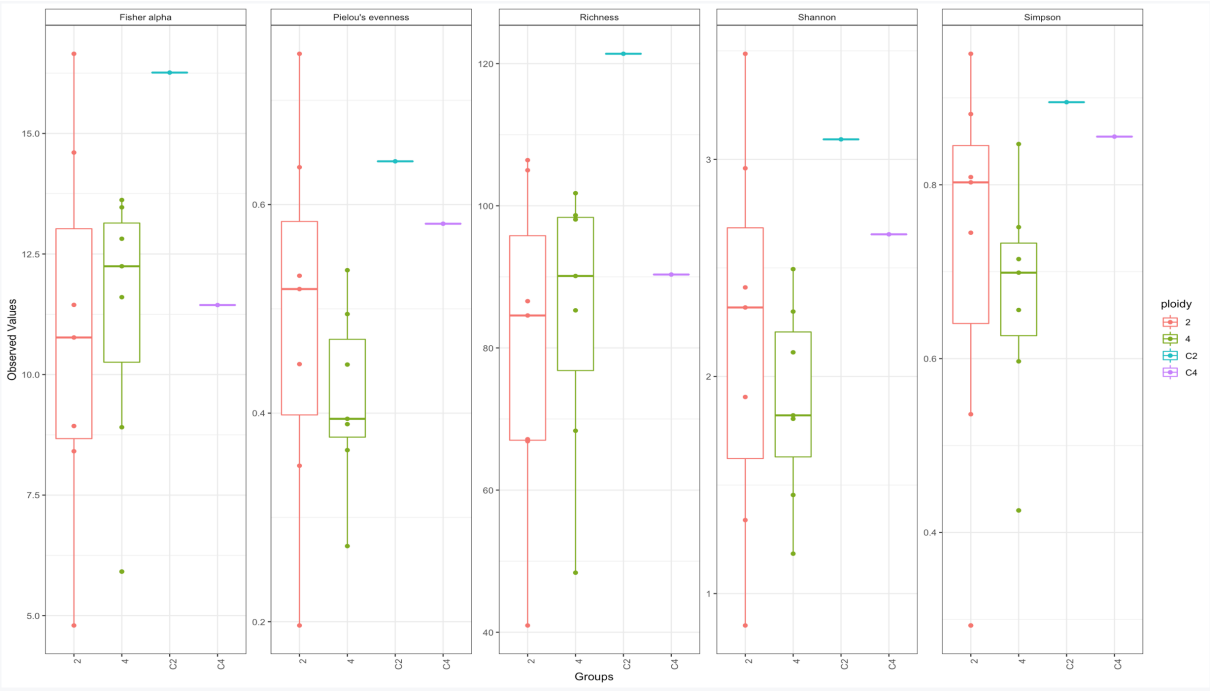


Figure 1.2: Fisher alpha, Pielou evenness, species richness, Simpson, and Shannon indices of alpha diversity grouped by ploidy for experimental samples. Pairwise analysis of variance in diversity was performed between groups and significance below the p value threshold of 0.05 is indicated by asterisks

Beta diversity

I used the `PERMDISP2` procedure to analyze multivariate homogeneity of group dispersions (variances) implemented in the `betadispr` function in the `vegan` package in `R`. For both bray and weighted and unweighted unifrac distances, I did not find any statistically significant differences in beta dispersion after computing Tukey's Honest Significant Differences test. This is further supported when I performed a PERMANOVA on a rarified dataset in order to test whether genotype or ploidy has a significant effect on beta diversity using weighted UniFrac distance, which accounts for differences in relative abundances. For both genotype and ploidy we fail to reject the null hypothesis that these treatments influence microbiome composition and structure (Figures 1.3,1.4).

Ordination plots were then used to arrange samples in two dimensions based on similarity calculated via various indices. The non-metric multidimensional scaling (NMDS) plots of both Bray and weighted UniFrac distances show that diploid and polyploid treatments overlap. However, polyploids have narrowed variation across both axes (Figure 1.4). These results correspond to the plots of relative abundance within the rarified dataset with chloroplast and mitochondrial DNA removed across genotypes and ploidies where there are few

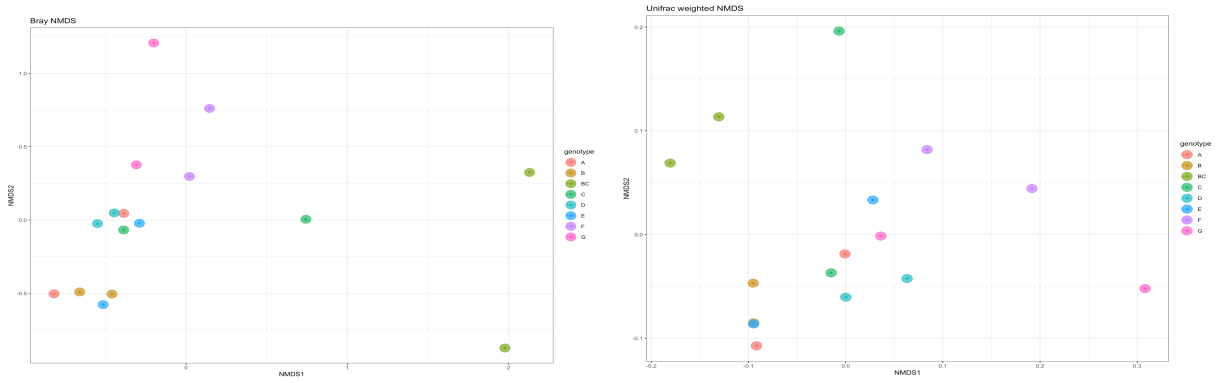


Figure 1.3: Non-metric multidimensional scaling (NMDS) plots based on a Bray-Curtis dissimilarity and weighted UniFrac distance. Experimental samples are colored by genotype

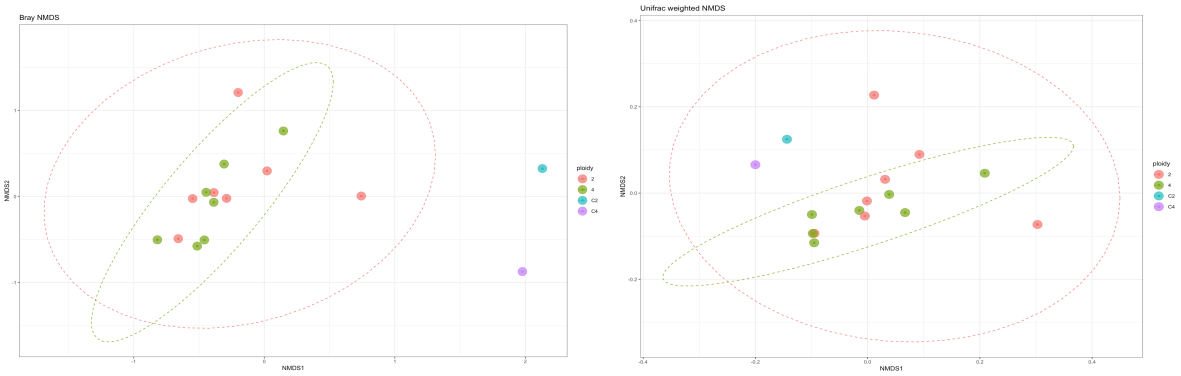


Figure 1.4: Non-metric multidimensional scaling (NMDS) plots based on a Bray-Curtis dissimilarity and weighted UniFrac distance. Experimental samples are colored by ploidy and fitted with normal confidence ellipses

discernible trends on the taxonomic level of Order (Figure 1.5).

Effect on the synthetic community

The potential effect that ploidy level had on the most abundant taxa (expected to contain the members of the synthetic community) was then examined. Using DESeq2, the log twofold change was calculated for the most relatively abundant taxa across samples, and three taxa were identified that significantly increased in abundance consistently in polyploids, however none of them were member of the synthetic community (Table 1.4. When I used the absolute abundances from qPCR and calculated the total bacterial load of each taxa from the relative abundances (Figure 1.7), I found that the synthetic community did not fully establish on

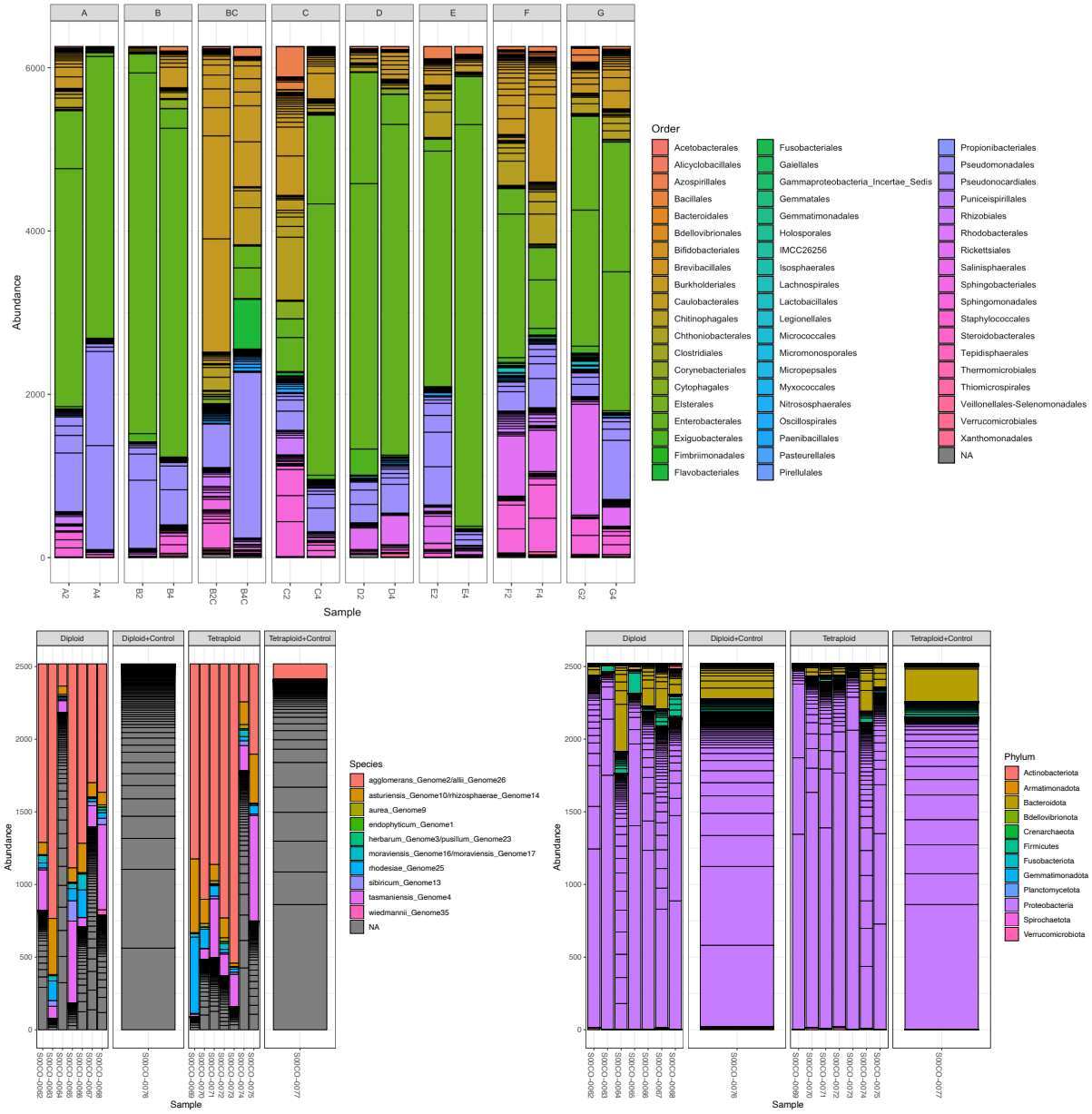


Figure 1.5: Relative abundance of taxa grouped on the level of order shared across rarified experimental samples across all genotypes and ploidy levels

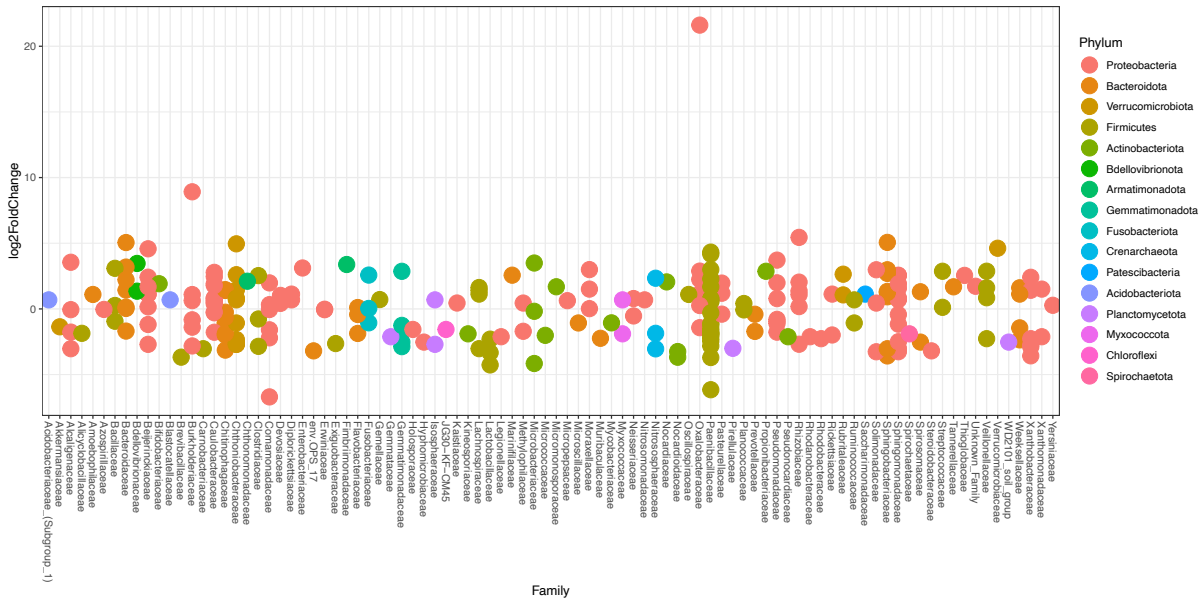


Figure 1.6: Log2-fold change in abundance based on ploidy level for genera based on pairwise comparisons between taxa observed in diploid and polyploid samples

three of the diploid samples. Therefore, I reran the analysis excluding C2, F2, and G2. When I did, I found similar results as without exclusion: ASV 46 (*Massilia* sp.) and ASV 59 (*Burkholdaria-Caballeronia-Paraburkholdaria* sp.) were significantly more abundant in the polyploid taxa, but I also found that ASV69 (*Novosphingobium* sp.) was significantly more abundant and ASV26, which is a member of the synthetic community *E*xiguobacterium sibiricum (Genome 13) was significantly less abundant in the polyploids than the diploids. I also reran the analysis excluding all C, F, and G samples and found that ASV68, a member of *A*llorhizobium-Neorhizobium-Pararhizobium-Rhizobium, was also found to be less abundant in the polyploids, but not *E.* sibiricum.

Absolute Abundance

Interestingly, we found that for our plants, the polyploids had a higher biomass with a mean of 0.44 g for diploids and 0.57 for polyploids (t = -3.4006, df = 60.157, p-value = 0.0012). Using qPCR, the absolute abundance of bacteria on the leaves of the plants was calculated one week following inoculation of the synthetic community. It was found that on average diploids had a higher total bacterial abundance than polyploids after controlling for sample weight, but that difference in means was not significant under a Welch Two-Sample t-test (Table 1.5, Figure 1.7). It was also found that there was no statistical difference when I performed a Welch Two-Sample t-test on the total bacterial abundance not controlling for sample weight.

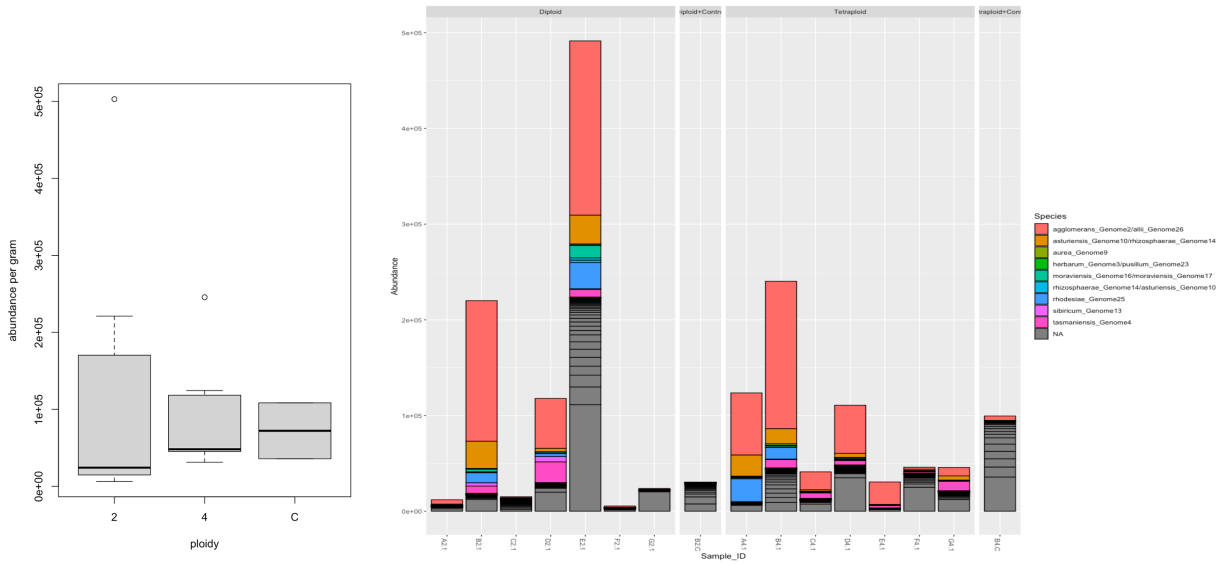


Figure 1.7: Average absolute bacterial load across ploidy levels (left). Total abundance of bacteria across samples (right)

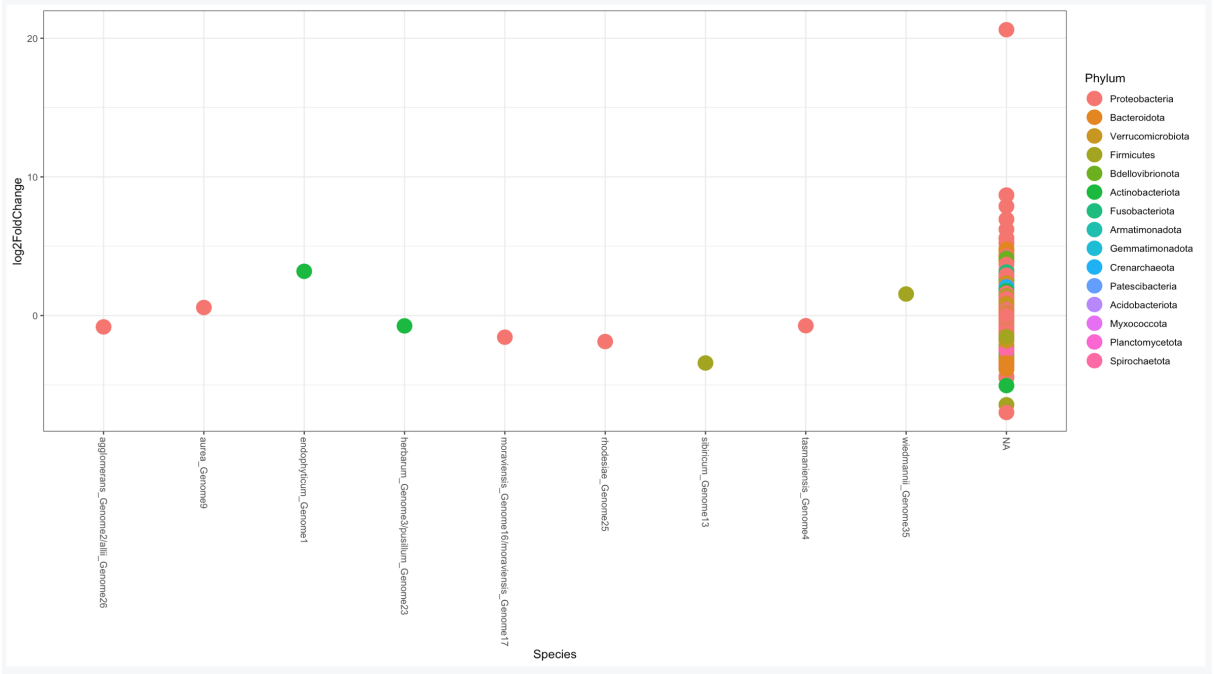


Figure 1.8: Log2-fold change in abundance based on ploidy level for synthetic community members based on pairwise comparisons between taxa observed in diploid and polyploid samples. N/A are taxa that are not in the synthetic community

Table 1.3: Significantly more or less abundant taxa between diploids and polyploids at the 0.1 p-value cut-off

Taxa	baseMean	log2FoldChange	lfcSE	padj	Phylum	Class	Family	Genus
ASV46	11.76168	21.610364	2.988731	1.235684e-10	Proteobacteria	Gammaproteobacteria	Oxalobacteraceae	Massilia
ASV59	46.59497	8.916989	2.161362	4.750829e-03	Proteobacteria	Gammaproteobacteria	Burkholderiaceae	Burkholderia-Caballeronia-Paraburkholderia
ASV128	45.36170	4.619460	1.420864	9.846008e-02	Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiaceae	NA

Table 1.4: Significantly more or less abundant taxa between diploids and polyploids at the 0.1 p-value cut-off after excluding samples where the syn com did not establish

Taxa	baseMean	log2FoldChange	lfcSE	padj	Phylum	Class	Family	Genus
ASV46	14.05766	20.616822	3.169792	1.625125e-08	Proteobacteria	Gammaproteobacteria	Oxalobacteraceae	Massilia
ASV59	57.86638	8.690829	2.463928	2.257471e-02	Proteobacteria	Gammaproteobacteria	Burkholderiaceae	Burkholderia-Caballeronia-Paraburkholderia
ASV69	32.99842	7.881186	2.119346	2.082702e-02	Proteobacteria	Alphaproteobacteria	Sphingomonadaceae	Novosphingobium
ASV26	594.5496	-3.420662	0.972216	0.02257471	Firmicutes	Bacilli	Exiguobacteraceae	Exiguobacterium

Table 1.5: Raw abundance and normalized abundance by sample weight for all samples in study. Values were calculated by qPCR.

Sample.ID	Sample.Type	Group	CT.mean	Quantity	Undiluted.Quantity	ploidy	Leaf Weight (g)	Quant/Weight
InocControl1	Inoculum Control	1	16.61353	225109.8447	225109.8447			
A2.1	Arabidopsis Leaf Epiphytes	2	24.15429	2001.2301	2001.2301	2	0.16	12507.6881
B2.1	Arabidopsis Leaf Epiphytes	2	19.78164	30950.357	30950.357	2	0.14	221073.979
C2.1	Arabidopsis Leaf Epiphytes	2	24.1127	2054.0466	2054.0466	2	0.12	17117.055
D2.1	Arabidopsis Leaf Epiphytes	2	20.45046	20358.3975	20358.3975	2	0.17	119755.279
E2.1	Arabidopsis Leaf Epiphytes	2	18.25592	80475.5961	80475.5961	2	0.16	502972.476
F2.1	Arabidopsis Leaf Epiphytes	2	25.85503	689.7524	689.7524	2	0.11	6270.47636
G2.1	Arabidopsis Leaf Epiphytes	2	23.30837	3399.2826	3399.2826	2	0.14	24280.59
A4.1	Arabidopsis Leaf Epiphytes	2	20.48653	19903.6289	19903.6289	4	0.16	124397.681
B4.1	Arabidopsis Leaf Epiphytes	2	18.82065	56500.8487	56500.8487	4	0.23	245655.864
C4.1	Arabidopsis Leaf Epiphytes	2	21.69644	9329.029	9329.029	4	0.22	42404.6773
D4.1	Arabidopsis Leaf Epiphytes	2	19.59118	34871.3756	34871.3756	4	0.31	112488.308
E4.1	Arabidopsis Leaf Epiphytes	2	22.48607	5635.7745	5635.7745	4	0.18	31309.8583
F4.1	Arabidopsis Leaf Epiphytes	2	21.42869	11069.6687	11069.6687	4	0.23	48128.9943
G4.1	Arabidopsis Leaf Epiphytes	2	21.17524	13013.9472	13013.9472	4	0.27	48199.8044
B2.C	Arabidopsis Leaf Negative Controls	2	22.10907	7169.4245	7169.4245	C	0.2	35847.1225
B4.C	Arabidopsis Leaf Negative Controls	2	20.53998	19523.0969	19523.0969	C	0.18	108461.649

1.4 Discussion

Effect of Polyploidy on microbiome diversity

I assessed whether there is a shared generalizable response of polyploidy on microbiome composition by assessing beta diversity across samples, which in this case I define as compositional dissimilarity between ploidy and between genotypes. By assaying seven different accessions of Arabidopsis, I attempt to distinguish general responses due to polyploidy from unique genotypic effects. Using both Bray-Curtis dissimilarity and weghted UnifFrac distance, I found there was no significant effect of ploidy on the community composition and structure. However, NMDS plots based on weighted UniFrac distance revealed that polyploid samples separated less across on NMDS axis 2 (Figure 1.4). This result leaves open the possibility that polyploidy may decrease variation within the microbiome if we were to look at differences within an accession, but that these differences are not generalizable across the accessions and more work would need to be done to test whether polyploidy reduces variation for a given genotype.

While polyploid plants tend to have larger biomass (Pacey et al., 2020), some autotetraploid Arabidopsis do not (Chen, 2010; Ng et al., 2012). Interestingly, I found that across all accessions polyploid plants weighed significantly more and yet there was no statistical difference between the means of absolute abundance of bacteria (Figures 1.1, 1.7), which could thus possibly be attributed to an increase in defense response of polyploids rather than a physiological limit of the amount of space bacteria could grow. These differences could also be explained by other phenotypes, however, such as stomata size and number (polyploids have larger guard cells than diploids, but fewer stomata per leaf area Robinson et al., 2018). This could potentially play a role in limiting gas exchange that would change the environment of the microbiome. Similarly, in Arabidopsis it has been shown that stomata play an active role in responding to bacterial invasion by closing up (Underwood et al., 2007), potentially further altering gas exchange processes. Likewise, the increase surface-area-to-volume ratio both has transcriptional and homeostatic consequences for the host plant, which could potentially effect microbial environment and recruitment. Some of these effects could be potentially compensated by the negative relationship between endopolyploidy and WGD, where polyploids have less endopolyploid cells than diploids (Pacey et al., 2020). Pacey et al. (2020) calculated the endopolyploidy index (EI, a measurement of mean number of endoreduplication cycles per nucleus) for the accessions Fei-0, Gd-1, HR-5, Sorbo, and Ws-2 used in this study, and on average, diploids had a leaf EI of 1.858 and tetraploids had a leaf EI of 1.128. This compensation could provide a partial physiological explanation for the lack of statistically significant differences between diploids and polyploids with respect to beta diversity metrics.

Phylosymbiosis, a pattern where the phylogeny of a host organism is recapitulated by similarities in microbial communities associated with the host, is a common pattern across many plants and animals, and can be caused by different evolutionary processes including natural selection for various microbiome functions for the host (reviewed in Kohl, 2020).

Although it is common in nature, phylosymbiosis can be disrupted by hybridization (Lim and Bordenstein, 2020), but it has not been demonstrated whether autopolyploidization, whereby an organism has the exact same genome as the diploid parent, exhibits patterns of phylosymbiosis–being genetically identical to the diploid–or diverges due to ecological or physiological changes. A necessary but not sufficient test in support of this phenomenon would be to show that closely related taxa have more similar microbiome compositions despite ploidy change. With a limited dataset, I found no statistically significant grouping by genotype between all samples regardless of ploidy level in both Bray-Curtis dissimilarity and weighted UniFrac distance metrics (Figures 1.3, 1.4 with a large divergence of many samples within genotypes.

Synthetic community changes in polyploid plants

Notably however, some differences between diploids and polyploids are revealed in the differential log2-fold increases and decreases of members of the synthetic community. Interestingly, the synthetic community more consistently established itself on polyploid plants than diploid plants. Across all 7 accessions, I found that *Exiguobacterium sibiricum*, a member of the synthetic community, was underabundant in the polyploids (Figure 1.8).

Members of *Exiguobacterium* are known to be extremely disparate and versatile, living across a large range of environments, and many members of this genus have been shown to have plant-growth promoting properties and/or stress-response genes which helps the bacteria adapt to difficult and changing environments (Kasana and Pandey, 2018). In particular, *Exiguobacterium sibiricum* was first isolated from permafrost in Siberia and its psychrotrophic properties are associate with the alleviation of cold stress in plants (Kasana and Pandey, 2018; Yadav et al., 2019). The significant decrease of this resilient member of the synthetic community on polyploid plants may indicate changing conditions of the leaf that hinders the establishment of certain beneficial members of the synthetic community potentially due to an increase in the production of defense genes as a function of a doubling of gene dosage. However, when all of the samples from C, F, and G were removed this difference goes away which could potentially be due to genotype-induced differences.

Polyploids recruit more bacteria from the environment

Our experiment also allows us to describe differences in how diploids and polyploids recruit bacteria from the environment. Across all samples, I found that polyploid plants had a consistent significant increase of three taxa that were not members of the synthetic community: *Burkholderia* sp., *Novosphingobium* sp., *Massilia* sp. Members of *Burkholderia* are known to be important facilitators of plant growth but others are plant and human pathogens (Eberl and Vandamme, 2016). *Novosphingobium* species have been likewise shown to promote plant growth through salt-stress alleviation (Vives-Peris et al., 2018) and *Massilia* also have been known to promote growth by indole acetic acid and siderophore production (Ofek et al., 2012; Poupin et al., 2013). However, since we do not have the resolution to discover the strains

of these taxa it is hard to generalize any patterns. Nonetheless, that these taxa, which were not members of the synthetic community (introduced at a much higher concentrations) were able to establish and grow much more on polyploids than diploids may likewise point towards differences between the leaf environments with the polyploids disrupting conditions for some commensal or beneficial bacteria and opening up niche space for other bacteria–some of which could be beneficial or potentially non-beneficial. This could be a major source of variation for polyploid plants, which are known to have expanded ranges and to thrive in disrupted or changing environments (Decanter et al., 2020; Karunarathne et al., 2018; Ficetola and Stöck, 2016). In conjunction with the measurements of higher biomass but similar bacterial load, these findings are consistent with the idea that the same mechanisms that increase defense in polyploids may alter the recruitment of phylosphere bacteria. Polyploidy is a mutation that changes both the output of gene products (e.g. several autopolyploids exhibit pronounced increases in alkaloids per unit dry weight, and synthetic Phlox autopolyploids synthesize flavanoids–compounds also critical in root nodule formation–that are not made by their diploid progenitors; (reviewed in Levin, 1983)), but also their basic physiology and is associated with larger cell and organ sizes (Coate et al., 2012). These findings imply that it is the changes in gene products that are driving the changes in recruitment, not changes on/to total area on which bacteria could establish.

Efficacy of the synthetic community approach

I assessed alpha diversity in order to validate whether our inoculation with a synthetic community was successful in establishing a population of bacteria that is more or less known with relatively little outside bacteria from the external environment. I found that the alpha diversity was consistent across ploidies and differed from the controls, which took up bacteria from the environment, probably from air contamination in a random manner (Figure 1.2). However, I identified around 400 ASVs across all the samples, many at very low abundances, which implies that the growing conditions did not entirely prevent external bacterial contamination. Similarly, I found that the synthetic community did not establish in three diploid samples after initial inoculation. Nonetheless, the inoculation was broadly successful, and by inoculating the synthetic community at a high abundance, I highlight the flexibility of this approach to accommodate modest levels of contamination. Not only is contamination and laboratory effects a huge problem for microbiome studies (Salter et al., 2014; Frankel-Bricker and Frankel, 2021), but previous studies that attempted to test whether polyploidy impacted microbiome composition in the wild (Wipf and Coleman-Derr, 2021) were unable to identify any trends which our experimental design where we can trace taxa and test explicit hypotheses allows us to discover. I encourage future studies to use synthetic communities to have hypothesis-driven microbiome studies in conjunction with the more common descriptive and observation driven approaches.

