Genomic and phenomic approaches for studying *Puccinia sorghi*-maize interactions

by

Katerina Louise Holan

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Biology

Specialization: Predictive Plant Phenomics

Program of Study Committee:
Steven A. Whitham, Major Professor
Michelle Graham
Carolyn Lawrence-Dill
Arti Singh
Justin Walley

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2023

DEDICATION

To Ashley, for the beginning, to Elli, for the middle, and to Chuck, Mae, and Freya, for the end. Thank you for your unending love and support. I love you all.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS

First off, I would like to thank my advisor Dr. Steve Whitham, for his mentorship and support throughout these years, for always supporting my research endeavors, and for guiding me along the way.

Thank you to everyone who has served on my committee, Dr. Baskar Ganapathysubramanian, Dr. Michelle Graham, Dr. Carolyn Lawrence-Dill, Dr. Arti Singh, and Dr. Justin Walley, for your guidance, feedback, and support.

Thank you to all my peers in the IPB and P3 programs, my family, and my friends, especially Devin, Ashley, Cam, B, and Paul for study sessions, game nights, and support.

Thank you to MaryAnn, Nicole, and Dai for making the transition to grad school life easier, answering last minute emails, and always being willing to help.

Lastly, thank you to all my friends and colleagues from the Whitham Lab. To Bliss, Mel, Ryan, Aline, Ekkachai, Manju, and Meiyu thank you for teaching, for being there to talk, to listen, to troubleshoot, to guide, and to help. I seriously wouldn't be here without you all!

All of you have helped me grow both as a person and as a scientist, and I will forever be grateful!

ABSTRACT

Rust fungal pathogens comprise the largest group of plant pathogenic fungi. Due to limitations of their study, like an inability to be cultured or difficulty in making genetic modifications, there are many gaps in the knowledge base of these organisms. One rust species, *Puccinia sorghi*, is a worldwide pathogen of maize that can cause significant yield losses. Much of the research for *P. sorghi* focuses on qualitative disease phenotypes of various isolates on different maize genetic backgrounds, with limited information regarding the key pathogenicity genes (effectors) required for a successful infection within this pathosystem. It is imperative to further develop the genomic and phenomic tools available for *P. sorghi* for use in effector characterization screens.

With the recent advent of long-read sequencing, rust genome assembly has transitioned from exceedingly fragmented contigs based on short-read sequencing to large, repeat-resolved scaffolds. More complete rust genomes have led to many discoveries about the true genome size, repeat content, and gene content of these organisms. Well-annotated assemblies also allow for the prediction of candidate effector proteins that function as pathogenicity and virulence determinants. In this work, the genomic resources for *P. sorghi* are expanded with a highly contiguous, long-read assembly of a previously undescribed isolate (IA16). Comprehensive annotation utilizing expressed sequence tags from several timepoints across the disease cycle in maize enabled the prediction of additional candidate effectors for this species. Comparison of these candidates to other *P. sorghi* isolates will lead to discoveries regarding a particular isolate's virulence.

We also report on the characterization of the members of a rust-specific candidate secreted effector protein family present in the *P. sorghi* IA16 isolate. Of eight candidates, we

were able to demonstrate that one is a weak suppressor of the plant hypersensitive immune response in the heterologous system *Nicotiana benthamiana*. This work also utilized an automated phenotyping setup to acquire time lapse images of leaves during experimental assays. By pairing effector characterization assays with automated phenotyping platforms, we can increase throughput, accuracy, and consistency in results.

Lastly, we detail a machine learning approach to quantifying common rust disease on maize leaves. Because plant-pathogen interactions are complex, and small changes to phenotype that are undetectable by human measurements may occur, the development of easy-to-use computer vision-based phenotyping platforms to provide consistent and quantitative results is essential. Additionally, a better understanding of the minimum requirements for a given phenotyping approach is useful for future development, as this can increase the speed at which new platforms are developed. This work demonstrates machine learning is a viable and accurate approach to the quantification of rust disease symptoms, corroborating ground truth experimental results. This work also provides extensive image and annotation data for use in future applications.

Overall, this dissertation presents a multi-disciplinary approach to the study of *P. sorghi* that provides both genomic resources and phenotyping pipelines for the study of candidate effectors and plant-pathogen interactions.

CHAPTER 1. GENERAL INTRODUCTION

Pucciniales, an order within Basidiomycota, contains the plant pathogenic rust fungi, well-known for their impacts on agriculturally important crop species (Figueroa et al. 2023; Avelino et al. 2015; Godoy et al. 2016). Their complex life cycles and genomes make these important plant pathogens difficult to study. Recent advances in genome sequencing technologies and assemblies, large-scale effector candidate screening, and computer vision-based phenotyping platforms have led to better understanding of these species (Petre and Duplessis 2022; Lorrain et al. 2019; Heineck et al. 2019).

General Biology of Rust Fungi

Rust fungi are obligate biotrophs, requiring living host tissue to survive and complete their life cycles. Additionally, they are highly specific parasites, infecting narrow ranges of host species (Duplessis et al. 2021). As a result, these species are not culturable on synthetic media, and their study is limited to within their host species. Rust fungi can produce up to five separate spore types, each infecting either a primary (telial) or an alternate (aecial) host, which can be the same (autoecious) or different (heteroecious) plant species. Almost all spore types are dikaryotic, containing two genetically distinct haploid nuclei within each cell. Aeciospores infect the telial host, giving rise to uredinia sori, which produce urediniospores. Urediniospores reinfect the telial host in a repetitive, cyclic fashion, where sporulating uredinia give rise to new urediniospores that infect new host tissues. This spore stage does not overwinter or survive on dead host tissue, and spring infections in colder climates rely on wind dispersal from pathogen reservoirs in warmer climates. Eventually, uredinia can transition into telia, producing teliospores, a durable spore stage able to overwinter. Upon germination, the nuclei within teliospores undergo karyogamy, followed by meiosis, forming haploid basidia. These structures give rise to haploid

basidiospores, which infect the aecial host and form pycnia. Receptive hyphae of compatible mating types merge, forming aecia followed by aeciospores. Diploid, dikaryotic aeciospores return to the telial host, which germinate to form uredinia (Aime et al. 2017). Most rust species are macrocyclic and produce all five spore stages, but some are demicyclic (no uredinial stage) or microcyclic (only telial and either basidial or pycnial stages). For autoecious rust species where only some spore stages have been described, it is difficult to determine if lack of those spore stages is due to loss of function, extinction of alternate hosts, or simply have yet to be discovered (Lorrain et al. 2019). *Puccinia striiformis* f. sp. *tritici* (wheat stripe rust, wheat yellow rust), previously believed to be an autoecious species, had its alternate host discovered only thirteen years ago, several decades after the species was originally described (Jin et al. 2010; Zhao et al. 2013). In other species, such as *Phakopsora pachyrhizi* (Asian soybean rust), the alternate hosts remain unknown (Vittal et al. 2012).

Research for many rust species has focused on the uredinial stage, as this stage is typically responsible for infecting crop species and causing disease epidemics. For most rust species, the urediniospores germinate on host plants, and germ tubes locate stomata through topographical and chemical signals (Allen et al. 1991), forming structures called appressoria over stomatal openings before entering plant tissue. Upon entry through the stoma, infection hyphae grow and form haustorial mother cells at mesophyll cells, which break through the plant cell wall to form structures known as haustoria. An extrahaustorial matrix and host-derived extrahaustorial membrane separates the fungal cell from the host cell, while still allowing for exchange of proteins, signaling molecules, and nutrients (Garnica et al. 2014). Additional hyphae continue to grow throughout intercellular spaces, developing more haustorial mother cells and haustoria before forming a uredinium and urediniospores that burst through the leaf epidermis. Haustoria

are of particular interest to rust researchers, as the close relationships they form with host cells facilitate nutrient uptake and the delivery of large quantities of secreted proteins (Voegele and Mendgen 2003; Catanzariti et al. 2006).

The *Puccinia sorghi*-Maize Pathosystem

Puccinia sorghi is a heteroecious and macrocyclic pathogen with maize and teosinte spp. serving as the telial hosts and *Oxalis* (wood sorrel) spp. serving as the aecial hosts. In maize, common rust develops as small, round to oval, brick-red to cinnamon brown pustules (the uredinia) on the adaxial and abaxial leaf surfaces, with cooler temperatures and higher humidity promoting fungal germination and growth. The sexual spore stages of *P. sorghi* are observed mostly in warmer climates (Guerra et al. 2016; Dunhin et al. 2004; Guerra et al. 2019), with occasional reports in temperate climates (Mahindapala 1978). In cooler climates, it is widely accepted that *P. sorghi* infection of maize is initiated in the spring each year from wind-blown urediniospores traveling from warmer regions, providing new inoculum to cycle in maize crops throughout the spring and summer. There are several qualitative "resistance to *Puccinia*" (*Rp*) genes in maize, commonly present in dent corn to confer resistance to *P. sorghi*. The majority of these genes are located at the *Rp1* locus on the short arm of chromosome 10 (Chavan et al. 2015; Hooker 1985). Additional genes have been located at loci on chromosomes 3, 4, 6, and 10 (Hulbert 1991; Hagan and Hooker 1965; Delaney et al. 1998). Quantitative resistance to common rust is present in maize as well, with some of this resistance presenting as adult plant resistance (Ren et al. 2021; Quade et al. 2021; Olukolu et al. 2016; Lübberstedt et al. 1998; Zheng et al. 2018). Despite the existence of these genetic resistance resources, *P. sorghi* still poses a threat to specialty maize varieties and has historically led to significant yield losses (Groth et al. 1983; Pataky 1987). Widespread use of the *Rp1-D* gene in sweet corn hybrids in the United States during the 1980s and 1990s significantly increased selection pressures for *P.*

sorghi populations, resulting in large virulent populations by 1999 (Pataky et al. 2001). Fungicides provide an alternative control method but can pose economic and environmental risks (Pataky and Eastburn 1993; Dey et al. 2012). The future outlook for *P. sorghi* is uncertain, as climate change and management factors will influence the pathogen's range and severity. Some estimates predict expansion of suitable conditions for *P. sorghi* in the Northern hemisphere, including the United States (Ramirez-Cabral et al. 2017; Figueroa et al. 2023). This expansion increases the risk of the development of resistant *P. sorghi* populations, both to innate resistance traits and fungicides, as the expansion of sexual phase habitat and more generations of urediniospores would provide more opportunities for genetic recombination and mutation. *P. sorghi* populations are already widely variable, with samples from within a given region having differing virulence patterns on various *Rp* maize lines (Darino et al. 2016; Quade et al. 2021).

Genomic Resources of Pucciniales Species

The earliest whole-genome assemblies for rust species were released less than 15 years ago (Cantu et al. 2011; Duplessis et al. 2011; Zheng et al. 2013). Since then, the number of genomic resources produced for rust fungi has quickly expanded, with over 80 assemblies currently available from GenBank, including multiple isolates of the same rust species and fully phased haploid assemblies representing both dikaryotic nuclei of a given isolate. Rust genomes are particularly large when compared to other Basidiomycetes, with sizes ranging from 60 Mb to over 2 Gb with extensive repetitive elements (35-90%) (Aime et al. 2017; Tobias et al. 2021). As a result, early assemblies relied on short-read sequences from Sanger or Illumina technologies, resulting in highly fragmented contigs and incomplete or truncated gene annotations. Short-read assemblies are additionally unable to phase apart the two nuclei present in dikaryotic spore stages, instead collapsing divergent haplotypic sequences. As a result, sequencing strategies quickly shifted towards long-read sequencing technologies such as PacBio SMRT (Single

Molecule, Real-Time) or Oxford Nanopore Technologies. Even though error rates were higher for long-read sequencing technologies relative to other next-generation sequencing platforms, distal sequences were inherently linked, which enabled increasing assembly contiguity at the single read level (Schwessinger et al. 2018; Miller et al. 2018; Xia et al. 2018). Additional chromatin interaction information was often generated with 10X Genomics or Hi-C reads, which cross-link spatially close DNA sequences before shearing, enabling linking of both proximal DNA within a strand or contigs contained within the same nuclei, i.e., phasing. Advances and tool development in bioinformatics have contributed to assembly contiguity and phasing as well, including SALSA2 (Scaffolding of Long-read Assemblies), FALCON-Phase, and NuclearPhaser (Ghurye et al. 2017; Kronenberg et al. 2021; Sperschneider 2022). NuclearPhaser seems promising for the correction of phase switching in phased assemblies, where sequences within contigs may be phased, but end up assigned to the wrong haplotype, meaning each haplotype results in a mixture of sequences from both nuclei. The pipeline has been used with promising results in *Puccinia coronata* f. sp. *avenae* (oat crown rust), *Puccinia triticina* (leaf rust of wheat, barley, and rye), and *Puccinia polysora* (southern rust of maize) resulting in chromosome-level assemblies (Henningsen et al. 2022; Duan et al. 2022; Liang et al. 2023). Advancements in sequencing technologies have also led to the assembly of notoriously difficult genomes. Assemblies of three isolates of *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust, were released in 2023, revealing a total genome size of 1.25 Gb, with repetitive elements accounting for 93% of the genome sequence (Gupta et al. 2023).

The number and quality of rust fungi genomes are only expected to increase. Long-read sequencing platforms continue to improve and are now on-par with highly accurate Illumina short-reads. HiFi (high-fidelity) libraries from PacBio and the newest flow cell chemistry and

basecaller from Oxford Nanopore Technologies both boast accuracy rates of 99.9% (Wenger et al. 2019). This higher accuracy is already being utilized for rust fungi, with a recent fully phased, gapless, and chromosome level assembly published for *P. triticina* (Li et al. 2023). Higher contiguity and better annotation will inform evolutionary and phylogenomic studies, both within and between species (Nandety et al. 2022).

The genomic resources of *P. sorghi* are limited to a draft assembly of the Argentine isolate RO10H11247 (Rochi et al. 2018). The assembly consists of 99.6 Mb contained in 15,722 scaffolds, assembled from short-read Illumina libraries. The assembly appears relatively complete, containing 85% Basidiomycete BUSCOs (Benchmarking Universal Single-Copy Orthologs, Manni et al. 2021) and approximately 21,000 gene models. However, due to high fragmentation, unknown bases within the assembly, and annotation with non-*P. sorghi* transcripts, gene models are likely incomplete, erroneous, or missing. As there is often considerable genetic variation between rust isolates and no other *P. sorghi* assemblies are available, it is unknown how similar the Argentine isolate is to North American isolates (Jochua et al. 2008; Anderson et al. 2010; Kolmer 2013). The sequencing of additional isolates could lead to insight into the evolutionary history of this pathogen, as has been shown in other rust species (Gupta et al. 2023).

Effector Proteins in Rust Species

In rust fungi, haustoria feeding structures are well-known for delivering massive amounts of proteins and signaling molecules to host plant cells (Voegele and Mendgen 2003; Garnica et al. 2014). Of particular interest are secreted effector proteins, pathogenicity factors involved in the suppression of plant immune responses or trafficking of nutrients (Rafiqi et al. 2012; Uhse and Djamei 2018). Effector proteins of rust fungi typically lack predicted functional annotation and homology to other annotated proteins, but are often small and cysteine-rich, with an N-

terminal secretion peptide for secretion from haustoria. The rapid expansion of genomic resources for rust fungi has led to the generation of extensive lists of candidate secreted effector proteins (CSEPs) (Gibriel et al. 2016) through computational tools like SignalP (Teufel et al. 2022) and EffectorP (Sperschneider and Dodds 2022). CSEPs are also predicted from haustoria-specific transcriptomes (Link et al. 2014; Garnica et al. 2013; Upadhyaya et al. 2015). For example, one study isolated haustoria from *P. pachyrhizi* and *Uromyces appendiculatus* (bean rust), identifying several CSEPs for each species, as well as families of secreted proteins common among rusts (Link et al. 2014).

The large numbers of CSEPs have spurred large functional characterization studies, where subsets of effectors are analyzed in various assays to elucidate their function (Gibriel et al. 2016; Aime et al. 2017). Genetic transformation of rust species is difficult and not routine, having only been described in two species to date (Lawrence et al. 2010; Djulic et al. 2011). As a result, characterization of CSEPs is often approached with transient assays like host-induced gene silencing (HIGS), by expressing candidates in heterologous systems to perform immune suppression or subcellular localization assays, or with yeast two-hybrid or co-immunoprecipitation to identify interaction with host factors (Lorrain et al. 2019). For example, after 156 *P. pachyrhizi* CSEPs were identified (Link et al. 2014), 82 were subsequently characterized for their ability to suppress plant basal defense, callose deposition, hypersensitive response, or yeast cell death, with 17 showing evidence for suppression (Qi et al. 2018). Of these 17, one, named *Pp*EC23, was also shown to interact with a soybean SQUAMOSA promoter-binding like transcription factor (Qi et al. 2016).