Chapter 2

Responses to pathogen infection differ between microbiome treated diploids and polyploids

Abstract

Polyploids have been theorized to be more resistant to pathogens, but empirical studies have generally been inconclusive. Using different accessions of Arabidopsis as biological replicates and a synthetic microbial community, I show that polyploids are more resistant to the the plant pathogen Pseudomonas syringae *pv. tomato DC3000 than diploids. I also show that there is a general pattern to the differences between diploids and polyploids and their response to pathogens in that the microbiome plays a protective role differentially between different ploidy levels. I found that there is a weak ploidy effect on pathogen growth over time and that there was no significant effect of the microbiome on polyploid pathogen response. However, I found that inoculation with a synthetic microbiome community can significantly arrest pathogen growth in diploids. I also found that the synthetic-community-derived microbiome changed over the course of pathogen infection differently between diploid and polyploid samples with evidence that it provides limited protection against DC3000 in diploids but none in polyploids.*

2.1 Introduction

Both the rhizosphere and the phyllosphere (reviewed in Berendsen et al. (2012) and Stone et al. (2018), respectively) microbiomes play important roles in plant health. In particular, beneficial microbes can trigger induced systemic resistance (ISR)–a mechanism that prepares a plant for accelerated plant defense, through ethylene and jasmonic acid signaling (Yan et al., 2002). At the same time, both beneficial and pathogenic bacteria often suppress the local plant defense responses to promote their own colonization, and it is not well characterized

how plants distinguish these good and bad bacteria (Berendsen et al., 2012; Zamioudis and Pieterse, 2012). As opposed to the bacteria and fungi in the rhizosphere, which are often thought to play a bigger role in nutrient uptake, bacteria that dominate the phyllosphere are thought to have important functions with regard to host stress tolerance and in mediating host-pathogen responses and the fungi are often pathogenic (Stone et al., 2018). The defense properties of specific common bacteria such as *Sphingomonas* have been demonstrated to act through the direct competition for resources (Innerebner et al., 2011). However, the phyllosphere also could have emergent defense properties whereby a plant may evolve to increase microbiome diversity on the leaf in order to increase the probability of associating with a bacteria that competes with a pathogen (Fargione and Tilman, 2005).

Likewise, both hybrid (reviewed in Fritz et al., 1999) and polyploid organisms (reviewed in Oswald and Nuismer (2007) and King et al. (2012)) have been theorized to be have a better response to pathogens. Oswald and Nuismer (2007) demonstrated, using both analytical and simulation methods, that well supported genetic models of pathogen resistance (such as the gene-for-gene model where host resistance alleles are generally dominant and pathogen virulence alleles are generally recessive or the inverse matching alleles model where the host can only fight pathogens that have alleles that are recognized by certain alleles of the host) can be mathematically extended to infer that neoautopolyploid populations should always be more resistant than diploid populations. Experimental evidence also points to increased resistance of polyploids in kiwifruit where the hexaploids are the most resistant to *Pseudomonas syringae* followed by tetraploids and then diploids (Saei et al., 2017). It has also been shown that inducing polyploidy in cultivated impatiens (*Impatiens walleriana*) also confers increased resistance to mildew due to changes to morphology such as thicker leaves and stems and fewer stomata (Wang et al., 2018). However, it has not yet been shown whether changes in ploidy are associated with microbiome changes that could improve their response to pathogens. I have shown in the previous chapter that there are ploidy-associated changes in microbiome which could potentially have an impact on pathogen resistance.

Importantly, WGD doubles the number to genes, i.e. doubles the gene dosage, which could potentially alter defense response either physiologically or due to transcriptional changes effected by dosage. Polyploidy is not only a macromutation that is associated with increased biomass and larger cell and organ sizes (Coate et al., 2012), but also fundamentally changes the output of gene products such as increased alkaloids and flavanoids (reviewed in Levin, 1983). These phenomena are related as WGD is characterized by larger cell size, which is intimately tied to cell-cycle timing and these associated changes have downstream physiological effects on the expression of defense genes or in the transit time of signaling molecules (reviewed in Doyle and Coate, 2019). Although there is not a large body of work on absolute dosage effects (per-genome, as opposed to relative dosage effects, where gene expression is measure on a relative per-transcriptome basis), there is evidence that there is variation across genes and across dosages, and that variation in absolute gene product abundance has fitness implications for the organism (reviewed in Doyle and Coate, 2019). The variable changes WGD has on the production of gene products whereby we may not expect a simple 1:1 doubling has potential implications on the trade-off between growth and defence in plants

since the same hormones are often involved in both pathways and these pathways are often tightly regulated (Karasov et al., 2017). For example, salicylic acid (SA) and jasmonic acid (JA) are both produced to defend against natural enemies, but they also antagonize both growth-related phytohormones and each other (Karasov et al., 2017). Likewise, the variation in gene products caused by WGD could lead to the recruitment of different bacteria via the production of different signaling compounds or secondary metabolites (reviewed in Powell and Doyle, 2015), which in turn could affect defense responses and could have lead to the evolution of complex symbiotic systems such as root nodule formation in legumes (Li et al., 2013).

Polyploidy may confer protection against fungal diseases due to the upregulation of defenese genes (Chen et al., 2017). However, in soybean it has been shown in natural populations that polyploids were more or less similar to diploids in their resistance to soybean leaf rust (Schoen et al., 1992). Likewise, allopolyploidy has been shown to lead to near-immediate changes in anti-herbivore defense systems (Pearse et al., 2006) although other studies have shown that polyploidy induces non-uniform effects on insect-herbivore interactions (Nuismer and Thompson, 2001). Nonetheless, there are very few studies that assess the phenotypic effects that WGD has on bacterial disease response and also few studies looking at autopolyploidy as a phenomenon in particular (Parisod et al., 2010; Soltis et al., 2007).

My study first examines whether there is a significant difference in defense response to the plant pathogen *Pseudomonas syringae* pv. tomato DC3000 across accession due to ploidy level in Arabidopsis and then examines whether these differences are attributable to changes in the interaction between the plants and a synthetic microbiome comprised of commensal bacteria. We further assess whether the microbiomes themselves change in response to pathogen exposure in any consistent way between ploidies or across genotypes using 16S sequencing of the known synthetic community.

2.2 Methods

Plant Material and Growth Conditions

Plant material and growth conditions are as described in Chapter 1.

Inoculation and Infection

Two weeks after germination, each plant was inoculated with either the synthetic community suspended in 10Mm MgCl buffer or just the 10Mm MgCl buffer as a control (Chapter 1). The plants were inoculated by spraying the plant until saturation. Three weeks after germination (one week post synthetic community inoculation), the plants were spray-inoculated with either the pathogen (*Pseudomonas syringae* pv. tomato DC3000) or a 10Mm MgCl buffer. The pathogen inoculation was at a density of .0001 at OD600 (Innerebner et al., 2011).

Table 2.1: Pseudomonas syringae pv. tomato DC3000 specific primer used for ddPCR amplification

Name	Sequence (5'->3')	Tm (C)	GC%	Length (bp)
Forward primer	GACCAAGGATGCAGCAGAAA	61	50	187
Reverse primer	GCCGTTACGGATATCAACGA	60	50	187

Sample Collection

Plants were collected at three time points: time-zero (immediately after being innoculated with the pathogen), 24 hours post pathogen inoculation, and 48 hours post pathogen inoculation. I removed the aerial portion of the plant by cutting it just above the roots and weighed the plant tissue and then transferred the samples into a tube containing 1.3 ml of 100 mM phosphate buffer (pH 7) with 0.2% Silwet L-77. Samples were bead homogenized using the FastPrep-24 Classic bead beating grinder and lysis system (MP Biomedicals, Inc., CA, USA).

ddPCR pathogen assay

Absolute bacterial abundance was estimated from 40 microliter samples using droplet digital PCR (ddPCR) with the BIO-RAD QX 200 Droplet Reader (Bio-Rad Laboratories, Inc., Hercules, CA, USA) and custom primers to specifically target and amplify a region in the rpoB gene (which encodes the β subunit of bacterial RNA polymerase) in *Pseudomonas syringae* pv. tomato DC3000 (Table 2.1), in runs randomized by position. The PCR protocol is as follows: 95° for 5 min., 95° for 30 sec., 60° for 1 min., return to step two 40 times., 4° for 5 min., 90° for 5 min., keep at 4° overnight. I compared the absolute abundance of DC3000 of polyploids and diploid accession pairs across each time point in order to assay how the pathogen interacted with ploidy and microbiome treatment.

16S sequencing

The synthetic community was then assayed at the final time of 48 hrs post pathogen inoculation. Samples were frozen and kept at -80°C and sent to Microbiome Insights for 16S V4 sequencing and qPCR analysis within one month.

Data Analysis

Forward and reverse paired-end reads were filtered and trimmed to 240 and 220 base pairs (bps), respectively, using the `DADA2` pipeline with default parameters (Callahan et al., 2016). Following denoising and merging reads and removing chimeras (Table 2), I used `DADA2` to infer amplicon sequence variants (ASVs), which are analogous to operational taxonomic

units (OTUs), and taxonomy was assigned to these ASVs using the DADA2-trained `SILVA` database (Version 132, https://benjjneb.github. io/dada2/training.html). A `GTR+I+G` maximum likelihood (ML) tree was then inferred using the `phangorn` package in R (Schliep, 2011) using a neighbor joining tree as the starting point. The ML tree, assigned ASVs, read count data, and sample metadata were combined in a `phyloseq` object (McMurdie and Holmes, 2013) for downstream analyses. Differential microbial changes (i.e. differences in the composition of taxa and abundances between samples) were calculated using `DESeq2` (Love et al., 2014) and the `phyloseq` (Ssekagiri et al., 2018) package, implemented in `R`, was used to calculate changes in alpha and beta diversity. For a permutational analysis of variance (PERMANOVA), the data were rarified to 90% of the read number of the sample with fewest reads and the test was performed using the `adonis` function in the `vegan` package (v2.5-2, Oksanen et al. (2007)) in `R` with 999 permutations to test whether ploidy or time point had an effect on beta diversity measures. The `betadisper` function was also used as implemented in the `vegan` package in `R` for the analysis of multivariate homogeneity of group dispersions (variances).

Accession number

Links to DNA sequencing data are available in Supplemental Table A.1.

2.3 Results

Pathogen assay

I first measured the abundance in DC3000 using ddPCR across all the samples and analyzed them using a general linear model of weight-normalized total counts of bacteria per sample as a function of the explanatory variables of ploidy, treatment, and time, as well as their interaction. I did not find a significant effect of time by treatment (Figure 2.1; Table 2.3), meaning that we cannot reject the hypothesis that the differences between microbiome-treated and untreated plants was due to chance. I found that there was a significant effect of ploidy by time (p-value = 0.045; Figure 2.1; Table 2.3). I also found that diploid samples are significantly different between treated and control groups having a notable decrease in pathogen growth at the 0.05 p-value threshold (Figure 2.2; Table 2.4). Polyploids did not exhibit this pattern.

I also found that at every single time point for both treated and control groups, the polyploids have a lower mean pathogen count than the diploids. However, these differences are not statistically significant at 24 hours and are only weakly significant at 48 hours for both treated and control groups (t = 2.1324, df = 7.151, p-value = 0.06959; t = 2.026, df = 7.2191, p-value = 0.08117).

Table 2.2: Sample metadata and counts of reads through each processing step

File	input	filtered	denoisedF	denoisedR	merged	nonchim	genotype	ploidy	timepoint
S00EJ-0111	28133	22100	21883	21929	21566	20822	WA	tetraploid	T0
S00EJ-0112	42658	33975	33197	33335	32127	30202	COL	diploid	T0
S00EJ-0113	38626	30660	30560	30568	30231	29060	COL	tetraploid	T0
S00EJ-0114	29911	23090	22348	22589	21383	20273	HR	diploid	T0
S00EJ-0115	52515	40996	40885	40875	40544	38759	HR	tetraploid	T0
S00EJ-0116	57844	45503	45319	45329	44730	43275	SOR	diploid	T0
S00EJ-0117	36605	28573	28490	28496	28261	26623	SOR	tetraploid	T0
S00EJ-0118	36152	28503	28413	28397	28131	26529	WS	diploid	T0
S00EJ-0119	27256	21501	21308	21382	21069	20449	WS	tetraploid	T0
S00EJ-0120	40514	32112	31972	32001	31766	30411	GD	diploid	T0
S00EJ-0121	48560	38414	38207	38331	37847	36347	GD	tetraploid	T0
S00EJ-0122	24731	19497	19356	19397	19150	18523	FEI	diploid	T0
S00EJ-0123	23859	18177	18076	18075	17858	17377	FEI	tetraploid	T0
S00EJ-0124	50262	39862	39689	39753	39111	35925	WA	diploid	T2
S00EJ-0125	46316	36078	35958	35969	35413	32124	COL	diploid	T2
S00EJ-0126	40685	31747	31629	31683	31229	29051	HR	diploid	T2
S00EJ-0127	34864	27062	26938	26963	26570	25827	SOR	diploid	T2
S00EJ-0128	30044	23726	23638	23662	23212	20809	WS	diploid	T2
S00EJ-0129	44721	35673	35590	35617	35116	33430	GD	diploid	T2
S00EJ-0130	24704	19460	19387	19423	19094	17857	FEI	diploid	T2
S00EJ-0131	26398	21165	21091	21119	20855	20006	WA	tetraploid	T2
S00EJ-0132	38987	30716	30642	30654	30370	28400	COL	tetraploid	T2
S00EJ-0133	30075	23968	23895	23905	23683	22430	HR	tetraploid	T2
S00EJ-0134	20747	16205	16151	16171	15996	15444	SOR	tetraploid	T2
S00EJ-0135	30328	24009	23919	23948	23604	22497	WS	tetraploid	T2
S00EJ-0136	28687	22784	22749	22753	22440	21120	GD	tetraploid	T2
S00EJ-0137	23585	18784	18742	18735	18464	18146	FEI	tetraploid	T2
S00EJ-0138	24499	19562	19512	19516	19292	17568	WA	diploid	T1
S00EJ-0139	30699	24692	24604	24643	24338	22611	COL	diploid	T1
S00EJ-0140	29987	23950	23858	23848	23484	21909	HR	diploid	T1
S00EJ-0141	42634	34160	34072	34094	33970	33485	SOR	diploid	T1
S00EJ-0142	14448	11457	11402	11421	11249	10519	WS	diploid	T1
S00EJ-0143	25020	20052	19988	19987	19741	18550	GD	diploid	T1
S00EJ-0144	16497	12943	12882	12922	12776	12363	FEI	diploid	T1
S00EJ-0145	15846	12551	12467	12475	12133	11349	WA	tetraploid	T1
S00EJ-0146	27702	22015	21950	21968	21760	20874	COL	tetraploid	T1
S00EJ-0147	37667	29660	29586	29617	29355	27112	HR	tetraploid	T1
S00EJ-0148	35594	27998	27938	27932	27539	25189	SOR	tetraploid	T1
S00EJ-0149	45917	36471	36351	36384	36022	34220	WS	tetraploid	T1
S00EJ-0150	36558	28828	28760	28782	28519	27822	GD	tetraploid	T1
S00EJ-0151	40541	31488	31446	31433	31214	30811	FEI	tetraploid	T1

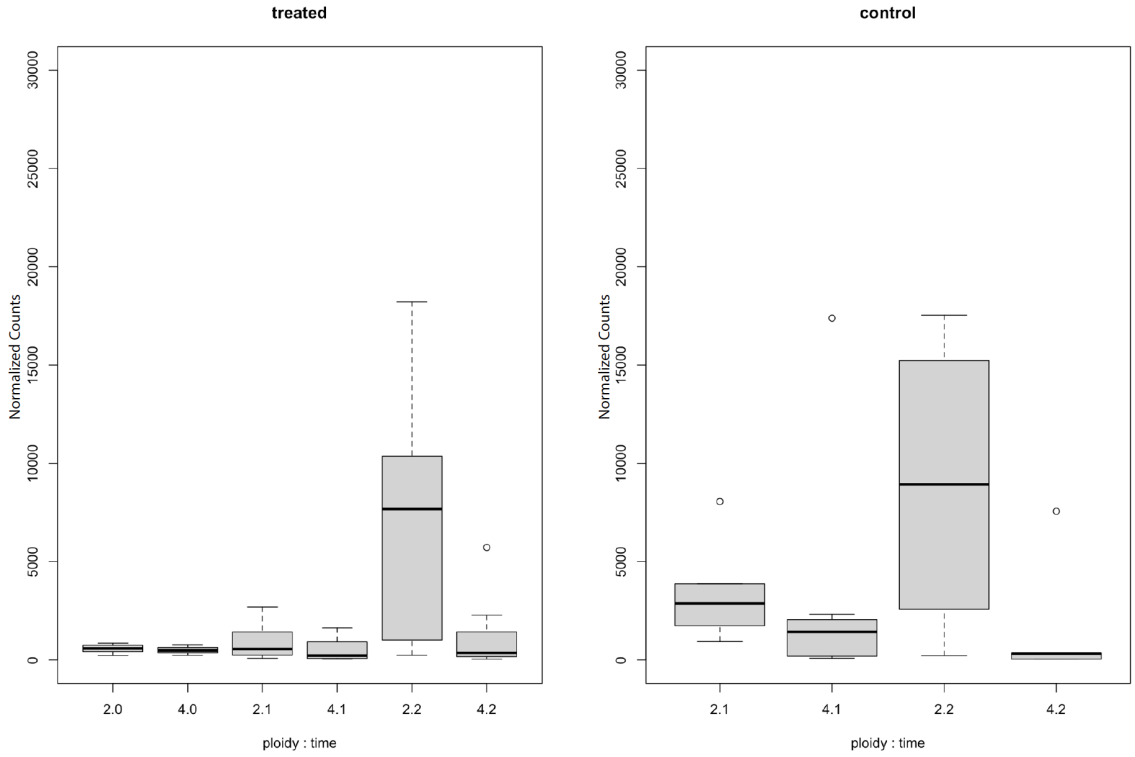


Figure 2.1: Box plots of weight-normalized total counts of bacteria per sample for each ploidy and each time point between the treated and control groups in the pathogen assay. T-tests between treated and control groups at each time point were performed and diploids at time 1 have significantly different means.

Table 2.3: General linear model of weight-normalized total counts of bacteria per sample as a function of the explanatory variables of ploidy, treatment, and time, as well as their interaction

glm(formula=log(norm)∼ploidy*time+ploidy*treatment, family=gaussian				
	estimate	std.error	t-value	p-value
(Intercept)	5.92268	1.05598	5.609	**5.47e-07**
ploidy	0.05522	0.34026	0.162	0.8716
time	1.84635	0.77920	2.370	**0.0210**
treatment	1.45326	1.20617	1.205	0.2330
ploidy:time	-0.51064	0.24951	-2.047	**0.0451**
ploidy:treatment	-0.25717	0.38160	-0.674	0.5029
Null deviance: 180.66 on 65 degrees of freedom				
Residual deviance: 123.42 on 60 degrees of freedom				

Table 2.4: Kruskal-Wallis rank sum tests for four comparisons between microbiome treated samples and controls at different ploidy levels and time points.

comparison	Kruskal-Wallis χ^2	df	p-value
Diploid t1: pathogen counts by treatment	5.898	1	**0.01516**
Diploid t2: pathogen counts by treatment	0.2	1	0.6547
Polyploid t1: pathogen counts by treatment	2.551	1	0.1102
Polyploid t2: pathogen counts by treatment	0.42346	1	0.5152

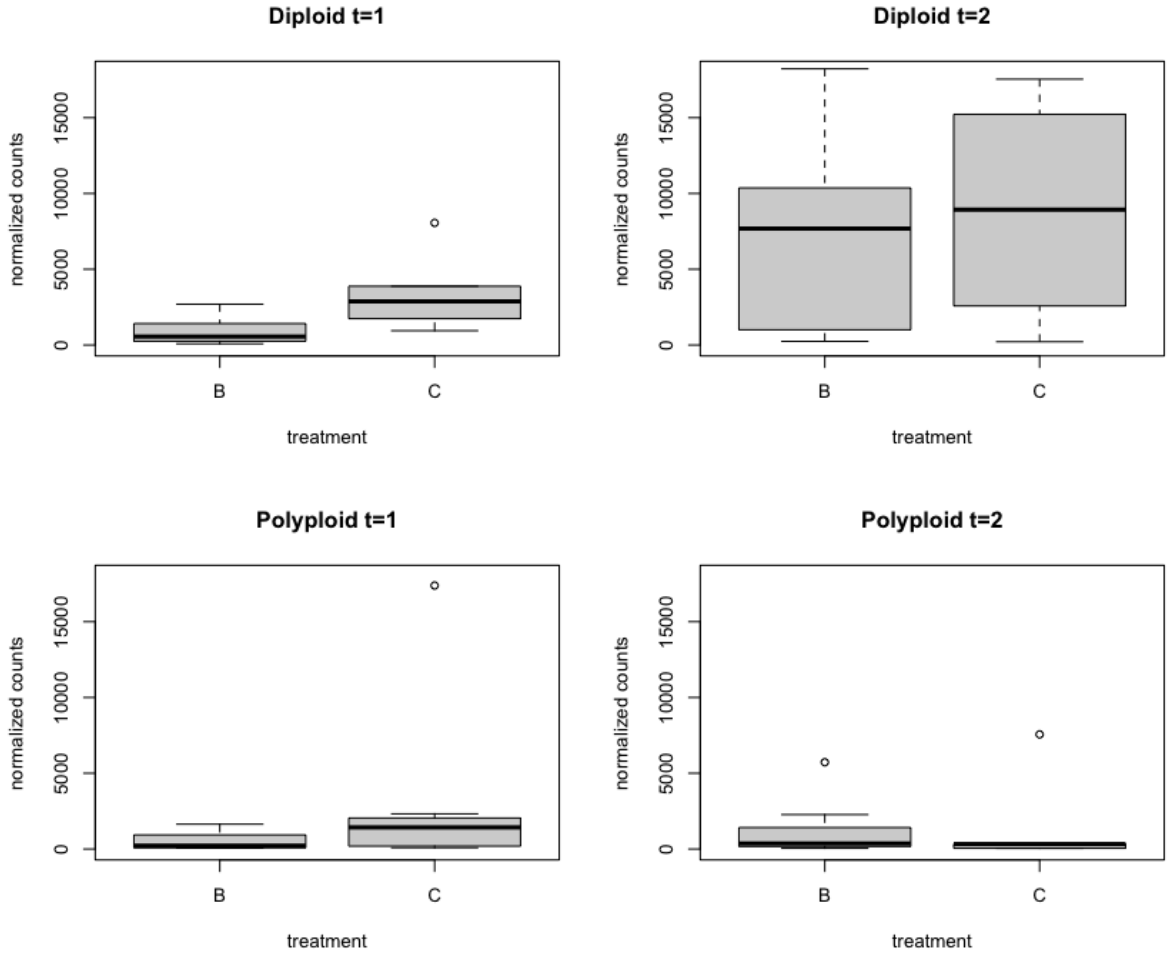


Figure 2.2: Box plots of weight-normalized total counts of bacteria per sample for each ploidy at each time point in the pathogen assay. Samples treated with the synthetic community are represented by the letter B and control samples that were treated with only a buffer are represented by the letter C.

Table 2.5: Permutational multivariate analyses of variance using Bray (top) and weighted unifrac (bottom) distance matrices

formula = dist ~ploidy*timepoint+genotype; permutations = 9999;method=bray

Term	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
ploidy	1	0.3964	0.39643	3.2650	0.04607	0.0405 *
timepoint	2	3.1963	1.59816	13.1625	0.37142	0.0001 ***
genotype	6	1.0597	0.17662	1.4547	0.12314	0.1551
ploidy:timepoint	2	0.4321	0.21606	1.7795	0.05021	0.1308
Residuals	29	3.5211	0.12142		0.40916	
Total	40	8.6057			1	

formula = dist~ploidy*timepoint+genotype; permutations=9999; method=wunifrac

Term	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
ploidy	1	0.0001741	0.00017415	1.3003	0.02910	0.2684
timepoint	2	0.0006162	0.00030810	2.3006	0.10298	0.0017 **
genotype	6	0.0008395	0.00013992	1.0448	0.14031	0.3168
ploidy:timepoint	2	0.0004698	0.00023492	1.7541	0.07852	0.0443 *
Residuals	29	0.0038837	0.00013392		0.64908	
Total	40	0.0059834			1.00000	

Table 2.6: DEseq comparisons of changes in taxa abundance across timepoints for diploid samples and polyploid samples

Comparison	Direction	baseMean	log2FoldΔ	lfcSE	padj	taxa
Dip t0 vs t1	↓ in t0	5407.3	-3.4	0.57	7.00E-07	*P. moraviensis*
Dip t0 vs t1	↓ in t0	914	-3.8	0.74	2.60E-05	*P. tasmaniensis*
Dip t1 vs t2	↑ in t1	5407.3	1.8	0.55	0.0729	*P. moraviensis*
Dip t1 vs t2	↓ in t1	6064	-2.8	0.77	0.057	*P. rhodesiae*
Poly t0 vs t1	↑ in t0	7632.3	3.4	0.99	0.03	*P. rhodesiae*

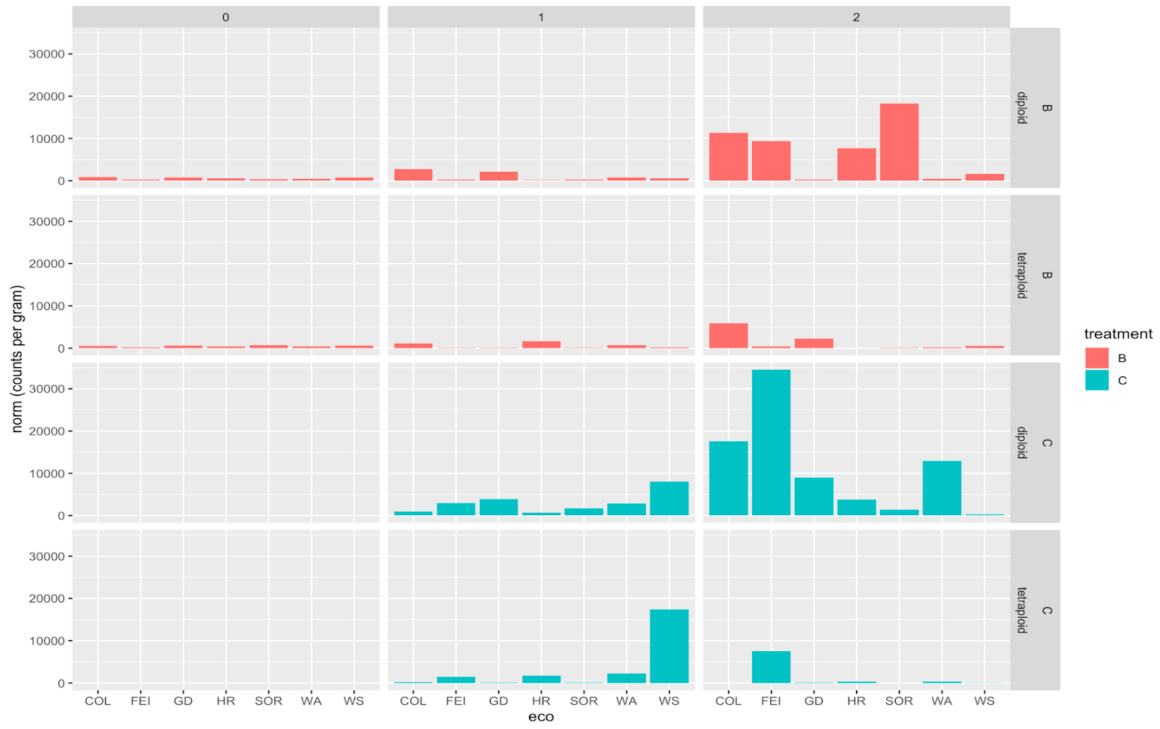


Figure 2.3: Bar chart of DC3000 absolute abundance for accession at each time point of the pathogen assay across ploidy levels ("diploid" vs "tetraploid") and treatment groups ("B" vs "C")

Changes in microbiome composition

The V4 16S region was sequenced for 41 total samples: seven accessions of *Arabidopsis thaliana* and their colchicine-induced autotetraploids at three time points. One sample was discarded due to human error. The sequencing generated an average of 34,000 reads per sample and 25,000 reads per sample following the combined filtering, denoising, merging, and removal of chimera steps (see Table 2.2 for metadata) and across all 40 samples 224 amplicon sequence variants were identified. Each plant was inoculated with a known synthetic microbial community of commensal bacteria known to associate with Tomato and to be broadly representative of bacteria naturally associated with the phylosphere. A pair-wise ANOVA of diversity measures between groups for Fisher Alpha, species richness, and Shannon indices was performed and we cannot reject the null hypotheses that there is not a significant differences between diversity measure for diploids and polyploids, between genotypes, or at different time points (Figure 2.4).

I used the PERMDISP2 procedure to analyze multivariate homogeneity of group dispersions (variances) implemented in the betadispr function in the vegan package in R. For

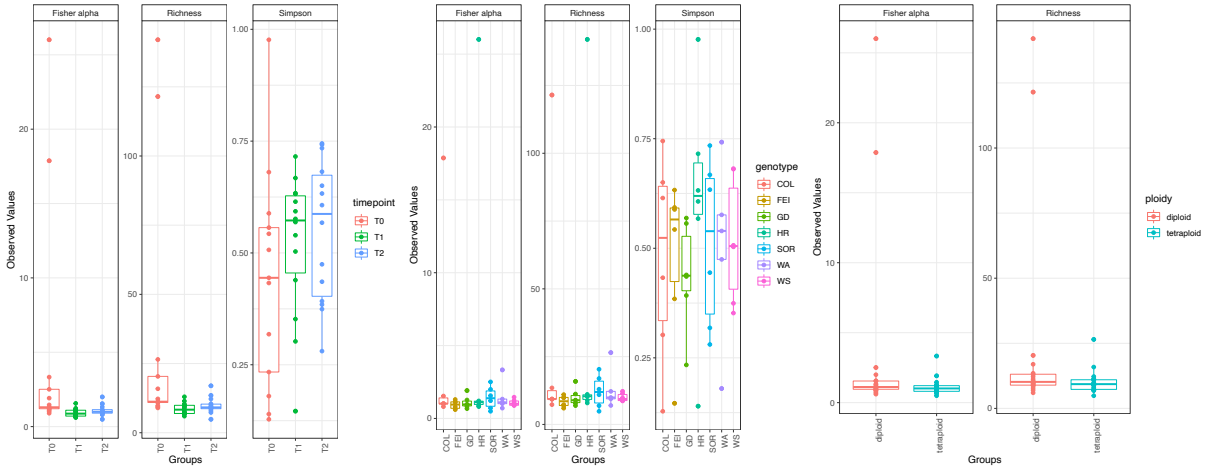


Figure 2.4: Fisher's alpha, species richness, and Simpson indices of alpha diversity grouped by time, genotype, and ploidy for experimental samples. Pairwise analysis of variance in diversity was performed between groups and significance below the p value threshold of 0.05 is indicated by asterisks

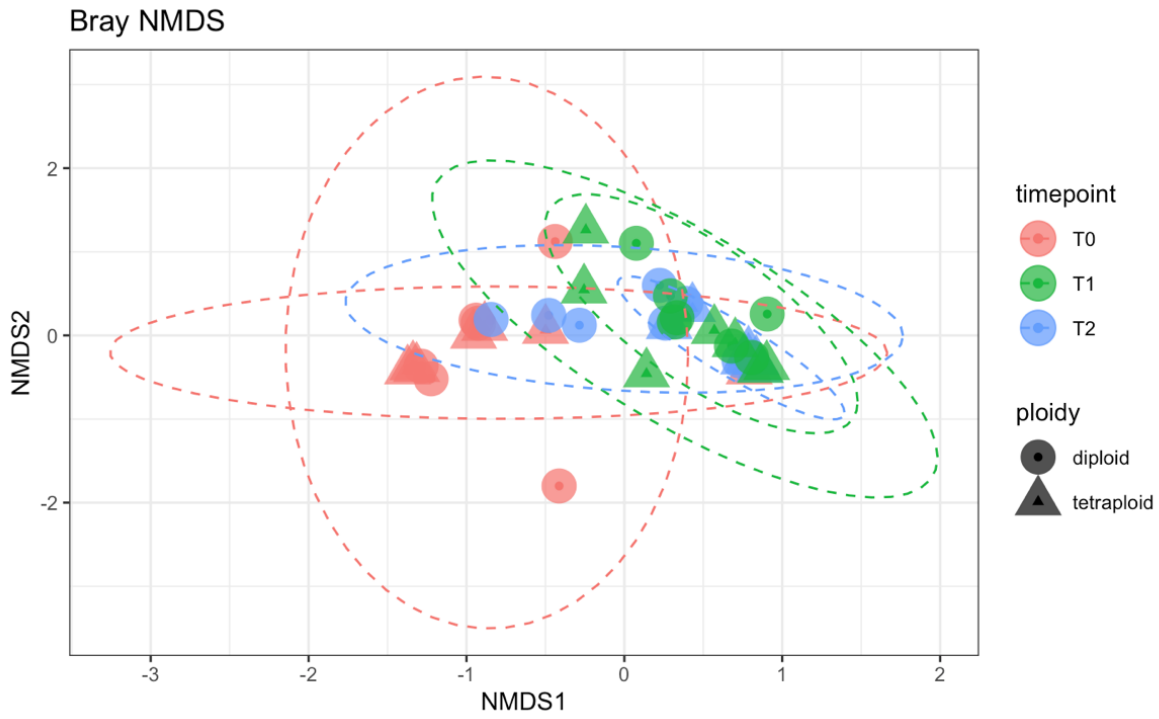


Figure 2.5: Non-metric multidimensional scaling plots based on a Bray-Curtis dissimilarity. Experimental samples are colored by time point and shapes correspond to ploidy levels

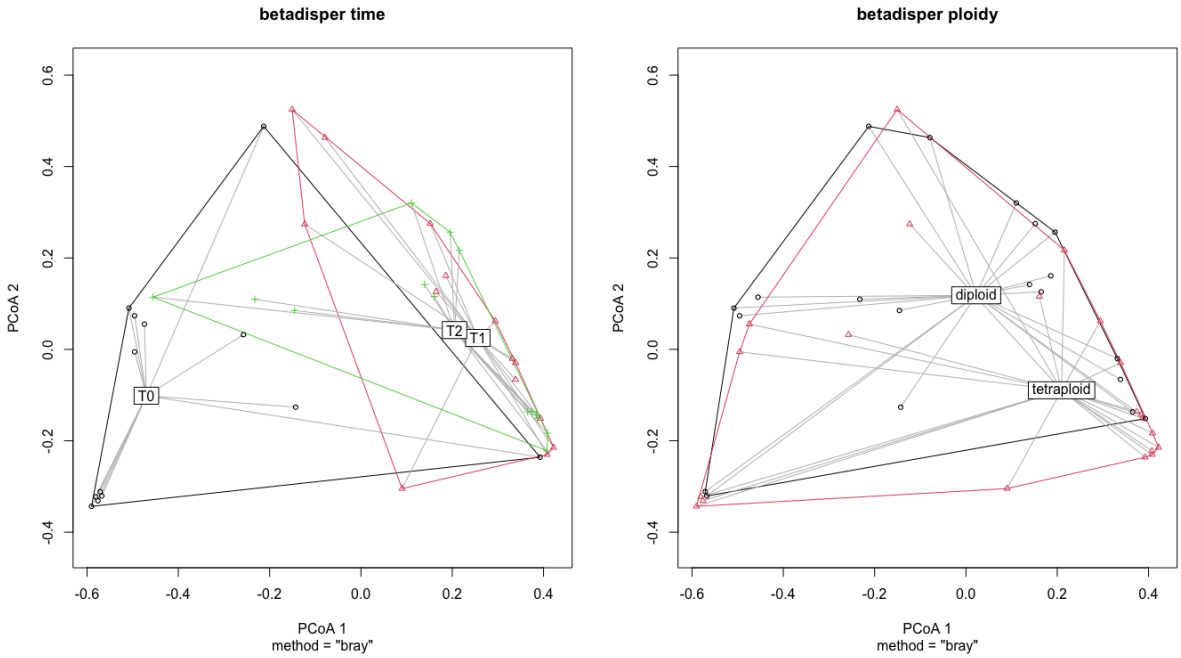


Figure 2.6: Plots of the Bray distances of time points (left) and ploidy (right) to centroids on the first two PCoA axes to assess the homogeneity of variance

Both bray and weighted and unweighted unifrac distance, I did not find any statistically significant differences in beta dispersion between time points or ploidy levels after computing Tukey's Honest Significant Differences test (Figure 2.6). I then performed a PERMANOVA on a rarified dataset in order to test whether time point or ploidy has a significant effect on beta diversity using bray distance and also weighted UniFrac distance which accounts for differences in relative abundances. For Bray distances, ploidy and timepoint were found to be significant and for weighted unifrac distances timepoint was significant and also the interaction between ploidy and timepoint (Figure 2.5).

Ordination plots were then used to arrange samples in two dimensions based on similarity calculated via various indices. The non-metric multidimensional scaling (NMDS) plots of Bray-Curtis distances show that diploid and polyploid treatments overlap at a given time point but that the time points shift along both axes (Figure 2.5). These results correspond to the plots of relative abundance within the rarified dataset with chloroplast and mitochondrial DNA removed across time points and ploidies where there are notable changes across time with a growing abundance of DC3000 and a reduction of Genome 25 *P*seudomonas rhodesiae (Figure 2.7).

These patterns were investigated further using DEseq to identify significant changes in abundance between treatment. When comparing the log two-fold change in relative abundance between time points and ploidy levels accounting for their interaction, we found that

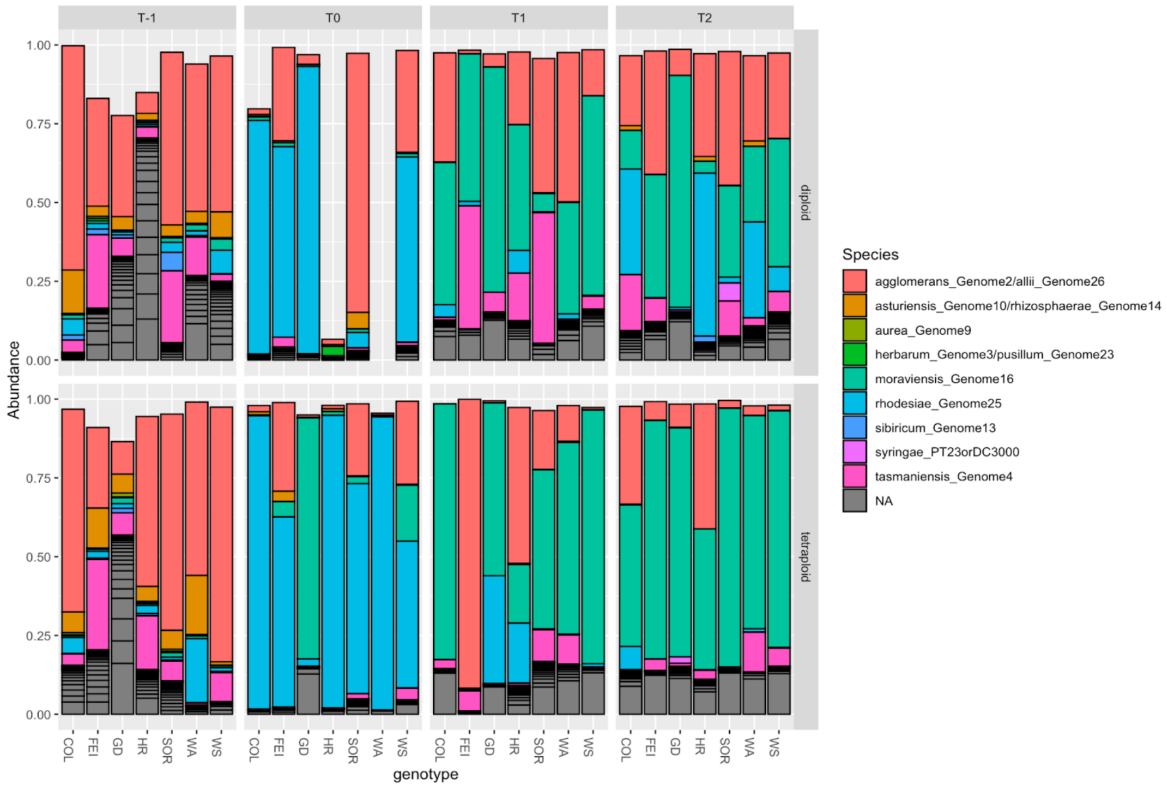


Figure 2.7: Relative abundance of taxa grouped on the level of species shared across experimental samples across all time points and ploidy levels

there was only a difference in three taxa (Table 2.6). For diploid samples, synthetic community member *Pseudomonas moraviensis* was significantly less abundant and synthetic community member *Pseudomonas tasmaniensis* was significantly more abundant at t0 than at t1 (Table 2.6). For polyploids, synthetic community member *Pseudomonas rhodesiae* was significantly more abundant t0 than at t1 (Table 2.6). Only diploid samples exhibited differences when comparing t1 with t2. Synthetic community member *Pseudomonas moraviensis* was found to be significantly more abundant and synthetic community member *Pseudomonas rhodesiae* was found to be significantly less abundant at t1 (Table 2.6).

2.4 Discussion

Until this study, there have been to date very little empirical evidence for a general effect of polyploidy on pathogen response. Although polyploids have been theorized to be more resistant to pathogens (Levin, 1983; Oswald and Nuismer, 2007), empirical studies have generally been inconclusive or non-uniform, i.e. finding evidence for increased resistance and increased susceptibility (Schoen et al., 1992; Nuismer and Thompson, 2001). My study uses

different accessions of Arabidopsis in order to discern any general pattern in the differences between diploids and polyploids in their response to pathogens and whether the microbiome plays a protective role differentially between different ploidy levels.

My experiment tested whether the treatment with a commensal microbiome impacts pathogen response differently between diploids and polyploids. I first compared whether the microbiome had a significant effect on the growth of the pathogen when compared with the control group using a general linear model of weight-normalized total counts of bacteria per sample as a function of the explanatory variables of ploidy, treatment, and time, as well as their interaction. I found that whereas there was no significant difference between treated and untreated samples, there was a significant time effect and a significant interaction effect between time and ploidy.

This difference highlights the major effect of microbiome treatment on the growth of the pathogen. While the microbiome does not significantly alter the progression of the pathogen in the polyploid samples, it does arrest pathogen growth on the diploids. Nonetheless, by t=2 (48 hours post inoculation) the pathogen has grown substantially on the diploids, whereas the polyploids see no notable growth and even a non-significant decrease in the mean number of bacteria. I then looked at accession specific trends between diploids and polyploids in treated and controls in order to assess whether there was any genotypic patterns driving the observed pathogen responses. Interestingly, the patterns seem to be similar across accessions, demonstrating a generalizable phenomenon. Nonetheless, there are some odd accessions, such as Col-0 diploids fairing worse with the microbiome treatment and an explosion of pathogen growth in Ws at t=1 in both the diploid and polyploid groups (Figure 2.3). However, further work will need to be done in order to investigate accession-specific patterns in a robust way.

The significant interaction of time and ploidy on pathogen growth was driven by polyploids exhibiting less growth of DC3000 at each time point regardless of treatment with the synthetic microbiome community or not. Compared with allopolyploids where it has been demonstrated that the merging of defense systems can lead to increased pathogen resistance due to an increase in the flexibility and adaptibility of defense systems (Pearse et al., 2006; Anssour and Baldwin, 2010), autopolyploids may be more resistant than diploids due to an upregulation of defense genes (King et al., 2012). For example, tetraploid Arabidopsis accessions acquired increased resistance to copper stress by having increased activation of antioxidative defense (Li et al., 2017). A buttressing of the antioxidant defense system was also found in colchicine-induced tetraploid plants of *Dioscorea zingiberensis* where antioxidant enzymes were over-produced and maintained at high concentration (Zhang et al., 2010). However, this comes with a fitness trade-off, as Chen (2010) found that there was not much growth vigor in *A. thaliana* autotetraploids, which have more or less the same leaf size and biomass than diploids even though they have much larger seed size, flower size, and stomata size. This phenomenon is elucidated by a study by Ng et al. (2012) which found that proteins associated with stimuli or stress responses were enriched in *A. thaliana autotetraploids*, and that the expression of these genes is associated with a fitness cost and slowed growth. However, our autotetraploids both exhibit greater biomass (Figure 1.1) and defensive capacity (Figure 2.3), which contradicts this general pattern. This is not to say that there does not

exist a growth/defense trade off, but does point to a scenario where polyploid plants can be both more robust and larger even given a growth/defense trade off. Importantly, we know that the growth/defense trade-off is due to antagonistic crosstalk between hormones rather than due to resource consumption, and plants can mitigate these trade-offs by controlling the expression of defense genes either through priming pathways or temporal and spacial concentration of defense (Karasov et al., 2017). Therefore, it makes sense that WGD, which doubles all of the genes, may differentially impact different pathways that are involved in growth and defense (e.g. doubling genes that are highly coordinated in gene complexes or complex pathways versus genes that do not interact with that many other gene products and whose impact on phenotype might scale directly with increasing absolute dosage), which may explain how polyploids one the one hand can still be liable to growth/defense constraints, but on the other hand can still be bigger and more robust.

Contrary to the finding that polyploids had no changes in pathogen response when treated with the synthetic community, I found that there is a significant effect for diploids treated with a microbiome compared to diploids that were not treated. The microbiome has been hypothesized to be implicated in pathogen defense, by allowing the host plant to "outsource" its defense in order to allocate more resources to growth and circumvent the growth-defence trade-off (Karasov et al., 2017). This hypothesis is supported by work demonstrating genotypic variation has an effect on microbiome composition in Arabidopsis (Lundberg et al., 2012). Diploids from natural populations may have evolved to recruit a diverse array of bacteria in order to get around the growth-defense trade-off which would be more protective, but polyploids may disrupt the pathways by which this occurs and instead inhibit the growth of bacteria by overexpressing defense genes and losing the ability to outsource their defense to the microbiome. When we consider how polyploids are often more fit in disturbed or changing habitats, we can see how this potentially deleterious loss of control of the microbiome by the polyploids is analogous to what we know about mutation. These large sources of variation, e.g. transposons or WGDs, are often deleterious but under certain circumstances this is the wellspring of variation upon which natural selection acts and therefore a the source of evolutionary innovation.

This study is consistent with my findings from Chapter 1 where I found a decrease in a member of the synthetic community and a increase of three non-synthetic community taxa in the polyploids and may be consistent with the empirical evidence that polyploids generally out-compete diploids in changing or disrupted habitats (Decanter et al., 2020; Karunarathne et al., 2018; Ficetola and Stöck, 2016) as potentially a function of their increased defensive output disrupting the growth of some bacteria and promoting the growth of other bacteria. When I assessed changes in relative abundance of the synthetic community over time across ploidies, I found that there was substantially more changes in the abundances of three synthetic community members *Pseudomonas moraviensis*, *Pseudomonas rhodesiae*, and *Pseudomonas tasmaniensis* across all the time points in the diploids samples, but that there was hardly any change in the synthetic community over time for the polyploid samples. However, when we take into account the absolute abundance, the polyploids have limited the growth of all bacteria relative to the diploids. Therefore, we may conclude that while

the competition for resources on the leaf between *Pseudomonas* bacteria may help defend against pathogens for diploids, it does not for polyploids. Yet, that matters little due to the polyploids' ability to limit all bacterial growth. Interestingly, DEseq analysis did not find that there was a statistically significant difference between the relative abundance of DC3000 between samples which can be possibly be attributed to the very low initial concentration of inoculate relative to the established synthetic community. In terms of the potential use of synthetic communities for agricultural purposes in plant defense, this finding suggests that *Pseudomonas rhodesiae* is a good competitor against the pathogen DC3000 and points to the efficacy of using commensal Pseudomonads against pathogenic Pseudomonads.

More work will need to be done in order to test whether the observed patterns are attributable to changes in transcriptional responses between diploids and polyploids and if the microbiome in any way mediates pathogen response differently between ploidy levels, either by priming defense or outsourcing defense and reallocating resources to growth, which I will address in the following chapter.

Chapter 3

Plant ploidy level impacts transcriptional response to microbiome colonization

Abstract

The microbiome plays many important roles for plants, including providing a source of protection against pathogens. In this study, I use RNA-seq to test whether there are transcriptional differences between diploids and polyploids when inoculated with a synthetic community of commensal bacteria. I find that treatment with a microbiome generally induces decreases in the expression of genes associated with defense, but that polyploids maintain more defense-related gene expression than diploids. Diploids being more responsive to treatment with the commensal community is consistent with the theory that polyploids are more resistant to pathogens in general. This study provides empirical evidence that supports the following potential consequences of whole genome duplication: polyploids are better defended against bacteria; polyploids potentially lose control of the ability to select their microbiome; polyploids can avoid a tradeoff or constraint that diploids can only deal with by off-loading defense onto the microbiome. While controlling the microbiome may be an adaptation for diploids as it allows for them to outsource the ability to defend against pathogens, my study demonstrates how the loss of this control in polyploids could also be a source of variation in the microbiome upon which natural selection can act.

3.1 Introduction

Both the root and shoot systems of plants are colonized by microbiota, including bacteria, fungi, and other non-fungi eukaryotes, many of whom play an important function in plant pathogen protection or nutrient acquisition (reviewed in Bulgarelli et al., 2013). These microbial associations are not random, rather plants associate with a subset of the possible

microbes by altering the environment that is encountered by these microbiota. This filtering of microbiota can be mediated by plants directly (e.g., salicylic acid modulates colonization of the root by specific bacteria (Lebeis et al., 2015)) or by the coordination of stress and immune system functions (e.g., phosphate stress responses interact with jasmonic acid signaling to control the assembly of a normal root microbiome (Castrillo et al., 2017)). However, studies of the proximate mechanisms by which plants shape their microbial community—the plants' transcriptional responses—traditionally focused on immune response to pathogens rather than response to commensal bacteria (Tao et al., 2003) and to identify the genes involved in defense response, especially in Arabidopsis (Mahalingam et al., 2003; Zhu et al., 2013). Recently, the impact that commensal and mutualistic bacteria—which constitute the vast majority of plant microbiota—have on plant transcriptomes have been studied, revealing both distinct and overlapping responses when compared to responses to pathogens (Vogel et al., 2016), but these studies have been limited to a small number of commensal microbial taxa. However, that there are shared and differing responses plants have to commensal and pathogenic microbiota demonstrates that plants are actively engaging with the microbiome to mediate pathogen encounter with potentially commensal-mediated prophylactic defense responses (Vogel et al., 2016). Notably, Vogel et al. (2016) found that plants usually respond differently to members of its natural phyllosphere microbiota than to pathogens, but some members of the natural microbiota prime the plant for defense by triggering the expression of defense-related genes. This study aims to answer: 1) how do plants of different ploidies manage their microbiota given that some of the taxa are mutalistic but others are pathogens by first looking at how they respond to commensal/mutalistic bacteria and 2) what role do the mutalistic microbes play in defending against the pathogenic ones?

It is well known that polyploidy contributes to the colonization of new niches in plants. In the *Dianthus broteri* complex, polyploids niches were able to diverge where some cytotypes exhibited niche expansion and ecological release, but that other cytotypes expanded into new niches and then had niche contraction due to specialization (López-Jurado et al., 2019). This spatial segregation of cytotypes has been also shown in *Saxifraga rosacea* where an increase in ploidy is associated with more extreme environmental conditions and an expansion of ecological niche (Decanter et al., 2020) (Decanter et al., 2020). Among the many changes to phenotype associated with polyploidy (Ni et al., 2009; Coate et al., 2012; Huang et al., 2007; Levin, 1983), it has been theorized that neopolyploids are more resistant to pathogens than their diploid progenitors (Oswald and Nuismer, 2007) although there has been very little empirical evidence that clearly supports this theory. Interestingly, there has been very few studies to date that look at biotic niche divergence in autopolyploids and how these biotic niches are consititued in part by microbial interactions. While it has been shown that autopolyploidy induces differential herbivory and pollination (Thompson et al., 2004) and association of specific mycorrhizal symbionts (Těšitelová et al., 2013), it has yet to show how these biotic niches diverge with respect to microbial interactions.

Nonetheless, even though coordinated transcriptional plant defense responses are well characterized in Arabidopsis (Schenk et al., 2000), there are few RNA-Seq or proteomics studies that assay the transcriptional response of new autopolyploids, especially to different

environmental factors (Wang et al., 2017). It has been shown, however, that polyploidy has an important impact in altering the relationship of symbiotic microorganisms to their host, especially mycorrhizal interactions, and also that this phenomenon could play an important role in the evolution of microbial symbiosis through the altered biosynthesis of signalling compounds and secondary metabolites, such as flavonoids (compounds that play a key role in root nodule formation (reviewed in Powell and Doyle, 2015)).

Additionally, autopolyploidy is thought to be underestimated and understudied in plants relative to allopolyploidy (Barker et al., 2016), and it remains unclear what the evolutionary advantages and disadvantages are of these two different mechanisms of genome doubling. However, autopolyploidy is an excellent lens to look at the consequences of polyploidy without the confounding effects of hybridization. Unlike allopolyploidy, there appears to be less genome restructuring and reorganization of gene expression immediately following autopolyploidy, though there is limited empirical support for this conclusion (Parisod et al., 2010), but also growing evidence that autopolyploidy has been underappreciated and, in fact, a large source of variation for adaption (Baduel et al., 2019; Monnahan et al., 2019). So, while many consider autopolyploidy to be an evolutionary dead-end or at least a nearly neutral process, it is also considered to be a mechanism that allows for enhanced colonization ability especially in changing habitats (Parisod et al., 2010) since it induces both physiological (e.g., altered growth rate or increased flower size) and genetic (e.g., different gene dosage) changes that themselves can interact in complex ways (Spoelhof et al., 2017). The short-term effects of this type of instantaneous, sympatric macromutation often include the colonization of new niches and the expansion of the range of the neopolyploid (Hijmans et al., 2007; Theodoridis et al., 2013; Molina-Henao and Hopkins, 2019). Therefore, we may wonder by which mechanisms are the biotic and abiotic interactions that constitute the realized niche altered by WGD, and if increased ability to interact with diverse microbes are an important factor. In particular, could the use of bacteria by plants be a way to mediate trade-offs between stress-response and growth by outsourcing the defence of the host to the promotion of non harmful bacteria instead of production of defensive compounds and would this be effected by a doubling of gene dosage following polyploidy (Karasov et al., 2017). The trade-offs between growth and defense in plants are often metabolic as well as, and more importantly, due to the negative interactions between hormones involved in both processes (Karasov et al., 2017). However, it is thought that plants, by shaping their own microbiome, could work around this trade-off, by outsourcing their defenses and thus have developed mechanisms for selecting their associated-microbiota (Karasov et al., 2017).

Vogel et al. (2016) found that *Sphingomonas melonis* upregulated several genes related to defense responses (some of which were shared and some distinct from those upregulated by the pathogen *Pseudomonas syringae DC3000*) and hypothesized that commensal bacteria have an important and active role in plant defense by triggering pathogen-related gene responses. However, they did not find a similar response in the other common phyllosphere bacterium they studied, *Methylobacterium extorquens*. Chen et al. (2020) used Arabidopsis quadruple mutants to show that the MIN7 vesicle-trafficking pathway and the gene *constitutively activated cell death1* (CAD1) are essential components of a genetic network that plants

use to control the bacterial in the phyllosphere and prevent disbiosis. My study attempts to discover if there is a more generalized response that commensal bacteria in the plant microbiome elicit and if there is variation in these responses between diploids and autotetraploids. I expand the scope of previous studies that looked at just one or two commensal bacteria or one or two genes to include a large community of known commensal bacteria, utilizing the same Arabidopsis diploid-polyploid system and synthetic community from Chapter 1 and 2, in order to assay differential transcriptional response between diploid and polyploid accessions treated with a synthetic community.

3.2 Methods

Plant material, growth conditions, and experimental design

Plant material and growth conditions are as described in Chapter 1. The seeds for the accessions used in this study—Columbia (Col-0), Wassilewskija (Ws-2), Sorbo (Sorbo)—were received from Luca Comai's lab, including both cytotypes for each accession (the tetraploids were induced using colchicine in Sarah Schaack's lab).

For each of the three accessions, I grew randomized blocks six plants of each ploidy level (diploids and induced autotetraploids) with three plants treated with the synthetic community and three treated with the control buffer, for a total of 36 plants.

Plant inoculation

Plant inoculation methods are as referenced in Chapter 1.

Sample collection and sequencing

I collected single leaves from the largest developmental node of plants at Stage 1.10 (ten rosette leaves >1 mm in length (Boyes et al., 2001)) and directly froze them in liquid nitrogen before subsequent storage at -80C. Tissue was homogenized using a Mini-BeadBeater 8 (BioSpec Products, Bartlesville, OK, USA) following the manufacturer's instructions. RNA was extracted using the Spectrum Plant Total RNA Kit (Merck / MilliporeSigma, MO, USA) according to the manufacturer's recommendations[1].

I pooled RNA from three different plants for each sample in order to account for variation across technical replicates and treated different accessions as biological replicates to account for accession-specific variation, attaining a conservative estimate for transcriptional changes given the treatment and ploidy. Increasing biological replicates, i.e., many samples per data point, has been known to increase power for detecting differentially expressed genes and for increasing accuracy relative to qRT-PCR (Liu et al. (2014), Dr. David L. Adelson personal communication) obviating the need for qRT-PCR validation. This is further supported by

[1]https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Bulletin/strn50bul.pdf

recent studies that have found very high correlation between RNA-Seq and qRT-PCR as well as the fact that probe bias or poor sensitivity may lead RT-qPCR to be more imprecise than RNA-Seq (Camarena et al., 2010; Nagalakshmi et al., 2008).

Samples were sent to Novogene USA Inc. (Sacramento, CA) for library prep (Poly(A) capture, ligation-based addition of adapters and indexes) and sequencing (Illumina NovaSeq 6000, paired-end reads of length 150 bps, 20M reads per sample). Links to raw RNA sequencing data are available in Supplemental Table A.1.

RNA-seq Data Processing and Analysis

Raw FASTQ files were trimmed and filtered to remove low-quality reads and Illumina adapters and other copyrighted technical sequences using `Trimmomatic` (Bolger et al., 2014) with the default settings. Filtered reads were aligned to the Arabidopsis reference sequence (TAIR10, Lamesch et al. (2012)) with `HISAT2` (Pertea et al., 2016). `HTSeq` (Kim et al., 2015) was used to determine read counts per gene for the test for euploidy. After removing genes with fewer than ten reads per gene, `DESeq2` was used to compare gene expression levels across treatments including a principal component analysis of variance stabilizing transformations and identification of differentially expressed genes (Love et al., 2014). `EdgeR` was also used to provide additional support for patterns of differential expression analysis. Enriched gene ontology (GO) terms were identified using `GOrilla` (Eden et al., 2009).

Mobile elements (transposable elements, TEs) comprise the majority of many eukaryotic genomes and a not insignificant amount of transcipts in transcriptomes and although they are known to play important regulatory functions and express accession-level variation, they are often ignored in many analyses (Underwood et al., 2017). I therefore also performed DESeq on the counts of transcripts that mapped to the Arabidopsis TE database from TAIR10. Transposable element transcription analysis using `DESeq2` was performed using the TAIR TE Database (TAIR10, Lamesch et al. (2012)). Further analyses were performed using `iDEP` (Ge et al., 2018) in order to assess patterns of differential gene expression and enrichment within Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa et al., 2017).

Test for euploidy

I tested if the tetraploid samples were aneuploid or euploid by calculating fold change in relative expression (transcripts per million; TPM) per gene for every pairwise comparison of biological replicates following the methods outlined in Song et al. (2020). If there is aneuploidy or a large segmental duplication, we would expect to see a coordinated increase. in TPM for genes on that chromosome, which would be reflected in a shift in fold change of expression relative to the other biological replicates. I plotted the TPMs per gene along all five chromosomes. If any one showed a stretch (or whole chromosome) of elevated or lowered TPM relative to any of the others it would suggest aneuploidy (chromosomal or segmental).

3.3 Results

Test for euploidy

No stretch was found and this was verified after performing a t-test of the mean log2 fold-changes between accessions for each chromosome with the mean log2 fold-change by chromosome being around one for all samples (Figure 3.1). Therefore, I conclude that all tetraploid individuals were euploid.

Overarching transcriptional differences

A principal component analysis of variance stabilizing transformations of the raw count data shows that the samples cluster together based on genotype rather than by treatment of microbiome or by ploidy level, although there is some separation of treatment along PC2 (Figure 3.2). This pattern can be further demonstrated through hierarchical clustering of the distance matrices, where samples of the same accession cluster closer together(Figure 3.5, panel A).

There were many significantly up- or down-regulated genes at the 0.05 p-value threshold: Col-0 vs Sorbo, 1451 up-regulated, 1825 down-regulated; Col-0 vs. Ws-2 1608 up-regulated, 2242 down-regulated; Sorbo vs Ws-2, 2137 up-regulated, 1960 down-regulated. However, our experimental design does not give us the statistical power to discern the sources (whether treatment or ploidy) of these differences conclusively. Nonetheless, generalizable patterns do emerge once a gene enrichment analysis is performed on the differentially expressed genes using GOrilla.

In pairwise comparisons, Col-0 is significantly enriched in GO terms pertaining to ADP binding (GO:0043531) compared to the other two accessions.Likewise, in terms of biological processes, when Col-0 is compared with Sorbo, there is significant enrichment for GO terms corresponding to translation (GO:0006412), regulation of cellular ketone metabolic process (GO:0010565), signal transduction (GO:0007165), and cellular response to oxidative stress (GO:0034599). In terms of cellular component, GO terms corresponding to the cytosolic ribosome (GO:0022625, GO:0022626) were enriched. When Col-0 was compared to Ws-2, there was no enrichment in terms of cellular component, but in terms of biological processes, photoperiodism/flowering (GO:0048573), import into cell (GO:0098657), signal transduction (GO:0007165), growth (GO:0040007), rhythmic process (GO:0048511), and circadian rhythm (GO:0007623) were found to be enriched. When Sorbo was compared to Ws-2, there was significant enrichment defense response (GO:0009814, GO:0009627) and metabolism/catabolism (GO:0044282, GO:0005976) in terms of process, ATP binding (GO:0043531) and NAD+ nucleosidase activity (GO:0003953, GO:0061809, GO:0050135, GO:0003824) in terms of function, and extracellular region (GO:0005576), preribosome (GO:0030684), and nucleolus components (GO:0005730).

Differential expression between plants treated with the synthetic community and those untreated

Between the plants treated with the synthetic community and the controls, I found 384 up- or down-regulated genes at the 0.1 p-value cut-off (a standard p-value in this discipline to increase the power to detect GO enrichment patterns e.g. the Functional Annotation Tool in DAVID (Sherman et al., 2007)). An analysis of the enriched GO terms revealed that that same biological pathway was significantly enriched for both treatments: cellular response to hypoxia (Figure 3.5, Table 3.4, 3.3). However, there were also significantly enriched GO terms associated with the secretory vesicle cellular component for the microbiome treatment experiment (Table 3.2) as well as significantly enriched GO terms with the biological process of nuclear-transcribed mRNA poly(A) tail shortening for the ploidy-level experiment (Table 3.3). `EdgeR` analyses detected only 65 up- or down-regulated genes at the 0.1 p-value cut-off; however the exact same GO terms were found to be enriched as in the `DESeq2` analyses (Supplemental Table A.1).

The Gene Set Enrichment Analysis (GSEA) analyses identified several other pathways that were enriched in the treated samples (Figure 3.6). Genes associated with ethylene signaling, response to nitrogen compounds, negative regulation of response to alcohol, and photosythesis were down-regulated Conversely, kinases and genes associated with negative regulation of growth were up-regulated (Figures 3.3, 3.6).

Comparisons of the effects of ploid (independent of microbiome) revealed 15 up- or down-regulated genes at the 0.1 p-value cut-off, there were no enriched GO terms with respect to biological process, molecular function, or cellular component (Supplemental Table A.1).

Effect of polyploidy on transcriptome changes

For the differences within the group treated with the synthetic community between diploids and polyploids, I found 16 up or down-regulated genes at the 0.1 p-value cut-off using DEseq (Table 3.3). When looking at the significant genes from this DEseq analysis, I found the major differences were between diploids and polyploid samples within the treated group (Figure 3.8). While most of the genes that were found to be significantly different between diploids and polyploids are hypothetical proteins (AT1G19020, AT1G23710) mostly associated with response to hypoxia, response to stress, and cellular response to chemical stimulus, for all of them the diploid was down-regulated and the polyploids were up-regulated relative to the diploids but not significantly different than the control group of untreated plants (Table 3.3). One gene of the set of down-regulated genes, AT5G22250, is well annotated; it encodes ccr4-associated factor 1b (ATCAF1B, CAF1B), a protein that plays a role in gene regulation by participating in the degradation of messenger RNA (mRNA) through deadenylation. Another gene, AT5G20230 (SENESCENCE ASSOCIATED GENE (SAG) 14), is a copper-binding protein senescence associated gene associated with reponse of oxidative stress and wounding. When I ran textttEdgeR to crosscheck these genes, only one gene was found to be up-regulated (SAG14).

These results were further corroborated when I assessed Generally Applicable Gene-set Enrichment (GAGE) and Parametric Gene Set Enrichment of KEGG pathways between tetraploid and diploid treated samples (Figure 3.5). I found that polyploids up-regulated several pathways related to defense such as ethylene signaling, defense response, response to nitrogen compounds, and disruption of cells of other organisms, whereas diploids were up-regulated in beta-glugan metabolism and genes associated with regulation of cell shape.