Other studies have utilized CSEPs and haustorial transcriptomes to identify avirulence (Avr) proteins, effector proteins recognized by host resistance (R) proteins typically encoded by

nucleotide-binding and leucine-rich repeat (NLR) genes (Catanzariti et al. 2006; Wu et al. 2019; Periyannan et al. 2017). The combination of genomic data of multiple isolates and haustorial transcripts can lead to interesting discoveries in factors responsible for virulence phenotypes. For example, one study compared various Australian *Puccinia graminis* f. sp. *tritici* (wheat stem rust) genome and transcriptome assemblies, including that of an ancestral isolate from 1954. When comparing more recent isolates to the ancestral isolate, the researchers discovered several mutations in haustoria-specific genes that may explain the development of virulence phenotypes on several wheat resistant backgrounds (Upadhyaya et al. 2015). In relevance to maize, two Avr proteins were recently discovered in *P. polysora* through screening of extensive CSEPs, and both proteins triggered strong HR phenotypes when co-expressed with the corresponding *Rpp* (resistance to *P. polysora*) maize gene (Deng et al. 2022; Chen et al. 2022). In *P. sorghi*, the draft genome identified 1,599 predicted secreted proteins, of which a subset are likely CSEPs (Rochi et al. 2018), but effector characterization studies or identified Avr proteins for *P. sorghi* have yet to be published.

Even with a plethora of options, it is often difficult to acquire meaningful data from effector characterization screens due to incomplete gene models, redundancy in functions, use of heterologous systems, or studies conducted in isolation of other effector proteins (Lorrain et al. 2019). Time and space constraints can often be a limiting factor in effector screens, because a low success rate is compensated for by assessing a large number of candidates. Speed is of importance as the virulence phenotypes of rust fungi change quickly. There is a lack of automatic data acquisition or computer vision-based methods for phenotyping of effector screens, although they may contribute to higher throughput and consistency in screening assays. Given lists of hundreds to thousands of candidate effectors in a single species, plus unique variants of those

proteins between isolates, higher throughput is needed to screen these effector candidates for functions.

Phenotyping Strategies for Rust Diseases

Historically, rust disease phenotyping has relied on standard area diagrams, graphical representations of disease symptoms at various severities (Peterson et al. 1948). Typically, manual quantification of disease symptoms is considered the "gold standard", but when assigning qualitative scores to disease incidence, human error is often high (Bade and Carmona 2011; Bock et al. 2021; Habib et al. 2022). Prior knowledge and experience, time allotted for scoring, and color discrepancies (such as those due to color-blindness) can all influence a particular scorer, and scores can vary both from the same person or between people. This is particularly evident at higher density disease phenotypes, as people tend to overestimate disease as disease percentage increases (Habib et al. 2022). These discrepancies can have impacts both on research conclusions and on control method applications, as an overestimation of disease might result in additional applications of expensive and environmentally harmful fungicides (Bock et al. 2021).

To reduce the inaccuracies and variabilities of human disease scoring, as well as generate phenotyping data not humanly possible, computer vision-based disease phenotyping platforms at all scales have been rapidly developed (Mutka and Bart 2015; Simko et al. 2017; Tanner et al. 2022). The advent of tools like PlantCV (Gehan et al. 2017) have made computer vision applications to plant phenotyping more accessible and customizable for a given problem. There are several approaches to image-based phenotyping, from fairly simple to particularly complex. One basic but powerful approach takes advantage of the distinct borders and contrasting colors of red to orange rust symptoms on green plant tissue. Some strategies utilize images of individual leaves (Cui et al. 2010) or whole plants at ground level (Agarwal and Samantaray 2016), while

others employ drones to take whole-plot aerial images (Ganthaler et al. 2018). Thresholding methods can then be applied to the images to segment rust symptoms from healthy leaf tissue and background based on HSI (hue saturation intensity) or RGB (red green blue) channels. One study assessing soybean rust symptoms, converted RGB images of soybean leaflets to HSI and identified rust symptoms based on sharp changes in hue and value (Cui et al. 2010). Another study obtained ground and aerial images of needle bladder rust diseased Norway spruce trees, and disease quantification as predicted by built-in functions of ImageJ correlated well to manually-generated ground truth disease coverage ($R2 = 0.87$-0.95) (Ganthaler et al. 2018).

Machine learning (ML) or deep learning is also becoming quite popular in plant disease phenotyping. ML models are often applied for feature extraction of disease symptoms coupled with classification or quantification in an autonomous manner (Mochida et al. 2018). Deep learning can be approached in many ways, but the most common methods utilize supervised learning, where manually generated labels, or annotations, inform model predictions. Many applications involve disease identification or classification in field settings, where multiple diseases can affect a single plot (Ullah et al. 2021; Paliwal and Joshi 2022; Mafukidze et al. 2022). A recently published study utilized convolutional neural networks (CNNs) to extract features of wheat leaves and classify them as healthy or infected with a particular disease. The CNNs were able to classify the five classes (healthy, powdery mildew, rust, blight, or septoria) with a 98.83% accuracy (Xu et al. 2023). Other approaches aim to quantify disease symptoms either on a plant (Heineck et al. 2019) or whole field level (DeSalvio et al. 2022), with applications in analysis of resistance trials or informing deployment of control methods.

The appeal of ML stems from fast and accurate phenotyping results, where classification or quantification of disease symptoms can be completed several times faster than manual

methods with potentially more consistent and accurate results (Habib et al. 2022). There are still many limitations to ML, however, as developed models have limited applications outside of their intended use case, and accuracy is largely dependent on the quality and amount of annotation data. For rust diseases in particular, the features (pustules) are very small and numerous, making annotation an exceptionally painstaking task. A better understanding of the types of data needed to train ML models and reduce the amount of annotation data required are necessary to fast-track ML applications in more rust pathosystems.

Dissertation Organization

The aim of this dissertation is to develop tools for the identification and characterization of effector proteins in the *P. sorghi*-maize pathosystem. In Chapter 2, I detail an improved genome assembly for *P. sorghi* utilizing long-read Oxford Nanopore sequencing technologies. This assembly provides not only a more complete view of this pathogen, but also expands on the current knowledge of the diversity within the species through the sequencing of a new isolate. In Chapter 3, I analyze eight CSEPs of *P. sorghi*, related to a previously characterized effector protein from *P. pachyrhizi*. In addition to traditional immune suppression assays, I detail the use of a phenotyping box to generate automated time course images throughout experiments. In Chapter 4, I discuss an ML pipeline for the quantification of *P. sorghi* on maize leaves. In addition to the development of a tool useful for future studies, particularly those involving silencing of maize proteins or pathogenicity factors, the chapter also discusses the effect that the type and amount of annotation data has on the ability to answer biologically meaningful questions with the resultant ML models.

References

Agarwal, C., and Samantaray, S. D. 2016. A Novel Image Processing based Approach for Identification of Yellow Rust in Wheat Plants. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 6:220-226.

Aime, M. C., McTaggart, A. R., Mondo, S. J., and Duplessis, S. 2017. Phylogenetics and Phylogenomics of Rust Fungi. In *Advances in Genetics*, Academic Press Inc. p. 267–307.

Allen, E. A., Hazen, B. E., Hoch, H. C., Kwon, Y., Leinhos, G. M. E., Staples, R. C., et al. 1991. Appressorium Formation in Response to Topographical Signals by 27 Rust Species. Phytopathology. 81:323.

Anderson, C. L., Kubisiak, T. L., Nelson, C. D., Smith, J. A., and Davis, J. M. 2010. Genome size variation in the pine fusiform rust pathogen *Cronartium quercuum* f.sp. *fusiforme* as determined by flow cytometry. Mycologia. 102:1295–1302.

Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., et al. 2015. The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. Food Secur. 7:303–321.

Bade, C. I. A., and Carmona, M. A. 2011. Comparison of methods to assess severity of common rust caused by *Puccinia sorghi* in maize. Trop. Plant Pathol. 36:264–266.

Bock, C. H., Chiang, K.-S., and Del Ponte, E. M. 2021. Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. Trop. Plant Pathol. 47:25–42.

Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K. K., et al. 2011. Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. PLoS One. 6.

Catanzariti, A.-M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A., and Ellis, J. G. 2006. Haustorially Expressed Secreted Proteins from Flax Rust Are Highly Enriched for Avirulence Elicitors. Plant Cell. 18:243–256.

Chavan, S., Gray, J., and Smith, S. M. 2015. Diversity and evolution of *Rp1* rust resistance genes in four maize lines. Theor. Appl. Genet. 128:985–998.

Chen, G., Zhang, B., Ding, J., Wang, H., Deng, C., Wang, J., et al. 2022. Cloning southern corn rust resistant gene *RppK* and its cognate gene *AvrRppK* from *Puccinia polysora*. Nat. Commun. 13:1–11.

Cui, D., Zhang, Q., Li, M., Hartman, G. L., and Zhao, Y. 2010. Image processing methods for quantitatively detecting soybean rust from multispectral images. Biosyst. Eng. 107:186–193.

Darino, M. A., Rochi, L., Lia, V. V., Kreff, E. D., Pergolesi, M. F., Ingala, L. R., et al. 2016. Virulence characterization and identification of Maize lines resistant to *Puccinia sorghi* Schwein. Present in the Argentine Corn Belt region. Plant Dis. 100:770–776.

Delaney, D. E., Webb, C. A., and Hulbert, S. H. 1998. A Novel Rust Resistance Gene in Maize Showing Overdominance. Mol. Plant-Microbe Interact. 11:242–245.

Deng, C., Leonard, A., Cahill, J., Lv, M., Li, Y., Thatcher, S., et al. 2022. The *RppC-AvrRppC* NLR-effector interaction mediates the resistance to southern corn rust in maize. Mol. Plant. 15:904–912.

DeSalvio, A. J., Adak, A., Murray, S. C., Wilde, S. C., and Isakeit, T. 2022. Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms. Sci. Rep. 12:1–14.

Dey, U., Harlapur, S. I., Dhutraj, D. N., Suryawanshi, A. P., Badgujar, S. L., Jagtap, G. P., et al. 2012. Spatiotemporal yield loss assessment in corn due to common rust caused by *Puccinia sorghi* Schw. African J. Agric. Res. 7:5265–5269.

Djulic, A., Schmid, A., Lenz, H., Sharma, P., Koch, C., Wirsel, S. G. R., et al. 2011. Transient transformation of the obligate biotrophic rust fungus *Uromyces fabae* using biolistics. Fungal Biol. 115:633–642.

Duan, H., Jones, A. W., Hewitt, T., Mackenzie, A., Hu, Y., Sharp, A., et al. 2022. Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. Genome Biol. 23:1–27.

Dunhin, B. J., Pretorius, Z. A., Bender, C. M., Kloppers, F. J., and Flett, B. C. 2004. Description of spore stages of *Puccinia sorghi* in South Africa. South African J. Plant Soil. 21:48–52.

Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., et al. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci. 108:9166–9171.

Duplessis, S., Lorrain, C., Petre, B., Figueroa, M., Dodds, P. N., and Aime, M. C. 2021. Host Adaptation and Virulence in Heteroecious Rust Fungi. Annu. Rev. Phytopathol. 59:403–422.

Figueroa, M., Dodds, P. N., Henningsen, E. C., and Sperschneider, J. 2023. Global Landscape of Rust Epidemics by *Puccinia* Species: Current and Future Perspectives. In *Plant Relationships, The Mycota*, Springer, Cham, p. 391–423.

Ganthaler, A., Losso, A., and Mayr, S. 2018. Using image analysis for quantitative assessment of needle bladder rust disease of Norway spruce. Plant Pathol. 67:1122–1130.

Garnica, D. P., Nemri, A., Upadhyaya, N. M., Rathjen, J. P., and Dodds, P. N. 2014. The Ins and Outs of Rust Haustoria. Joseph Heitman. PLoS Pathog. 10:e1004329.

Garnica, D. P., Upadhyaya, N. M., Dodds, P. N., and Rathjen, J. P. 2013. Strategies for Wheat Stripe Rust Pathogenicity Identified by Transcriptome Sequencing. PLoS One. 8.

Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., et al. 2017. PlantCV v2: Image analysis software for high-throughput plant phenotyping. PeerJ. 2017:e4088.

Ghurye, J., Pop, M., Koren, S., Bickhart, D., and Chin, C. S. 2017. Scaffolding of long-read assemblies using long range contact information. BMC Genomics. 18:527.

Gibriel, H. A. Y., Thomma, B. P. H. J., and Seidl, M. F. 2016. The Age of Effectors: Genome-Based Discovery and Applications. Phytopathology. 106:1206–1212.

Godoy, C. V., Seixas, C. D. S., Soares, R. M., Marcelino-Guimarães, F. C., Meyer, M. C., and Costamilan, L. M. 2016. Asian soybean rust in Brazil: Past, present, and future. Pesqui. Agropecu. Bras. 51:407–421.

Groth, J. V, Zeyen, R. J., Davis, D. W., and Christ, B. J. 1983. Yield and quality losses caused by common rust (*Puccinia sorghi* Schw.) in sweet corn (*Zea mays*) hybrids. Crop Prot. 2:105–111.

Guerra, F. A., Brücher, E., De Rossi, R. L., Plazas, M. C., Guerra, G. D., and Ducasse, D. A. 2016. First report of *Oxalis conorrhiza* as alternate host of *Puccinia sorghi*, causal agent of common rust of Maize. Plant Dis. 100:519.

Guerra, F. A., De Rossi, R. L., Brücher, E., Vuletic, E., Plazas, M. C., Guerra, G. D., et al. 2019. Occurrence of the complete cycle of *Puccinia sorghi* Schw. in Argentina and implications on the common corn rust epidemiology. Eur. J. Plant Pathol. 154:171–177.

Gupta, Y. K., Marcelino-Guimarães, F. C., Lorrain, C., Farmer, A., Haridas, S., Ferreira, E. G. C., et al. 2023. Major proliferation of transposable elements shaped the genome of the soybean rust pathogen *Phakopsora pachyrhizi*. Nat. Commun. 2023 141. 14:1–16.

Habib, A., Abdullah, A., and Puyam, A. 2022. Visual Estimation: A Classical Approach for Plant Disease Estimation. In *Trends in Plant Disease Assessment*, Springer, Singapore, p. 19–45.

Hagan, W. L., and Hooker, A. L. 1965. Genetics of reaction to *Puccinia sorghi* in eleven corn inbred lines from Central and South America. Phytopathology. 55:193–197.

Heineck, G. C., McNish, I. G., Jungers, J. M., Gilbert, E., and Watkins, E. 2019. Using R-Based Image Analysis to Quantify Rusts on Perennial Ryegrass. Plant Phenome J. 2:1–10.

Henningsen, E. C., Hewitt, T., Dugyala, S., Nazareno, E. S., Gilbert, E., Li, F., et al. 2022. A chromosome-level, fully phased genome assembly of the oat crown rust fungus *Puccinia coronata* f. sp. *avenae*: a resource to enable comparative genomics in the cereal rusts. G3 Genes, Genomes, Genet. 12.

Hooker, A. L. 1985. Corn and Sorghum Rusts. Dis. Distrib. Epidemiol. Control. :207–236.

Hulbert, S. H. 1991. Reactions of Maize Lines Carrying *Rp* Resistance Genes to Isolates of the Common Rust Pathogen, *Puccinia sorghi*. Plant Dis. 75:1130.

Jin, Y., Szabo, L. J., and Carson, M. 2010. Century-old mystery of *Puccinia striiformis* life history solved with the identification of berberis as an alternate host. Phytopathology. 100:432–435.

Jochua, C., Amane, M. I. V, Steadman, J. R., Xue, X., and Eskridge, K. M. 2008. Virulence Diversity of the Common Bean Rust Pathogen Within and Among Individual Bean Fields and Development of Sampling Strategies. Plant Dis. 92:401–408.

Kolmer, J. 2013. Leaf rust of wheat: Pathogen biology, variation and host resistance. Forests. 4:70–84.

Kronenberg, Z. N., Rhie, A., Koren, S., Concepcion, G. T., Peluso, P., Munson, K. M., et al. 2021. Extended haplotype-phasing of long-read *de novo* genome assemblies using Hi-C. Nat. Commun. 12:1–10.

Lawrence, G. J., Dodds, P. N., and Ellis, J. G. 2010. Transformation of the flax rust fungus, *Melampsora lini*: Selection via silencing of an avirulence gene. Plant J. 61:364–369.

Li, C., Qiao, L., Lu, Y., Xing, G., Wang, X., Zhang, G., et al. 2023. Gapless Genome Assembly of *Puccinia triticina* Provides Insights into Chromosome Evolution in Pucciniales. Microbiol. Spectr. 11.

Liang, J., Li, Y., Dodds, P. N., Figueroa, M., Sperschneider, J., Han, S., et al. 2023. Haplotype-phased and chromosome-level genome assembly of *Puccinia polysora*, a giga-scale fungal pathogen causing southern corn rust. Mol. Ecol. Resour. 23:601–620.

Link, T. I., Lang, P., Scheffler, B. E., Duke, M. V., Graham, M. A., Cooper, B., et al. 2014. The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. Mol. Plant Pathol. 15:379–393.

Lorrain, C., Gonçalves dos Santos, K. C., Germain, H., Hecker, A., and Duplessis, S. 2019. Advances in understanding obligate biotrophy in rust fungi. New Phytol. 222:1190–1206.

Lübberstedt, T., Klein, D., and Melchinger, A. E. 1998. Comparative Quantitative Trait Loci Mapping of Partial Resistance to *Puccinia sorghi* Across Four Populations of European Flint Maize. Phytopathology. 88:1324–1329.

Mafukidze, H. D., Owomugisha, G., Otim, D., Nechibvute, A., Nyamhere, C., and Mazunga, F. 2022. Adaptive Thresholding of CNN Features for Maize Leaf Disease Classification and Severity Estimation. Appl. Sci. 12:8412.

Mahindapala, R. 1978. Occurrence of maize rust, *Puccinia sorghi*, in England. Trans. Br. Mycol. Soc. 70:393–399.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. 38:4647–4654.

Miller, M. E., Zhang, Y., Omidvar, V., Sperschneider, J., Schwessinger, B., Raley, C., et al. 2018. *De novo* Assembly and Phasing of Dikaryotic Genomes from Two Isolates of *Puccinia coronata* f. sp. *avenae*, the Causal Agent of Oat Crown Rust. MBio. 9:e01650-17.

Mochida, K., Koda, S., Inoue, K., Hirayama, T., Tanaka, S., Nishii, R., et al. 2018. Computer vision-based phenotyping for improvement of plant productivity: A machine learning perspective. Gigascience. 8:1–12.

Mutka, A. M., and Bart, R. S. 2015. Image-based phenotyping of plant disease symptoms. Front. Plant Sci. 5.

Nandety, R. S., Gill, U. S., Krom, N., Dai, X., Dong, Y., Zhao, P. X., et al. 2022. Comparative Genome Analyses of Plant Rust Pathogen Genomes Reveal a Confluence of Pathogenicity Factors to Quell Host Plant Defense Responses. Plants. 11:1962.

Olukolu, B. A., Tracy, W. F., Wisser, R., De Vries, B., and Balint-Kurti, P. J. 2016. A genome-wide association study for partial resistance to maize common rust. Phytopathology. 106:745–751.

Paliwal, J., and Joshi, S. 2022. An Overview of Deep Learning Models for Foliar Disease Detection in Maize Crop. J. Artif. Intell. Syst. 4:1–21.

Pataky, J. K. 1987. Quantitative Relationships Between Sweet Corn Yield and Common Rust, *Puccinia sorghi*. Phytopathology. 77:1066.

Pataky, J. K., and Eastburn, D. M. 1993. Comparing Partial Resistance to *Puccinia sorghi* and Applications of Fungicides for Controlling Common Rust on Sweet Corn. Phytopathology. 83:1046.

Pataky, J. K., Gonzalez, M., Brewbaker, J. L., and Kloppers, F. J. 2001. Reactions of *Rp*-resistant, processing sweet corn hybrids to populations of *Puccinia sorghi* virulent on corn with the *Rp1-D* gene. HortScience. 36:324–327.

Periyannan, S., Milne, R. J., Figueroa, M., Lagudah, E. S., and Dodds, P. N. 2017. An overview of genetic rust resistance: From broad to specific mechanisms. PLoS Pathog. 13:e1006380.

Peterson, R. F., Campbell, A. B., and Hannah, A. E. 1948. A Diagrammatic Scale for Estimating Rust Intensity on Leaves and Stems of Cereals. Can. J. Res. 26c:496–500.

Petre, B., and Duplessis, S. 2022. A decade after the first Pucciniales genomes: A bibliometric snapshot of (post) genomics studies in three model rust fungi. Front. Microbiol. 13:3480.

Qi, M., Grayczyk, J. P., Seitz, J. M., Lee, Y., Link, T. I., Choi, D., et al. 2018. Suppression or Activation of Immune Responses by Predicted Secreted Proteins of the Soybean Rust Pathogen *Phakopsora pachyrhizi*. Mol. Plant-Microbe Interact. 31:163–174.

Qi, M., Link, T. I., Müller, M., Hirschburger, D., Pudake, R. N., Pedley, K. F., et al. 2016. A Small Cysteine-Rich Protein from the Asian Soybean Rust Fungus, *Phakopsora pachyrhizi*, Suppresses Plant Immunity ed. Peter N Dodds. PLoS Pathog. 12:e1005827.

Quade, A., Ash, G. J., Park, R. F., and Stodart, B. 2021. Resistance in Maize (*Zea mays*) to Isolates of *Puccinia sorghi* from Eastern Australia. Phytopathology. 111:1751–1757.

Rafiqi, M., Ellis, J. G., Ludowici, V. A., Hardham, A. R., and Dodds, P. N. 2012. Challenges and progress towards understanding the role of effectors in plant-fungal interactions This review comes from a themed issue on Biotic interactions. Curr. Opin. Plant Biol. 15:477–482.

Ramirez-Cabral, N. Y. Z., Kumar, L., and Shabani, F. 2017. Global risk levels for corn rusts (*Puccinia sorghi* and *Puccinia polysora*) under climate change projections. J. Phytopathol. 165:563–574.

Ren, J., Li, Z., Wu, P., Zhang, A., Liu, Y., Hu, G., et al. 2021. Genetic Dissection of Quantitative Resistance to Common Rust (*Puccinia sorghi*) in Tropical Maize (*Zea mays* L.) by Combined Genome-Wide Association Study, Linkage Mapping, and Genomic Prediction. Front. Plant Sci. 12.

Rochi, L., Diéguez, M. J., Burguener, G., Darino, M. A., Pergolesi, M. F., Ingala, L. R., et al. 2018. Characterization and comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize common rust. Fungal Genet. Biol. 112:31–39.

Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., et al. 2018. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high interhaplotype diversity. MBio. 9.

Simko, I., Jimenez-Berni, J. A., and Sirault, X. R. R. 2017. Phenomic approaches and tools for phytopathologists. Phytopathology. 107:6–17.

Sperschneider, J. 2022. NuclearPhaser. Available at: https://github.com/JanaSperschneider/NuclearPhaser.git.

Sperschneider, J., and Dodds, P. N. 2022. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. Mol. Plant-Microbe Interact. 35:146–156.

Tanner, F., Tonn, S., de Wit, J., Van den Ackerveken, G., Berger, B., and Plett, D. 2022. Sensor-based phenotyping of above-ground plant-pathogen interactions. Plant Methods. 18:1–18.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 40:1023–1025.

Tobias, P. A., Schwessinger, B., Deng, C. H., Wu, C., Dong, C., Sperschneider, J., et al. 2021. *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. G3 Genes, Genomes, Genet. 11.

Uhse, S., and Djamei, A. 2018. Effectors of plant-colonizing fungi and beyond. PLOS Pathog. 14:e1006992.

Ullah, K., Jan, M. A., and Sayyed, A. 2021. Automatic Diseases Detection and Classification in Maize Crop using Convolution Neural Network. Int. J. Adv. Trends Comput. Sci. Eng. 10:675–679.

Upadhyaya, N. M., Garnica, D. P., Karaoglu, H., Sperschneider, J., Nemri, A., Xu, B., et al. 2015. Comparative genomics of australian isolates of the wheat stem rust pathogen *Puccinia graminis* f. Sp. *Tritici* reveals extensive polymorphism in candidate effector genes. Front. Plant Sci. 5.

Vittal, R., Yang, H. C., and Hartman, G. L. 2012. Anastomosis of germ tubes and migration of nuclei in germ tube networks of the soybean rust pathogen, *Phakopsora pachyrhizi*. Eur. J. Plant Pathol. 132:163–167.

Voegele, R. T., and Mendgen, K. 2003. Rust haustoria: Nutrient uptake and beyond. New Phytol. 159:93–100.

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. 37:1155–1162.

Wu, W., Nemri, A., Blackman, L. M., Catanzariti, A. M., Sperschneider, J., Lawrence, G. J., et al. 2019. Flax rust infection transcriptomics reveals a transcriptional profile that may be indicative for rust Avr genes. PLoS One. 14:e0226106.

Xia, C., Wang, M., Yin, C., Cornejo, O. E., Hulbert, S. H., and Chen, X. 2018. Genome Sequence Resources for the Wheat Stripe Rust Pathogen (*Puccinia striiformis* f. sp. *tritici*) and the Barley Stripe Rust Pathogen (*Puccinia striiformis* f. sp. *hordei*). Mol. Plant-Microbe Interact. 31:1117–1120.

Xu, L., Cao, B., Zhao, F., Ning, S., Xu, P., Zhang, W., et al. 2023. Wheat leaf disease identification based on deep learning algorithms. Physiol. Mol. Plant Pathol. 123:101940.

Zhao, J., Wang, L., Wang, Z., Chen, X., Zhang, H., Yao, J., et al. 2013. Identification of eighteen berberis species as alternate hosts of *Puccinia striiformis* f. sp. *Tritici* and virulence variation in the pathogen isolates from natural infection of barberry plants in China. Phytopathology. 103:927–934.

Zheng, H., Chen, J., Mu, C., Makumbi, D., Xu, Y., and Mahuku, G. 2018. Combined linkage and association mapping reveal QTL for host plant resistance to common rust (*Puccinia sorghi*) in tropical maize. BMC Plant Biol. 18.

Zheng, W., Huang, L., Huang, J., Wang, X., Chen, X., Zhao, J., et al. 2013. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. Nat. Commun. 4.

CHAPTER 2. LONG-READ GENOME ASSEMBLY OF AN IOWAN *PUCCINIA SORGHI* ISOLATE

Katerina L. Holan[1], Manjula Elmore[2], and Steven A. Whitham[1]

[1]Department of Plant Pathology, Entomology, and Microbiology, Iowa State University, Ames, IA

[2]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Modified from a manuscript to be submitted to *BMC Genomics*

Abstract

With the recent advance in long-read sequencing, methods for generating linkage information between reads, and computational methods for assembly and phasing, rust fungal genomes have gone from highly fragmented with limited information on repeat regions, to scaffolded, nearly complete to complete genome resources for various species and isolates. In this work, we report on a long-read-based genome assembly, scaffolded with Hi-C reads, for the *Puccinia sorghi* isolate IA16, totaling 902 scaffolds. We additionally include pseudophased haplotypes, with 1,277 and 1,262 scaffolds. The assembled genome has a haploid size of 174 Mb, including significant repeat content of 76% and 16,336 predicted genes, of which 742 are predicted to code for effector proteins. This assembly provides a more complete view on the genomic content of *P. sorghi*, as well as providing additional information as to the effector content and evolution of *P. sorghi* and other rust pathogens.

Introduction

Fungal rust pathogens of the order Pucciniales are responsible for some of the most impactful crop diseases. One member of this order, *Puccinia sorghi*, causes common rust of maize and is a global threat to maize production. Climate change models suggest suitable environmental conditions for common rust disease are expected to expand in the Northern

hemisphere in the next 100 years (Ramirez-Cabral et al. 2017). There are several *Rp* genes in maize that provide resistance to various *P. sorghi* populations (Hulbert 1997; Chavan et al. 2015; Hooker 1985; Delaney et al. 1998; Hagan and Hooker 1965), but many isolates are known to overcome individual resistances (Hulbert 1991; Quade et al. 2021). Genomic resources for maize are abundant, but high-quality rust fungi genome assemblies have historically been difficult to generate (Duplessis et al. 2014). Rust fungi are obligate biotrophs, making it difficult to acquire enough tissue for sequencing, and metabolite carryover in genomic DNA isolation methods can interfere with sequencing chemistry (Jones and Schwessinger 2021). The genomes of rust fungi are large and complex when compared to other fungi, with some haploid sizes reaching over 1 Gb (Ramos et al. 2015; Tobias et al. 2021; Tavares et al. 2014), and with repetivie sequences accounting for up to 90% of the genome (Xia et al. 2022; Liang et al. 2022). Rust fungi are also dikaryotic during the majority of their life cycles, with two genetically distinct haploid nuclei within each cell, and there have been many recent attempts to fully phase the separate nucleic sequences present in the two nuclei (Aime et al. 2017).

The earliest sequencing projects for rust fungi genomes utilized Illumina short-reads (typically 50-300 bp), often resulting in highly fractured assemblies composed of 20,000-50,000 contigs or scaffolds (Aime et al. 2017). A lack of contiguity limits the gene order information required for phylogenetic and evolutionary analyses of these genomes and complicates phasing of the two nuclei into two distinct haplotypes. Recent assembly strategies have utilized long-read sequencing (typically 5,000-30,000 bp) technologies such as those from Oxford Nanopore Technologies and PacBio. Pairing long-read assemblies with linked read information, such as from 10X Genomics or Hi-C libraries, can further aid scaffolding and phasing efforts. Presently, there are 60 *Puccinia* genome assemblies available from GenBank, which includes alternate

haplotypes of some isolates, with ten containing chromosome-level sequences, and two considered complete assemblies. Some strategies for phasing utilized PacBio long-reads in conjunction with the FALCON-Unzip pipeline (Chin et al. 2016) to recover haplotypic information (Schwessinger et al. 2018; Vasquez-Gross et al. 2020). The development of the NuclearPhaser pipeline (Sperschneider 2022) has also aided in the production of phased haplotypes in rust fungi assemblies (Henningsen et al. 2022; Liang et al. 2022; Duan et al. 2022). The two currently complete assemblies represent the fully phased haplotypes for *Puccinia triticina* isolate Pt15, generated from PacBio HiFi long-reads, Illumina short-reads, and Hi-C linked reads (Li et al. 2023). The assembly of additional rust fungal genomes of species with no previous assembly and isolates of previously sequenced species will allow for comparative genomics between rust pathogens and isolates of the same species (Nandety et al. 2022).

The current genomic resources for *P. sorghi* consist of a draft assembly and annotation of an Argentine isolate, RO10H11247, originally collected in 2010 (Rochi et al. 2018). The *de novo* assembly was generated from a 200 bp paired-end (PE) library and scaffolded with a 5000 bp mate-paired library. The final assembly consists of 15,722 scaffolds with an estimated size of 99.6 Mb, with 28.6% consisting of unknown bases. The assembly was annotated with whole transcriptomic data from *Puccinia graminis* f. sp. *tritici*, *Puccinia striiformis* f. sp. *tritici*, *Puccinia triticina*, and *Ustilago maydis*, resulting in ~21,000 predicted protein-coding genes and 1,599 candidate secreted effector proteins (CSEPs).

Sequencing technologies and protocols have rapidly improved since the original draft *P. sorghi* genome was reported, thus we aimed to improve on genomic structure information and resolve repeat regions of an isolate of *P. sorghi* from the Midwestern United States. In support of this aim, we utilized long-read Oxford Nanopore sequences and short-read Hi-C sequences to

assemble a pseudohaplotypic, long-read draft genome for the Iowan *P. sorghi* isolate IA16. After annotation with comprehensive IA16-specific transcriptome, the assembly reported here provides another resource for rust fungi phylogenomics and effector identification and evolution.

Material and Methods

IA16 isolation

The IA16 isolate was generated from a *P. sorghi* sample gathered from a maize field in Boone County, Iowa, in 2016. To ensure homogeneity, the isolate was purified with four rounds of single pustule isolation, where one well-isolated pustule from the previous round was used to inoculate new plants.

Maize and *P. sorghi* growth and maintenance

The sweet corn variety 'Golden Bantam' was used to maintain and accumulate *P. sorghi* urediniospores. All plants were grown under 16-hour days in either a greenhouse or growth room (22-25˚C). Sweet corn seeds were sown in a peat-based growing medium in plastic 10"x20" greenhouse trays inset with 48-well inserts, with two to three seeds per well. When grown in a growth room, trays were covered with a transparent plastic dome for a week to aid germination. Trays grown in greenhouse conditions were left uncovered. Seedlings were grown until they had two to four leaves, or approximately 10-14 days, before inoculation. Initial inoculation utilized urediniospores from frozen storage (-80˚C), which were prepared by thawing in a 42˚C water bath for five minutes before direct inoculation onto sweet corn seedlings by dusting the thawed urediniospores onto leaves. Subsequent inoculations were conducted by brushing sporulating plants onto new seedlings. After inoculation, seedlings were moved to a dew chamber constructed from PVC pipe and a waterproof tarp, sprayed thoroughly with water, and left overnight. Alternatively, plants were sprayed thoroughly with water and covered with a tall, transparent plastic cover overnight. The seedlings were then returned to normal growth

conditions and left uncovered. Disease symptoms began appearing at four days after inoculation (DAI), with uredinia visible after six to seven DAI, and significant spore production at nine to ten DAI. Urediniospores were collected either by tapping sporulating leaves over a piece of paper and pouring into a 1.5 mL microcentrifuge tube or via vacuum collection. To vacuum collect spores, a cyclone spore collector (Pretorius et al. 2019) connected via tubing to a vacuum pump was systematically run across sporulating maize leaves. The 1.5 mL microcentrifuge tubes containing the collected spores were placed directly into a -80˚C freezer with no prior flash freezing in liquid nitrogen.