I compared untreated diploids vs untreated polyploids using a GAGE analysis for enriched biological processes and found that the same pathways were enriched in the polyploids vs the diploids as the comparison just looking at the treated samples: down regulation of photosynthesis in polyploid relative to diploids and up regulation of defense associated pathways (Figure 3.4, 3.3).

Effect within ploidy levels between treated and untreated samples

Comparing treated vs control diploid samples using a GAGE analysis, I found that the significant cellular components and biological processes were mostly down-regulated in the treated samples. These cellular components are related to photosynthesis: either Photosystem I and II or the thylakoid. The biological processes were either photosynthesis related or related to ethylene-mediated defense response or other defense response such as defense response to bacteria, response to chitin and drugs, and response to UV-B (Figure 3.5).

DESeq analyses showed that both ccr4-associated factor 1b (CAF1b, AT5G22250) and putative ccr4-associated factor 1 (CAF1a, AT3G44260) were down-regulated in the treated samples, as was lipoxygenase 4 (AT1G72520), while a galactose-binding protein (AT1G22882) was up-regulated (Figure 3.9).

Comparing treated vs control polyploid samples using a GAGE analysis, I found that polyploids up-regulated genes associated with negative regulation of growth, signal transduction, DNA replication initiation, negative regulation of dephosphorylation, and disruption to cells of other organisms; whereas they down-regulated genes associated with the negative regulation of response to alcohol, regulation of secondary metabolism, trehalose metabolism, regulation of leaf development and senescence, response to nitrogen compounds and chitin, and ion transport (Figure 3.5).

Differential transcription of transposable elements

There was no significant differential TE expression between diploids and polyploids, surprisingly, either compared across the microbiome-treated samples or all samples together. However, I found three significantly under-expressed TEs when comparing the treated to untreated samples, although only one remained significant following adjustment for multiple testing using Benjamini-Hochberg method (padj, Love et al. (2014)). The significant TE was identified as ATMUN1, a nonautonomous DNA MuDR transposon. Between accessions, I found 90 differentially expressed TEs between Col-0 and Ws-2 and 87 differentially expressed TEs between Col-0 and Sorbo, which demonstrates that there are many elements that are

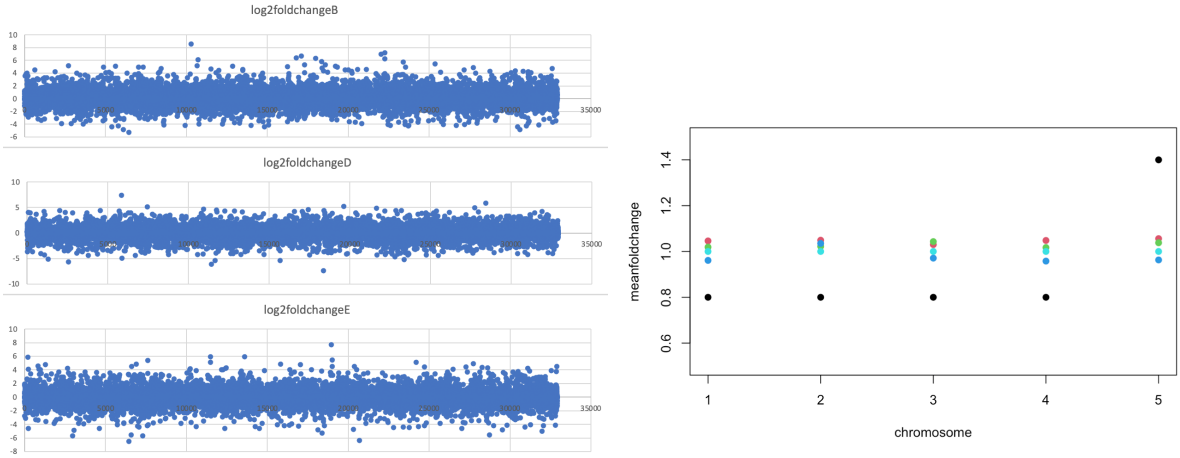


Figure 3.1: Assessment of polyploid lines for aneuploidy. Left: Transcripts per million (TPM) per gene among the biological replicates of an accession (Col-0 (top), Ws-2 (middle), or Sorbo (bottom)) and plotted along the length of all five chromosomes. If any one showed a stretch (or whole chromosome) of elevated or lowered TPM relative to any of the others it would suggest aneuploidy (chromosomal or segmental). Right: Blue, green, and red dots represent the mean fold change per gene per chromosome for Col-0, Sorbo, and Ws-2, respectively. Cyan dots represent the expected pattern for an euploid (all 1.0) and black dots represent the expected pattern for an aneuploid where there is a coordinated transcriptional increase due to a segmental or chromosome duplication.

transcriptionally active and that vary across genotypes (especially Copia and VANDAL elements).

3.4 Discussion

Genotype not microbiome treatment or ploidy drives main transcriptional differences

I found that the broad-scale transcriptional profiles of the samples grouped strongly together by accession. Although the differences among genotypes cannot be reliable attributed to one source of variation (i.e., accession, ploidy, treatment) or their interaction, the large differences between accessions demonstrates that any pattern observed using accessions as replicates should be more generalizable than a pattern using individuals within a single accession as replicates.

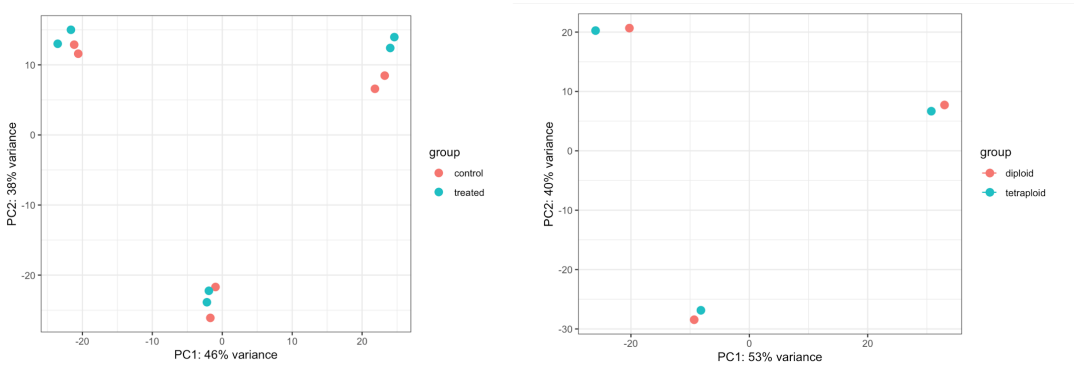


Figure 3.2: Principal component plot spanned by their first two principal components for samples grouped by (a) condition and (b) ploidy

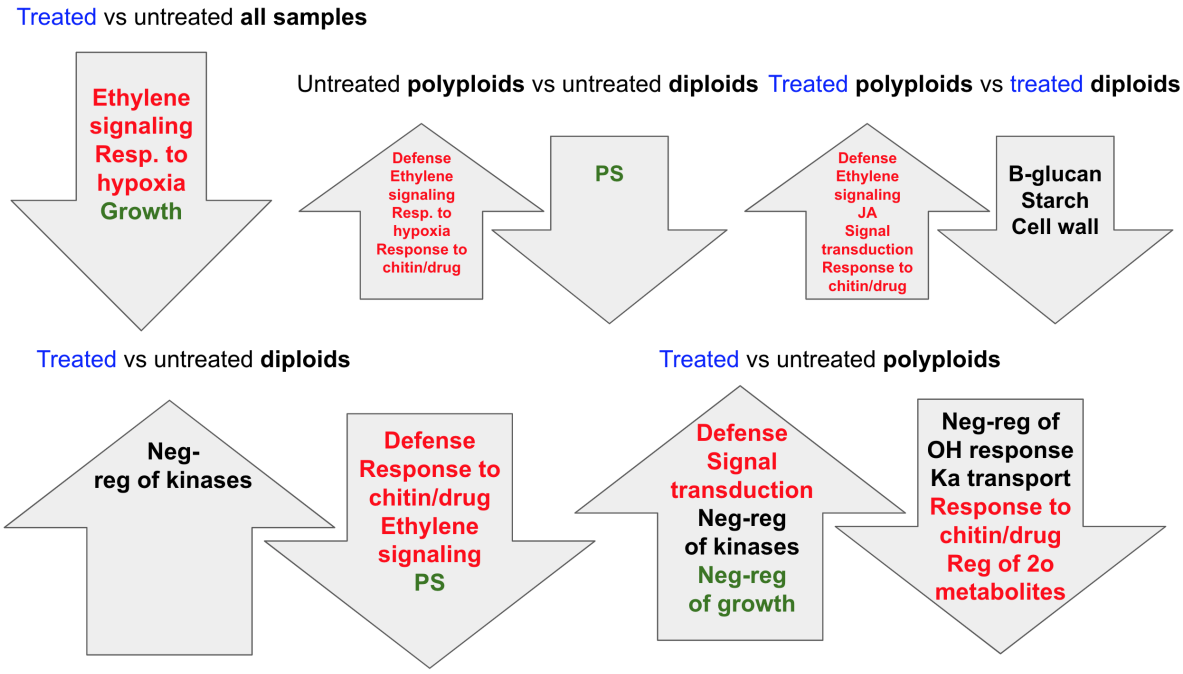


Figure 3.3: Summary of main RNA-seq findings between samples

Table 3.1: Statistics and metadata for RNA-seq analysis

Sample	Library_Flowcell_Lane	Raw reads	Raw data(G)	Effective(%)	Error(%)	Q20(%)	Q30(%)	GC(%)
B2B	CRRA200014390-1a_H3LCGDSXY_L2	42255022	6.3	98.91	0.02	98.65	95.53	46.76
B2C	CRRA200014391-1a_H3LCGDSXY_L2	46450100	7	98.89	0.02	98.65	95.55	46.9
B4B	CRRA200014392-1a_H3LCGDSXY_L2	41759630	6.3	98.69	0.02	98.6	95.38	46.19
B4C	CRRA200014393-1a_H3LCGDSXY_L2	44558440	6.7	98.93	0.02	98.63	95.42	46.81
D2B	CRRA200014394-1a_H3LCGDSXY_L2	42809026	6.4	98.55	0.02	98.64	95.49	46.88
D2C	CRRA200014395-1a_H3LCGDSXY_L2	46897160	7	98.83	0.02	98.69	95.63	46.61
D4B	CRRA200014396-1a_H3LCGDSXY_L3	60918608	9.1	98.79	0.02	98.38	94.96	46.46
D4C	CRRA200014397-1a_H3LCGDSXY_L2	59517730	8.9	98.69	0.02	98.54	95.22	46.4
E2B	CRRA200014398-1a_H3LCGDSXY_L2	59039034	8.9	98.94	0.02	98.66	95.54	46.44
E2C	CRRA200014399-1a_H3LCGDSXY_L2	48807568	7.3	98.42	0.02	98.68	95.69	46.95
E4B	CRRA200014400-1a_H3LCGDSXY_L2	50121980	7.5	98.27	0.02	98.73	95.81	46.8
E4C	CRRA200014401-1a_H3LCGDSXY_L2	55007296	8.3	98.64	0.02	98.62	95.58	46.03

Table 3.2: enriched GO terms and associated genes in cellular components between microbiome-treated and untreated samples

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0099503	secretory vesicle	5.26E-06	3.93E-03	13.45 (12597,77,73,6)	AT1G03220 - aspartyl protease-like protein AT2G39360 - putative receptor-like protein kinase AT2G43620 - chitinase AT3G07470 - hypothetical protein AT5G48540 - receptor-like protein kinase-related family protein AT1G78830 - curculin-like (mannose-binding) lectin-like protein

Table 3.3: enriched GO terms and associated genes in biological processes between diploids and polyploids

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0071453	cellular response to oxygen levels	2.68E-04	1.00E+00	70.07 (12193,116,3,2)	AT1G19020 - hypothetical protein AT1G23710 - hypothetical protein
GO:0071456	cellular response to hypoxia	2.68E-04	5.39E-01	70.07 (12193,116,3,2)	AT1G19020 - hypothetical protein AT1G23710 - hypothetical protein
GO:0036294	cellular response to decreased oxygen levels	2.68E-04	3.59E-01	70.07 (12193,116,3,2)	AT1G19020 - hypothetical protein AT1G23710 - hypothetical protein
GO:0001666	response to hypoxia	2.77E-04	2.79E-01	68.89 (12193,118,3,2)	AT1G19020 - hypothetical protein AT1G23710 - hypothetical protein
GO:0036293	response to decreased oxygen levels	2.82E-04	2.27E-01	68.31 (12193,119,3,2)	AT1G19020 - hypothetical protein AT1G23710 - hypothetical protein
GO:0070482	response to oxygen levels	2.86E-04	1.92E-01	67.74 (12193,120,3,2)	AT1G19020 - hypothetical protein AT1G23710 - hypothetical protein
GO:0006950	response to stress	3.65E-04	2.10E-01	13.98 (12193,872,3,3)	AT1G19020 - hypothetical protein AT1G23710 - hypothetical protein AT5G22250 - ccr4-associated factor 1b
GO:0070887	cellular response to chemical stimulus	6.58E-04	3.31E-01	44.66 (12193,182,3,2)	AT1G19020 - hypothetical protein AT1G23710 - hypothetical protein
GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	9.84E-04	4.40E-01	1,016.08 (12193,4,3,1)	AT5G22250 - ccr4-associated factor 1b

Table 3.4: enriched GO terms and associated genes in biological processes between microbiome-treated and untreated samples

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
					AT1G03220 - aspartyl protease-like protein
					AT3G43700 - putative protein phosphatase 2c 27
					AT1G23710 - hypothetical protein
					AT3G60520 - hypothetical protein
GO:0009628	response to abiotic stimulus	4.97E-05	2.01E-01	3.89 (12597,532,73,12)	AT1G72430 - saur-like auxin-responsive protein family
					AT4G24800 - ma3 domain-containing protein
					AT1G22810 - ethylene-responsive transcription factor erf019
					AT2G22880 - vq motif-containing protein
					AT5G48540 - receptor-like protein kinase-related family protein
					AT1G33590 - leucine-rich repeat (lrr) family protein
					AT3G55880 - vq motif-containing protein
					AT4G17250 - hypothetical protein
					AT1G03220 - aspartyl protease-like protein
					AT1G23710 - hypothetical protein
GO:0070887	cellular response to chemical stimulus	1.08E-04	2.19E-01	6.39 (12597,189,73,7)	AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT1G22810 - ethylene-responsive transcription factor erf019
					AT2G33580 - lysm-containing receptor-like kinase
					AT3G56880 - vq motif-containing protein
					AT5G44680 - putative 3-methyladenine glycosylase i
					AT1G03220 - aspartyl protease-like protein
					AT2G33700 - putative protein phosphatase 2c 27
GO:0051716	cellular response to stimulus	3.58E-04	4.84E-01	4.04 (12597,384,73,9)	AT1G23710 - hypothetical protein
					AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT1G22810 - ethylene-responsive transcription factor erf019
					AT2G33580 - lysm-containing receptor-like kinase
					AT3G56880 - vq motif-containing protein
					AT1G03220 - aspartyl protease-like protein
					AT1G23710 - hypothetical protein
GO:0071453	cellular response to oxygen levels	6.46E-04	6.56E-01	7.19 (12597,120,73,5)	AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT3G56880 - vq motif-containing protein
					AT1G03220 - aspartyl protease-like protein
					AT3G60520 - hypothetical protein
GO:0071456	cellular response to hypoxia	6.46E-04	5.24E-01	7.19 (12597,120,73,5)	AT1G23710 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT3G56880 - vq motif-containing protein
					AT1G03220 - aspartyl protease-like protein
					AT1G23710 - hypothetical protein
GO:0036294	cellular response to decreased oxygen levels	6.46E-04	4.37E-01	7.19 (12597,120,73,5)	AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT3G56880 - vq motif-containing protein
					AT1G03220 - aspartyl protease-like protein
					AT1G23710 - hypothetical protein
GO:0001666	response to hypoxia	6.97E-04	4.04E-01	7.07 (12597,122,73,5)	AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT3G56880 - vq motif-containing protein
					AT5G44680 - putative 3-methyladenine glycosylase i
					AT1G03220 - aspartyl protease-like protein
GO:0033554	cellular response to stress	7.14E-04	3.62E-01	4.10 (12597,337,73,8)	AT2G33700 - putative protein phosphatase 2c 27
					AT1G23710 - hypothetical protein
					AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT1G22810 - ethylene-responsive transcription factor erf019
					AT3G56880 - vq motif-containing protein
					AT1G03220 - aspartyl protease-like protein
					AT1G23710 - hypothetical protein
GO:0036293	response to decreased oxygen levels	7.23E-04	3.26E-01	7.01 (12597,123,73,5)	AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT3G56880 - vq motif-containing protein
					AT1G03220 - aspartyl protease-like protein
					AT1G23710 - hypothetical protein
GO:0070482	response to oxygen levels	7.50E-04	3.04E-01	6.96 (12597,124,73,5)	AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
					AT3G56880 - vq motif-containing protein
					AT1G03220 - aspartyl protease-like protein
					AT2G33700 - putative protein phosphatase 2c 27
					AT1G23710 - hypothetical protein
					AT3G60520 - hypothetical protein
					AT1G72430 - saur-like auxin-responsive protein family
GO:0042221	response to chemical	7.53E-04	2.78E-01	3.10 (12597,613,73,11)	AT4G24800 - ma3 domain-containing protein
					AT1G22810 - ethylene-responsive transcription factor erf019
					AT2G33580 - lysm-containing receptor-like kinase
					AT3G56880 - vq motif-containing protein
					AT1G64380 - ethylene-responsive transcription factor erf061
					AT5G01830 - arm repeat superfamily protein

Table 3.5: Differentially expressed biological pathways between treated vs. untreated diploids

	Direction	adj.Pval	nGenes	Pathways	Genes
1	Down regulated	4.32E-25	22	Response to chitin	AT1G27730 AT3G55980 AT5G04340 AT1G42990 AT1G51700 AT1G80840 AT2G23320 AT2G43380 AT2G35930 AT4G38470 AT3G15210 AT3G16720 AT3G18710 AT3G46600 AT4G17490 AT4G23810 AT4G31550 AT4G37040 AT4G38470 AT4G39780 AT5G47220 AT5G51190 AT5G59550 AT5G64660 AT4G23810 AT4G31550 AT5G47220 AT5G51190 AT5G59550
2	Down regulated	5.46E-17	28	Response to drug	AT1G27730 AT2G22500 AT3G55980 AT4G24570 AT5G04340 AT1G18570 AT1G42990 AT1G51700 AT1G72520 AT1G80840 AT2G23320 AT2G33580 AT2G35930 AT2G38470 AT3G15210 AT3G16720 AT3G18710 AT3G46600 AT4G17490 AT4G23810 AT4G31550 AT5G08790 AT5G27420 AT5G47230 AT5G51190 AT5G59550 AT5G62470 AT5G64660
3	Down regulated	5.77E-17	42	Response to oxygen-containing compound	AT1G07430 AT1G27730 AT3G11410 AT3G55980 AT3G57530 AT4G26080 AT5G04340 AT1G10560 AT1G19770 AT1G22810 AT1G28370 AT1G53170 AT1G72520 AT1G78850 AT1G80840 AT2G23320 AT2G33580 AT2G38470 AT3G15210 AT3G46600 AT3G47640 AT3G53260 AT4G11280 AT4G17490 AT4G23810 AT4G31550 AT4G37040 AT5G08790 AT5G27420 AT5G45340 AT5G47230 AT5G51190 AT5G52890 AT5G59550 AT5G62470 AT5G64660 AT1G01720
4	Down regulated	1.28E-15	53	Response to chemical	AT1G07430 AT1G27730 AT3G11410 AT3G55980 AT3G57530 AT4G26080 AT5G04340 AT4G31550 AT5G26920 AT1G10560 AT1G19770 AT1G22810 AT1G28370 AT1G53170 AT1G72520 AT1G78850 AT1G80840 AT2G23320 AT2G33580 AT2G38470 AT3G15210 AT3G46600 AT3G47640 AT3G53260 AT4G11280 AT4G17490 AT4G23810 AT4G31550 AT4G37040 AT5G08790 AT5G27420 AT5G45340 AT5G47230 AT5G51190 AT5G52890 AT5G59550 AT5G62470 AT5G64660 AT1G01720 AT1G69530
5	Down regulated	1.13E-14	43	Response to organic substance	AT1G07430 AT1G27730 AT3G11410 AT3G55980 AT3G57530 AT4G26080 AT5G04340 AT4G31550 AT5G26920 AT1G10560 AT1G19770 AT1G22810 AT1G28370 AT1G53170 AT4G17490 AT4G39780 AT5G25190 AT5G47230 AT5G51190 AT5G59550 AT5G62470 AT1G18570 AT1G42990 AT1G51700 AT1G80840 AT2G23320 AT2G33580 AT2G35930 AT2G38470 AT3G16720 AT3G18710 AT3G46600 AT4G11280 AT4G23060 AT4G23810 AT5G08790 AT5G27420 AT5G64660 AT1G01720 AT1G69530
6	Down regulated	2.43E-11	50	Response to stress	AT1G21326 AT1G29690 AT3G55980 AT3G16720 AT1G42990 AT2G35930 AT2G38470 AT2G41010 AT3G11840 AT3G15210 AT3G56880 AT4G11280 AT4G17490 AT4G23810 AT4G31550 AT4G33465 AT5G26920 AT5G27420 AT5G47230 AT5G57220 AT5G62470 AT1G01720 AT1G03220 AT1G18570 AT1G27730 AT1G32920 AT1G55450 AT1G72520 AT1G80840 AT2G37040 AT3G11410 AT3G28340 AT3G44260 AT3G53260 AT3G57530 AT4G25470 AT4G25490 AT5G08790 AT5G22250 AT3G62400 AT1G19020 AT1G22810 AT1G52890 AT1G78410 AT2G33580 AT4G26080 AT5G03210 AT5G45340 AT5G59550 AT5G63790
7	Down regulated	1.76E-10	29	Response to acid chemical	AT3G11410 AT3G57530 AT4G26080 AT4G31550 AT1G10560 AT2G41010 AT3G56880 AT5G26920 AT5G59550 AT5G62470 AT1G18570 AT1G27730 AT1G72520 AT1G80840 AT2G38470 AT4G11280 AT4G23060 AT4G23810 AT5G08790 AT5G27420 AT1G01720 AT1G22810 AT1G52890 AT2G35930 AT2G37040 AT3G15210 AT5G45340 AT1G69530
8	Down regulated	5.87E-09	32	Response to endogenous stimulus	AT1G07430 AT3G11410 AT3G57530 AT4G26080 AT4G31550 AT1G10560 AT1G19770 AT1G22810 AT1G28370 AT1G53170 AT2G22200 AT3G15210 AT4G17490 AT4G39780 AT5G25190 AT5G26920 AT5G47230 AT5G51190 AT5G59550 AT5G62470 AT1G18570 AT1G27730 AT1G78410 AT2G33580 AT4G23060 AT4G23810 AT5G08790 AT5G27420 AT1G01720 AT1G52890 AT1G69530
9	Down regulated	5.87E-09	38	Regulation of cellular macromolecule biosynthetic process	AT1G07430 AT3G11410 AT3G57530 AT4G26080 AT4G31550 AT4G33465 AT5G04260 AT5G22250 AT5G47640 AT1G27730 AT3G21460 AT3G62950 AT4G23810 AT4G25490 AT5G08790 AT1G01720 AT1G18570 AT1G22810 AT1G26945 AT1G28370 AT1G53170 AT1G69490 AT1G72520 AT1G78410 AT2G22200 AT2G23320 AT2G38470 AT3G15210 AT4G17490 AT4G23800 AT4G25470 AT4G31550 AT4G37040 AT4G39780 AT5G04340 AT5G25190 AT5G47230 AT5G51190 AT5G63790 AT5G26920 AT5G52890 AT5G59980
10	Down regulated	6.50E-09	36	Regulation of transcription, DNA-templated	AT1G07430 AT1G27730 AT3G11410 AT3G57530 AT4G26080 AT4G31550 AT4G33465 AT5G04260 AT5G22250 AT5G47640 AT1G27730 AT3G21460 AT3G62950 AT4G23810 AT5G08790 AT1G01720 AT1G18570 AT1G22810 AT1G26945 AT1G28370 AT1G53170 AT1G69490 AT1G72520 AT1G78410 AT2G22200 AT2G23320 AT2G38470 AT3G15210 AT4G17490 AT4G23800 AT4G25470 AT4G31550 AT4G39780 AT5G04340 AT5G25190 AT5G47230 AT5G51190 AT5G63790 AT5G26920 AT5G52890 AT5G59980
11	Up regulated	0.00281321	9	Cell wall organization or biogenesis	AT1G03870 AT3G02570 AT1G24170 AT1G70090 AT2G43620 AT4G03210 AT4G13340 AT4G30270 AT5G62360

Table 3.6: Differentially expressed biological pathways between treated vs. untreated diploids

	Direction	adj.Pval	nGenes	Pathways	Genes
1	Down regulated	2.34E-06	5	Negative regulation of signal transduction	ATTG07430 AT3G11410 AT4G26080 ATTG01720 ATTG19770
2	Down regulated	4.30E-06	4	Negative regulation of abscisic acid-activated signaling pathway	AT1G07430 AT3G11410 AT4G26080 AT1G01720
3	Down regulated	2.45E-05	10	Response to endogenous stimulus	ATTG07430 AT3G11410 AT4G26080 AT4G31550 ATTG19770 ATTG53170 ATTG01720 AT1G05680 AT4G27260 AT4G27410
4	Down regulated	2.45E-05	6	Regulation of signal transduction	ATTG07430 AT3G11410 AT4G26080 AT4G31550 ATTG01720 ATTG19770
5	Down regulated	2.45E-05	8	Cellular response to hormone stimulus	ATTG07430 AT3G11410 AT4G26080 AT4G31550 ATTG19770 ATTG53170 ATTG01720 AT1G05680
6	Down regulated	2.45E-05	4	Indole-containing compound metabolic process	AT4G37370 AT5G57220 ATTG05680 AT4G27260

Table 3.7: Differentially expressed biological pathways between treated polyploids vs. treated diploids

	Direction	adj.Pval	nGenes	Pathways	Genes
1	Up regulated	2.97E-07	12	Response to stress	ATTG02930 AT2G26020 AT2G44840 AT4G02380 ATTG21850 AT5G47230 AT5G57220 ATTG18570 AT5G22250 ATTG19020 AT5G20230
2	Up regulated	1.36E-05	8	Defense response	ATTG02930 AT2G26020 AT2G44840 AT4G02380 AT5G47230 AT5G57220 AT5G22250 AT1G18570
3	Up regulated	3.63E-05	7	Response to biotic stimulus	AT2G26020 ATTG02930 ATTG18570 AT4G02380 AT5G20230 AT5G22250 AT5G57220
4	Up regulated	4.10E-05	5	Defense response to bacterium	ATTG02930 AT5G22250 ATTG18570 AT4G02380 AT5G57220
5	Up regulated	6.07E-05	5	Response to oxidative stress	ATTG02930 AT4G21850 AT4G02380 ATTG19020 AT5G20230
6	Up regulated	6.07E-05	2	Indole glucosinolate biosynthetic process	ATTG18570 AT5G57220
7	Up regulated	9.19E-05	6	Defense response to other organism	AT2G26020 ATTG02930 AT5G22250 ATTG18570 AT4G02380 AT5G57220
8	Up regulated	0.00013271	7	Response to oxygen-containing compound	AT5G04340 ATTG02930 ATTG18570 AT2G44840 AT4G02380 AT5G47230 AT2G26650
9	Up regulated	0.00014233	5	Response to drug	AT5G04340 ATTG18570 AT2G44840 AT4G02380 AT5G47230
10	Up regulated	0.00014343	4	Response to ethylene	AT2G44840 AT5G47230 ATTG18570 AT4G02380

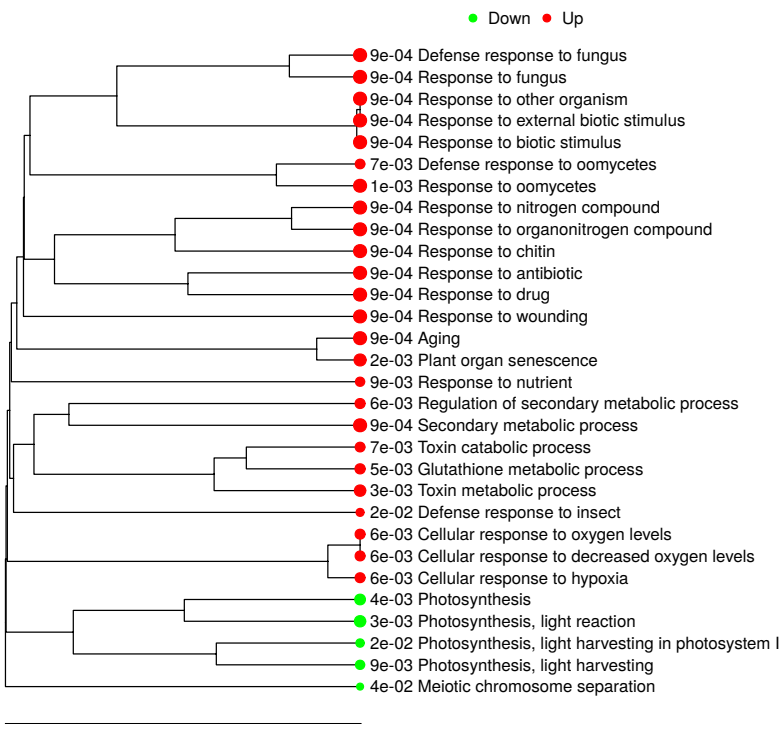


Figure 3.4: Enriched KEGG pathways of differentially expressed genes between polyploid and diploid untreated samples

Both polyploids and diploids down regulate defense pathways

Defense-associated genes

Using the most conservative approach (DESeq) to identify differentially expressed genes, I found that several defense-associated genes were down-regulated in the microbiome treated plants compared to the controls (Figure 3.8). In particular, I found that AT1G03220, which encodes secreted aspartic protease 2 (SAP2), was down-regulated, which is involved in defense against bacterial pathogens by cleaving the conserved bacterial protein MucD in order to inhibit growth and incurs no penalty to plant growth (Wang et al., 2019). Similarly, AT2G33580, which encodes LYSM-containing receptor-like kinase 5 (ATLYK5) was found to be down-regulated, which is associated with protection against fungi, as it forms a chitin induced complex with the main chitin receptor in Arabidopsis, AtLYK5, in order to induce immune response (Cao et al., 2014). In a seemingly opposite manner, jasmonate (JA)-signaling pathway protein VQ12, which is encoded by AT2G22880 was also found to be down regulated and it plays a role as a negative regulator in plant basal defense to fungi (Wang et al., 2015). However, other VQ proteins serve as positive regulators of plant basal defense and it is hypothesized that these proteins act to fine-tune defense signaling pathways in order to mediate the trade-off between defense and growth (Wang et al., 2015). I

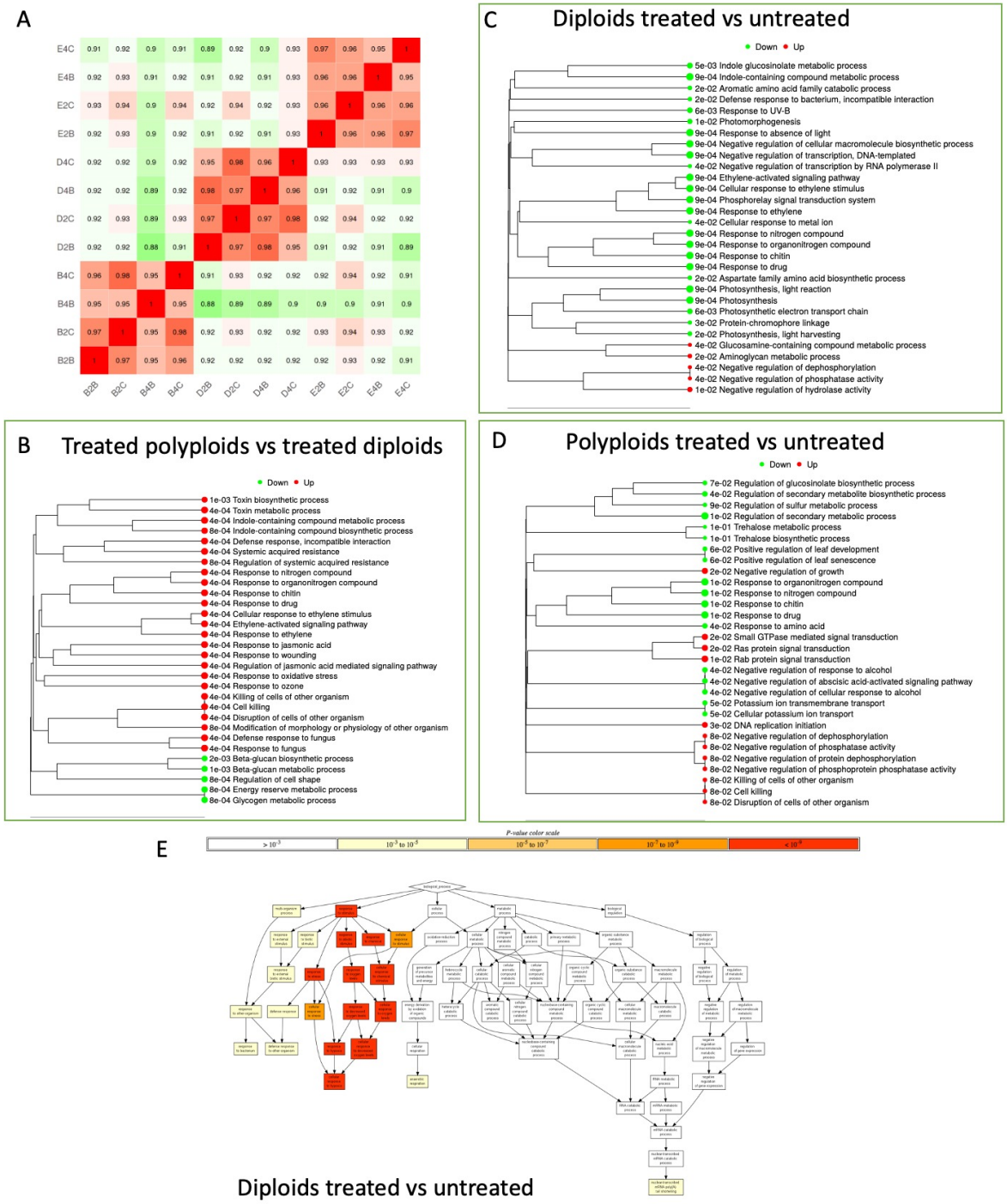


Figure 3.5: Heatmap of the sample-to-sample distances based on hierarchical clustering of the variance-stabilized distance matrices for samples (A). Enriched KEGG pathways of differentially expressed genes between (B) microbiome-treated polyploids and treated diploids, (C) treated diploids and untreated diploids, (D) treated poylploids and untreated polyploids. (E) GOrilla-generated visualisation of enriched GO terms in biological processes for target genes significantly up or down regulated between microbiome treatments within diploid samples.

Direction	Pathways	NES	nGenes	adj.Pval
Down	Response to chitin	-2.4694	122	0.0092
Down	Response to organonitrogen compound	-2.2748	232	0.0092
Down	Response to nitrogen compound	-2.2742	288	0.0092
Down	Photosynthesis	-1.9855	229	0.0092
Down	Cellular response to ethylene stimulus	-1.9724	159	0.0092
Down	Response to drug	-1.9354	566	0.011
Down	Response to ethylene	-1.9141	244	0.0092
Down	Ethylene-activated signaling pathway	-1.9076	146	0.0092
Down	Indole-containing compound metabolic process	-1.8779	80	0.014
Down	Photosynthesis, light reaction	-1.8651	116	0.011
Down	Phosphorelay signal transduction system	-1.8556	183	0.0092
Down	Protein-chromophore linkage	-1.8368	40	0.056
Down	Negative regulation of abscisic acid-activated signaling pathway	-1.8265	47	0.05
Down	Negative regulation of response to alcohol	-1.8265	47	0.05
Down	Negative regulation of cellular response to alcohol	-1.8265	47	0.05
Down	Photosynthetic electron transport chain	-1.8245	47	0.05
Down	Negative regulation of transcription by RNA polymerase II	-1.8028	39	0.073
Down	Glycosyl compound catabolic process	-1.8021	37	0.073
Down	Response to water deprivation	-1.6917	343	0.0092
Up	Negative regulation of phosphatase activity	1.931	17	0.014
Up	Negative regulation of dephosphorylation	1.931	17	0.014
Up	Negative regulation of phosphoprotein phosphatase activity	1.9119	16	0.014
Up	Negative regulation of protein dephosphorylation	1.9119	16	0.014
Up	Negative regulation of growth	1.8379	31	0.03
Up	Regulation of phosphoprotein phosphatase activity	1.816	38	0.034
Up	Regulation of phosphatase activity	1.809	42	0.034
Up	Regulation of protein dephosphorylation	1.8063	40	0.032
Up	Regulation of dephosphorylation	1.7926	45	0.038
Up	Regulation of meristem growth	1.7475	41	0.068
Up	Negative regulation of catalytic activity	1.7011	128	0.014

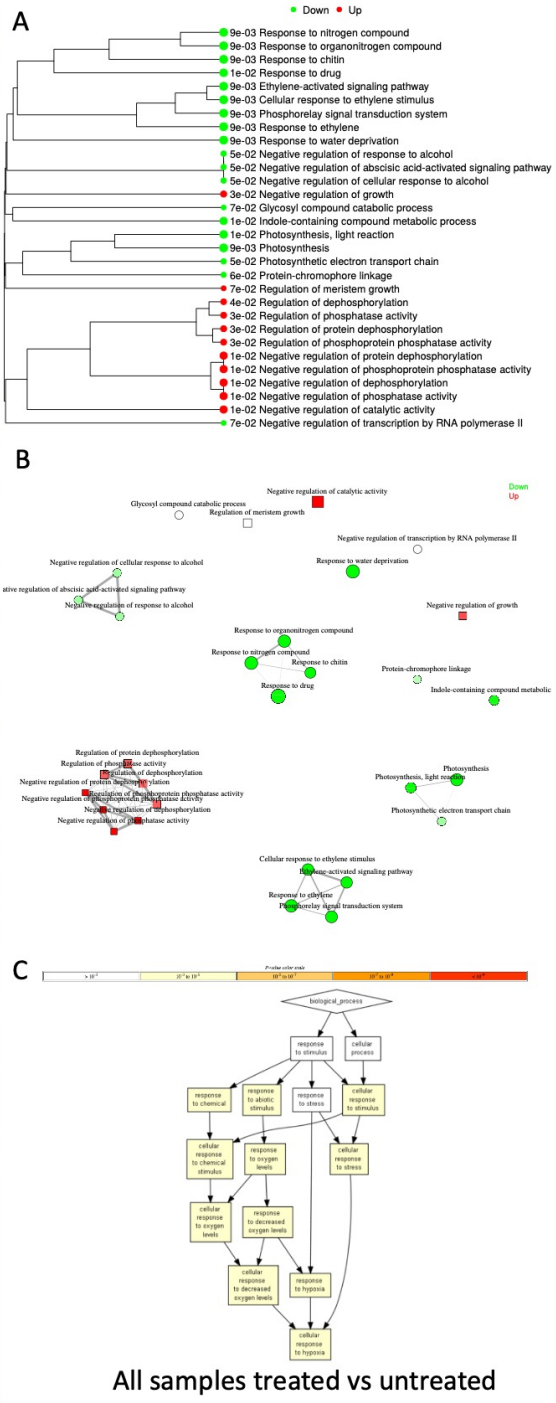
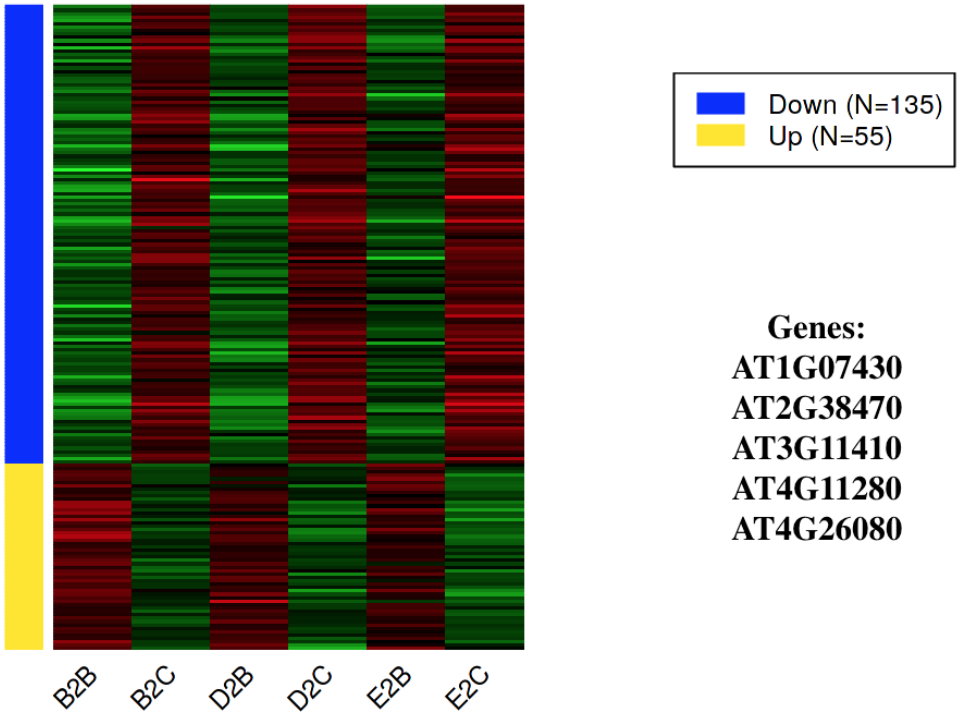


Figure 3.6: Comparisons of microbiome-treated plants with non-treated plants. (Right) Summary statistics of up and down regulated pathways. Enriched KEGG pathways of differentially expressed genes maped by similarity (A) and network analysis (B). (C) GOrilla analysis of enriched genes from the set of significantly up- or down-expressed genes discovered using DEseq2.



Enriched pathways in DEGs for the selected comparison:

Direction	adj.Pval	nGenes	Pathways
Down regulated	3.2e-03	5	MAPK signaling pathway

Figure 3.7: Enriched pathways in differentially expressed genes in the KEGG pathway database for microbiome-treated versus control diploid samples.

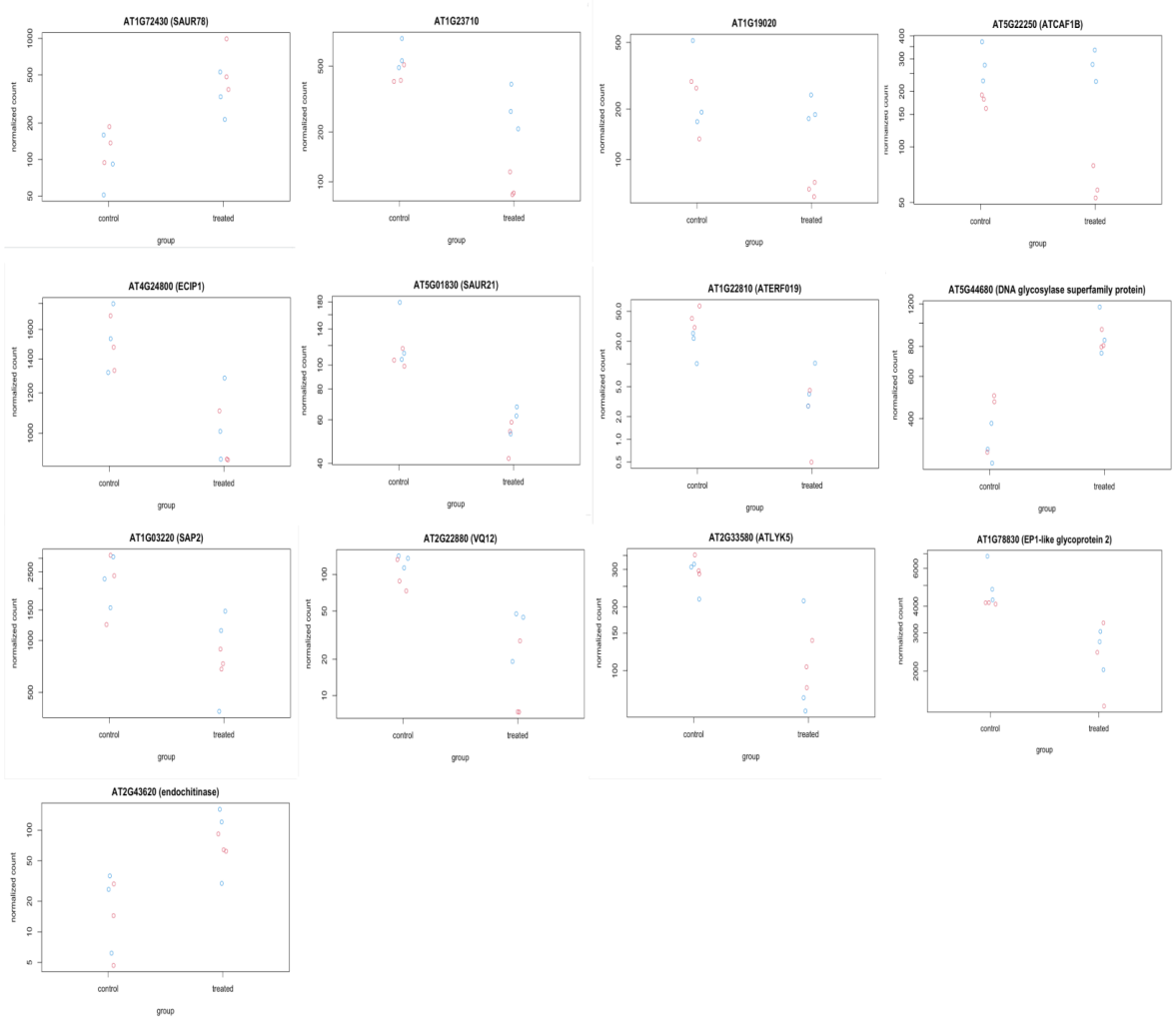


Figure 3.8: Plots of counts for each sample in microbiome-treated or non-treated groups. Ploidy is colored: diploids in red, tetraploids in blue.

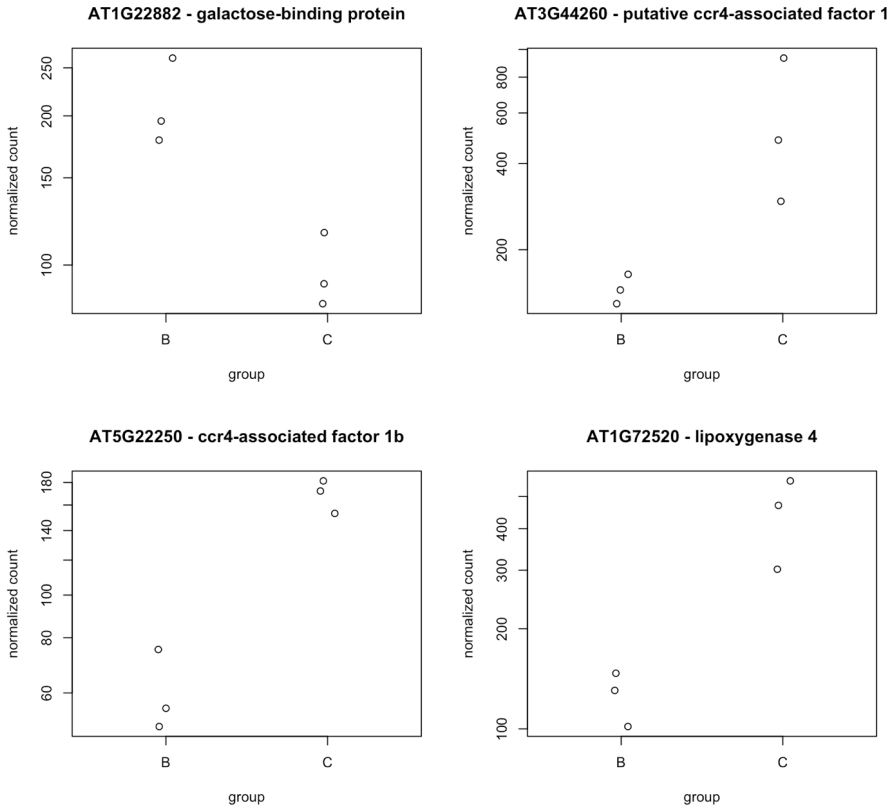


Figure 3.9: Plots of counts of genes associated with response to bacteria for each sample in microbiome-treated or non-treated diploids.

also looked at Constructively Activated Cell Death 1 (CAD1, AT1G29690, Figure 3.10), a negative regulator of SA mediated cell death defense (Morita-Yamamuro et al., 2005), which was found by Chen et al. (2020) to control phyllosphere microbiota and prevent dysbiosis in the phylosphere and found that treated groups had significantly less expression (adjusted p-value = 0.0065, DESeq) than untreated plants–another example of microbiome treated plants increasing their defenses, but maybe at the cost of controlling other bacteria that could potentially aid themselves in controlling pathogenic taxa. However, when I investigated if this pattern was driven more by samples of one ploidy versus another, I found that while this general pattern held true, diploids were down regulating gene networks associated with defenses while polyploids were maintaining them. This is of interest as plants have been hypothesized to have evolved to associate with specific bacteria in order to get around trade-offs that stem from growth-defense coregulation by outsourcing their defense to the beneficial bacteria. The results here point to a scenario where diploids are potentially doing this outsourcing, but polyploids are not.

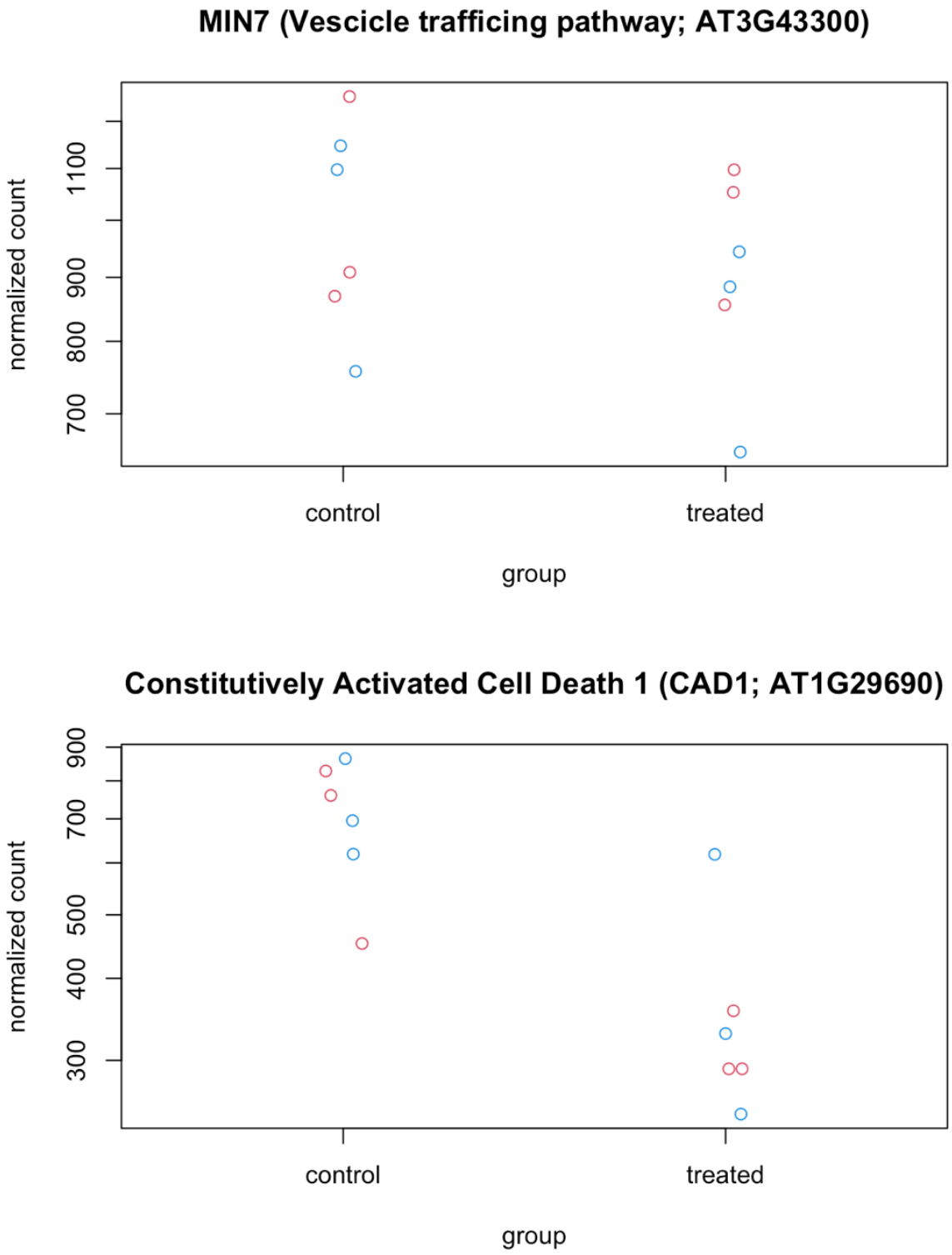


Figure 3.10: Plots of counts for each sample in microbiome-treated or non-treated groups for the genes in Chen et al. (2020). Ploidy is colored: diploids in red, tetraploids in blue.

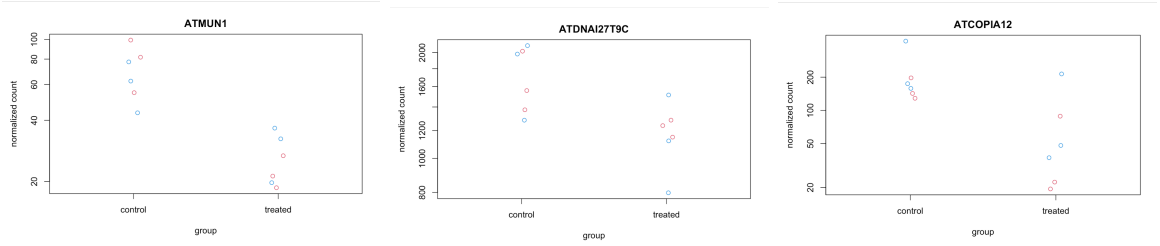


Figure 3.11: Plots of transposon counts for each sample in microbiome-treated or non-treated groups. Ploidy is colored: diploids in red, tetraploids in blue.

Growth associated genes

Ethylene-signaling genes were found by DESeq to be highly enriched in this comparison as well. I found that AT1G72430 which encodes Small Auxin Upregulated RNA 78 (SAUR78) was significantly up-regulated in plants that were inoculated with the synthetic microbiome. SAUR78 is one of three SAUR proteins (SAUR 76-78) that are associated with ethylene receptor signaling and are hypothesized to promote plant growth in Arabidopsis (Li et al., 2015). Another gene, AT1G22810 (ATERF019), associated with ethylene response was found to be significantly down-regulated in treated samples, and is a member of the APETALA2/ethylene response factor superfamily (Jofuku et al., 1994). The overexpression of AtERF019 was found to delay both flowering time and senescence time and to improve drought tolerance (Scarpeci et al., 2017). AT4G24800 which encodes ECIP1 was also found to be down-regulated, and has been found to increase ethylene response and salt tolerance in loss-of-function mutants (Lei et al., 2011). However, AT5G01830 which encodes armadillo/beta-catenin repeat family protein (SAUR21) is down-regulated in the treatment samples, which is in a group of genes (SAUR19–24) known to be associated with auxin mediated increases to hypocotyl and leaf size (Spartz et al., 2012). These results together demonstrate that inoculation with commensal microbiome elicits a strong transcriptional response of the hormones that regulate both growth and senescence (Figure 3.8). This could potentially be associated with the down-regulation of defense genes across ploidies when treated with the synthetic community as there is significant growth-defense coregulation via similar molecular pathways. Therefore, by lowering defense, these plants are potentially able to increase the transcription and regulation of growth processes.

Polyploids consistently maintain more defenses than diploids

An interaction between stress response and nutrient response mediated by preferential plant associations with microbes has been demonstrated in phosphate stress response in Arabidopsis (Castrillo et al., 2017). I found two main results: 1) polyploids have more defenses in general, and 2) polyploids don't change their defence allocation in response to the microbiome.

However, this does not necessarily point to a scenario where diploids living in a "safe-space" lower their defenses to bacteria to prioritize nutrient accumulation, as the GAGE analysis between treated and control diploid samples identified that genes involved in photosynthesis were significantly down-regulated in the treated samples (Figure 3.5). However, considering the importance of sugar as energy and structure for plant defense and its role in signaling and priming of defense pathways (Morkunas and Ratajczak, 2014), this could be another example of diploids lowering intrinsic defensive capacity at the same time that polyploids are maintaining it. When comparing between diploid and polyploid treated samples, the diploids have significantly up-regulated starch and glycogen metabolism pathways, potentially pointing to differential resource allocation leaning more towards defense in the polyploids (Figure 3.5) as a consequence of the limitations due to growth-defense coregulation (i.e. less defense means those pathways can be used for growth). Similarly, enriched pathways in differentially expressed gene sets of the KEGG pathway database for microbiome treated versus control diploid samples show us that the mitogen-activated protein kinase (MAPK) signaling pathway is downregulated in treated samples (Figure 3.7) and MAPKs are extremely important signaling factors in plant defense signaling playing an role in both jasmonic acid and salicylic acid signaling (Jagodzik et al., 2018).

Similarly, I found that with respect to bacterial pathogen response, diploids significantly down regulated ATCAF1B, but polyploids maintained a similar expressional profile to the untreated samples. ATCAF1B has been found to be involved with multistress resistance (biotic and abiotic) (Walley et al., 2010), in particular resistance to the pathogen *Pseudomonas syringae* (Liang et al., 2009). In fact, the GAGE and parametric GAGE enrichment analysis of KEGG pathways found that the plant-pathogen interaction and defense and signaling pathway were both significantly up-regulated in polyploids than diploids. One possible explanation for this pattern could be that polyploids are less responsive to the microbiome or that they are less selective in selecting their microbiome due to a disruption of fine-tuned pathways as a consequences of a doubling of gene dosage across the genome. Indeed, when comparing the untreated polyploids with the untreated diploids, I found similar patterns of increase polyploid defense transcription to diploids as when I compared treated polyploids with treated diploids. These results for the first time provide empirical transcriptional evidence that polyploids have a more active defense response to bacteria and supports the theoretical models that hypothesize that they are more resistant to pathogens (Oswald and Nuismer, 2007).

A similar explanation is that diploids have a more active role in promoting a community of commensal bacteria. Diploids being more responsive to treatment with the commensal community is consistent with the theory that polyploids are more resistant to pathogens and therefore have transcription patterns consistent with bacterial defense even when the bacteria are potentially beneficial. This could be attributable to an increased insensitivity to the growth-defense trade-off due to the general robustness of some polyploids even given increased dosage of defense genes (Ng et al., 2012), whereas diploids may be able to lower defense expression in order to allocate more resources to growth in the presence of beneficial bacteria. Therefore, a consequence of whole genome duplication maybe a loss of control over

the composition of the microbiome. While controling the microbiome may be an adaptation for diploids, my study demonstrates how the loss of this control in polyploids, which would be a spandrel of increased resistant to pathogens, could also be a source of variation in the microbiome upon which natural selection can act.

Transposable elements are active following neopolyploidy

I also demonstrate that the induction of a microbial environment on Arabidopsis plants has generalizable effects across accessions and ploidies on the transcription of at least one family of Mutator (Mu) regulatory element (MuDR). MuDR/Mu transposable elements are highly active in plants, especially in Maize, and plants have developed many developmental or epigenetic strategies to control their proliferation (Walbot, 1992). The decreased transcription of ATMUN1, which is a non-automonous element (i.e. reliant on the transcription mechanisms of other TEs to move), could potentially be due to ways by which the broad scale changes of plants in response to the microbiome have interactions with pathways involved with the regulation and silencing of TEs.

Conclusion

My study provides transcriptional evidence for a model where neopolyploids have increased resistance to pathogens by having an increase in the expression of defense genes regardless of the surrounding bacteria, but that diploids have potentially evolved to down regulate defense genes in the presence of protective bacteria. Therefore one potential immediate consequence of whole genome duplication may be a loss of control over the composition of the defensive microbiome and may explain why diploids may outcompete polyploids in the parent range but that polyploids often outcompete diploids in disturbed or changing environments (López-Jurado et al., 2019; Decanter et al., 2020).

Chapter 4

Understanding endopolyploidy: realistic next steps

Abstract

Polyploidy is a ubiquitous phenomenon across the tree of life. While the ecological and evolutionary consequences of this phenomenon as a large-scale, germline mutation have been well-studied, polyploidy also occurs in somatic cells—such genome duplications are referred to as endopolyploidy. This review surveys what is known about somatic polyploidy and frames the open questions in the field to assess whether endopolyploidy and germline whole-genome duplication have shared underlying evolutionary rules.

4.1 Introduction

There have been recent calls to welcome in a new decade of polyploidy research ("Polyploidy 2030") by integrating knowledges across different fields that have hitherto remained separate and which span different levels of organization, diversity, and applications in health and agriculture (Fox et al., 2020). Importantly, these authors note that polyploidy occurs on both organismal and sub-organismal levels and argue that genome doubling can be conceptualized as a unified biological phenomenon with emergent commonalities. However, it is not readily explicable how somatic polypolyploidy (endopolyploidy) relates to germline polyploidy (whole-genome duplication or WGD[1]) from an evolutionary perspective since there are many obvious differences in both the proximate and ultimate consequences of the two processes. In fact, the nascent field of endopolyploidy studies has not been thoroughly reviewed and there remains many open questions about its fundamental biology. If we are to seriously embark on "joint research ventures across diverse disciplines [that] will promote a better understanding of polyploidy at both the cellular and organismal levels", it is incumbent on

[1] in this chapter WGD will refer solely to germline polyploidy

us to clearly articulate the differences and similarities between these two phenomena. This review surveys what is known about somatic polyploidy and frames the open questions in the field around necessary first steps that we are to take to meaningfully address the question whether endopolyploidy and whole-genome duplication have shared underlying rules.

There has been no broad survey of endopolyploidy research to date. Therefore this review will do the following: first, provide a background on ways that scientists study endopolyploidy in different groups; second, compare the major findings and interpretations between animals, plants, and cancer research, which has hiterto remained separate fields; third summarize the evolutionary implications of a potential synthesis of these researches and the open questions that remain if we are to unite the germline and somatic polyploidy research and uncover "rules of life." I argue that it will be important to clarify the traits we are talking about when talking about polyploidy and that a helpful trait to consider may not just be polyploidy of a cell or of all the cells in an organism, but rather the ability to transition between diploid and polyploid states. I also argue that the possible overlap in the ways that lineages have evolved to make use of polyploidy through both endopolyploidy and WGD cannot yet be inferred until more basic research is done into understanding how common these phenomena are across the tree of life and encourage more work in the descriptive sciences.

4.2 What are the methods used to study endopolyploidy?

Unlike germline polyploidy that has an clearly observed and historical suite of traits (e.g. chromosome counts, synteny) that have been used to map the extant and historical distribution of this trait across various branches on the tree of life, not much is known about endopolyploidy and it has only recently become an object of inquiry.

There are several ways to assess the amount of endopolyploidy in a tissue by identifying the number of endopolyploid cells. DNA content of individual fluorescently labelled cells can be measured using flow cytometry and used to identify different ploidy levels by the nuclei peaks on a histogram (e.g., 2C, 4C, 8C, etc.). The most basic description of endopolyploidy is simply the nuclei number for each ploidy level and individuals can be compared using the mean C-value (picogram amount of DNA contained within a haploid nucleus). However, these methods pose statistical problems in measuring variance and accounting for the outsized weighting of higher ploidy levels (Gegas et al., 2014; Barow and Meister, 2003). To account for these issues, the cycle value or Endopolyploidy Index (EI) has been proposed by Barow and Meister (2003) as a standard measurement and has been adopted by many recent studies (Pacey et al., 2020; Zedek et al., 2020; Bainard et al., 2020). EI can be calculated using the following equation:

$$EI = \frac{0 * n_{2C} + 1 * n_{4C} + 2 * n_{8C} + 3 * n_{8C} + ...}{n_{2C} + n_{4C} + n_{8C} + n_{16C} + ...} \tag{4.1}$$

DNA densitometry (Feulgen densitometry) is another method that is used to measure the amount of endopolyploidy in a cell (how many copies of its genome) or tissue and has been used to discover endopolyploid cells that are uncommon or uncommonly high in DNA content (Neiman et al., 2017). These methods are much more labor intensive than flow cytometry and have been more commonly used to asses endopolyploidy levels in animals than in plants (Rasch and Wyngaard, 2008; Carella et al., 2017).

In the last couple of years, we have begun to see endopolyploidy being studied using phylogenetic comparative methods, a historical framework that allows scientists to study the evolution of characters across a phylogeny and/or the process of branching itself to infer diversification rates. Bainard et al. (2020) tested whether there was phylogenetic signal (related species having similar traits) for genome size or EI across the moss phylogeny and if genome size and EI were correlated. These scientists found that in mosses there is phylogenetic signal for genome size but not for EI and that EI and genomic size are not correlated when accounting for the phylogeny. This pivotal study demonstrates the potential for endopolyploidy to be studied in a phylogenetic framework and the production of more datasets with better sampling and trait information will allow for questions pertaining to joint estimation of trait evolution and diversification rates to be addressed in the future.