IA16 differential characterization

Maize inbred lines with various *Rp* genes were inoculated with IA16 urediniospores and scored at seven to ten DAI. Disease development was scored as either virulent (+), with heavy pustule coverage, avirulent (-), with either chlorotic flecking or no visible disease development, or intermediate (I), with variable or indeterminate symptoms.

DNA isolation and sequencing

High molecular weight DNA was extracted from urediniospores according to fungi-specific protocols (Schwessinger and Rathjen 2017; Schwessinger 2019; Jones et al. 2019; Jones and Schwessinger 2021; Duan et al. 2022). DNA samples were sent to the Iowa State University DNA Facility for quality analysis, size selection, library preparation, and sequencing. Oxford Nanopore Technologies libraries were created with the SQK-LSK109 kit and run on a GridIONx5 using FLO-MIN106 flow cells. A total of three libraries were created and run on three separate flow cells. Base-calling was done with Guppy 2.1.3, 3.2.10, and 5.1.13 respectively.

For Hi-C sequencing, 200 mg of urediniospores were ground in a liquid nitrogen-cooled mortar and resuspended in 5 mL of 1% formaldehyde. The mixture was incubated at room

temperature for 20 minutes, with periodic vortexing. Glycine was added to a final concentration of 125 mM and incubated at room temperature for 15 minutes, again vortexing periodically. The spores were centrifuged for one minute at 1000 × g and rinsed with ddH_2O after removal of the supernatant. The spores were again spun down and the supernatant was removed. The cross-linked tissue was transferred to a 1.5 mL microcentrifuge tube and stored at -80˚C. The cross-linked sample was shipped to Phase Genomics (Seattle, WA, USA) on dry ice for DNA extraction, Hi-C library generation, and sequencing. The 150-bp paired end library was created according to the Proximo Hi-C (Fungal) 3.0 protocol and sequenced with Illumina to a read depth of 100 million read pairs (RPs).

RNA isolation and sequencing

Seedlings of the maize inbred H95 were inoculated with the IA16 isolate and samples were taken at 18 hours post inoculation (HPI), 24 HPI, 3 DAI, 40 HPI, 5 DAI, and 7 DAI, each with four biological replicates. Resting spores and germinated spores were also sampled, with five biological replicates each. Three 250-bp PE libraries were created from pooled RNA; the first included all 28 H95 samples, the second included all five resting spore samples, and the third included all five germinated spore samples. The resulting libraries were each sequenced on two SP flow cells on a NovaSeq 6000.

Genome assembly, cleaning, and polishing

Genome assembly was conducted with Flye 2.9.1 (Lin et al. 2016; Kolmogorov et al. 2019), using the '--scaffold' flag for the collapsed haploid assembly and the '--keep-haplotypes' and '--min-overlap 10000' flags for the haplotype assemblies. Mitochondrial and contaminant contigs were identified by BLAST+ 2.13 (Sperschneider 2021). Briefly, mitochondrial contigs were searched against the NCBI mitochondrial database and moved to a separate file. To identify

contaminant contigs for removal, contigs were searched against the NCBI nucleotide database. Each assembly was polished twice with medaka 1.5 (https://github.com/nanoporetech/medaka).

Haplotype phasing and scaffolding

To create the pseudohaplotypes, HapDup 0.12 was applied to the Flye haplotype assembly (Kolmogorov et al. 2019; Shafin et al. 2021). The Hi-C reads were aligned and processed with the Arima Genomics Hi-C mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline). Briefly, Hi-C reads were mapped individually to each assembly with BWA-MEM from BWA 0.7.17 (https://github.com/lh3/bwa), chimeric regions were removed from the 5' end of each read, and read pairs were matched. Assemblies were scaffolded with SALSA 2.2, using the '--clean yes' flag to correct any identified mis-assemblies in the input (Ghurye et al. 2017, 2019). BUSCO analysis and identification was conducted with BUSCO 5.4.3 with the basidiomycota_odb10 database (Manni et al. 2021).

Annotation of genome assembly

De novo repeat libraries were predicted with RepeatModeler 2.0.3 with the '-LTRStruct' option (Flynn et al. 2020). Repeats were soft-masked by running RepeatMasker 4.1.2 (Smit et al. 2015) twice in succession, once with the parameter '-species fungi' and again with the '-lib' option directing to the RepeatModeler *de novo* library output. The resultant soft-masked assemblies were used for gene annotation. The sequencing output from the six previously described RNA sequencing flow cells was combined and reads were trimmed and cleaned with Fastp 0.12.4 (Chen et al. 2018; Chen 2023). Cleaned reads were aligned to each assembly with HISAT2 2.2.1 (Kim et al. 2019) with the parameters '--max-intronlen 3000' and '--dta'. Transcripts were assembled with Trinity 2.15.1 (Grabherr et al. 2011; Haas et al. 2013) in genome-guided mode with the parameters '--jaccard_clip', '--genome_guided_bam', and '--

genome_guided_max_intron 3000'. Annotation was performed on the soft-masked assemblies with funannotate 1.8.15 (Palmer and Stajich 2022). Funannotate train was run first with the Trinity assembly using the parameters '--jaccard_clip', '--no_normalize_reads', and '--no_trimmomatic' as reads were previously cleaned. Funannotate predict was run next with the parameter '--optimize_augustus'. Erroneous gene models detected by funannotate predict were manually removed, and funannotate fix was run to update the gene models. Funannotate update was run next with the parameters '--jaccard_clip', '--no_trimmomatic', and '--no_normalize_reads'. Using the "proteins.fa" files from funannotate update, gene model annotations were predicted through several methods. InterProScan 5 (Jones et al. 2014; Blum et al. 2021) was run through funannotate. The fungal version of antiSMASH 6.1.1 (Blin et al. 2021) was run locally. Phobius 1.01 (Käll et al. 2004, 2007), eggNOG-mapper 2.1.9 using the eggNOG 5.0 database (Huerta-Cepas et al. 2019; Cantalapiedra et al. 2021), and SignalP 6.0 (Teufel et al. 2022), with the options 'Eukarya' and 'Slow', were run through their respective web services. Finally, funannotate annotate was run, using the aforementioned predicted annotations as input. To predict effector proteins, the combined unique predicted secreted proteins from SignalP and Phobius were run through EffectorP 3.0 (Sperschneider and Dodds 2022). Genomic feature distribution was visualized with karyoploteR 1.27.0, with metrics calculated in 100 kb non-overlapping sliding windows (Gel and Serra 2017).

Comparison between IA16 and RO10H11247 isolates

The online interactive D-GENIES (Cabanettes and Klopp 2018) application was used with Minimap2 2.24 (Li 2018) to generate a dot plot to compare the haploid assembly and the RO10H11247 assembly. OrthoFinder 2.4.5 (Emms and Kelly 2015) was used to predict orthogroups between assemblies.

Results

Virulence of IA16 on various *Rp* maize lines

To determine the virulence profile of the IA16 *P. sorghi* isolate, we inoculated various maize inbred lines, sweet corn, or maize H95 lines each carrying a different *Rp* (resistance to *Puccinia sorghi*) gene. IA16 was virulent (produced sporulating pustules) on the majority of *Rp* genes tested, developing similar pustule coverages as maize lines containing no *Rp* genes (Table 1). Two *Rp* lines, *Rp4B* and *Rp*5, had intermediate virulence phenotypes, as fewer pustules developed, suggesting partial resistance to IA16. IA16 was avirulent on two lines, *Rp1-I* and *RpG*, on which some chlorotic flecking developed, indicating a resistance response (Table 1).

The assembled genome is highly contiguous

Approximately 3.3 Gb of quality sequences were acquired from three Nanopore flow cells, totaling ~20x genome coverage in 271,978 reads (Figure 1). The reads were used to generate two assemblies with Flye, one with the '--keep-haplotypes' flag to retain haplotypic information and another without, to create a collapsed haploid genome. Both assemblies were screened for mitochondrial contigs, which were identified by BLAST matches to the NCBI mitochondrial database, high sequencing fold coverage, and low GC content (Sperschneider 2021). Four mitochondrial contigs were identified for the haploid assembly. Upon closer inspection, we found they were two sets of nearly identical sequences, so the shorter contig of each set was removed, leaving us with two mitochondrial contigs totaling 80,903 bps. One mitochondrial contig was identified in the haplotypic assembly, with 80,857 total bps. Given that other *Puccinia* species have mitochondrial genomes of around 80 kb (Cuomo et al. 2017; Vasquez-Gross et al. 2020) and the previously reported *P. sorghi* genome estimated a mitochondrial genome size of 83.8 kb (Darino et al. 2016), we estimate this captures the full mitochondrial sequence for IA16.

After removing the mitochondrial contigs, both assemblies were assessed for contaminant contigs by BLAST against the NCBI nt database (Sperschneider 2021). Nineteen contaminant contigs were removed from the haploid assembly and seven were removed from the haplotypic assembly, mostly containing maize and *Enterobacter* matches. Many rust genome assembly strategies also remove contigs with less than two-fold coverage. However, as we had lower read coverage to begin with, all genomic contigs were retained. Once assemblies were free of contaminants, the remaining contigs were polished with the long-reads in two successive rounds of medaka. The haplotypic assembly was converted to a diploid assembly with HapDup (Kolmogorov et al. 2019; Shafin et al. 2021), which produced two full length pseudohaplotypes, hereafter referred to as haplotype A and haplotype B. They are expected to be chimeric haplotypes. All three assemblies were scaffolded separately with the Hi-C reads using SALSA2 (Ghurye et al. 2017) and Basidiomycota BUSCOs were identified in each scaffolded assembly (Table 2) (Manni et al. 2021).

The haploid assembly is composed of 902 total scaffolds, totaling 174 Mb. The majority of the genome is contained within very few scaffolds, with an L50 of 16 scaffolds and an L90 of 57 scaffolds (Table 2). The haplotype assemblies are less contiguous due to contig breaking by HapDup but are similar in size to the haploid genome, each containing approximately 170 Mb in 1,277 (A) and 1,262 (B) scaffolds (Table 2). The L50 and L90 scores of the haplotypes are also larger, but still much less than their respective total number of scaffolds, with an L50 of 32 (A) or 26 (B) and an L90 of 197 (A) or 190 (B). The haploid assembly has a BUSCO content of 80% complete BUSCOs, with an additional 5% of fragmented BUSCOs. The haplotype assemblies have fewer BUSCO members and more members are fragmented (Table 2). The haplotypes seem to contain unique BUSCOs, and when assessing both haplotypes together, they have

approximately 80% of members represented, with 75% complete (Table 3). The GC content is approximately 45% in all three assemblies.

The genome of IA16 is highly repeat-rich

Repeats were annotated with RepeatMasker using assembly-specific model libraries built by RepeatModeler for all three IA16 assemblies. We similarly reannotated the RO10H11247 assembly to enable a direct comparison between repeat annotation of the two isolates, resulting in 34% total repeats for the RO10H11247 isolate, which is nearly identical to the previously reported 32.53% (Table 4). This is in stark contrast to the three IA16 assemblies, which all have repeat contents of approximately 76%, more than double the previously published assembly, but still in line with more recent rust fungi assemblies (Liang et al. 2022; Duan et al. 2022). The largest proportion of sequence is composed of retroelements, with ~30% of repeat sequence belonging to the Ty3/DIRS1 class and ~6% belonging to the Ty1/Copia class in each IA16 assembly, followed by ~28% of unclassified repeats (Table 4). Approximately 9% of IA16 sequence was classified as DNA transposons. The expanded LTR retrotransposon class is commonly cited for genome expansion in rust fungi (Tobias et al. 2021; Liang et al. 2022). Despite the expanded genome assembly of IA16, the assemblies of the two isolates seem largely correlated according to a dot plot (Figure 2). The amount of non-repeat coverage in the genome is similar for the two assemblies, with IA16 containing 41 Mb and RO10H11247 containing 37 Mb (Table 4).

The genome of IA16 contains similar genic content to other rust fungal species

To ensure a comprehensive transcriptome, RNAseq data from multiple time points of a *P. sorghi*-maize disease time course, as well as from germinated and resting spores, were used to generate transcriptome libraries. Each assembly was annotated with funannotate in the order train, predict, fix, update, and annotate, with annotate run with predictions from InterProScan5,

antiSMASH, eggNOG mapper, SignalP 6.0, and Phobius. This resulted in 16,336 predicted genes in the haploid assembly and ~19,500 within each haplotype assembly (Table 5). However, the average gene length is much smaller with the haplotype assemblies, and gene models may have been split when constructing the haplotypes, artificially increasing gene number. This is supported by the fact that although total gene models are increased for the haplotypes as compared to the haploid, there are 11 to 18% fewer predicted secreted proteins, which rely on an N-terminus secretion signal. Additionally, gene models are shorter and total gene coverage is ~16 Mb in each haplotype as compared to the 20.85 Mb in the haploid assembly (Table 5). To have a direct comparison, we reannotated the RO10H11247 assembly with the same funannotate pipeline using the IA16 ESTs. This resulted in 10,922 gene models, approximately 10,000 fewer models than the previously reported 21,087 genes (Table 5) (Rochi et al. 2018). Although the number of genes drastically differs between the two methods, the total amount of annotated sequence is similar between our annotation (19.32 Mb) and the previous annotation (22.39 Mb). The discrepancy may be due to the high number of unknown bases in the reference genome paired with the genome-guided approach to transcriptome library generation. Despite this, the funannotate pipeline annotated 419 tRNAs, similar to the 405 previously reported (Supplemental Table 1) (Rochi et al. 2018). The IA16 assemblies all have approximately 830 tRNAs, but the proportions of specific tRNAs is similar between IA16 and RO10H11247 (Supplemental Table 1).

To identify predicted secreted proteins, all predicted proteins were analyzed with SignalP 6.0 and Phobius to identify predicted secretion signals. The combined unique proteins from both lists were used as input in EffectorP 3.0, which resulted in 1,845, 1,635, and 1,511 predicted secretion proteins for the haploid assembly, haplotype A, and haplotype B, respectively. Of

these, 742, 655, and 616 are predicted to be effector proteins (Table 5). We repeated this process similarly for both the versions of RO10H11247 annotations (previously reported and funannotate), predicting a total of 1,471 secretion proteins and 563 effector proteins for the funannotate annotations and 1,609 secretion proteins and 615 effector proteins for the previously reported RO10H11247 annotations (Table 5). There are varying numbers of unique orthogroups both for all predicted proteins and for effector proteins between the IA16 haplotypes and between the haploid IA16 and RO10H11247 proteins, with the majority of differences found between the previously reported RO10H11247 proteins and either the IA16 or RO10H11247 funannotate annotated proteins (Supplemental Table 2). When analyzing the largest 16 scaffolds (L50), genomic features appear evenly distributed (Figure 3). In particular, there do not appear to be large trends between gene density and CSEP density, similar to other rust species (Tobias et al. 2021; Schwessinger et al. 2018; Miller et al. 2018).

Discussion

Long-read sequencing technology has been used to great success in the generation of rust fungi genome assemblies. Here, we present an assembly for the IA16 isolate of *P. sorghi* through the use of Oxford Nanopore technologies long-read sequences and Hi-C linked reads. Our assembly is highly contiguous, with half of the sequence contained in only 16 scaffolds. It is a significant improvement over current *P. sorghi* resources and provides additional information in the form of resolved repeat regions and predicted protein coding genes of a different *P. sorghi* isolate. The 742 CSEPs from this assembly will undoubtedly provide exciting avenues for future studies, both in regard to characterization studies and comparison between other *P. sorghi* isolates or other rust species.

Many of the challenges we encountered with the genome assembly likely stem from lower Nanopore output than expected coupled with the relatively high error rates of the flow

cells used for sequencing. Although fungal-specific high molecular weight gDNA isolation protocols were used, it is noted that unidentified contaminants often carry-over and interfere with sequencing (Jones and Schwessinger 2021). Presumably as a result of these contaminants, each of the three Nanopore flow cells resulted in sub-optimal outputs of 1-1.5 Gb of raw data. The large amount of urediniospore tissue required for each Nanopore flow cell limited the number of flow cells we were able to run. As a result, the phasing of the assembly presented here is minimal and phase switching could not be corrected. We were also unable to utilize the promising NuclearPhaser pipeline (Sperschneider 2022), as it is designed to find exactly two haplotypic groups, and we consistently had more than two groups with various initial long-read assembly strategies and settings. NuclearPhaser relies on haplotigs as input, but it seems the majority of contigs for our various assembly attempts are largely collapsed.