Table 4.1: Recent studies of endopolyploidy in animals

Taxa	Paper	Summary	Germline WGD?
Ants	Scholes et al. (2013)	abdomens and males have higher rates of endoreduplication	Not likely in the lineages studied (Tsutsui et al. (2008))
Honey Bee	Rangel et al. (2015)	many types of endopolyploidy cells have a reduction in ploidy levels with age	Not likely (Roelofs et al. (2020))
Termites	Nozaki and Matsuura (2019)	fat body of queens have more endopolyploidy than non-queens and foragers more than wood-dwelling	Not likely
Ribbon worms, Nemertea	Mulligan et al. (2014)	endopolyploidy is common	Unknown
Drosophila	Losick et al. (2013)	endopolyploidy is induced by wounding	Not likely
Mouse	Hu and Cross (2009)	endopolyploidy is essential for reproduction	Not likely
Mammals	Gandarillas et al. (2018)	endoreplication is widespread in mammalian developing tissues	Not likely

Table 4.2: Recent studies of endopolyploidy in plants

Taxa	Paper	Summary	Germline WGD?
Arabidopsis	Zedek et al. (2020)	endopolyploidy is a response to UV stress	Yes (Blanc and Wolfe (2004))
Corydalis	Kolarčik et al. (2020)	endopolyploidy common especially in geophytes and negatively correlated with genome size	Unknown but found in other members of Papaveraceae (Cui et al. (2006))
Arabidopsis	Mesa et al. (2019)	endopolyploidy is a response to herbivory	Yes
Conophytum	Powell et al. (2020)	endopolyploidy is extremely common in flower and leaf tissue and negatively correlated with genome size	Yes up to 17
Mesembryanthemum crystallinum	Barkla et al. (2018)	endopolyploidy increases with leaf size and in response to salt stress	Yes
Arabidopsis	Pacey et al. (2020)	induced whole genome duplication reduced endopolyploidy levels	Yes
Bryophyta	Bainard et al. (2020)	mosses consistently have high levels of endopolyploidy and small genome sizes	Most likely (Rensing et al. (2007))
Chenopodioideae	Skaptsov et al. (2017)	endopolyploidy is common and more common in diploids than polyploids	Yes
Phalaenopsis	Kwon et al. (2018)	endopolyploidy is induced by light stress	Yes (Cai et al. (2015))
Maize	Li et al. (2019)	endopolyploidy is common in many organs and tissues	Yes (Gaut et al. (2000))
Barley	Zeng et al. (2017)	endopolyploidy is reduced in response to phosphorus stress	Yes (Thiel et al. (2009))
Tomato	Musseau et al. (2020)	endopolyploidy plays an important role in the production of fruit	Yes (Consortium et al. (2012))
Bee orchid	Bateman et al. (2018)	endopolyploidy negative associated with polyploid speciation	Yes
Opuntia	Palomino et al. (2016)	endopolyploidy was common in many tissues	Yes

4.3 Mechanisms of genome doubling

While much of the current research into endopolyploidy has been studying plant taxa (Table 4.1, 4.2), endopolyploidy has evolved several times across the tree of life (Figure 4.2) through roughly the same mechanisms: endocycling, whereby the S and G phases of the cell cycle are alternated without mitosis, or endomitosis, whereby mitosis occurs in a cell but cytokynesis is not carried out (Figure 4.1, Lee et al. (2009); Donné et al. (2018)). These processes are a form of somatic mutation but have also evolved to be programmed developmentally and to exhibit heritable variation across tissues and taxa (Galbraith et al., 1991; Barkla et al., 2018; Jalal et al., 2015; Barow, 2006).

The cellular effects of whole genome duplication especially on cell and nuclear volume and other organismal processes are driven in the same way as endopolyploid cells through the increase of copies of the nuclear genome (Scholes and Paige, 2015). WGD is formed, nonetheless, as a germ line mutation that is heritable and expressed somatically throughout the entirety of an organisms, whereas endopolyploidy happens on a cell-by-cell basis. The two broad categories of WGD formation are autopolyploidy and allopolyploidy and these have important cellular and organismal differences. First, the problem of improper segregation of sister chromosomes in meiosis can lead to reduced fertility. Since the homologous chromosomes of allotetraploids (which arise from hybridization of different species) often segregate by subgenome, multivalents are more of a problem for autotetraploids and are resolved evolutionarily by reducing cross-over rates (Lloyd and Bomblies, 2016).

4.4 Is endopolyploidy specific to plants?

In animals (for a review see Neiman et al. (2017)), endopolyploidy is hypothesized to play two main functional roles: increasing cell size and/or altering gene expression, both of which have phenotypic consequences that are visible to selection such as growth rate, organ/body size, or organ/body modifications. There is limited evidence in animals that endopolyploidy has evolved to increase protein production in certain organs (e.g., in silk-producing animals; Neiman et al. (2017); Xia et al. (2009)). There is also evidence that increases in RNA polymerase II are associated with cell ploidy level in both Arabidopsis and tomato and have evolved to increase the transcriptional activity of cells (Schubert, 2014; Bourdon et al., 2012). Similarly, endopolyploidy modifies the accessibility of chomatin within a cell that could lead to increases in gene expression (Barow, 2006). However, while the increased transcriptional and metabolic output caused by endopolyploidy could have an important role in stress response by producing more chemicals that can defend against herbiovry or UV stress (Leitch and Dodsworth, 2007), the main role that endopolyploidy is thought to have in plants is by the positive correlation between DNA content and cell size and the associated changes in growth and cell function with cell size (Barow, 2006).

Endopolyploidy is very common in insects where most of the observational work has been performed. Endopolyploidy has been found to be common in ants especially in the

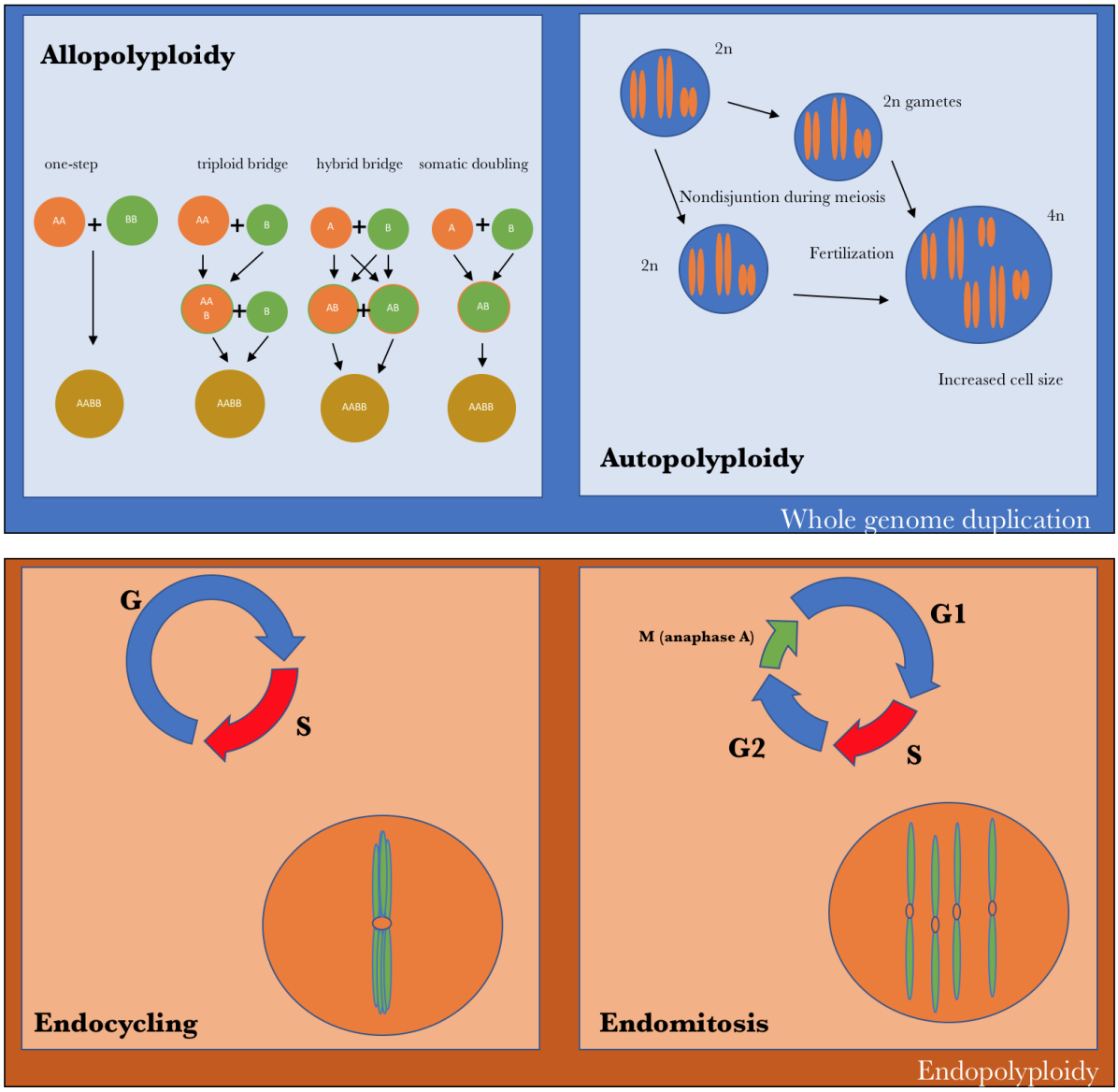


Figure 4.1: Different mechanisms of somatic and germline polyploidy. For endopolyploidy (orange), either endocycling occurs whereby the S and G phases of the cell cycle are alternated without mitosis or endomitosis occurs whereby mitosis occurs in a cell but cytokynesis is not carried out. WGD (blue) can occur either via allopolyploidy or autopolyploidy. Multiple sources of allopolyploid formation are presented.

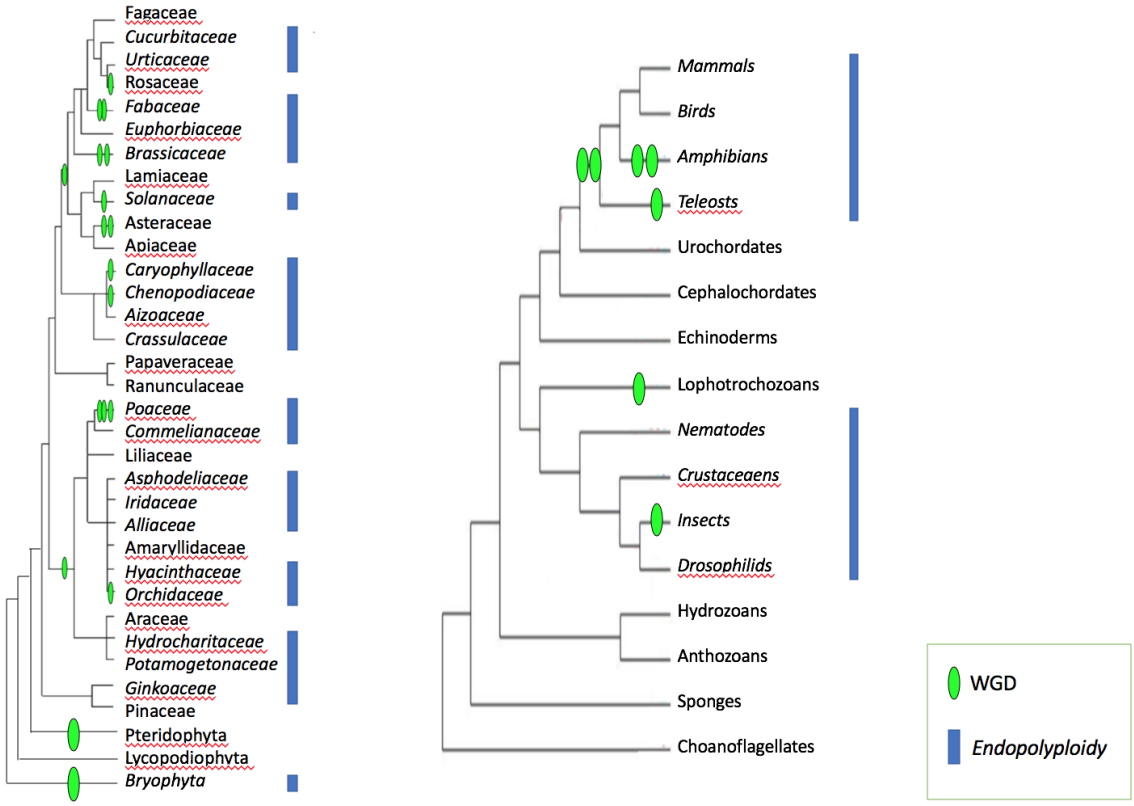


Figure 4.2: Cladograms of major lineages of plants (left) and animals (right) with putative stem whole-genome duplications (WGD) marked by green ovals and lineages with evidence of endopolyploidy indicated by blue bars and italicized taxa names. Cladograms modified from Barow (2006) and Maeso et al. (2012). WGD data mapped from CoGepedia Wiki and Maeso et al. (2012) and endopolyploidy data mapped from Barow (2006) and Neiman et al. (2017). However, endopolyploid has not been widely investigated and it remains to be known whethere these are false or true negatives.

abdomen and among males (Scholes et al., 2013). In honeybees, where females exhibit polyethism, levels of endopolyploidy was found to decrease in most tissues over time and with more complex and age-associated social roles such as foraging (Rangel et al., 2015). In termites, the fat bodies of forager species were found to have more endopolyploidy than wood-dwelling species and queens were found to have more endopolyploidy in their fat bodies than non-queens (Nozaki and Matsuura, 2019). Endopolyploidy has been found to be common in *Drosophila* (Losick et al., 2013) and ribbon worms (Mulligan et al., 2014), and also in mammals, where endoreplication is widespread in developing tissues and found to be essential to the development of the rodent placenta (Gandarillas et al., 2018; Hu and Cross, 2009).

Nonetheless, there is a substantially more research into endopolyploidy in plants including a large body of experimental work in addition to natural history. The phenomenon is common in many plant taxa such as *Opuntia*, Maize, Arabidopsis and *Conophytum* where it is found in many different tissues and organs (Palomino et al., 2016; Li et al., 2019; Powell et al., 2020; Pacey et al., 2020; Zedek et al., 2020) and endopolyploidy has been found to be induced in response to UV stress (Zedek et al., 2020), herbivory (Mesa et al., 2019), salt stress (Barkla et al., 2018), light stess (Kwon et al., 2018), and phosphorous stress (Zeng et al., 2017). Just as ancient WGD is now considered to be ubiquitous across the evolution of plants (Cui et al., 2006), it is now accepted that endopolyploidy is also a ubiquitous and widespread phenomenon (Figure 4.2).

4.5 Are there tissue-specific patterns?

Endopolyploidy allows both for cells to have an alternative strategy for growth and to support specialized function for differentiated cells (reviewed in Lee et al., 2009). For example, the growth of the young hypocotyl in plants is driven by endoreplication (Lee et al., 2009) as well as the growth of larvae in *Caenorhabditis elegans* and Drosophila (Edgar and Orr-Weaver, 2001; Lozano et al., 2006) and both plants and animals use endopolyploidy for nutrient uptake and storage (Lee et al., 2009). Specialized structures such as the trichomes in plants or nurse and follicle cells in Drosophila are also endopolyploid and have evolved specialized and essential functions within lineages (Lee et al., 2009). Likewise, endopolyploidy plays an important role in arbuscular mycorrhizas where endopolyploidy is induced in anticipation of fungal colonization (Carotenuto et al., 2019). Endopolyploidy has been found to be common especially in geophytes and fruit production and negatively correlated with genome size Kolarčik et al. (2020); Musseau et al. (2020).

Importantly, endopolyploidy allows both plants and animals to respond to stress by increasing growth when growth is prevented via mitosis (Weigmann et al., 1997; Cookson et al., 2006). Changing cell size can lead to different structural and physiological properties such as storing more salt, having more copies of DNA to buffer the effects of UV-B damage, or increasing surface area for photosynthesis (reviewed in Leitch and Dodsworth, 2007). However, Neiman et al. (2017) points out that while plants almost always see a direct correlation between increased cell size due to polyploidy and increased body size, animals often do not

and may maintain the same body size leading to a trade-off in organ complexity, e.g., in salamander brains (Vernon and Butsch, 1957).

Endopolyploidy has also been studied extensively in the context of cancer, but these contributions have been rarely examined in the context of ecology and evolution. In organs that have evolved to have a large amount of endopolyploid cells, such as the human liver (where around 30% of cells are polyploid (Donne et al., 2020)), it has been found that there is a positive correlation between the proportion of polyploid cells in the organ and the suppression of tumorigenesis, i.e., the diploid state was more susceptible to genomic aberrations in the form of tumor-suppression loss or oncogene activation (Zhang et al., 2018). This putative protective role of endopolyploidy is further supported by the observation that a reduction of ploidy is characteristic of liver cancer (Delgado et al., 2020) although this association is correlative and not causal. Although the polyploid state of the liver may act as a gatekeeper of tumeriogenesis by increasing the dosage of tumour suppressor genes and wild type copies of Tumor protein p53, it may also be a tumor-promoting mechanism since interfering with the p53 pathway is one way for polyploid cells to resume the cell-cycle and increase genomic instability often leading to a worse prognosis (Donne et al., 2020).

Moreover, there is also evidence that tetraploidy and associated aneuploidy play an important role in the development of cancer as the increased number of chromosomes buffers the negative consequences of evolving a mutator phenotype and that the increases in the rate of chromosomal missegregation will also increase the probability of evolving these mutator phenotypes (reviewed in Davoli and de Lange, 2011). Polyploid cancer cells have been found in the most common cancers (e.g., pancreatic, lung, colon, etc.) and many of the most difficult to treat cancers are characterized by polyploid cells (reviewed in Donne et al., 2020). Cancer stem-like cells are often derived from Polyploid Giant Cancer Cells (PGCCs) have been shown to occur due to a conserved evolutionary response to hypoxia stress that leads to endopolyploidy and also to play an integral role in regulating and growth and heterogeneity in other cancer cells (Zhang et al., 2014). This co-opting of a stress response pathway to induce polyploidy is thought to be an exaptation leading to the evolution of PGCCs as an adaptive response of cancers to chemotherapy (Lin et al., 2018).

4.6 Is endopolyploidy beneficial or deleterious?

Endopolyploidy is considered a potentially adaptive mechanism by which an organism can respond to environmental stress. In plants, which have been known to induce increased endopolyploidy in cells following myriad stress treatments including UV-B, temperature changes, drought, pH, and herbivory (for a more comprehensive list of demonstrated responses to environmental stress and hypothesized adaptive explanations see Leitch and Dodsworth (2007)), it has been hypothesized that endoreduplication has evolved as an integrated part of the general plant stress response pathway (Scholes and Paige, 2015). Mesa et al. (2019) found that both mammalian herbivory, which often removes the apical meristem, and insect herbivory, which does not, trigger endoreduplication. In animals, endopolyploid

cells in the epidermal cells of *Daphnia* helmets, a defense structure on the head which is mostly diploid, is thought to act as storage sites of dopamine to modulate the formation of bigger helmets in the presence of predators (Neiman et al., 2017). Likewise, damaged liver cells in animals are able to regain functional activity through growth via endoreplication instead of cell replication (Denchi et al., 2006; Lee et al., 2009). Due to increases in cellular imaging, it is now thought that more than fifty per cent of mammalian cells are developmentally programmed polyploid cells which have essential roles in tissue repair, but may come with a latent cost of being a driver of disease in old age (reviewed in Gjelsvik et al., 2019).

Like whole-genome duplication, endopolyploidy is also considered to be a mechanism that has evolved for plants to accelerate growth in environments that benefit from faster development times (Barow and Meister, 2003). However, although endopolyploidy is known to have a heritable basis (Neiman et al., 2017), that traits correlated with endopolyploidy are potentially visible to natural selection, that there is variation in levels of endopolyploidy across closely related taxa, and that endopolyploidy as a trait has high phylogenetic signal, little experimental work has been done to assess the evolutionary or ecological consequences of endopolyploidy. Some work on cellular polyploidy and niche differentiation has been studied in cancer, however. In a way analogous to the on-off switch metaphor of polyploidy-diploidy or asex-sex transitions within a lineage in response to stressful vs optimal environments, Pienta et al. (2020) also propose that polyploid cancer cells, in particular poly-aneuploid cancer cells (PACCs), can be used in a conditional evolutionary strategy of facultative evolvability. During times of stress, PACCs can cause cancer cells to accelerate evolution by proliferating and thereby causing more heritable variation, but, as they are induced by stress, they are less common during optimal conditions of cancer growth leading to a decrease in heritable variation and increased stability which avoids the disadvantageous consequences of high mutation rates (Pienta et al., 2020). Therefore, the trait that is important to evolution may not be polyploidy of a cell or of all the cells in an organism, but rather the ability to transition between diploid and polyploid states or a propensity to become polyploid. In fact, this directly is analogous to one way that scientists frame polyploidy as a propensity for plants to transition between diploid and polyploid states throughout a lineage (Freeling, 2017) and could be a potential way to incorporate insights between somatic and germline polyploid studies.

4.7 Is there a correlation between endopolyploidy and germline polyploidy?

WGD has many of the same effects on the cell that endopolyploid does. Namely, increasing the number of copy of genomes in a cell increases the volume of the cell and decreases the surface-to-volume ratio (Figure 4.1). This ratio holds despite the differences in WGD cells and those endopolyploid cells produced by endocycling which produces polytene chromosomes (very long chromosomes formed by the successive duplication of each chromatid

without segregation (Zykova et al., 2018)). Like endopolyploid cells, polyploid cells derived from WGD have more chromatin resulting in larger nuclei and longer cell cycles, leading to larger cells with larger or more organelles—these modifications to the organization of chromatin may change gene expression in both plants and animals resulting in differences in the movement of and timing of translation of mRNA (reviewed in Doyle and Coate, 2019).

There is an inverse relationship between endopolyploidy and WGD in plants. Pacey et al. (2020) have shown that there is a trade off between endopolyploidy and WGD in autotetraploid accessions of Arabidopsis with a negative correlation between endopolyploidy index and ploidy level in conjunction with strong cytotype associations of these traits. Similarly, research done on members of the Chenopodioideae has shown that endopolyploidy is more common in diploid species than polyploids (Skaptsov et al., 2017). However, many lineages that are known to have endopolyploidy are also lineages that are derived from ancient WGD and the occurrence of neo-WGD may also be a common trait to these clades (Figure 4.2, Barow (2006)).

Both germline and endopolyploidy are mutations that have been co-opted to perform similar roles in plants. Endopolyploidy is thought to be an important mechanism in plant hedging strategies to mediate the trade-off between fitness when damaged and tolerance following damage (Scholes et al., 2017). WGD is also known to affect the evolution of trade-offs as studies into groups in Asteraceae have shown that the transition between diploidy and polyploidy is associated with both competition-colonization and competition-defence trade-offs and are attributed to the success of polyploids in being able to colonize new niches (Thébault et al., 2011). The flexibility of growth and defense strategies as well as the developmental flexibility of the plant lineage points to possible overlap in the ways that lineages have evolved to make use of polyploidy through both endopolyploidy and WGD.

4.8 What are the long-term macroevolutionary consequences of endopolyploidy?

The flexibility that polyploidy affords plant lineages has raised the hypothesis that genera that have high levels of endopolyploidy and WGD incidence would have higher speciation rates. However, a recent study looking at different genera of orchids found that endopolyploidy was negatively associated with polyploid speciation (Bateman et al., 2018). Nonetheless, there are very few researches into the long-term consequences of endopolyploidy, although some patterns have emerged. The most important factor in predicting endopolyploidy in a lineage is phylogeny (Barow and Meister, 2003). Endopolyploidy is very common in flowering plants and mosses but not in ferns, gymnosperms, liverworts, or lycophytes (Barlow (1978); Barow and Jovtchev (2007); Bainard and Newmaster (2010), Figure 4.2). Recent phylogenetic studies have shown that there is phylogenetic signal for both endoreduplication index and genome size (higher endoreduplication is associated with smaller genome size to accommodate the reduction in genetic content through specialization) in plants and

that these traits are correlated with many important phenotypes including life history and flowing time and even mycorrhizal association (Bainard et al., 2012). However, this correlation was not found in moss, where it was hypothesized that the strong phylogenetic signal in the clade is due to evolutionary pressure to maintain small genomes by also maintaining high levels of endopolyploidy (Bainard et al., 2020). There is some evidence that endopolyploidy allows for genome reduction, however. For example, the relatively small genome size conserved in the Aizoaceae is also associated with large amounts of endoreduplication found across the clade and is thought to be an important feature in the diversification of this highly diverse succulent clade (Powell et al., 2020). However, the importance of endopolyploidy as a trait (or suite of traits e.g. degree od endopolyploidy, propensity of endopolyploidy, or presence/absence of endopolyploidy) that influences diversification rates across a phylogeny has not yet been studied in a robust statistical phylogenetic framework.

Nagl (1976) understands endopolyploidy as one of many evolutionary strategies to increase nuclear DNA content and can make up for a lack of increase of DNA content across a given lineage through tissue specific gene expression. While we are arriving at an understanding of the evolution of regulatory mechanisms and of ways that complexity and phenotypic diversity are driven by other means than a simple increase in number of genes, it is clear that these processes evolve in a manner governed by developmental constraints of a given lineage, such as sex chromosomes. Spoelhof et al. (2020) point out that there besides the plant clade, essentially nothing is known about chromosome counts, genome size estimates, or amount of endopolyploidy in the Vertebrata, Cnidaria, Arthropoda, Fungi, or Spiralia clades. Notably, animals possess the same or analogous diversity in reproductive strategies as plants from sexual to asexual to vegetative and it remains unknown how these macroevolutionary dynamics play out in groups that resemble plant life-styles, e.g., sponges or Cnidaria.

4.9 What remains unknown and next steps?

Several of the same questions that phylogeneticts have been studying with respect to WGD have also been asked about "endopolyploid species" (i.e., species that may have developmentally programmed and abundant endoreduplication often measured using the endoreduplication index (EI)), such as: is endopolyploidy an adaptive trait that allows for fast development in plant lineages and contributes to the maintenance of small genomes and large cells or is endopolyploidy an adaptation to high altitude or cold environment that arrest mitosis but not cell expansion (Barow and Meister, 2003). The observation that EI is relatively conserved within a species and similar across a genus and yet is vastly different within families, demonstrates its potential to be studied as a species level trait and it would be interesting to perform phylogenetic tests to see whether there is an association of EI with increased diversification rates, environmental traits, or other natural history traits. In this way, we can imagine expanding the phylogenetic apparatus developed for investigating the importance of WGD in the diversification of lineages across a phylogeny in plants to investigating

endopolyploidy as a driver of diversity as well. More measurements of EI across families would be required but there is no reason that it could not be treated as a quantitative trait and modeled in a Bayesian framework.

Likewise, there is also a considerable amount of work to be done from an experimental standpoint as well. The classic experiment to test the ecological importance of WGD was done by Ramsey (2011) using transplant experiments of natural and synthetic polyploids of different ploidy levels to demonstrate how WGD confers a fitness advantage to polyploids in novel environments. In Arabidopsis it is been recently shown that there is an association between endopolyploidy (which varies significantly between accessions) and leaf functional traits and climate variation (Pacey et al., 2020) and therefore it is a candidate for future studies looking at local adaptation in a similar way that WGD has been in the past. If endopolyploidy is an adaptive trait then these experiments would be a fruitful and inexpensive avenue to test these hypotheses.

I propose that the decade of "Polyploidy 2030" must attempt to answer basic questions of endopolyploidy biology before we can realistically uncover potential shared "rules of life" between germline and somatic polyploidy. These open questions include:

- How common is endopolyploidy across the tree of life?
- How does endopolyploidy as a somatic mutation arise and become integrated into the developmental programme of plants and animals?
- Is endopolyploidy more common than WGD across the tree of life? Or do lineages that that more endopolyploidy have less WGD, as is hypothesized in animals?
- To what extent is there an inverse relationship between endopolyploidy and WGD and does this pattern also extend to animals and plants other than Arabidopsis? Does endopolyploidy and WGD have similar effects on the expression of gene networks and are there shared transcriptional patterns for a given phenomenon (e.g. stress response)?
- Is there an association between endopolyploidy and diversification rates across the tree of life and are these patterns the same as those assessing the role WGDs have on diversification?

4.10 Concluding Remarks

Many of the possibilities to arrive at a synthesis about the "rules of life" when it comes to understanding polyploidy have been made possible by the "old school" work of scientists past, namely observational science. Chromosome counts and genome size estimates, as well as the broad early sequencing of whole genomes of diverse clades of plants have opened a window into how polyploidy shapes evolution, but mainly only for plants— a small clade in the tree of life. In order to expand our knowledge of polyploidy as an extremely important source of genetic variation in evolution we need to encourage observational sciences in other realms of

biology other than plants, while still promoting it in plants. This project will include getting a better understanding of chromosome number evolution, genome size evolution, and EI, and increasing taxonomic sampling in not only animals but also fungi and other clades as well, However, the next decade of polyploid research is looking to be promising as many unanswerable questions are now becoming in reach and with it an truly integrative approach to understanding evolution.

Bibliography

Anssour, S. and I. T. Baldwin, 2010. Variation in antiherbivore defense responses in synthetic nicotiana allopolyploids correlates with changes in uniparental patterns of gene expression. Plant Physiology 153:1907–1918.

Baduel, P., L. Quadrana, B. Hunter, K. Bomblies, and V. Colot, 2019. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nature communications 10:1–10.

Bainard, J. D., L. D. Bainard, T. A. Henry, A. J. Fazekas, and S. G. Newmaster, 2012. A multivariate analysis of variation in genome size and endoreduplication in angiosperms reveals strong phylogenetic signal and association with phenotypic traits. New Phytologist 196:1240–1250.

Bainard, J. D. and S. G. Newmaster, 2010. Endopolyploidy in bryophytes: Widespread in mosses and absent in liverworts. Journal of Botany .

Bainard, J. D., S. G. Newmaster, and J. M. Budke, 2020. Genome size and endopolyploidy evolution across the moss phylogeny. Annals of Botany 125:543–555.

Barker, M. S., N. Arrigo, A. E. Baniaga, Z. Li, and D. A. Levin, 2016. On the relative abundance of autopolyploids and allopolyploids. New Phytologist 210:391–398.

Barkla, B. J., T. Rhodes, K.-N. T. Tran, C. Wijesinghege, J. C. Larkin, and M. Dassanayake, 2018. Making epidermal bladder cells bigger: developmental-and salinity-induced endopolyploidy in a model halophyte. Plant physiology 177:615–632.

Barlow, P. W., 1978. Endopolyploidy: towards an understanding of its biological significance. Acta biotheoretica 27:1–18.

Barow, M., 2006. Endopolyploidy in seed plants. Bioessays 28:271–281.

Barow, M. and G. Jovtchev, 2007. Endopolyploidy in plants and its analysis by flow cytometry. Flow cytometry with plant cells: analysis of genes, chromosomes and genomes Pp. 349–372.

Barow, M. and A. Meister, 2003. Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. Plant, Cell & Environment 26:571–584.

Bateman, R. M., J. J. Guy, P. J. Rudall, I. J. Leitch, J. Pellicer, and A. R. Leitch, 2018. Evolutionary and functional potential of ploidy increase within individual plants: somatic ploidy mapping of the complex labellum of sexually deceptive bee orchids. Annals of botany 122:133–150.

Beer, S., J. Rundle, and R. Wodzinski, 1983. Interaction between erwinia amylovora and erwinia herbicola in vitro, in immature pear fruits and in apple blossoms. *in* III International Workshop on Fire Blight 151, Pp. 203–204.

Berendsen, R. L., C. M. Pieterse, and P. A. Bakker, 2012. The rhizosphere microbiome and plant health. Trends in plant science 17:478–486.

Bhardwaj, V., S. Meier, L. N. Petersen, R. A. Ingle, and L. C. Roden, 2011. Defence responses of arabidopsis thaliana to infection by pseudomonas syringae are regulated by the circadian clock. PloS one 6.

Blanc, G. and K. H. Wolfe, 2004. Functional divergence of duplicated genes formed by polyploidy during arabidopsis evolution. The Plant Cell 16:1679–1691.

Bolger, A. M., M. Lohse, and B. Usadel, 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30:2114–2120.

Bourdon, M., J. Pirrello, C. Cheniclet, O. Coriton, M. Bourge, S. Brown, A. Moïse, M. Peypelut, V. Rouyère, J.-P. Renaudin, et al., 2012. Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. Development 139:3817–3826.

Boyes, D. C., A. M. Zayed, R. Ascenzi, A. J. McCaskill, N. E. Hoffman, K. R. Davis, and J. Görlach, 2001. Growth stage–based phenotypic analysis of arabidopsis: a model for high throughput functional genomics in plants. The Plant Cell 13:1499–1510.