The reduced Nanopore data output also likely contributed to reduction of Basidiomycota BUSCOs and gene models when compared to the RO10H11247 assembly. Regardless, the majority of BUSCOs present in the haploid assembly are complete, and the assembly is likely nearly complete. The most significant difference between the two isolates' assemblies remains the quantity of repeat regions, where IA16 has more than double the identified repeats when compared to RO10H11247. The increased number of repeats within the IA16 assembly, and particularly retrotransposon repeats, are the likely the main contributors to the size discrepancy between the assemblies, as non-repeat length is similar, with 41.1 Mb of non-repeat space in IA16 and 37 Mb in RO10H11247. The 28% of unknown sequence within the RO10H11247 genome are likely largely repeat-rich, hence why they were not assembled from the short Illumina reads, but some percentage is likely non-repetitive as well, and may make up for some of the discrepancy between the two isolates. Although it is difficult to determine synteny when

comparing to a largely fragmented genome, there is a strong correlation between the two assemblies (Figure 2). It is likely that the true genome size of RO10H11247 is larger than the reported 99.6 Mb due to collapsing of repeat regions, but it is difficult to conclude from the available data whether RO10H11247 is a similar size to IA16 or if IA16 has undergone significant repeat expansion.

When specifically comparing predicted protein-coding genes, the haploid assembly of IA16 has fewer genes than the published RO10H11247 assembly, but more predicted secreted and effector proteins. We also repeated the annotation process with the Funannotate pipeline for the RO10H11247 assembly to have a direct comparison, but this strategy resulted in significantly fewer genes than either the IA16 assembly or the previous report. We believe this is an artifact of the high proportion of unknown bases in the RO10H11247 genome paired with genome-guided annotation, as many of the RO10H11247 genes predicted with Funannotate spanned stretches of unknown bases, artificially increasing gene size. Both the funannotate annotations and the original annotations for RO10H11247 resulted in fewer predicted secreted and effector proteins, despite the original having more overall annotations. The reduction in predicted secreted and effector proteins likely stems from incomplete gene models, as missing N-terminal sequence, where secretion signals are located, would erroneously label secreted proteins as non-secreted, and thus as non-effector proteins.

Currently, no Avr gene has been identified in *P. sorghi* for any of the described *Rp* genes in maize. The sequencing of this additional *P. sorghi* isolate may aid in the search for Avr genes, particularly in isolates with differing virulence phenotypes on the same *Rp* lines. For example, the IA16 isolate is avirulent on *RpG* and virulent on *Rp1-D* while IN2, an isolate from Indiana, has the opposite phenotype (Richter et al. 1995).

Conclusions

As high molecular weight DNA extraction methods, sequencing technologies, and computational pipelines and strategies continue to develop, we expect more repeat-resolved genomes for both additional *P. sorghi* isolates and other Pucciniales species to be assembled. Here we showed that even minimal long-read sequencing can be used to generate a contiguous and relatively complete *P. sorghi* genome. We found that the IA16 isolate contains significantly more repeats than previously reported for *P. sorghi*, with the majority being LTR retroelements, similar to other rust fungal species with high percentages of repeat regions (Liang et al. 2022; Tobias et al. 2021), with a total haploid genome size of 174 Mb contained within 902 scaffolds. The pseudohaplotypes were of a similar size, each with ~170 Mb contained within 1,277 or 1,262 scaffolds. Using IA16 specific whole transcriptome data, we were able to annotate 16,336 protein coding genes on the haploid assembly, of which 742 are predicted effectors. This assembly appears to be relatively complete and provides both another resource for *P. sorghi* and Pucciniales, as well as a new resource for IA16 and similar Midwestern isolates or races.

Acknowledgements

We thank Scot Hulbert for the H95 and *Rp* lines, Peter Balint-Kurti for the *Rp1-D21* lines, and Nick Lauter for the B104, B73, and W22 lines. This study was supported by the Iowa State University Predictive Plant Phenomics graduate training program funded by the National Science Foundation (DGE #1545453) and by Agricultural and Food Research Initiative grant no. 2019-07318 from the USDA National Institute of Food and Agriculture. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders. We also

received support from the Plant Sciences Institute at Iowa State University, the Lois H. Tiffany Scholarship from Iowa State University, and the Gilman Scholarship from Iowa State University.

We would additionally like to thank Andrew Severin for advice regarding genome assembly and analysis, Phase Genomics and Ivan Liachko for advice on Hi-C and generation of the Hi-C libraries and sequencing, and the Iowa State University DNA Facility, specifically Tanya Murtha and Michael Baker for Nanopore library preparation and sequencing.

Author Contributions

ME collected and isolated the IA16 *P. sorghi* isolate. KH and SW designed the sequencing plan. ME generated the RNAseq data. KH assembled and annotated the genome and conducted all analyses. KH wrote the manuscript and SW edited the manuscript.

References

Aime, M. C., McTaggart, A. R., Mondo, S. J., and Duplessis, S. 2017. Phylogenetics and Phylogenomics of Rust Fungi. In *Advances in Genetics*, Academic Press, p. 267–307..

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Van Wezel, G. P., Medema, M. H., et al. 2021. antiSMASH 6.0: Improving cluster detection and comparison capabilities. Nucleic Acids Res. 49:W29–W35.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49:D344–D354.

Cabanettes, F., and Klopp, C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 2018:e4958.

Cantalapiedra, C. P., Herņandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol. Biol. Evol. 38:5825–5829.

Chavan, S., Gray, J., and Smith, S. M. 2015. Diversity and evolution of *Rp1* rust resistance genes in four maize lines. Theor. Appl. Genet. 128:985–998.

Chen, S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2:e107.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. fastp: An ultra-fast all-in-one FASTQ preprocessor. In *Bioinformatics*, Oxford Academic, p. i884–i890..

Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods. 13:1050–1054.

Cuomo, C. A., Bakkeren, G., Khalil, H. B., Panwar, V., Joly, D., Linning, R., et al. 2017. Comparative Analysis Highlights Variable Genome Content of Wheat Rusts and Divergence of the Mating Loci. G3 Genes, Genomes, Genet. 7:361–376.

Darino, M. A., Rochi, L., Lia, V. V., Kreff, E. D., Pergolesi, M. F., Ingala, L. R., et al. 2016. Virulence characterization and identification of Maize lines resistant to *Puccinia sorghi* Schwein. Present in the Argentine Corn Belt region. Plant Dis. 100:770–776.

Delaney, D. E., Webb, C. A., and Hulbert, S. H. 1998. A Novel Rust Resistance Gene in Maize Showing Overdominance. Mol. Plant-Microbe Interact. 11:242–245.

Duan, H., Jones, A. W., Hewitt, T., Mackenzie, A., Hu, Y., Sharp, A., et al. 2022. Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. Genome Biol. 23:1–27.

Duplessis, S., Bakkeren, G., and Hamelin, R. 2014. Advancing knowledge on biology of rust fungi through genomics. In *Advances in Botanical Research*, Academic Press, p. 173–209.

Emms, D. M., and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:1–14.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117:9451–9457.

Gel, B., and Serra, E. 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. 33:3088–3090.

Ghurye, J., Pop, M., Koren, S., Bickhart, D., and Chin, C.-S. 2017. Scaffolding of long-read assemblies using long range contact information. BMC Genomics. 18:1–11.

Ghurye, J., Rhie, A., Walenz, B. P., Schmitt, A., Selvaraj, S., Pop, M., et al. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLOS Comput. Biol. 15:e1007273.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Hagan, W. L., and Hooker, A. L. 1965. Genetics of Reaction to *Puccinia sorghi* in Eleven Corn Inbred Lines from Central and South America. Phytopathology. 55:193–197.

Henningsen, E. C., Hewitt, T., Dugyala, S., Nazareno, E. S., Gilbert, E., Li, F., et al. 2022. A chromosome-level, fully phased genome assembly of the oat crown rust fungus *Puccinia coronata* f. sp. *avenae*: a resource to enable comparative genomics in the cereal rusts. G3 Genes, Genomes, Genet. 12.

Hooker, A. L. 1985. Corn and Sorghum Rusts. In *Diseases, Distribution, Epidemiology, and Control*, Academic Press, p. 207–236.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47:D309–D314.

Hulbert, S. H. 1991. Reactions of Maize Lines Carrying *Rp* Resistance Genes to Isolates of the Common Rust Pathogen, *Puccinia sorghi*. Plant Dis. 75:1130.

Hulbert, S. H. 1997. Structure and evolution of the *Rp1* complex conferring rust resistance in maize. Annu. Rev. Phytopathol. 35:293–310.

Jones, A., Naga, R., Sharp, A., and Schwessinger, B. 2019. High-molecular weight DNA extraction from challenging fungi using CTAB and gel purification. protocols.io.

Jones, A., and Schwessinger, B. 2021. High-molecular weight DNA extraction from challenging fungi using CTAB for lysis and precipitation. protocols.io.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236–1240.

Käll, L., Krogh, A., and Sonnhammer, E. L. L. 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338:1027–1036.

Käll, L., Krogh, A., and Sonnhammer, E. L. L. 2007. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. Nucleic Acids Res. 35:W429–W432.

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37:907–915.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37:540–546.

Li, C., Qiao, L., Lu, Y., Xing, G., Wang, X., Zhang, G., et al. 2023. Gapless Genome Assembly of *Puccinia triticina* Provides Insights into Chromosome Evolution in Pucciniales. Microbiol. Spectr. Evol. 11.

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34:3094–3100.

Liang, J., Li, Y., Dodds, P. N., Figueroa, M., Sperschneider, J., Han, S., et al. 2022. Haplotype-phased and chromosome-level genome assembly of *Puccinia polysora*, a giga-scale fungal pathogen causing southern corn rust. Mol. Ecol. Resour. 23:601–620.

Lin, Y., Yuan, J., Kolmogorov, M., Shen, M. W., Chaisson, M., and Pevzner, P. A. 2016. Assembly of long error-prone reads using de Bruijn graphs. Proc. Natl. Acad. Sci. U. S. A. 113:E8396–E8405.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. 38:4647–4654.

Miller, M. E., Zhang, Y., Omidvar, V., Sperschneider, J., Schwessinger, B., Raley, C., et al. 2018. *De novo* Assembly and Phasing of Dikaryotic Genomes from Two Isolates of *Puccinia coronata* f. sp. *avenae*, the Causal Agent of Oat Crown Rust. MBio. 9:e01650-17.

Nandety, R. S., Gill, U. S., Krom, N., Dai, X., Dong, Y., Zhao, P. X., et al. 2022. Comparative Genome Analyses of Plant Rust Pathogen Genomes Reveal a Confluence of Pathogenicity Factors to Quell Host Plant Defense Responses. Plants. 11:1962.

Palmer, J. M., and Stajich, J. E. 2022. Funannotate. Available at: https://github.com/nextgenusfs/funannotate.

Pretorius, Z. A., Booysen, G. J., Boshoff, W. H. P., Joubert, J. H., Maree, G. J., and Els, J. 2019. Additive Manufacturing of Devices Used for Collection and Application of Cereal Rust Urediniospores. Front. Plant Sci. 10:639.

Quade, A., Ash, G. J., Park, R. F., and Stodart, B. 2021. Resistance in Maize (*Zea mays*) to Isolates of *Puccinia sorghi* from Eastern Australia. Phytopathology. 111:1751–1757.

Ramirez-Cabral, N. Y. Z., Kumar, L., and Shabani, F. 2017. Global risk levels for corn rusts (*Puccinia sorghi* and *Puccinia polysora*) under climate change projections. J. Phytopathol. 165:563–574.

Ramos, A. P., Tavares, S., Tavares, D., Silva, M. D. C., Loureiro, J., and Talhinhas, P. 2015. Flow cytometry reveals that the rust fungus, *Uromyces bidentis* (Pucciniales), possesses the largest fungal genome reported-2489Mbp. Mol. Plant Pathol. 16:1006–1010.

Richter, T. E., Pryor, T. J., Bennetzen, J. L., and Hulbert, S. H. 1995. New Rust Resistance Specificities Associated with Recombination in the *Rp1* complex in Maize. Genetics. 141:373–81.

Rochi, L., Diéguez, M. J., Burguener, G., Darino, M. A., Pergolesi, M. F., Ingala, L. R., et al. 2018. Characterization and comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize common rust. Fungal Genet. Biol. 112:31–39.

Schwessinger, B. 2019. High quality DNA from Fungi for long-read sequencing e.g. PacBio. protocols.io.

Schwessinger, B., and Rathjen, J. P. 2017. Extraction of high molecular weight DNA from fungal rust spores for long-read sequencing. In *Methods in Molecular Biology*, Humana Press, New York, NY, p. 49–57..

Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., et al. 2018. A Near-Complete Haplotype-Phased Genome of the Dikaryotic Wheat Stripe Rust Fungus *Puccinia striiformis* f. sp. *tritici* Reveals High Interhaplotype Diversity. MBio. 9:e02275-17.

Shafin, K., Pesout, T., Chang, P.-C., Nattestad, M., Kolesnikov, A., Goel, S., et al. 2021. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. Nat. Methods. 18:1322–1332.

Smit, A., Hubley, R., and Green, P. 2015. RepeatMasker Open-4.0..

Sperschneider, J. 2021. ContaminantScreening. Available at: https://github.com/JanaSperschneider/GenomeAssemblyTools/tree/master/ContaminantScreening.

Sperschneider, J. 2022. NuclearPhaser. Available at: https://github.com/JanaSperschneider/NuclearPhaser.git.

Sperschneider, J., and Dodds, P. N. 2022. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. Mol. Plant-Microbe Interact. 35:146–156.

Tavares, S., Ramos, A. P., Pires, A. S., Azinheira, H. G., Caldeirinha, P., Link, T., et al. 2014. Genome size analyses of Pucciniales reveal the largest fungal genomes. Front. Plant Sci. 5:422.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 40:1023–1025.

Tobias, P. A., Schwessinger, B., Deng, C. H., Wu, C., Dong, C., Sperschneider, J., et al. 2021. *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. G3 Genes, Genomes, Genet. 11.

Vasquez-Gross, H., Kaur, S., E*Pst*ein, L., and Dubcovsky, J. 2020. A haplotype-phased genome of wheat stripe rust pathogen *Puccinia striiformis* f. sp. *tritici*, race *PST*-130 from the Western USA. PLoS One. 15:e0238611.

Xia, C., Qiu, A., Wang, M., Liu, T., Chen, W., and Chen, X. 2022. Current Status and Future Perspectives of Genomics Research in the Rust Fungi. Int. J. Mol. Sci. 23.

Figures

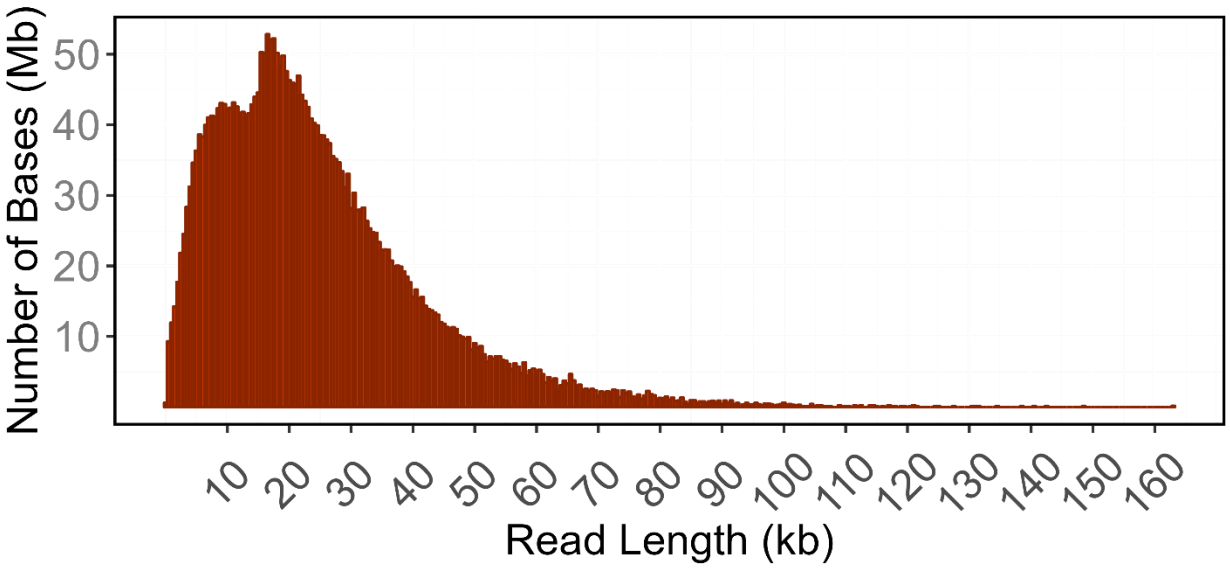


Figure 1. Weighted histogram of read lengths obtained from the output from three Nanopore flow cells. Each bar indicates total bases contained within reads of a given length. Bin size is 500 bp.

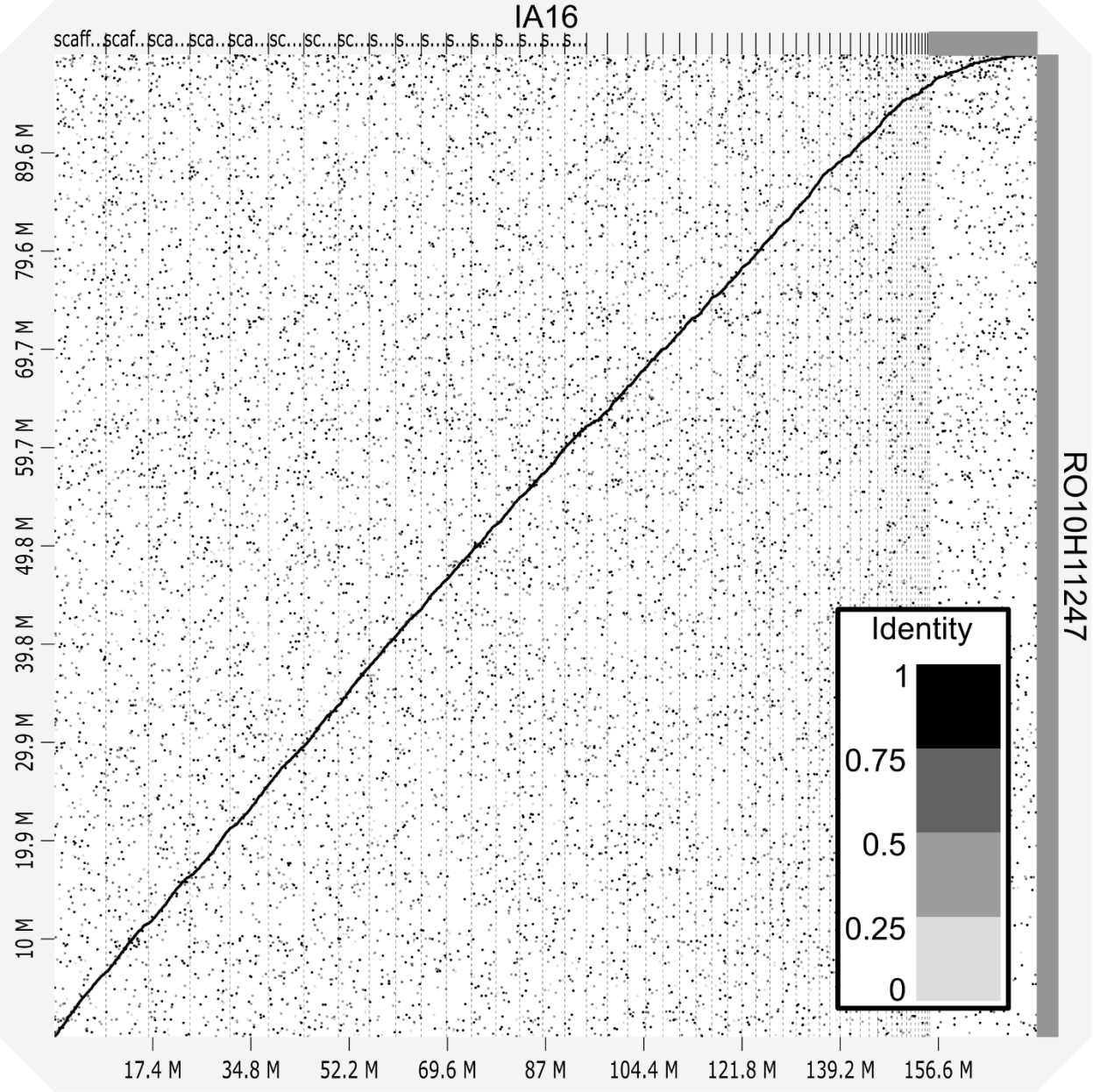


Figure 2. A D-GENIES dot plot showing mapping matches between the haploid IA16 assembly and the RO10H11247 assembly. M indicates millions of bases.

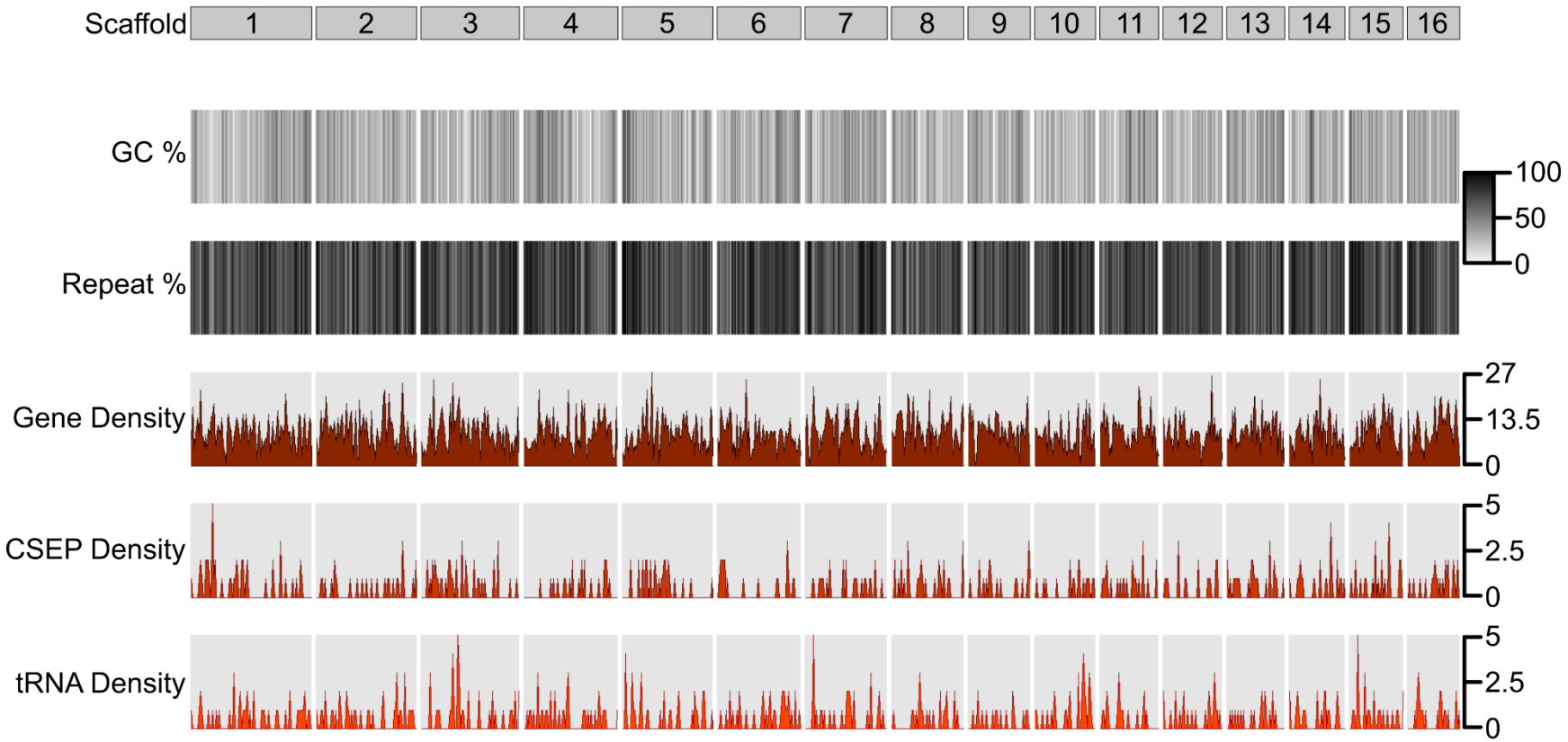


Figure 3. The genomic features of the largest 16 scaffolds of the haploid assembly of IA16. Each track is in vertical alignment with its respective scaffold at the top of the figure, and all scaffold representations are proportional to their size. Feature content was calculated in 100 kb non-overlapping sliding windows. The tracks are of GC density and repeat density as the percentage of sequence and gene density, CSEP density, and tRNA density as the number of features per 100 kb.

Tables

Table 1. Virulence phenotype of the IA16 *P. sorghi* isolate on various maize genotypes and *Rp* resistance genes and alleles. Maize lines were inoculated with IA16 urediniospores and scored 7-10 days later. "+" indicates virulence, "-" indicates avirulence, and "I" indicates an intermediate or indeterminable virulence. *Rp1-D21* is an autoactive mutant of *Rp1-D.*

Background	*Rp* Gene	Virulence Phenotype
H95	-	+
B104	-	+
B73	-	+
W22 bz1-mum9	-	+
Sweet Corn (Golden Bantam)	-	+
H95	*Rp1-A*	+
H95	*Rp1-B*	+
H95	*Rp1-C*	+
H95	*Rp1-I*	-
H95	*Rp1-J*	+
H95	*Rp1-Kn1*	+
H95	*Rp1-M*	+
H95	*Rp3-A*	+
H95	*Rp4-A*	+
H95	*Rp4-B*	I
H95	*Rp5*	I
H95	*Rp7*	+
H95	*RpG*	-
H95	*RpGA*	-
H95	*Rp1-D*	+
A632	*Rp1-D21*	+
B73	*Rp1-D21*	+
H95	*Rp1-D21*	+

Table 2. Genome assembly metrics for all three IA16 assemblies plus the previously reported RO10H11247 assembly. The BUSCO database basidiomycete_odb10 (1,764 total members) was used to calculate BUSCO scores for all four assemblies.

Assembly	Haploid	Haplotype A	Haplotype B	RO10H11247
Contigs	1444	2757	2763	28117
Scaffolds	902	1277	1262	15715
Total Length (Mb)	174.05	170.45	170.59	99.53
Mean Length Scaffold (kb)	192.96	133.48	135.18	6.33
N50 (Mb)	3.93	1.56	1.83	0.019
L50	16	32	26	1530
N90 (kb)	200.11	63.70	62.79	3.84
L90	57	197	190	5788
GC Content	45.18%	44.98%	44.97%	43.15%
All BUSCOs	84.64%	73.53%	72.17%	88.49%
Complete BUSCOs	79.59%	63.32%	62.59%	85.37%
Complete	1404	1117	1104	1506
Complete and Single Copy	1376	1091	1075	1497
Complete and Duplicated	28	26	29	9
Fragmented	89	180	169	55
Missing	271	467	491	203

Table 3. The basidiomycete_odb10 BUSCO results for the combined haplotype A and haplotype B assemblies of IA16. The basidiomycete_odb10 database contains 1,764 members.

Assembly	A + B
All BUSCOs	80.16%
Complete BUSCOs	75.40%
Complete	1330
Complete and Single Copy	818
Complete and Duplicated	512
Fragmented	84
Missing	350

Table 4. Repeat coverage results for the three IA16 assemblies, plus recalculated statistics for the RO10H11247 assembly. Percentages shown are of total assembly sequence. Ty1/Copia and Ty3/DIRS1 represent the majority of annotated retroelements within each assembly. *This metric does not include the 29% of unknown sequence within the RO10H11247 assembly and true values are expected to be higher.

	Haploid	Haplotype A	Haplotype B	RO10H11247
Total Repeats	76.38%	76.34%	76.51%	34.10%
Repeat Coverage (Mb)	132.94	130.12	130.52	33.94*
Non-Repeat Coverage (Mb)	41.11	40.33	40.07	36.96*
Retroelements	37.33%	36.22%	37.18%	7.48%
Ty1/Copia	6.51%	6.00%	6.03%	3.17%
Ty3/DIRS1	30.50%	29.81%	30.42%	3.92%
DNA transposons	9.67%	8.98%	9.07%	2.68%
Unclassified	27.73%	29.79%	28.26%	22.93%

Table 5. Annotation metrics for the three IA16 assemblies and the RO10H11247 assembly. Genome assemblies were annotated using funannotate and the IA16 ESTs. The previously reported RO10H11247 annotation metrics are shown for comparative purposes. Predicted secreted and candidate secreted effector protein results were recalculated with the updated programs for the previously reported RO10H11247 predicted proteins. "fun." indicates funannotate results and "PR" indicates results from the previously reported annotations. *Value is from previously reported metrics from Rochi, et al 2018.

Assembly	Haploid	Haplotype A	Haplotype B	RO10H11247 (fun.)	RO10H11247 (PR)
Genes	16336	19487	19432	10992	21087*
Average Gene Length (kb)	1277	838	801	1757	1062*
Gene Coverage (Mb)	20.85	16.33	15.6	19.32	22.39*
% of Assembly Covered by Genes	11.98%	9.58%	9.13%	19.41%	22.50%*
tRNAs	839	829	835	419	405*
CDS Transcripts	16458	19035	18923	11908	-
SignalP 6.0 Predicted Secretion Proteins	1128	803	728	957	753
Phobius Predicted Secretion Proteins	1775	1580	1471	1414	1609
Transmembrane Proteins	2552	2390	2378	2223	5399
Total Unique Predicted Secretion Proteins	1845	1635	1511	1471	1690
Predicted Effectors	742	655	616	563	615
Predicted Cytoplasmic Effectors	527	498	456	393	449
Predicted Apoplastic Effectors	215	157	160	170	166
Cytoplasmic Effectors Predicted Dual-Localized	10.40%	7.40%	7.00%	9.90%	7.60%
Apoplastic Effectors Predicted Dual-Localized	15.30%	13.40%	16.90%	17.10%	9.00%

Supplemental Tables

Supplemental Table 1. Breakdown of tRNAs identified within the three IA16 assemblies, plus funannotate metrics for the RO10H11247 assembly.

tRNA	Haploid	Haplotype A	Haplotype B	RO10H11247 (fun.)
Ala	43	42	39	27
Arg	15	18	16	12
Asn	13	15	14	8
Asp	8	9	9	7
Cys	2	4	5	3
Gln	10	7	8	6
Glu	30	29	27	10
Gly	45	53	47	19
His	14	16	15	11
Ile	17	21	16	12
iMet	3	4	3	2
Leu	31	32	30	17
Lys	10	13	11	7
Met	8	7	10	7
Phe	6	6	5	4
Pro	95	91	97	43
Ser	256	250	269	128
Thr	214	193	194	78
Trp	4	3	3	2
Tyr	3	4	6	6
Val	12	12	11	10
Total	839	829	835	419

Supplemental Table 2. OrthoFinder orthogroup results for comparisons between each assembly's annotations. Each orthogroup may include orthologues and paralogs and includes all genes that descend from a gene within the last common ancestor. All predicted protein sequences for each assembly were compared. "fun." indicates funannotate results and "PR" indicates results from previously reported annotations.

	Haplotype A	Haplotype B	Haploid	RO10H11247 (fun.)	Haploid	RO10H11247 (PR)	RO10H11247 (fun.)	RO10H11247 (PR)
All Genes								
Total Genes	37958		28366		37536		32986	
Total Orthogroups	11349		8809		8564		8237	
Genes	19035	18923	16458	11908	16458	21078	11908	21078
Genes in Orthogroups	14082	13855	12605	10713	11913	19694	10277	19835
Unassigned Genes	4953	5068	3853	1195	4545	1384	1631	1243
Assembly-Specific Orthogroups	287	211	317	108	461	514	135	431
Genes in Assembly-Specific Orthogroups	833	632	1034	332	1605	4364	383	4243
Predicted Effectors								
Total Genes	1271		1305		1357		1178	
Total Orthogroups	226		379		289		292	
Genes	655	616	742	563	742	615	563	615
Genes in Orthogroups	321	302	531	448	439	374	369	389
Unassigned Genes	334	314	208	115	303	241	194	226
Assembly-Specific Orthogroups	11	9	17	7	46	40	30	40
Genes in Assembly-Specific Orthogroups	29	22	47	15	155	121	74	122

CHAPTER 3. IDENTIFICATION AND CHARACTERIZATION OF CANDIDATE SECRETED EFFECTOR PROTEINS OF *PUCCINIA SORGHI*

Katerina L. Holan[1], Manjula Elmore[2], and Steven A. Whitham[1]

[1]Department of Plant Pathology, Entomology, and Microbiology, Iowa State University, Ames, IA

[2]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Modified from a manuscript to be submitted to *PhytoFrontiers*

Abstract

Rust fungi, such as *Puccinia sorghi* (common rust of maize), secrete hundreds of effector proteins during their colonization of host plant tissues, altering immune responses and interfering with nutrient trafficking. Due to long candidate lists and poor predicted functional annotations, few rust effectors have been fully characterized. Rust genomes encode several members of a rust protein family named cluster 112, which are small proteins with secretion signal peptides and a distinctive 10-cysteine motif. At least one family member from the Asian soybean rust fungus, *Phakopsora pachyrhizi*, has been shown to be involved in plant immune suppression. Here, we characterize eight predicted *Pp*EC23 homologs from an Iowan isolate of *P. sorghi* in plant immune response assays. To increase throughput of immune assay experiments, we utilized an inexpensive time lapse phenotyping setup to acquire multiple images during the course of a single experiment. Our data show that one of the *Pp*EC23 homologs in *P. sorghi*, 930g11, was able to suppress hypersensitive immune response, but does not seem to have a function in suppressing basal immunity.

Introduction

Fungal rust species (Pucciniales) remain consistent and impactful plant pathogens worldwide, causing significant yield losses on many crop species (Figueroa et al. 2023). The

study of the molecular mechanisms involved in rust diseases is imperative to better understand how these fungi manipulate their host plants. However, many research approaches available to other pathosystems are not available for rust fungi. Rust fungi are obligate biotrophs that each have highly specific host ranges, resulting in an inability to be cultured on media or non-host plant species. Basic information and genomic resources regarding rust fungal species are often lacking due to their complicated life cycles and relatively large, repeat-rich genomes.