Braun-Kiewnick, A., B. J. Jacobsen, and D. C. Sands, 2000. Biological control of pseudomonas syringae pv. syringae, the causal agent of basal kernel blight of barley, by antagonistic pantoea agglomerans. Phytopathology 90:368–375.

Brooks, A. W., K. D. Kohl, R. M. Brucker, E. J. van Opstal, and S. R. Bordenstein, 2016. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. PLoS biology 14:e2000225.

Bulgarelli, D., K. Schlaeppi, S. Spaepen, E. V. L. Van Themaat, and P. Schulze-Lefert, 2013. Structure and functions of the bacterial microbiota of plants. Annual review of plant biology 64:807–838.

Cai, J., X. Liu, K. Vanneste, S. Proost, W.-C. Tsai, K.-W. Liu, L.-J. Chen, Y. He, Q. Xu, C. Bian, et al., 2015. The genome sequence of the orchid phalaenopsis equestris. Nature genetics 47:65–72.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes, 2016. Dada2: high-resolution sample inference from illumina amplicon data. Nature methods 13:581.

Camarena, L., V. Bruno, G. Euskirchen, S. Poggio, and M. Snyder, 2010. Molecular mechanisms of ethanol-induced pathogenesis revealed by rna-sequencing. PLoS Pathog 6:e1000834.

Cao, Y., Y. Liang, K. Tanaka, C. T. Nguyen, R. P. Jedrzejczak, A. Joachimiak, and G. Stacey, 2014. The kinase lyk5 is a major chitin receptor in arabidopsis and forms a chitin-induced complex with related kinase cerk1. elife 3:e03766.

Carella, F., G. De Vico, and G. Landini, 2017. Nuclear morphometry and ploidy of normal and neoplastic haemocytes in mussels. Plos one 12:e0173219.

Carotenuto, G., V. Volpe, G. Russo, M. Politi, I. Sciascia, J. de Almeida-Engler, and A. Genre, 2019. Local endoreduplication as a feature of intracellular fungal accommodation in arbuscular mycorrhizas. New Phytologist 223:430–446.

Castrillo, G., P. J. P. L. Teixeira, S. H. Paredes, T. F. Law, L. de Lorenzo, M. E. Feltcher, O. M. Finkel, N. W. Breakfield, P. Mieczkowski, C. D. Jones, et al., 2017. Root microbiota drive direct integration of phosphate stress and immunity. Nature 543:513–518.

Chen, M., F. Wang, Z. Zhang, J. Fu, and Y. Ma, 2017. Characterization of fungi resistance in two autotetraploid apple cultivars. Scientia Horticulturae 220:27–35.

Chen, T., K. Nomura, X. Wang, R. Sohrabi, J. Xu, L. Yao, B. C. Paasch, L. Ma, J. Kremer, Y. Cheng, et al., 2020. A plant genetic network for preventing dysbiosis in the phyllosphere. Nature 580:653–657.

Chen, Z. J., 2010. Molecular mechanisms of polyploidy and hybrid vigor. Trends in plant science 15:57–71.

Coate, J. E., A. K. Luciano, V. Seralathan, K. J. Minchew, T. G. Owens, and J. J. Doyle, 2012. Anatomical, biochemical, and photosynthetic responses to recent allopolyploidy in glycine dolichocarpa (fabaceae). American Journal of Botany 99:55–67.

Consortium, T. G. et al., 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635.

Cookson, S. J., A. Radziejwoski, and C. Granier, 2006. Cell and leaf size plasticity in arabidopsis: what is the role of endoreduplication? Plant, Cell & Environment 29:1273–1283.

Cui, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis, J. J. Doyle, P. S. Soltis, J. E. Carlson, K. Arumuganathan, A. Barakat, et al., 2006. Widespread genome duplications throughout the history of flowering plants. Genome research 16:738–749.

Davoli, T. and T. de Lange, 2011. The causes and consequences of polyploidy in normal development and cancer. Annual review of cell and developmental biology 27:585–610.

Decanter, L., G. Colling, N. Elvinger, S. Heimarsson, and D. Matthies, 2020. Ecological niche differences between two polyploid cytotypes of saxifraga rosacea. American journal of botany 107:423–435.

Delgado, E. R., E. C. Stahl, N. Roy, P. D. Wilkinson, and A. W. Duncan, 2020. Polyploidy in liver function, mitochondrial metabolism, and cancer. The Liver: Biology and Pathobiology Pp. 603–613.

Delmotte, N., C. Knief, S. Chaffron, G. Innerebner, B. Roschitzki, R. Schlapbach, C. von Mering, and J. A. Vorholt, 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proceedings of the National Academy of Sciences 106:16428–16433.

Denchi, E. L., G. Celli, and T. De Lange, 2006. Hepatocytes with extensive telomere deprotection and fusion remain viable and regenerate liver mass through endoreduplication. Genes & development 20:2648–2653.

Dermen, H., 1940. Colchicine polyploidy and technique. The botanical review 6:599–635.

Donné, R., M. B. Nader, and C. Desdouets, 2018. Cellular and molecular mechanisms controlling ploidy .

Donne, R., M. Saroul-Aïnama, P. Cordier, S. Celton-Morizur, and C. Desdouets, 2020. Polyploidy in liver development, homeostasis and disease. Nature Reviews Gastroenterology & Hepatology Pp. 1–15.

Doyle, J. J. and J. E. Coate, 2019. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. International Journal of Plant Sciences 180:1–52.

Eberl, L. and P. Vandamme, 2016. Members of the genus burkholderia: good and bad guys. F1000Research 5.

Eden, E., R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, 2009. Gorilla: a tool for discovery and visualization of enriched go terms in ranked gene lists. BMC bioinformatics 10:1–7.

Edgar, B. A. and T. L. Orr-Weaver, 2001. Endoreplication cell cycles: more for less. Cell 105:297–306.

Fargione, J. E. and D. Tilman, 2005. Diversity decreases invasion via both sampling and complementarity effects. Ecology Letters 8:604–611.

Ficetola, G. F. and M. Stöck, 2016. Do hybrid-origin polyploid amphibians occupy transgressive or intermediate ecological niches compared to their diploid ancestors? Journal of Biogeography 43:703–715.

Fitzpatrick, C. R., J. Copeland, P. W. Wang, D. S. Guttman, P. M. Kotanen, and M. T. Johnson, 2018. Assembly and ecological function of the root microbiome across angiosperm plant species. Proceedings of the National Academy of Sciences 115:E1157–E1165.

Foster, K. R., J. Schluter, K. Z. Coyte, and S. Rakoff-Nahoum, 2017. The evolution of the host microbiome as an ecosystem on a leash. Nature 548:43–51.

Fox, D. T., D. E. Soltis, P. S. Soltis, T.-L. Ashman, and Y. Van de Peer, 2020. Polyploidy: A biological force from cells to ecosystems. Trends in Cell Biology .

Frankel-Bricker, J. and L. K. Frankel, 2021. Re-analysis of 16s rrna gene sequence data sets uncovers disparate laboratory-specific microbiomes associated with the yellow fever mosquito (aedes aegypti). Microbial Ecology Pp. 1–15.

Freeling, M., 2017. Picking up the ball at the k/pg boundary: the distribution of ancient polyploidies in the plant phylogenetic tree as a spandrel of asexuality with occasional sex. The Plant Cell 29:202–206.

Fritz, R. S., C. Moulia, and G. Newcombe, 1999. Resistance of hybrid plants and animals to herbivores, pathogens, and parasites. Annual review of ecology and systematics 30:565–591.

Galbraith, D. W., K. R. Harkins, and S. Knapp, 1991. Systemic endopolyploidy in arabidopsis thaliana. Plant Physiology 96:985–989.

Gandarillas, A., R. Molinuevo, and N. Sanz-Gómez, 2018. Mammalian endoreplication emerges to reveal a potential developmental timer. Cell Death & Differentiation 25:471–476.

Gaut, B. S., M. L. T. d'Ennequin, A. S. Peek, and M. C. Sawkins, 2000. Maize as a model for the evolution of plant nuclear genomes. Proceedings of the National Academy of Sciences 97:7008–7015.

Ge, S. X., E. W. Son, and R. Yao, 2018. idep: an integrated web application for differential expression and pathway analysis of rna-seq data. BMC bioinformatics 19:534.

Gegas, V. C., J. J. Wargent, E. Pesquet, E. Granqvist, N. D. Paul, and J. H. Doonan, 2014. Endopolyploidy as a potential alternative adaptive strategy for arabidopsis leaf size variation in response to uv-b. Journal of Experimental Botany 65:2757–2766.

Gjelsvik, K., R. Besen-McNally, and V. Losick, 2019. Solving the polyploid mystery in health and disease. Trends in Genetics 35:6–14.

Hijmans, R. J., T. Gavrilenko, S. Stephenson, J. Bamberg, A. Salas, and D. M. Spooner, 2007. Geographical and environmental range expansion through polyploidy in wild potatoes (solanum section petota). Global ecology and biogeography 16:485–495.

Hu, D. and J. C. Cross, 2009. Development and function of trophoblast giant cells in the rodent placenta. International Journal of Developmental Biology 54:341–354.

Huang, M.-L., X.-P. Deng, Y.-Z. Zhao, S.-L. Zhou, S. Inanaga, S. Yamada, and K. Tanaka, 2007. Water and nutrient use efficiency in diploid, tetraploid and hexaploid wheats. Journal of Integrative Plant Biology 49:706–715.

Innerebner, G., C. Knief, and J. A. Vorholt, 2011. Protection of arabidopsis thaliana against leaf-pathogenic pseudomonas syringae by sphingomonas strains in a controlled model system. Appl. Environ. Microbiol. 77:3202–3210.

Jagodzik, P., M. Tajdel-Zielinska, A. Ciesla, M. Marczak, and A. Ludwikow, 2018. Mitogen-activated protein kinase cascades in plant hormone signaling. Frontiers in plant science 9:1387.

Jalal, M., T. Andersen, and D. O. Hessen, 2015. Temperature and developmental responses of body and cell size in drosophila; effects of polyploidy and genome configuration. Journal of Thermal Biology 51:1–14.

Jofuku, K. D., B. Den Boer, M. Van Montagu, and J. K. Okamuro, 1994. Control of arabidopsis flower and seed development by the homeotic gene apetala2. The Plant Cell 6:1211–1225.

Kanehisa, M., M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, 2017. Kegg: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research 45:D353–D361.

Karasov, T. L., E. Chae, J. J. Herman, and J. Bergelson, 2017. Mechanisms to mitigate the trade-off between growth and defense. The Plant Cell 29:666–680.

Karunarathne, P., M. Schedler, E. J. Martínez, A. I. Honfi, A. Novichkova, and D. Hojsgaard, 2018. Intraspecific ecological niche divergence and reproductive shifts foster cytotype displacement and provide ecological opportunity to polyploids. Annals of botany 121:1183–1196.

Kasana, R. C. and C. Pandey, 2018. Exiguobacterium: an overview of a versatile genus with potential in industry and agriculture. Critical reviews in biotechnology 38:141–156.

Kim, D., B. Langmead, and S. Salzberg, 2015. hisat2. Nature methods .

King, K., O. Seppälä, and M. Neiman, 2012. Is more better? polyploidy and parasite resistance. Biology letters 8:598–600.

Kohl, K. D., 2020. Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities. Philosophical Transactions of the Royal Society B 375:20190251.

Kolarčik, V., V. Fráková, V. Kocová, L. Koprivỳ, and P. Mártonfi, 2020. Endopolyploidy pattern in corydalis early spring geophytes. Flora 270:151651.

Kwon, A.-R., M.-M. Oh, K.-Y. Paek, and S.-Y. Park, 2018. The effect of light quality on growth and endopolyploidy occurrence of in vitro-grown phalaenopsis 'spring dancer'. Horticulture, Environment, and Biotechnology 59:179–188.

Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander, M. Garcia-Hernandez, et al., 2012. The arabidopsis information resource (tair): improved gene annotation and new tools. Nucleic acids research 40:D1202–D1210.

Lebeis, S. L., S. H. Paredes, D. S. Lundberg, N. Breakfield, J. Gehring, M. McDonald, S. Malfatti, T. G. Del Rio, C. D. Jones, S. G. Tringe, et al., 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science 349:860–864.

Lee, H. O., J. M. Davidson, and R. J. Duronio, 2009. Endoreplication: polyploidy with purpose. Genes & development 23:2461–2477.

Lei, G., M. Shen, Z.-G. LI, B. Zhang, K.-X. DUAN, N. Wang, Y.-R. CAO, W.-K. ZHANG, B. Ma, H.-Q. LING, et al., 2011. Ein2 regulates salt stress response and interacts with a ma3 domain-containing protein ecip1 in arabidopsis. Plant, Cell & Environment 34:1678–1692.

Leitch, I. J. and S. Dodsworth, 2007. Endopolyploidy in plants. eLS Pp. 1–10.

Levin, D. A., 1983. Polyploidy and novelty in flowering plants. The American Naturalist 122:1–25.

Li, M., G. Xu, X. Xia, M. Wang, X. Yin, B. Zhang, X. Zhang, and Y. Cui, 2017. Deciphering the physiological and molecular mechanisms for copper tolerance in autotetraploid arabidopsis. Plant Cell Reports 36:1585–1597.

Li, Q.-G., L. Zhang, C. Li, J. M. Dunwell, and Y.-M. Zhang, 2013. Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the papilionoideae. Molecular Biology and Evolution 30:2602–2611.

Li, S., L. Liu, T. Li, T. Lan, Y. Wang, Z. Zhang, J. Liu, S. Xu, X. Zhang, J. Zhu, et al., 2019. The distribution pattern of endopolyploidy in maize. Theoretical and Applied Genetics 132:1487–1503.

Li, Z.-G., H.-W. Chen, Q.-T. Li, J.-J. Tao, X.-H. Bian, B. Ma, W.-K. Zhang, S.-Y. Chen, and J.-S. Zhang, 2015. Three saur proteins saur76, saur77 and saur78 promote plant growth in arabidopsis. Scientific reports 5:12477.

Liang, W., C. Li, F. Liu, H. Jiang, S. Li, J. Sun, X. Wu, and C. Li, 2009. The arabidopsis homologs of ccr4-associated factor 1 show mrna deadenylation activity and play a role in plant defence responses. Cell research 19:307–316.

Lim, S. J. and S. R. Bordenstein, 2020. An introduction to phylosymbiosis. Proceedings of the Royal Society B 287:20192900.

Lin, K.-C., G. Torga, J. Sturm, K. Pienta, and R. Austin, 2018. The emergence of polyploid giant cancer cells as the reservoir of theraputic resistance. APS 2018:V46–003.

Liu, Y., J. Zhou, and K. P. White, 2014. Rna-seq differential expression studies: more sequence or more replication? Bioinformatics 30:301–304.

Lloyd, A. and K. Bomblies, 2016. Meiosis in autopolyploid and allopolyploid arabidopsis. Current opinion in plant biology 30:116–122.

López-Jurado, J., E. Mateos-Naranjo, and F. Balao, 2019. Niche divergence and limits to expansion in the high polyploid dianthus broteri complex. New Phytologist 222:1076–1087.

Losick, V. P., D. T. Fox, and A. C. Spradling, 2013. Polyploidization and cell fusion contribute to wound healing in the adult drosophila epithelium. Current Biology 23:2224–2232.

Love, M. I., W. Huber, and S. Anders, 2014. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome biology 15:550.

Lozano, E., A. G. Sáez, A. J. Flemming, A. Cunha, and A. M. Leroi, 2006. Regulation of growth by ploidy in caenorhabditis elegans. Current biology 16:493–498.

Lundberg, D. S., S. L. Lebeis, S. H. Paredes, S. Yourstone, J. Gehring, S. Malfatti, J. Tremblay, A. Engelbrektson, V. Kunin, T. G. Del Rio, et al., 2012. Defining the core arabidopsis thaliana root microbiome. Nature 488:86–90.

Maeso, I., S. W. Roy, and M. Irimia, 2012. Widespread recurrent evolution of genomic features. Genome biology and evolution 4:486–500.

Mahalingam, R., A. Gomez-Buitrago, N. Eckardt, N. Shah, A. Guevara-Garcia, P. Day, R. Raina, and N. V. Fedoroff, 2003. Characterizing the stress/defense transcriptome of arabidopsis. Genome biology 4:R20.

McMurdie, P. J. and S. Holmes, 2013. phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data. PloS one 8.

Mercado-Blanco, J. and P. A. Bakker, 2007. Interactions between plants and beneficial pseudomonas spp.: exploiting bacterial traits for crop protection. Antonie van Leeuwenhoek 92:367–389.

Mesa, J. M., J. A. Juvik, and K. N. Paige, 2019. Individual and interactive effects of herbivory on plant fitness: endopolyploidy as a driver of genetic variation in tolerance and resistance. Oecologia 190:847–856.

Molina-Henao, Y. F. and R. Hopkins, 2019. Autopolyploid lineage shows climatic niche expansion but not divergence in arabidopsis arenosa. American journal of botany 106:61–70.

Monnahan, P., F. Kolář, P. Baduel, C. Sailer, J. Koch, R. Horvath, B. Laenen, R. Schmickl, P. Paajanen, G. Šrámková, et al., 2019. Pervasive population genomic consequences of genome duplication in arabidopsis arenosa. Nature ecology & evolution 3:457–468.

Morita-Yamamuro, C., T. Tsutsui, M. Sato, H. Yoshioka, M. Tamaoki, D. Ogawa, H. Matsuura, T. Yoshihara, A. Ikeda, I. Uyeda, et al., 2005. The arabidopsis gene cad1 controls programmed cell death in the plant immune system and encodes a protein containing a macpf domain. Plant and cell physiology 46:902–912.

Morkunas, I. and L. Ratajczak, 2014. The role of sugar signaling in plant defense responses against fungal pathogens. Acta Physiologiae Plantarum 36:1607–1619.

Mulligan, K. L., T. C. Hiebert, N. W. Jeffery, and T. R. Gregory, 2014. First estimates of genome size in ribbon worms (phylum nemertea) using flow cytometry and feulgen image analysis densitometry. Canadian journal of zoology 92:847–851.

Musseau, C., J. Jorly, S. Gadin, I. Sørensen, C. Deborde, S. Bernillon, J.-P. Mauxion, I. Atienza, A. Moing, M. Lemaire-Chamley, et al., 2020. The tomato guanylate-binding protein slgbp1 enables fruit tissue differentiation by maintaining endopolyploid cells in a non-proliferative state. The Plant Cell 32:3188–3205.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder, 2008. The transcriptional landscape of the yeast genome defined by rna sequencing. Science 320:1344–1349.

Nagl, W., 1976. Dna endoreduplication and polyteny understood as evolutionary strategies. Nature 261:614–615.

Neiman, M., M. J. Beaton, D. O. Hessen, P. D. Jeyasingh, and L. J. Weider, 2017. Endopolyploidy as a potential driver of animal ecology and evolution. Biological Reviews 92:234–247.

Ng, D. W., C. Zhang, M. Miller, Z. Shen, S. Briggs, and Z. Chen, 2012. Proteomic divergence in arabidopsis autopolyploids and allopolyploids and their progenitors. Heredity 108:419–430.

Ngou, B. P. M., H.-K. Ahn, P. Ding, and J. D. Jones, 2021. Mutual potentiation of plant immunity by cell-surface and intracellular receptors. Nature 592:110–115.

Ni, Z., E.-D. Kim, M. Ha, E. Lackey, J. Liu, Y. Zhang, Q. Sun, and Z. J. Chen, 2009. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature 457:327–331.

Nozaki, T. and K. Matsuura, 2019. Evolutionary relationship of fat body endoreduplication and queen fecundity in termites. Ecology and Evolution 9:11684–11694.

Nuismer, S. L. and J. N. Thompson, 2001. Plant polyploidy and non-uniform effects on insect herbivores. Proceedings of the Royal Society of London. Series B: Biological Sciences 268:1937–1940.

Ofek, M., Y. Hadar, and D. Minz, 2012. Ecology of root colonizing massilia (oxalobacteraceae). PloS one 7:e40117.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens, M. J. Oksanen, and M. Suggests, 2007. The vegan package. Community ecology package 10:631–637.

Oswald, B. P. and S. L. Nuismer, 2007. Neopolyploidy and pathogen resistance. Proceedings of the Royal Society B: Biological Sciences 274:2393–2397.

Pacey, E. K., H. Maherali, and B. C. Husband, 2020. The influence of experimentally induced polyploidy on the relationships between endopolyploidy and plant function in arabidopsis thaliana. Ecology and Evolution 10:198–216.

Palomino, G., J. Martínez, I. Méndez, A. Muñoz-Urías, V. Cepeda-Cornejo, and E. Pimienta-Barrios, 2016. Nuclear genome size, ploidy level and endopolyploidy pattern in six species of opuntia (cactaceae). Caryologia 69:82–89.

Parisod, C., R. Holderegger, and C. Brochmann, 2010. Evolutionary consequences of autopolyploidy. New phytologist 186:5–17.

Pearse, I. S., T. Krügel, and I. T. Baldwin, 2006. Innovation in anti-herbivore defense systems during neopolypoloidy–the functional consequences of instantaneous speciation. The Plant Journal 47:196–210.

Van de Peer, Y., E. Mizrachi, and K. Marchal, 2017. The evolutionary significance of polyploidy. Nature Reviews Genetics 18:411.

Pertea, M., D. Kim, G. M. Pertea, J. T. Leek, and S. L. Salzberg, 2016. Transcript-level expression analysis of rna-seq experiments with hisat, stringtie and ballgown. Nature protocols 11:1650.

Pienta, K. J., E. U. Hammarlund, R. Axelrod, J. S. Brown, and S. R. Amend, 2020. Polyaneuploid cancer cells promote evolvability, generating lethal cancer. Evolutionary Applications .

Poupin, M. J., T. Timmermann, A. Vega, A. Zuñiga, and B. González, 2013. Effects of the plant growth-promoting bacterium burkholderia phytofirmans psjn throughout the life cycle of arabidopsis thaliana. PLoS One 8:e69435.

Powell, A. F. and J. J. Doyle, 2015. The implications of polyploidy for the evolution of signalling in rhizobial nodulation symbiosis. Pp. 149–190, *in* Advances in botanical research, vol. 75. Elsevier.

Powell, R. F., L. Pulido Suarez, A. R. Magee, J. S. Boatwright, M. V. Kapralov, and A. J. Young, 2020. Genome size variation and endopolyploidy in the diverse succulent plant family aizoaceae. Botanical Journal of the Linnean Society .

Ramsey, J., 2011. Polyploidy and ecological adaptation in wild yarrow. Proceedings of the National Academy of Sciences 108:7096–7101.

Rangel, J., K. Strauss, K. Seedorf, C. E. Hjelmen, and J. S. Johnston, 2015. Endopolyploidy changes with age-related polyethism in the honey bee, apis mellifera. PLoS One 10:e0122208.

Rasch, E. M. and G. A. Wyngaard, 2008. Endopolyploidy in cyclopoid copepods. Journal of Crustacean Biology 28:412–416.

Rensing, S. A., J. Ick, J. A. Fawcett, D. Lang, A. Zimmer, Y. Van de Peer, and R. Reski, 2007. An ancient genome duplication contributed to the abundance of metabolic genes in the moss physcomitrella patens. BMC evolutionary biology 7:130.

Robinson, D. O., J. E. Coate, A. Singh, L. Hong, M. Bush, J. J. Doyle, and A. H. Roeder, 2018. Ploidy and size at multiple scales in the arabidopsis sepal. The Plant Cell 30:2308–2329.

Roelofs, D., A. Zwaenepoel, T. Sistermans, J. Nap, A. A. Kampfraath, Y. Van de Peer, J. Ellers, and K. Kraaijeveld, 2020. Multi-faceted analysis provides little evidence for recurrent whole-genome duplications during hexapod evolution. BMC Biology 18:1–13.

Rosenberg, E. and I. Zilber-Rosenberg, 2018. The hologenome concept of evolution after 10 years. Microbiome 6:78.

Saei, A., K. Hoeata, A. Krebs, P. Sutton, J. Herrick, M. Wood, and L. Gea, 2017. The status of pseudomonas syringae pv. actinidiae (psa) in the new zealand kiwifruit breeding programme in relation to ploidy level. *in* IX International Symposium on Kiwifruit 1218, Pp. 293–298.

Salter, S. J., M. J. Cox, E. M. Turek, S. T. Calus, W. O. Cookson, M. F. Moffatt, P. Turner, J. Parkhill, N. J. Loman, and A. W. Walker, 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC biology 12:1–12.

Scarpeci, T. E., V. S. Frea, M. I. Zanor, and E. M. Valle, 2017. Overexpression of aterf019 delays plant growth and senescence, and improves drought tolerance in arabidopsis. Journal of experimental botany 68:673–685.

Schenk, P. M., K. Kazan, I. Wilson, J. P. Anderson, T. Richmond, S. C. Somerville, and J. M. Manners, 2000. Coordinated plant defense responses in arabidopsis revealed by microarray analysis. Proceedings of the National Academy of Sciences 97:11655–11660.

Schliep, K. P., 2011. phangorn: phylogenetic analysis in r. Bioinformatics 27:592–593.

Schnorr, S. L., K. Sankaranarayanan, C. M. Lewis Jr, and C. Warinner, 2016. Insights into human evolution from ancient and contemporary microbiome studies. Current opinion in genetics & development 41:14–26.

Schoen, D., J. Burdon, and A. Brown, 1992. Resistance of glycine tomentella to soybean leaf rust phakopsora pachyrhizi in relation to ploidy level and geographic distribution. Theoretical and Applied Genetics 83:827–832.

Scholes, D. R. and K. N. Paige, 2015. Plasticity in ploidy: a generalized response to stress. Trends in Plant Science 20:165–175.

Scholes, D. R., E. N. Rasnick, and K. N. Paige, 2017. Characterization of arabidopsis thaliana regrowth patterns suggests a trade-off between undamaged fitness and damage tolerance. Oecologia 184:643–652.

Scholes, D. R., A. V. Suarez, and K. N. Paige, 2013. Can endopolyploidy explain body size variation within and between castes in ants? Ecology and Evolution 3:2128–2137.

Schubert, V., 2014. Rna polymerase ii forms transcription networks in rye and arabidopsis nuclei and its amount increases with endopolyploidy. Cytogenetic and genome research 143:69–77.

Sherman, B. T., Q. Tan, J. R. Collins, W. G. Alvord, J. Roayaei, R. Stephens, M. W. Baseler, H. C. Lane, R. A. Lempicki, et al., 2007. The david gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome biology 8:1–16.

Skaptsov, M. V., M. N. Lomonosova, M. G. Kutsev, S. V. Smirnov, and A. I. Shmakov, 2017. The phenomenon of endopolyploidy in some species of the chenopodioideae (amaranthaceae). Botany Letters 164:47–53.

Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson, B. C. Husband, and W. S. Judd, 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56:13–30.

Song, M. J., B. I. Potter, J. J. Doyle, and J. E. Coate, 2020. Gene balance predicts transcriptional responses immediately following ploidy change in arabidopsis thaliana. The Plant Cell 32:1434–1448.

Spartz, A. K., S. H. Lee, J. P. Wenger, N. Gonzalez, H. Itoh, D. Inzé, W. A. Peer, A. S. Murphy, P. J. Overvoorde, and W. M. Gray, 2012. The saur19 subfamily of small auxin up rna genes promote cell expansion. The Plant Journal 70:978–990.

Spoelhof, J. P., R. Keeffe, and S. F. McDaniel, 2020. Does reproductive assurance explain the incidence of polyploidy in plants and animals? New Phytologist 227:14–21.

Spoelhof, J. P., P. S. Soltis, and D. E. Soltis, 2017. Pure polyploidy: closing the gaps in autopolyploid research. Journal of Systematics and Evolution 55:340–352.

Ssekagiri, A., W. Sloan, and U. Z. Ijaz, 2018. microbiomeseq: an r package for microbial community analysis in an environmental context.

Stone, B. W., E. A. Weingarten, and C. R. Jackson, 2018. The role of the phyllosphere microbiome in plant health and function. Annual Plant Reviews online Pp. 533–556.

Tao, Y., Z. Xie, W. Chen, J. Glazebrook, H.-S. Chang, B. Han, T. Zhu, G. Zou, and F. Katagiri, 2003. Quantitative nature of arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen pseudomonas syringae. The Plant Cell 15:317–330.

Těšitelová, T., J. Jersáková, M. Roy, B. Kubátová, J. Těšitel, T. Urfus, P. Trávníček, and J. Suda, 2013. Ploidy-specific symbiotic interactions: divergence of mycorrhizal fungi between cytotypes of the g ymnadenia conopsea group (o rchidaceae). New Phytologist 199:1022–1033.

Thébault, A., F. Gillet, H. Müller-Schärer, and A. Buttler, 2011. Polyploidy and invasion success: trait trade-offs in native and introduced cytotypes of two asteraceae species. Plant Ecology 212:315–325.

Theodoridis, S., C. Randin, O. Broennimann, T. Patsiou, and E. Conti, 2013. Divergent and narrower climatic niches characterize polyploid species of european primroses in primula sect. aleuritia. Journal of Biogeography 40:1278–1289.

Thiel, T., A. Graner, R. Waugh, I. Grosse, T. J. Close, and N. Stein, 2009. Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. BMC Evolutionary Biology 9:1–18.

Thompson, J. N., S. L. Nuismer, and K. Merg, 2004. Plant polyploidy and the evolutionary ecology of plant/animal interactions. Biological Journal of the Linnean Society 82:511–519.

Tsutsui, N. D., A. V. Suarez, J. C. Spagna, and J. S. Johnston, 2008. The evolution of genome size in ants. BMC Evolutionary Biology 8:64.

Turner, T. R., E. K. James, and P. S. Poole, 2013. The plant microbiome. Genome biology 14:209.

Underwood, C. J., I. R. Henderson, and R. A. Martienssen, 2017. Genetic and epigenetic variation of transposable elements in arabidopsis. Current opinion in plant biology 36:135–141.

Underwood, W., M. Melotto, and S. Y. He, 2007. Role of plant stomata in bacterial invasion. Cellular microbiology 9:1621–1629.

Vernon, J. A. and J. Butsch, 1957. Effect of tetraploidy on learning and retention in the salamander. Science 125:1033–1034.

Vives-Peris, V., A. Gómez-Cadenas, and R. M. Pérez-Clemente, 2018. Salt stress alleviation in citrus plants by plant growth-promoting rhizobacteria pseudomonas putida and novosphingobium sp. Plant cell reports 37:1557–1569.

Vogel, C., N. Bodenhausen, W. Gruissem, and J. A. Vorholt, 2016. The arabidopsis leaf transcriptome reveals distinct but also overlapping responses to colonization by phyllosphere commensals and pathogen infection with impact on plant health. New Phytologist 212:192–207.

Walbot, V., 1992. Strategies for mutagenesis and gene cloning using transposon tagging and t-dna insertional mutagenesis. Annual review of plant physiology and plant molecular biology 43:49–82.

Walley, J. W., D. R. Kelley, G. Nestorova, D. L. Hirschberg, and K. Dehesh, 2010. Arabidopsis deadenylases atcaf1a and atcaf1b play overlapping and distinct roles in mediating environmental stress responses. Plant physiology 152:866–875.

Wang, H., Y. Hu, J. Pan, and D. Yu, 2015. Arabidopsis vq motif-containing proteins vq12 and vq29 negatively modulate basal defense against botrytis cinerea. Scientific reports 5:14185.

Wang, W., Y. He, Z. Cao, and Z. Deng, 2018. Induction of tetraploids in impatiens (impatiens walleriana) and characterization of their changes in morphology and resistance to downy mildew. HortScience 53:925–931.

Wang, Y., R. Garrido-Oter, J. Wu, T. M. Winkelmüller, M. Agler, T. Colby, T. Nobori, E. Kemen, and K. Tsuda, 2019. Site-specific cleavage of bacterial mucd by secreted proteases mediates antibacterial resistance in arabidopsis. Nature communications 10:1–12.

Wang, Z., G. Fan, Y. Dong, X. Zhai, M. Deng, Z. Zhao, W. Liu, and Y. Cao, 2017. Implications of polyploidy events on the phenotype, microstructure, and proteome of paulownia australis. PloS one 12.

Weigmann, K., S. M. Cohen, and C. F. Lehner, 1997. Cell cycle progression, growth and patterning in imaginal discs despite inhibition of cell division after inactivation of drosophila cdc2 kinase. Development 124:3555–3563.

Wipf, H. M. and D. Coleman-Derr, 2021. Evaluating domestication and ploidy effects on the assembly of the wheat bacterial microbiome. Plos one 16:e0248030.

Xia, Q., Y. Guo, Z. Zhang, D. Li, Z. Xuan, Z. Li, F. Dai, Y. Li, D. Cheng, R. Li, et al., 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (bombyx). Science 326:433–436.