Much of the current research on rust fungi focuses on characterization of candidate secreted effector proteins (CSEPs) (Lorrain et al. 2019), which are secreted from specialized structures call haustoria. Haustoria form close relationships to plasma membranes of host cells, allowing for the transfer of nutrients, signaling molecules, and effector proteins (Garnica et al. 2014). Effector proteins are small, secreted proteins that influence and alter host cell processes to aid in fungal colonization and proliferation (Uhse and Djamei 2018; Figueroa et al. 2021). These proteins have a wide variety of functions, from modulation of plant immune responses to trafficking of nutrients. Although many rust effector candidates lack functional annotation, features such as a small size, cysteine-rich, and the presence of an N-terminal secretion peptide are commonly used to predict CSEPs within rust genomes. However, many of the effectors are encoded by orphan genes or are rust-specific, making identification difficult. As genomic and transcriptomic data becomes available for a given rust species, they are mined for these potentially unique effector candidates. A typical strategy identifies proteins with predicted secretion signals with SignalP (Teufel et al. 2022) followed by effector prediction with EffectorP (Sperschneider and Dodds 2022). Such analyses identify a few hundred to 1,000 or more effector candidates, many of which have redundant functions.

One common characterization method for confirming rust effectors involves prioritizing candidates, often through homology to CSEPs from other species, and then conducting assays to identify recognition by host resistance proteins (Avirulence effectors) or effector-like functions such as immune suppression (de Carvalho et al. 2017; Ramachandran et al. 2017; Liu et al. 2016). These assays are typically conducted *in planta* in heterologous systems, such as Arabidopsis or *Nicotiana benthamiana*, to identify potential function, localization, and interacting host proteins (Lorrain et al. 2018). One effector screen identified 156 predicted CSEPs from *Phakopsora pachyrhizi* (Asian soybean rust) based on haustorial expression and predicted secretion signal peptides (Link et al. 2014). Eighty-two of these candidates were cloned, and various experiments were conducted to identify localization and impact on both pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI) (Qi et al. 2018). The success rate for identifying effectors shown to influence phenotype is often low, even for studies that look at 10s to 100s of effectors (Lorrain et al. 2019). Given this low success rate, the throughput of screens for effector functions needs to be increased. The utilization of high-throughput phenotyping methods may increase the number of effectors that can be characterized in a given experiment, reduce the variability between experiments, and allow for more flexible timing during assays. One example of such a system involves a platform for imaging the development of *Sclerotinia sclerotiorum* on detached *Arabidopsis thaliana* leaves and subsequent automatic phenotyping of those images, dubbed Navautron (Barbacci et al. 2020). Although not specifically utilized for effector characterization immune suppression assays, this system can be easily modified to suit other use cases.

In *Puccinia sorghi* (common rust of maize), no CSEP effector screens have been published to date, despite the availability of nearly 1,600 putative effectors (Rochi et al. 2018).

There are several families of proteins previously identified in various rust genomes, including the rust specific cluster 112 family of CSEPs identified from *P. pachyrhizi* and *Uromyces appendiculatus* (common bean rust) (Link et al. 2014). The cluster 112 proteins contain a 94-amino acid motif with 10 conserved cysteine residues, which are particularly important for the formation of disulfide bridges during protein folding (Zhang et al. 2017; Wiedemann et al. 2020). At least one cluster 112 member, *Pp*EC23 from *P. pachyrhizi*, which contains two tandem 10-cysteine motifs, has been characterized and shown to both suppress hypersensitive response (HR) involved in ETI and interact with the soybean SQUAMOSA promotor binding protein-like (SPL) transcription factor GmSPL121 (Qi et al. 2016). From the predicted proteins identified by Rochi et al. (2018), we have identified 11 *P. sorghi* CSEPs (isolate IA16) with homology to *Pp*EC23. To characterize these homologs, we conducted immune assays in *N. benthamiana* and built a phenotyping box similar to the cabinet in the Navautron system (Barbacci et al. 2020) to acquire time lapse images of HR immune assays. The use of a phenotyping box with a Raspberry Pi operated camera allowed us to capture time lapse images of up to 96 leaves at once during immune assay experiments. Of the eight *P. sorghi* CSEPs we were able to amplify from cDNA, one, namely 930g11, was found to have a small, suppressive effect on HR.

Materials and Methods

Identification of *Pp*EC23 homolog targets in *P. sorghi*

To identify *Pp*EC23 homologs in *P. sorghi*, a BLASTP search (E-value threshold 5e-02) was conducted using the amino acid sequence of *Pp*EC23 (Qi et al. 2016). The predicted CDSs were downloaded from the genome annotation information available for the *P. sorghi* isolate RO10H11247 (Rochi et al. 2018).

Amplification and cloning of *Pp*EC23 homologs from the IA16 *P. sorghi* isolate

Maize leaf samples were collected seven days after inoculation with *P. sorghi* isolate IA16 and immediately frozen in liquid nitrogen. Using the TRIzol® method, RNA was extracted from these samples. cDNA was generated using oligodT primers and Superscript III reverse transcriptase, according to the manufacturer's protocol. SignalP 3.0 (Bendtsen et al. 2004) and 4.1 (Petersen et al. 2011) were used to predict secretion signals in the available CDS sequence for each candidate. To amplify the coding sequences minus the signal peptides (CDS_{ns}) from the cDNA, specific primers for each candidate were designed using the previously identified sequence (Supplemental Table 1). Each PCR product was cloned via a TOPO reaction into pCR8 and transformed into *E. coli* One Shot® TOP 10 competent cells. Colonies were confirmed by sequencing and plasmids were extracted via miniprep from *E. coli* cultures. A Gateway™ LR Clonase II™ reaction was performed according to the manufacturer's protocol to recombine each CSEP CDS_{ns} from pCR8 into the Effector Detector Vector pEDV6, and the resultant plasmids were transformed into DH5α *E. coli* competent cells. Colonies were sequenced again to confirm the CDS and ensure it was in-frame with the AvrRps4 type III secretion signal on the pEDV6 plasmid. Triparental mating was conducted to conjugate each pEDV6-CDS_{ns} plasmid into *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000, using *E. coli* carrying pRK2013 as the helper strain. Mating was confirmed via selective plating on LB plates containing gentamicin and rifampin and by conducting colony PCR. The 10-cysteine motif alignment of each cloned candidate was performed in Clustal Omega (McWilliam et al. 2013; Goujon et al. 2010; Sievers et al. 2011) and visualization of residue conservation was executed in Jalview 2.11 (Waterhouse et al. 2009). A *Pst* DC3000 pEDV6-GFP strain was created similarly.

Pst DC3000 pEDV6-CSEP$_{ns}$ HR immune assays

The *Pst* DC3000 pEDV6-CDS$_{ns}$ strains and the *Pst* DC3000 pEDV6-GFP strain were grown overnight at 28˚C in liquid LB media containing gentamycin and rifampin with shaking ~200 RPM. The bacteria were pelleted for 10 minutes at 4,000 RPM and washed once with ddH_2O. The bacteria were pelleted again, and the pellet was resuspended in 10 mM $MgCl_2$ and diluted to an OD_{600} of 0.2, 0.02, and 0.004. Each bacterial suspension was infiltrated into the underside of 4-5 week old *N. benthamiana* leaves with a needleless syringe. One half of each leaf was infiltrated with pEDV6-GFP at each OD_{600} concentration while the other half was infiltrated with pEDV6-CSEP$_{ns}$ at each OD_{600} concentration, resulting in six total infiltration spots per leaf. The infiltrated regions on each leaf were dried, marked with a permanent marker, and assessed for HR at 18 to 22 hours post infiltration (HPI) on a scale of 0-3, with 0 being no HR and 3 being complete HR.

N. benthamiana transformation with pBI121-3XFLAG-930g11

Primers designed to add a 3XFLAG tag to the 5’ end of the 930g11$_{ns}$ CDS were used to amplify the previously cloned pEDV6-930g11$_{ns}$. A forward primer containing a BamHI restriction site, a 3XFLAG tag, and partial homology to the attB1 site and a reverse primer with homology to 930g11, a stop codon, and a SacI restriction site were used to PCR amplify the 930g11$_{ns}$ CDS from the pEDV6-930g11$_{ns}$ vector. Both pBI121, a binary Agrobacterium transformation vector, and the PCR product were digested with BamHI and SacI restriction enzymes (New England Biolabs) and ligated together with T4 DNA ligase (New England Biolabs), according to the manufacturer’s protocol. DH5α *E. coli* was transformed with the resulting ligation reaction and miniprepped to isolate the plasmid. After confirmation via sequencing, the dried plasmid was sent to the University of Nebraska-Lincoln Plant Transformation Core Research Facility to be transformed into *N. benthamiana*.

Rooted plants (T0) were received and transplanted into peat-based substrate, covered with a clear plastic dome, and placed in a growth room with 16 hour days for two weeks. Over the next two weeks, the plastic dome was gradually removed. When plants were fully hardened, they were transplanted to larger pots and moved to a greenhouse. Pots were watered with Peter's fertilizer once a week. A western blot was conducted to confirm transgene expression and protein accumulation. Briefly, total protein was extracted from T0 plants and run on a Criterion™ TGX™ precast gel (BioRad) in SDS-PAGE buffer. The proteins were transferred to PVDF membrane via wet transfer and blocked with 5% milk/TBST prior to blocking with anti-FLAG-HRP antibody. The membrane was incubated with luminol and imaged. After imaging, the membrane was stained with Coomassie brilliant blue (CBB) to visualize all transferred protein. T1 seed from the six lines with the strongest $930g11_{ns}$ expression was planted, and seedlings were genotyped for the transgene by CTAB extraction of genomic DNA (gDNA) and amplification of the transgene using $930g11_{ns}$ specific primers. One plant without transgene expression for each line was saved to serve as an azygous negative control. T2 seed was collected from these plants. T2 seed from each transgenic and azygous plant was plated on ½ MS media plus kanamycin at 50 µg/µL to test for transgene presence and estimate gene copy number. Four lines and their respective azygous counterparts were used in the experiments moving forward. All transgenic and azygous plants in each experiment were genotyped as previously described.

Time lapse phenotyping box setup

The phenotyping boxes made for this project were based on the Navautron system (Barbacci et al. 2020). More information on the phenotyping box construction used in this manuscript is available at https://github.com/katholan/timelaspe_phenotyping_boxes.git. As an overview, two plastic storage bins were used as the main cabinet, where a taller bin composed

the bottom of the cabinet and a shorter bin of the same opening size served as the top. A small hole was cut into the roof of the box to mount a Raspberry Pi-operated 70˚ field of view (FOV) Arducam M12 low distortion wide-angle camera lens. The low distortion lenses cut down substantially on fisheye effect. White LED strip lights were mounted to the top of the phenotyping box and remained on during the entire course of each experiment. The Raspberry Pi was operated from a laptop with a headless connection. Experiments were run in a room with minimal external light.

Time lapse immune assay experiments

Pst DC3000 was grown overnight at 30˚C in liquid LB culture with rifampin with shaking at ~200 RPM. The bacterial culture was rinsed once with ddH_2O and resuspended in 10 mM $MgCl_2$ to an OD_{600} of 0.02. Four to five week old *N. benthamiana* plants that were approximately the same size were chosen for each experiment. One side of leaf 3, when counting down from the top of the plant, was infiltrated with 10 mM $MgCl_2$ buffer and the other side was infiltrated with the previously described *Pst* DC3000 bacterial suspension. After infiltration, the leaves were blotted with paper towels and then allowed to dry for approximately one to two hours. A layer of paper towels was laid in the bottom of the phenotyping box, and a small amount of water was added to the paper towels so that they were fully soaked but there was minimal free-standing water. Three layers of plastic mesh were laid on top of the paper towels to prevent the leaves from directly touching the water, as direct contact will prevent the infiltration spots from fully drying out, thus confounding results. After drying, leaves were cut and laid adaxial side down onto the mesh as flat as possible. The abaxial side is lighter in color, and contrast between HR symptoms and the rest of the leaf underside is higher, making it easier to phenotype from images. The camera lens was focused manually, and the “lid” was secured to the bottom box with twist ties to prevent camera movement. The ‘camera.py’ script from the Python

package PiCamera (https://picamera.readthedocs.io/en/release-1.13/) was used to take photos of the leaves every hour, starting at 0 hours, for two days.

After each experiment, photos were copied to a GoogleDrive folder. Leaf regions for a single image were acquired with ImageJ/Fiji (Schindelin et al. 2012). These were used to extract a particular leaf from each photo in the time lapse. The resultant cropped images were rearranged into a grid with 0 HPI at the upper left and the latest HPI (47 HPI) in the lower right for phenotyping. Infiltration locations for each leaf at each timepoint were scored on a scale of 0-7, with 0 being no discernable HR and 7 being complete HR. The first few hours of all experiments were discarded, as many of the infiltration were not dry yet at this point. Leaves with infiltration or buffer control sites that never fully dried were discarded. All plants in each experiment, regardless of background, were genotyped for the presence of the transgene via CTAB genomic DNA extraction and subsequent PCR amplification to determine transgene presence.

ROS burst assays

Transgenic and azygous *N. benthamiana* plants were grown until 4-5 weeks of age. Reactive oxygen species (ROS) burst assays were conducted to determine ROS burst response to a PAMP (Bredow et al. 2019). Briefly, the second leaf of individual plants was sampled with a 4 mm biopsy punch and added to a 96-well plate containing 100 µL of ddH_2O, adaxial side up. The plate was kept in the dark overnight at room temperature. The ddH_2O was replaced with 100 µL of an assay solution containing 100 mM luminol, 10 µg/mL HRP, and 100 nM flagellin-22 (flg22). The plates were immediately placed into a GloMax® microplate reader and light emission data was gathered for 30 cycles, with two minutes per cycle. The experiment was repeated twice.

Results

*Pp*EC23 homologs in the *P. sorghi* isolate IA16

Using BLASTP and the protein sequence of PpEC23 as the query, eleven candidates in *P. sorghi* were identified in the RO10H11247 genome assembly (Link et al. 2014; Qi et al. 2016; Rochi et al. 2018). All eleven *P. sorghi* genes have largely uninformative predicted annotations, namely “hypothetical protein”, “uncharacterized protein”, or “putative signal peptide protein”. The percentage of cysteine residues ranges from 3.9% to 10.3% of the total predicted protein. The predicted size of each member ranges from 97 to 216 amino acids. The number of cysteines in the 10-cysteine motif varies from 7 to 15, and all but one contains the conserved tyrosine residue at the third amino acid after the first conserved cysteine residue (Table 1). All but two predicted CDSs had predicted secretion signals according to SignalP 3.0 and 4.1 (Bendtsen et al. 2004; Petersen et al. 2011).

Of the eleven candidates, eight were successfully amplified from cDNA generated from *P. sorghi* inoculated maize leaves (isolate IA16) collected at 7 days after inoculation (DAI). The coding sequences minus the signal peptides were subsequently cloned into the pCR8 TOPO TA-cloning vector and sequenced. The predicted coding sequences from the reference genomes were highly similar to the sequences cloned from the IA16 isolate. However, additional or missing stretches of nucleotides and SNPs were observed (Supplemental File 1). The 10-cysteine motif of cluster 112 is largely conserved among the 8 cloned members (Figure 1). The length of the motif among the members varies from 85-92 base pairs (bps), but all ten cysteine residues are present in each motif (Figure 1, Table 1). There are several other conserved residues present in the IA16 members that are also found in both PpEC23 10-cysteine motifs, such as a tyrosine at the fourth residue and a [GxA]xC motif at the ninth cysteine. The previously identified [AFY]xC motif at

cysteine two is present in all but one member, and the [YFW]xC motif at cysteine eight is present in all eight cloned members (Figure 1).

Hypersensitive response immune assays

The CDS_{ns} of the eight amplified candidates were transferred into Effector Detector Vector pEDV6 (Fabro et al. 2011) in frame with the secretion signal for the AvrRps4 gene, enabling secretion via the bacterial type III secretion system. The resulting plasmids were transformed via triparental mating into *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000, which contains a functional bacterial type-III secretion system. *Pst* DC3000 is avirulent on *N. benthamiana* and contains several effectors that trigger HR (Wei et al. 2007). The full-length GFP CDS was likewise cloned and used as a negative control for each assay. The previously identified immune suppressor $PpEC23_{ns}$ was used as a positive control.