Yadav, A. N., D. Kour, S. Sharma, S. G. Sachan, B. Singh, V. S. Chauhan, R. Sayyed, R. Kaushik, and A. K. Saxena, 2019. Psychrotrophic microbes: biodiversity, mechanisms of adaptation, and biotechnological implications in alleviation of cold stress in plants. Pp. 219–253, *in* Plant growth promoting rhizobacteria for sustainable stress management. Springer.

Yan, Z., M. Reddy, C.-M. Ryu, J. A. McInroy, M. Wilson, and J. W. Kloepper, 2002. Induced systemic protection against tomato late blight elicited by plant growth-promoting rhizobacteria. Phytopathology 92:1329–1333.

Yao, L., Z. Wu, Y. Zheng, I. Kaleem, and C. Li, 2010. Growth promotion and protection against salt stress by pseudomonas putida rs-198 on cotton. European Journal of Soil Biology 46:49–54.

Zamioudis, C. and C. M. Pieterse, 2012. Modulation of host immunity by beneficial microbes. Molecular Plant-Microbe Interactions 25:139–150.

Zedek, F., K. Plačková, P. Veselỳ, J. Šmerda, P. Šmarda, L. Horová, and P. Bureš, 2020. Endopolyploidy is a common response to uv-b stress in natural plant populations, but its magnitude may be affected by chromosome type. Annals of Botany 126:883–889.

Zeng, Z., H. Huang, N. Han, C. Y. Huang, P. Langridge, H. Bian, and M. Zhu, 2017. Endopolyploidy levels in barley vary in different root types and significantly decrease under phosphorus deficiency. Plant Physiology and Biochemistry 118:11–21.

Zhang, S., I. Mercado-Uribe, Z. Xing, B. Sun, J. Kuang, and J. Liu, 2014. Generation of cancer stem-like cells through the formation of polyploid giant cancer cells. Oncogene 33:116–128.

Zhang, S., K. Zhou, X. Luo, L. Li, H.-C. Tu, A. Sehgal, L. H. Nguyen, Y. Zhang, P. Gopal, B. D. Tarlow, et al., 2018. The polyploid state plays a tumor-suppressive role in the liver. Developmental cell 44:447–459.

Zhang, X.-Y., C.-G. Hu, and J.-L. Yao, 2010. Tetraploidization of diploid dioscorea results in activation of the antioxidant defense system and increased heat tolerance. Journal of plant physiology 167:88–94.

Zhu, Q.-H., S. Stephen, K. Kazan, G. Jin, L. Fan, J. Taylor, E. S. Dennis, C. A. Helliwell, and M.-B. Wang, 2013. Characterization of the defense transcriptome responsive to fusarium oxysporum-infection in arabidopsis using rna-seq. Gene 512:259–266.

Zykova, T. Y., V. G. Levitsky, E. S. Belyaeva, and I. F. Zhimulev, 2018. Polytene chromosomes–a portrait of functional organization of the drosophila genome. Current genomics 19:179–191.

Appendix A

Supplemental Data and Code

A.1 Links to Raw Data

See Table A.1.

A.2 16S Analysis

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

Table A.1: Links to the raw and processed data used in the dissertation

Description	Link
Chapter 1: 16S Raw Reads	https://drive.google.com/drive/folders/1L9qWDJrltgPLD0hkfiC3L4-XHXHdvduL?usp=sharing
Chapter 2: 16S Raw Reads	https://drive.google.com/drive/u/1/folders/1qmi5AxOTOQqjTk5fG2L9NtuKxxh4XDlN
Chapter 3: RNASeq Raw Reads	https://drive.google.com/drive/folders/1Cxwr8XiJKIKrOxIeYWIrTFVtqJc-zt7a?usp=sharing
Chapter 3: DESeq2/EdgeR comparisons	https://drive.google.com/drive/folders/1SN_FHNEv3NJBc50203C2n4g6E3xcb9X_?usp=sharing
```

```
library(dada2)
library(ape)
library(phyloseq)
library(Biostrings)
library(ggplot2)
library(tidyverse)
library(DESeq2)
library(DECIPHER)
library(phyloseq)
library(ade4)
library(adespatial)
library(microbiomeSeq)
library(vegan)
library(phangorn)
library(reshape2)
theme_set(theme_bw())
```

Set path to fies

```{r}
# path <- "~/Downloads/raw"
# list.files(path)
#
# fnFs <- sort(list.files(path, pattern="_R1_001.fastq.gz"
, full.names = TRUE))
# fnRs <- sort(list.files(path, pattern="_R2_001.fastq.gz"
, full.names = TRUE))
# # Extract sample names, assuming filenames have format
: SAMPLENAME_XXX.fastq
# sample.names <- sapply(strsplit(basename(fnFs), "_"), `[`, 1)
```

Inspect Read Quality Profiles

```{r}
# plotQualityProfile(fnFs[1:2])
# plotQualityProfile(fnRs[1:2])
```

Filter and trim

```{r}
# # Place filtered files in filtered/ subdirectory
# filtFs <- file.path(path, "filtered",
```

```
paste0(sample.names, "_F_filt.fastq.gz"))
# filtRs <- file.path(path,
"filtered", paste0(sample.names, "_R_filt.fastq.gz"))
# names(filtFs) <- sample.names
# names(filtRs) <- sample.names
#
#
# out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, truncLen=c(230,160),
#               maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,
#               compress=TRUE, multithread=TRUE)
# On Windows set multithread=FALSE
# head(out)
```

```
# Learn the Error Rates
```{r}
# errF <- learnErrors(filtFs, multithread=TRUE)
# errR <- learnErrors(filtRs, multithread=TRUE)
# plotErrors(errF, nominalQ=TRUE)
```

```
# Sample Inference
```{r}
# dadaFs <- dada(filtFs, err=errF, multithread=TRUE)
# dadaRs <- dada(filtRs, err=errR, multithread=TRUE)
# dadaFs[[1]]

```

```
# Merge paired reads
```{r}
# mergers <- mergePairs(dadaFs, filtFs, dadaRs, filtRs, verbose=TRUE)
# # Inspect the merger data.frame from the first sample
# head(mergers[[1]])

```

```
# Construct sequence table
```{r}
# seqtab <- makeSequenceTable(mergers)
# dim(seqtab)
# table(nchar(getSequences(seqtab)))
```

```
```

Remove chimeras
```{r}
# seqtab.nochim <- removeBimeraDenovo(seqtab,
method="consensus", multithread=TRUE, verbose=TRUE)
# dim(seqtab.nochim)
# sum(seqtab.nochim)/sum(seqtab)
```

Track reads through the pipeline
```{r}
# getN <- function(x) sum(getUniques(x))
# track <- cbind(out, sapply(dadaFs, getN), sapply(dadaRs,
getN), sapply(mergers, getN), rowSums(seqtab.nochim))
# # If processing a single sample, remove the
sapply calls: e.g. replace sapply(dadaFs, getN) with getN(dadaFs)
# colnames(track) <- c("input", "filtered", "denoisedF",
"denoisedR", "merged", "nonchim")
# rownames(track) <- sample.names
# head(track)

```

Assign taxonomy
```{r}
# taxa <- assignTaxonomy(seqtab.nochim,
"~/Downloads/raw/silva_nr_v138_train_set.fa",
multithread=TRUE)
#
# taxa <- addSpecies(taxa,
"~/Downloads/raw/silva_species_assignment_v138.fa.gz")
#
# taxa.print <- taxa # Removing sequence rownames for display only
# rownames(taxa.print) <- NULL
# head(taxa.print)
```

Build Tree
```{r}
# seqs <- getSequences(taxa)
```

```
# names(seqs) <- seqs
# alignment <- AlignSeqs(DNAStringSet(seqs), anchor=NA)
# phang.align <- phyDat(as(alignment, "matrix"), type="DNA")
# dm <- dist.ml(phang.align)
# treeNJ <- NJ(dm)
#
# fit = pml(treeNJ, data=phang.align)
# fitGTR <- update(fit, k=4, inv=0.2)
# fitGTR <- optim.pml(fitGTR, model="GTR", optInv=TRUE, optGamma=TRUE,
#                     rearrangement = "stochastic", control =
# pml.control(trace = 0))
#
#
# detach("package:phangorn", unload=TRUE)
```

```
# Handoff to phyloseq
```{r}
# samdf <- read.csv("~/Downloads/raw/samdf.csv", row.names = 1)
#
# ps <- phyloseq(otu_table(seqtab.nochim, taxa_are_rows=FALSE),
#                sample_data(samdf),
#                tax_table(taxa),
#                phy_tree(fitGTR$tree))
#
# #save file
# saveRDS(ps, "~/Downloads/raw/ps.rds")

# #remove chloroplast
# ps <- subset_taxa(ps, (Class!="Chloroplast") | is.na(Class))

ps <- readRDS("~/Downloads/raw/ps.rds")
ps <- subset_taxa(ps, (Class!="mitochondria") | is.na(Class))
ps <- subset_taxa(ps, (Class!="Chloroplast") | is.na(Class))
```

```
# Visualize alpha-diversity
```{r}
plot_richness(ps, x="ploidy", measures=c("Shannon", "Simpson"),
color="samples")
```

```
ps.a.plot <- plot_anova_diversity(ps, method = c("richness",
"fisher", "simpson", "shannon", "evenness"),
    grouping_column = "ploidy", pValueCutoff = 0.05)
ps.a.plot
```

```
# Ordinate
```{r}
ps.prop <- transform_sample_counts(ps, function(otu) otu/sum(otu))
ord.nmds.bray <- ordinate(ps.prop, method="NMDS", distance="bray")
plot_ordination(ps.prop, ord.nmds.bray, color="ploidy", title="Bray NMDS")



pORD <- plot_ordination(ps.prop, ord.nmds.bray, color="ploidy",
title="Bray NMDS")
plot420 <- pORD + geom_point(size=7, alpha=0.75)
plot420 +  stat_ellipse(type = "norm", linetype = 2)

ord.nmds.uni <- ordinate(ps.prop, method="NMDS",
distance="unifrac", weighted=TRUE)
pORD.uni <- plot_ordination(ps.prop, ord.nmds.uni,
color="ploidy", title="Unifrac weighted NMDS")
plot420 <- pORD.uni + geom_point(size=7, alpha=0.75)
plot420 +  stat_ellipse(type = "norm", linetype = 2)

ord.nmds.uni <- ordinate(ps.prop, method="NMDS",
distance="unifrac", weighted=TRUE)
pORD.uni.g <- plot_ordination(ps.prop, ord.nmds.uni,
color="genotype", title="Unifrac weighted NMDS")
plot420.g <- pORD.uni.g + geom_point(size=7, alpha=0.75)
plot420.g +  stat_ellipse(type = "norm", linetype = 2)


```

```{r}
ps <- subset_taxa(ps, (Order!="Chloroplast") | is.na(Order))
ps.rarefied = rarefy_even_depth(ps, rngseed=1,
sample.size=0.9*min(sample_sums(ps)), replace=F)

wunifrac_dist = phyloseq::distance(ps.rarefied,
```
```

```
method="unifrac", weighted=F)
plot_bar(ps.rarefied, fill="Order") + facet_wrap(~genotype,
scales="free_x", nrow=1)

adonis(wunifrac_dist ~ sample_data(ps.rarefied)$ploidy)
adonis(wunifrac_dist ~ sample_data(ps.rarefied)$genotype)
```

```
# Bar plot
```{r}
top20 <- names(sort(taxa_sums(ps), decreasing=TRUE))[1:20]
ps.top20 <- transform_sample_counts(ps, function(OTU) OTU/sum(OTU))
ps.top20 <- prune_taxa(top20, ps.top20)
plot_bar(ps.top20, x="ploidy", fill="Order")
```

```
# DeSeq
```{r}
psr <- subset_samples(ps, ploidy != "C2")
psr <- subset_samples(ps, ploidy != "C4")
diagdds = phyloseq_to_deseq2(psr, ~ ploidy)
gm_mean = function(x, na.rm=TRUE){
  exp(sum(log(x[x > 0]), na.rm=na.rm) / length(x))
}
geoMeans = apply(counts(diagdds), 1, gm_mean)
diagdds = estimateSizeFactors(diagdds, geoMeans = geoMeans)
diagdds = DESeq(diagdds, fitType="local")

res = results(diagdds)
res = res[order(res$padj, na.last=NA), ]
alpha = 0.10
sigtab = res[(res$padj < alpha), ]
sigtab = cbind(as(sigtab, "data.frame"),
as(tax_table(psr)[rownames(sigtab), ], "matrix"))

posigtab = sigtab[sigtab[, "log2FoldChange"] > 0, ]
posigtab = posigtab[, c("baseMean", "log2FoldChange", "lfcSE",
"padj", "Phylum", "Class", "Family", "Genus", "Species")]
posigtab

negitab = sigtab[sigtab[, "log2FoldChange"] < 0, ]
negitab = tab[, c("baseMean", "log2FoldChange", "lfcSE", "padj",
```

```
"Phylum", "Class", "Family", "Genus", "Species")]
negitab
```

```{r}

# Phylum order
x = tapply(sigtab$log2FoldChange, sigtab$Phylum, function(x) max(x))
x = sort(x, TRUE)
sigtab$Phylum = factor(as.character(sigtab$Phylum), levels=names(x))
# Genus order
x = tapply(sigtab$log2FoldChange, sigtab$Genus, function(x) max(x))
x = sort(x, TRUE)
sigtab$Genus = factor(as.character(sigtab$Genus), levels=names(x))
ggplot(sigtab, aes(x=Genus, y=log2FoldChange, color=Phylum)) +
geom_point(size=6) +
  theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust=0.5))

```

A.3 Pathogen assay

A.4 RNA Seq analysis

```
Outline of RNA-seq pipeline

java -jar Trimmomatic-0.39/trimmomatic-0.39.jar PE
RNAseqData/raw_data/B2B/B2B_CRRA200014390-1a_H3LCGDSXY_L2_1.fq.
gz RNAseqData/raw_data/B2B/B2B_CRRA200014390-1a_H3LCGDSXY_L2_
2.fq.gz  TrimOut/B2B_output_forward_paired.fq.gz
TrimOut/B2B_output_forward_unpaired.fq.gz
TrimOut/B2B_output_reverse_paired.fq.gz
TrimOut/B2B_output_reverse_unpaired.fq.gz
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3
TRAILING:3 MINLEN:36

hisat2 --phred33  -x TAIR10_TE -1
1TrimOut/B2B_output_forward_paired.fq -2
1TrimOut/B2B_output_reverse_paired.fq -S 2SamOutTE/
B2B_trim.sam
```

Table A.2: Supplemental data for data normalization for ddPCR

plate	code	code2	eco	ploidy	treatment	time	CopiesPer20uLWell	weight (g)	norm (counts per gram)
C8	6*	F2.B.T2	GD	2	B	1	932	0.45	2071.11111
D7	4*	D2.B.T2	SOR	2	B	1	160	0.64	250
F2	2*	B2.B.T2	COL	2	B	1	1240	0.46	2695.65217
D5	3*	C2.B.T2	HR	2	B	1	44	0.53	83.0188679
C7	13	G2.B.T0	FEI	2	B	0	110	0.48	229.166667
F6	6x	F2.B.T3	GD	2	B	2	102	0.42	242.857143
G5	7*	G2.B.T2	FEI	2	B	1	104	0.42	247.619048
G3	7	D2.B.T0	SOR	2	B	0	156	0.42	371.428571
E8	1x	A2.B.T3	WA	2	B	2	156	0.36	433.333333
C6	1	A2.B.T0	WA	2	B	0	136	0.28	485.714286
E9	5*	E2.B.T2	WS	2	B	1	240	0.43	558.139535
A2	5	C2.B.T0	HR	2	B	0	152	0.26	584.615385
D10	9	E2.B.T0	WS	2	B	0	200	0.27	740.740741
G6	11	F2.B.T0	GD	2	B	0	320	0.42	761.904762
A4	1*	A2.B.T2	WA	2	B	1	130	0.17	764.705882
E7	3	B2.B.T0	COL	2	B	0	298	0.35	851.428571
F8	5x	E2.B.T3	WS	2	B	2	792	0.5	1584
C5	3x	C2.B.T3	HR	2	B	2	3380	0.44	7681.81818
A9	7x	G2.B.T3	FEI	2	B	2	3740	0.4	9350
G9	2x	B2.B.T3	COL	2	B	2	6600	0.58	11379.3103
B2	4x	D2.B.T3	SOR	2	B	2	11480	0.63	18222.2222
B1	19x	E2.C.T2	WS	2	C	2	168	0.77	218.181818
E1	17*	C2.C.T1	HR	2	C	1	214	0.34	629.411765
C2	16*	B2.C.T1	COL	2	C	1	512	0.54	948.148148
F4	18x	D2.C.T2	SOR	2	C	2	812	0.59	1376.27119
A1	18*	D2.C.T1	SOR	2	C	1	870	0.5	1740
E3	15*	A2.C.T1	WA	2	C	1	756	0.27	2800
F9	21*	G2.C.T1	FEI	2	C	1	1154	0.39	2958.97436
B3	17x	C2.C.T2	HR	2	C	2	2080	0.55	3781.81818
D2	20*	F2.C.T1	GD	2	C	1	1940	0.5	3880
F7	19*	E2.C.T1	WS	2	C	1	2740	0.34	8058.82353
A10	20x	F2.C.T2	GD	2	C	2	5180	0.58	8931.03448
B6	15x	A2.C.T2	WA	2	C	2	4260	0.33	12909.0909
B4	16x	B2.C.T2	COL	2	C	2	6840	0.39	17538.4615
F10	21x	G2.C.T2	FEI	2	C	2	18620	0.54	34481.4815
B8	8	D4.B.T0	SOR	4	B	0	378	0.49	771.428571
E2	9*	B4.B.T2	COL	4	B	1	660	0.58	1137.93103
G2	10	E4.B.T0	WS	4	B	0	372	0.58	641.37931
C3	11x	D4.B.T3	SOR	4	B	2	34	0.68	50
C4	14*	G4.B.T2	FEI	4	B	1	40	0.73	54.7945206
B7	11*	D4.B.T2	SOR	4	B	1	36	0.48	75
A5	13*	F4.B.T2	GD	4	B	1	62	0.75	82.6666667
D4	10x	C4.B.T3	COL	4	B	2	112	0.73	153.424658
B10	8x	A4.B.T3	WA	4	B	2	84	0.46	182.608696
D1	12*	E4.B.T2	WS	4	B	1	136	0.61	222.95082
E5	14	G4.B.T0	FEI	4	B	0	114	0.48	237.5
G4	2	A4.B.T0	WA	4	B	0	224	0.62	361.290323
C10	14x	G4.B.T3	FEI	4	B	2	208	0.57	364.912281
F5	6	C4.B.T0	HR	4	B	0	158	0.39	405.128205
C1	12x	E4.B.T3	WS	4	B	2	288	0.51	564.705882
D9	4	B4.B.T0	COL	4	B	0	322	0.57	564.912281
D8	12	F4.B.T0	GD	4	B	0	312	0.49	636.734694
F1	8*	A4.B.T2	WA	4	B	1	146	0.2	730
G8	10*	C4.B.T2	HR	4	B	1	1000	0.61	1639.34426
E10	13x	F4.B.T3	GD	4	B	2	1410	0.62	2274.19355
B5	9x	B4.B.T3	COL	4	B	2	4580	0.8	5725
F3	27x	F4.C.T2	GD	4	C	2	6	0.12	50
A6	26x	E4.C.T2	WS	4	C	2	40	0.84	47.6190476
G1	25*	D4.C.T1	SOR	4	C	1	52	0.6	86.6666667
E4	27*	F4.C.T1	GD	4	C	1	106	0.72	147.222222
A7	23*	B4.C.T1	COL	4	C	1	94	0.4	235
A3	24x	C4.C.T2	HR	4	C	2	276	0.84	328.571429
G7	22x	A4.C.T2	WA	4	C	2	190	0.57	333.333333
D3	28*	G4.C.T1	FEI	4	C	1	600	0.42	1428.57143
E6	24*	C4.C.T1	HR	4	C	1	1310	0.74	1770.27027
G10	22*	A4.C.T1	WA	4	C	1	1186	0.51	2325.4902
B9	28x	G4.C.T2	FEI	4	C	2	4160	0.55	7563.63636
D6	26*	E4.C.T1	WS	4	C	1	6780	0.39	17384.6154

```
samtools view -bS 2SamOut/B2B_trim.sam | samtools sort - -o
3BamOut/B2B_trim_sort.bam

htseq-count -f bam -r pos -m intersection-nonempty -t gene -i
Name 3BamOut/B2B_trim_sort.bam
Araport11_GFF3_genes_transposons.201606.gff >
4CountOut/B2B_Araport11_HTseq.txt



We are comparing the gene expression of one species (Arabidopsis
thaliana) of 2 ploidies (2x vs 4x) under 2 conditions
(treated with microbiome vs nontreated. We have the count tables
for 3 technical replicates for each ploidy by each treatment.

First we are going to import the library DESeq2.
‘‘‘{r}
library("DESeq2")
library("pheatmap")
library("vsn")
library("RColorBrewer")
library("ggplot2")
‘‘‘

Now we are going to make a table that describes the file names
by their treatment and ploidy. We are going to be doing
pair-wise comparisons between Treated with a microbiome vs
treated with a control and also Treated diploid vs Treated
polyploid to test whether there is any reponse to treatment with
a microbiome and if so, if there are differences in that
response between diploid and polyploids. To that end we are
going to have three different sampleTables we can make either
using all of the data, or separating them out by ploidy
‘‘‘{r}
directory <- "../data/rnaseqCountData"
sampleFiles <- c("B2B.txt", "B2C.txt", "B4B.txt", "B4C.txt",
"D2B.txt", "D2C.txt", "D4B.txt", "D4C.txt", "E2B.txt",
"E2C.txt", "E4B.txt", "E4C.txt")
condition <- c("treated", "control", "treated", "control",
"treated", "control", "treated", "control", "treated",
```

```
"control", "treated", "control")
ploidy <- c("diploid", "diploid", "tetraploid", "tetraploid",
"diploid", "diploid", "tetraploid", "tetraploid", "diploid",
"diploid", "tetraploid", "tetraploid" )

#here we are separating them out just to compare the files
ending with B (bacteria) and their assosiate ploid (the
ploidy.B)
sampleFiles.B <- c("B2B.txt", "B4B.txt", "D2B.txt", "D4B.txt",
"E2B.txt", "E4B.txt")
sampleFiles.C <- c("B2C.txt", "B4C.txt", "D2C.txt", "D4C.txt",
"E2C.txt", "E4C.txt")
ploidy.B <- c("diploid",  "tetraploid",  "diploid",
"tetraploid",  "diploid",  "tetraploid")
ploidy.C <- c("diploid",  "tetraploid",  "diploid",
"tetraploid",  "diploid",  "tetraploid")


sampleTable <- data.frame(sampleName = sampleFiles,
                      fileName = sampleFiles,
                      condition = condition,
                      ploidy = ploidy)


sampleTable.treatment <- data.frame(sampleName = sampleFiles,
                      fileName = sampleFiles,
                      condition = condition)

sampleTable.ploidy <- data.frame(sampleName = sampleFiles.B,
                      fileName = sampleFiles.B,
                      ploidy = ploidy.B)

ddsHTSeq.treatment <- DESeqDataSetFromHTSeqCount(sampleTable =
sampleTable.treatment,
                                    directory = directory,
                                    design= ~  condition)

ddsHTSeq.ploidy <- DESeqDataSetFromHTSeqCount(sampleTable =
sampleTable.ploidy,
                                    directory = directory,
                                    design= ~  ploidy)
```

```
ddsHTSeq.interact <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable,
                                     directory = directory,
                                     design= ~  condition +
                                     ploidy + condition*ploidy)

```

```
Pre-filtering out genes that have less than 10 reads. The
running DESeq on the filtering the data and summarizing the
results. Here let's do it for treatment.
```{r}
keep.b <- rowSums(counts(ddsHTSeq.treatment)) >= 10
dds.b <- ddsHTSeq.treatment[keep.b, ]

dds.b <- DESeq(dds.b)
res.b <- results(dds.b, alpha=0.1)
res.b

resOrdered.b <- res.b[order(res.b$pvalue),]
summary(res.b)
sum(res.b$padj < 0.1, na.rm=TRUE)

ntd.b <- normTransform(dds.b)
vsd.b <- vst(dds.b, blind=FALSE)

# write.csv(resOrdered.b, "~/Desktop/resOrdered.b.csv")
```
We found 384 up or down-reuglated genes at the 0.05 p-value
cut-off. We can export the DESeq analysis using write.csv and
paste the genes into a GO term enrichment analysis program like
[GOrilla](http://cbl-gorilla.cs.technion.ac.il).

Now we can do it for ploidy.
```{r}
keep.p <- rowSums(counts(ddsHTSeq.ploidy)) >= 10
dds.p <- ddsHTSeq.ploidy[keep.p, ]

dds.p <- DESeq(dds.p)
res.p <- results(dds.p, alpha=0.1)
res.p

resOrdered.p <- res.p[order(res.p$pvalue),]
```
```

```
summary(res.p)
sum(res.p$padj < 0.1, na.rm=TRUE)

ntd.p <- normTransform(dds.p)
vsd.p <- vst(dds.p, blind=FALSE)

write.csv(resOrdered.p, "~/Desktop/resOrdered.p.csv")
```
```
And now for interaction
```{r}
keep.i <- rowSums(counts(ddsHTSeq.interact)) >= 10
dds.i <- ddsHTSeq.interact[keep.i, ]

dds.i <- DESeq(dds.i)
res.i <- results(dds.i, alpha=0.1)
res.i

resOrdered.i <- res.i[order(res.i$pvalue),]
summary(res.i)
sum(res.i$padj < 0.1, na.rm=TRUE)

ntd.i <- normTransform(dds.i)
vsd.i <- vst(dds.i, blind=FALSE)

resultsNames(dds.i)
#Shows you the differnt comparisons you can look at

DiffAbundByInteract <- results(dds.i,
                      name= "ploidy_tetraploid_vs_diploid")
#Pulls out the specified comparison

```


I've download the GOrilla results in this directory. We can see
that not only are there several enriched GO terms between the
treated and non treated plants but that there are also enriched
GO terms between the treated ploidies. We can use these gene
lists to look at gene annotations of these genes too.
```

```
Now we can try our hand and visualizing the DESeq analysis in
other ways. Here by condition (treated vs non treated)
```{r}
sampleDists <- dist(t(assay(vsd.b)))
sampleDistMatrix <- as.matrix(sampleDists)
rownames(sampleDistMatrix) <- paste(vsd.b$condition, vsd.b$type, sep="-")
colnames(sampleDistMatrix) <- NULL
colors <- colorRampPalette( rev(brewer.pal(9, "Blues")) )(255)
pheatmap(sampleDistMatrix,
         clustering_distance_rows=sampleDists,
         clustering_distance_cols=sampleDists,
         col=colors)


plotCounts(dds.b, gene=c("AT1G72416"), intgroup="condition")
plotCounts(dds.b, gene=c("AT3G05660"), intgroup="condition")
plotCounts(dds.b, gene=c("AT5G63790"), intgroup="condition")
plotCounts(dds.b, gene=c("AT1G53170"), intgroup="condition")
plotCounts(dds.b, gene=c("AT1G04833"), intgroup="condition")
plotCounts(dds.b, gene=c("AT2G17230"), intgroup="condition")
plotCounts(dds.b, gene=c("AT3G11410"), intgroup="condition")
plotCounts(dds.b, gene=c("AT4G31550"), intgroup="condition")
plotCounts(dds.b, gene=c("AT2G28500"), intgroup="condition")
plotCounts(dds.b, gene=c("AT4G01870"), intgroup="condition")

plotPCA(vsd, intgroup=c("condition"))
```
These heatmaps and PCAs tell us that the transcriptome profile
actual clusers around genotype since we used three different
accessions of Arabidopsis as our biological replicates. However,
there is separation between control and treated within the
genotype. Nonetheless, we can see which genes across all
genotypes are being upregulated by treatment with the microbiome
such as AT1G72416, a Chaperone DnaJ-domain superfamily protein.

Here by ploidy within treated condition
```{r}
sampleDists <- dist(t(assay(vsd.p)))
sampleDistMatrix <- as.matrix(sampleDists)
rownames(sampleDistMatrix) <- paste(vsd.p$ploidy, vsd.p$type, sep="-")
colnames(sampleDistMatrix) <- NULL
colors <- colorRampPalette( rev(brewer.pal(9, "Blues")) )(255)
```
```

```
pheatmap(sampleDistMatrix,
         clustering_distance_rows=sampleDists,
         clustering_distance_cols=sampleDists,
         col=colors)

plotCounts(dds.p, gene="AT5G04340", intgroup="ploidy")
plotPCA(vsd.p, intgroup=c("ploidy")) +  stat_ellipse(type = "norm", linetype = 2)
```
```
Looking at ploidy, we find that the transcriptome as a whole
also separates our mostly by genotype. Nonetheless, we can find
different genes that are differentially expressed between the
two ploidies such as AT5G20230, which is a copper-binding
protein senescence associated gene associated with reponse of
oxidative stress and wounding. '

```{r}
plotCounts(dds.b, gene=c("AT1G72430"), intgroup="condition",
main="AT1G72430 (SAUR78)", col=c("2","2", "4", "4","2","2", "4",
"4","2","2", "4", "4"))
plotCounts(dds.b, gene=c("AT1G22810"), intgroup="condition",
main ="AT1G22810 (ATERF019)", col=c("2","2", "4", "4","2","2",
"4", "4","2","2", "4", "4"))
plotCounts(dds.b, gene=c("AT5G44680"), intgroup="condition",
main ="AT5G44680 (DNA glycosylase superfamily protein)",
col=c("2","2", "4", "4","2","2", "4", "4","2","2", "4", "4"))
plotCounts(dds.b, gene=c("AT1G03220"), intgroup="condition",
main ="AT1G03220 (SAP2)", col=c("2","2", "4", "4","2","2", "4",
"4","2","2", "4", "4"))
plotCounts(dds.b, gene=c("AT2G22880"), intgroup="condition",
main ="AT2G22880 (VQ12)", col=c("2","2", "4", "4","2","2", "4",
"4","2","2", "4", "4"))
plotCounts(dds.b, gene=c("AT2G33580"), intgroup="condition",
main ="AT2G33580 (ATLYK5)",col=c("2","2", "4", "4","2","2", "4",
"4","2","2", "4", "4"))
plotCounts(dds.b, gene=c("AT1G78830"), intgroup="condition",
main ="AT1G78830 (EP1-like glycoprotein 2)", col=c("2","2", "4",
"4","2","2", "4", "4","2","2", "4", "4"))
plotCounts(dds.b, gene=c("AT2G43620"), intgroup="condition",
main ="AT2G43620 (endochitinase)", col=c("2","2", "4",
"4","2","2", "4", "4","2","2", "4", "4"))

plotCounts(dds.b, gene=c("AT5G01830"), intgroup="condition",
```
```

```
main ="AT5G01830 (SAUR21)", col=c("2","2", "4", "4","2","2",
"4", "4","2","2", "4", "4"))

plotCounts(dds.b, gene=c("AT4G24800"), intgroup="condition",
main ="AT4G24800 (ECIP1)", col=c("2","2", "4", "4","2","2", "4",
"4","2","2", "4", "4"))

plotCounts(dds.b, gene=c("AT5G22250"), intgroup="condition",
main = "AT5G22250 (ATCAF1B)", col=c("2","2", "4", "4","2","2",
"4", "4","2","2", "4", "4"))
plotCounts(dds.b, gene=c("AT1G19020"), intgroup="condition",
col=c("2","2", "4", "4","2","2", "4", "4","2","2", "4", "4"))

plotCounts(dds.b, gene=c("AT1G23710"), intgroup="condition",
col=c("2","2", "4", "4","2","2", "4", "4","2","2", "4", "4"))

plotCounts(dds.b, gene=c("AT1G23710"), intgroup="condition",
col=c("2","2", "4", "4","2","2", "4", "4","2","2", "4", "4"))

plotCounts(dds.b, gene=c("AT3G60520"), intgroup="condition",
col=c("2","2", "4", "4","2","2", "4", "4","2","2", "4", "4"))


plotCounts(dds.p, gene=c("AT5G22250"), intgroup="ploidy", main =
"AT5G22250 (ATCAF1B)")

plotCounts(dds.p, gene=c("AT1G19020"), intgroup="ploidy", main =
"AT1G19020")
plotCounts(dds.p, gene=c("AT1G23710"), intgroup="ploidy", main =
"AT1G23710")


‘‘‘
```

Colophon

Non cuïvis homini contingit adire Corinthum - Jona Mensch

This document was typeset using LaTeX
based on a template maintained by Dr. Paul Vojta.
The typefaces used are in the Computer Modern family.