Three concentrations, namely OD_{600} 0.2, 0.02, and 0.004, of *Pst* DC3000 pEDV6-CSEPns were infiltrated into one half of 4-5 week old *N. benthamiana* leaves, with up to three separate leaves per plant. The other half was infiltrated with the *Pst* Dc3000 pEDV6-GFP control (Figure 2a). After 18-22 HPI, the 6 infiltrated regions on each leaf were phenotyped on a scale of 0-3, with zero being no HR and 3 being complete HR (Figure 2b). *Pst* DC3000 pEDV6-$PpEC23_{ns}$ was included as an immune suppression positive control, as it has been previously shown to inhibit plant immune response in this assay (Qi et al. 2016). For each bacterial concentration, the CSEP score was subtracted from the GFP control score, and these three differences were added together for one value per leaf, called the sum of differences. A positive sum of differences indicates supression of HR and a negative sum indicates activation of HR. As expected, the positive control $PpEC23_{ns}$ showed significant suppression of HR (Figure 2c). Of the eight *P. sorghi* CSEPs tested, only $930g11_{ns}$ showed significant suppression of HR, with all other CSEPs either showing no significant impact on HR ($2734g2_{ns}$ and $1483g10_{ns}$) or an

increased HR (α=0.05, Figure 2c). Although some leaves infiltrated with 930g11$_{ns}$ indicated no impact or increased induction of HR, the majority indicated suppression, with a mean sum of differences of 0.83. The remaining CSEPs had mean sums of differences ranging from 0.057 to -1.48, with more infiltration regions indicating increased induction of HR. Although the other CSEPs may have functions outside of suppression of ETI, 930g11$_{ns}$ was the only CSEP that showed promise in plant immune suppression in this assay.

HR immune suppression assays in transgenic *N. benthamiana* using time lapse images

As 930g11$_{ns}$ was the only of our eight CSEPs with evidence of immune suppression, we cloned it into the pBI121 Agrobacterium binary vector, adding a 3XFLAG tag to the 5' end of the coding sequence. The subsequent vector was used to transform *N. benthamiana* plants, and after two rounds of selfing, T2 seed was collected from four lines showing strong expression of the transgene and used for all subsequent experiments (Supplemental Figure 1). Azygous seed was also gathered from each line and used as a negative control.

A similar immune suppression assay to the previous experiment was conducted in the 930g11$_{ns}$ *N. benthamiana* lines. However, instead of leaving infiltrated plants in a growth room or chamber, detached leaves were placed into a phenotyping box modeled after the Navautron cabinet (Barbacci et al. 2020). The boxes were created from two plastic storage bins and LED strip lighting, with a wide-angled camera lens controlled by a Raspberry Pi mounted to the top of the box (Figure 3a). The boxes are almost fully enclosed, maintaining humidity to prevent detached leaves from wilting, and can hold up to 96 individual leaves.

To conduct these immune suppression assays, *Pst* DC3000 at an OD_{600} of 0.02 was infiltrated into one side of the first full leaf of 4-5 week old transgenic or azygous plants and a mock buffer control ($MgCl_2$) was infiltrated into the other side. Leaves were allowed to dry for 1-2 hours, then detached and placed in the phenotyping cabinets adaxial side down (Figure 3b).

Images were taken every hour, starting at 0 HPI, for two days, for a total of 48 images per leaf. Each leaf at each timepoint was scored on a scale of 0-7, with zero being no HR, and 7 being complete HR. The area under the curve (AUC) for each leaf was calculated by summing the scores for all time points. Although no *N. benthamiana* lines had significant differences in mean AUC, each transgenic line mean AUC was lower than the respective azygous mean AUC, both when all lines were compared separately (Figure 3c) and together (Figure 3d). Additionally, the mean HR score at each time point for transgenic leaves was consistently lower than the mean for azygous leaves (Figure 3e). Although not statistically significant, transgenic $930g11_{ns}$ plants consistently show a reduction in HR phenotype when compared to azygous control plants, both when analyzing AUC and mean HR score by time point. Given that in the previous HR experiments, $930g11_{ns}$ appeared to have less of an impact on plant immunity than PpEC23ns, it is likely that it has a small impact on ETI suppression and thus is harder to detect.

ROS burst assays

As the first two assays both were designed to detect suppression of HR/ETI, we also conducted reactive oxygen species (ROS) burst assays to investigate PTI response in the presence of flagellin-22 (flg22). Leaf samples were taken from 4-5 week old transgenic or azygous plants. After an overnight incubation in water at room temperature to stabilize the leaf samples, a luminol-based assay solution containing 100 nM flg22 was added and light emission was promptly analyzed using a microplate reader. The total photon count between the transgenic and azygous plants for each line was not significantly different, save for line 1C, where the transgenic plants had a greater ROS burst (α=0.05, Figure 4a). When examining all lines pooled together, there was no significant difference seen between azygous and transgenic plants (α=0.05, Figure 4b). For each cycle, transgenic plants consistently have a higher response to flg22, as indicated by mean photon count (Figure 4c). Coupling these results together, CSEP

$930g11_{ns}$ does not seem to have a direct effect on PTI suppression in the transgenic *N. benthamiana* plants and may be a slight activator of PTI.

Discussion

Effector screens often begin with large numbers of CSEPs for characterization in order to identify and characterize those which significantly contribute to rust diseases. Typically, few CSEPs yield significant phenotypes in these screens (Lorrain et al. 2019). This is due to several reasons, such as uninformative predicted annotations or highly redundant functions or phenotypes. Additionally, rust genomes contain large and expanding gene families that evolve quickly, often leaving behind non-functional gene copies or paralogous proteins with different gene functions (Aime et al. 2017). This may be why only one of the eight cloned *Pp*EC23 homologs in *P. sorghi* showed a plant immune suppression response in our assays, even though they all contain the intact 10-cysteine motif of the cluster 112 family. Interestingly, five of the cloned CSEPs showed an increased activation of HR in *N. benthamiana* leaves, which may indicate functionality as avirulence factors. There are many functions of effectors outside of plant immunity regulation, and it is likely that some CSEPs we identified have functions that were not tested for here. Furthermore, although many studies investigate effectors singly for ease of interpretation and reduction of confounding factors, no protein functions in isolation. As a result, valid functional characterizations will inherently be missed or may be misinterpreted.

Another interesting aspect was the variation of the *Pp*EC23 homologs between the RO10H11247 isolate and the IA16 isolate, illustrating the potential sequence divergence between the same CSEPs in different isolates of the same rust species. Isolates developed from temporally and spatially different wild populations often have different effector repertoires and differentials (Richter et al. 1995; Quade et al. 2021), and additional genomic resources are imperative to further dissecting these discrepancies.

As additional CSEPs or transcript variants are identified, and if chances of functional characterization of effector candidates with an observable phenotype remains low, higher throughput methods for conducting and analyzing effector screens is needed. In this manuscript, we implemented an inexpensive phenotyping box with a small footprint to generate time lapse imaging data during an HR immune suppression assay. During the initial immune suppression assay, we found that leaves needed to be checked and phenotyped at multiple time points, as the timing of HR varied between experiments. The utilization of a hands-off imaging setup meant that during the second immune suppression assay, no valuable data was lost if HR progressed quicker than expected. Furthermore, after the initial box setup, the experiment is entirely autonomous until complete, and the number of time points and length of experiment can be easily scaled up or down as needed without additional labor. Due to a simple design, the boxes can be easily modified for different use-cases, such as alternative or additional camera lenses that image in infra-red or hyperspectral wavelengths. Currently, the biggest drawback is a lack of a compatible automated phenotyping pipeline for this application, meaning all phenotyping is manual. However, being able to image up to 96 leaves at a time with no additional input was extremely useful for this study.

Conclusions

Functional characterization of CSEPs in rust fungi remains a top priority for rust researchers, particularly as the quantity and quality of rust genomes increases. Although many effector screens in various rust species have been conducted, thousands more CSEPs remain without any functional annotation, neither experimentally shown nor predicted. Further complicating effector research, conservation of a given effector's function is not guaranteed between rust species, or even between isolates of the same species.

To investigate a subset of CSEPs in the common rust pathogen *P. sorghi*, we identified eleven members homologous to *Pp*EC23, a member of the large, rust specific gene family cluster 112 shown to suppress plant immune response, from predicted *P. sorghi* proteome data. Of those eleven members, eight were successfully cloned and tested in various immune suppression assays. Although all eight appear to have intact 10-cysteine motifs, only one member was shown to inhibit plant immunity. The other seven candidates either had no effect on plant immunity or were found to induce plant immune responses in the experiments presented here. The candidate that showed an immune suppression phenotype, 930g11, was found to have a small effect on HR immune suppression but did not seem directly involved in PTI suppression in *N. benthamiana* leaves. Additionally, the use of a phenotyping box enabled high throughput of leaves with limited hands-on experimental time, as in-person phenotyping and manual photo-taking at the end of an experiment were eliminated, while simultaneously supplying additional data in the form of time lapse images.

Acknowledgements

This study was supported by the Iowa State University Predictive Plant Phenomics graduate training program funded by the National Science Foundation (DGE #1545453) and by Agricultural and Food Research Initiative grant no. 2019-07318 from the USDA National Institute of Food and Agriculture. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders. We also received support from the Plant Sciences Institute, the Lois H. Tiffany Scholarship, and the Gilman Scholarship at Iowa State University. We would like to thank the University of Nebraska at Lincoln Plant Transformation Core Research Facility for the generation of the transgenic pBI121-930g11*ns N. benthamiana*

lines, Hannah Craven for sampling, processing samples, and plant care, Peng Liu for statistics advice, and Melissa Bredow and Bliss Beernink for help with protocols.

Author Contributions

ME identified the *Pp*EC23 homologs from *P. sorghi*. KH conducted all experiments, data analyses, and interpretations. KH wrote the manuscript and SW edited the manuscript.

References

Aime, M. C., McTaggart, A. R., Mondo, S. J., and Duplessis, S. 2017. Phylogenetics and Phylogenomics of Rust Fungi. In *Advances in Genetics*, Academic Press, p. 267–307..

Barbacci, A., Navaud, O., Mbengue, M., Barascud, M., Godiard, L., Khafif, M., et al. 2020. Rapid identification of an Arabidopsis NLR gene as a candidate conferring susceptibility to *Sclerotinia sclerotiorum* using time-resolved automated phenotyping. Plant J. 103:903–917.

Bendtsen, J. D., Nielsen, H., Von Heijne, G., and Brunak, S. 2004. Improved Prediction of Signal Peptides: SignalP 3.0. J. Mol. Biol. 340:783–795.

Bredow, M., Sementchoukova, I., Siegel, K., and Monaghan, J. 2019. Pattern-triggered oxidative burst and seedling growth inhibition assays in Arabidopsis thaliana. J. Vis. Exp. :e59437.

de Carvalho, M. C. da C. G., Costa Nascimento, L., Darben, L. M., Polizel-Podanosqui, A. M., Lopes-Caitar, V. S., Qi, M., et al. 2017. Prediction of the *in planta Phakopsora pachyrhizi* secretome and potential effector families. Mol. Plant Pathol. 18:363–377.

Fabro, G., Steinbrenner, J., Coates, M., Ishaque, N., Baxter, L., Studholme, D. J., et al. 2011. Multiple Candidate Effectors from the Oomycete Pathogen *Hyaloperonospora arabidopsidis* Suppress Host Plant Immunity ed. Frederick M. Ausubel. PLoS Pathog. 7:e1002348.

Figueroa, M., Dodds, P. N., Henningsen, E. C., and Sperschneider, J. 2023. Global Landscape of Rust Epidemics by *Puccinia* Species: Current and Future Perspectives. In Plant Relationships, The Mycota, Springer, Cham, p. 391–423..

Figueroa, M., Ortiz, D., and Henningsen, E. C. 2021. Tactics of host manipulation by intracellular effectors from plant pathogenic fungi. Curr. Opin. Plant Biol. 62:102054.

Garnica, D. P., Nemri, A., Upadhyaya, N. M., Rathjen, J. P., and Dodds, P. N. 2014. The Ins and Outs of Rust Haustoria ed. Joseph Heitman. PLoS Pathog. 10:e1004329.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., et al. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 38:W695–W699.

Link, T. I., Lang, P., Scheffler, B. E., Duke, M. V., Graham, M. A., Cooper, B., et al. 2014. The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. Mol. Plant Pathol. 15:379–393.

Liu, C., Pedersen, C., Schultz-Larsen, T., Aguilar, G. B., Madriz-Ordeñana, K., Hovmøller, M. S., et al. 2016. The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases. New Phytol.

Lorrain, C., Gonçalves dos Santos, K. C., Germain, H., Hecker, A., and Duplessis, S. 2019. Advances in understanding obligate biotrophy in rust fungi. New Phytol. 222:1190–1206.

Lorrain, C., Petre, B., and Duplessis, S. 2018. Show me the way: rust effector targets in heterologous plant systems. Curr. Opin. Microbiol. 46:19–25.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., et al. 2013. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 41:W597–W600.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods. 8:785–786.

Qi, M., Grayczyk, J. P., Seitz, J. M., Lee, Y., Link, T. I., Choi, D., et al. 2018. Suppression or Activation of Immune Responses by Predicted Secreted Proteins of the Soybean Rust Pathogen *Phakopsora pachyrhizi*. Mol. Plant-Microbe Interact. 31:163–174.

Qi, M., Link, T. I., Müller, M., Hirschburger, D., Pudake, R. N., Pedley, K. F., et al. 2016. A Small Cysteine-Rich Protein from the Asian Soybean Rust Fungus, *Phakopsora pachyrhizi*, Suppresses Plant Immunity ed. Peter N Dodds. PLoS Pathog. 12:e1005827.

Quade, A., Ash, G. J., Park, R. F., and Stodart, B. 2021. Resistance in Maize (*Zea mays*) to Isolates of *Puccinia sorghi* from Eastern Australia. Phytopathology. 111:1751–1757.

Ramachandran, S. R., Yin, C., Kud, J., Tanaka, K., Mahoney, A. K., Xiao, F., et al. 2017. Effectors from Wheat Rust Fungi Suppress Multiple Plant Defense Responses. Phytopathology. 107:75–83.

Richter, T. E., Pryor, T. J., Bennetzen, J. L., and Hulbert, S. H. 1995. New Rust Resistance Specificities Associated with Recombination in the *Rp1* complex in Maize. Genetics. 141:373–81.

Rochi, L., Diéguez, M. J., Burguener, G., Darino, M. A., Pergolesi, M. F., Ingala, L. R., et al. 2018. Characterization and comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize common rust. Fungal Genet. Biol. 112:31–39.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. 2012. Fiji: An open-source platform for biological-image analysis. Nat. Methods. 9:676–682.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539.

Sperschneider, J., and Dodds, P. N. 2022. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. Mol. Plant-Microbe Interact. 35:146–156.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 40:1023–1025.

Uhse, S., and Djamei, A. 2018. Effectors of plant-colonizing fungi and beyond. PLOS Pathog. 14:e1006992.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 25:1189–1191.

Wei, C.-F., Kvitko, B. H., Shimizu, R., Crabill, E., Alfano, J. R., Lin, N.-C., et al. 2007. A *Pseudomonas* syringae pv. tomato DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*. Plant J. 51:32–46.

Wiedemann, C., Kumar, A., Lang, A., and Ohlenschläger, O. 2020. Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. Front. Chem. 8:280.

Zhang, X., Nguyen, N., Breen, S., Outram, M. A., Dodds, P. N., Kobe, B., et al. 2017. Production of small cysteine-rich effector proteins in Escherichia coli for structural and functional studies. Mol. Plant Pathol. 18:141–151.

Figures

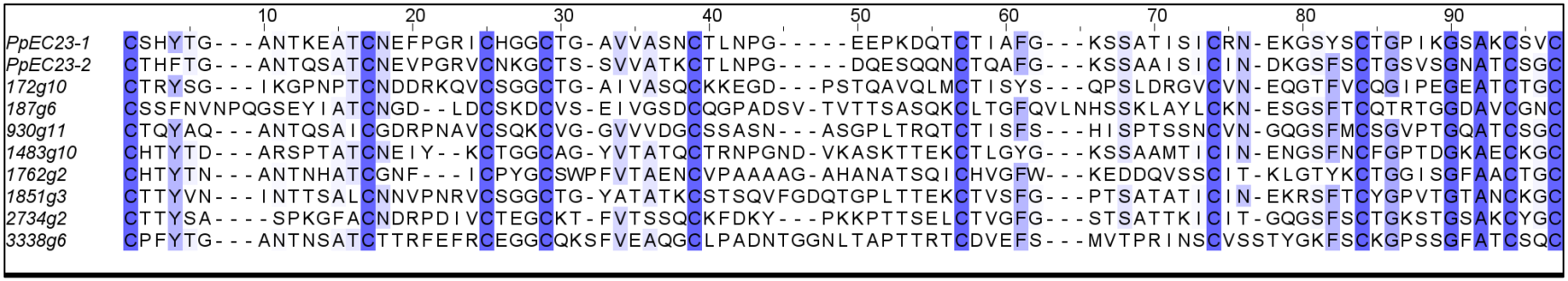


Figure 1. Conservation of the 10-cysteine motif in the eight cloned *Pp*EC23 homologs in *P. sorghi* isolate IA16 and the two motifs of *Pp*EC23. Figure was made in Jalview 2.11.2.6 from a Clustal Omega alignment. Color saturation is based on percentage identity by conservation with a threshold of 30%.

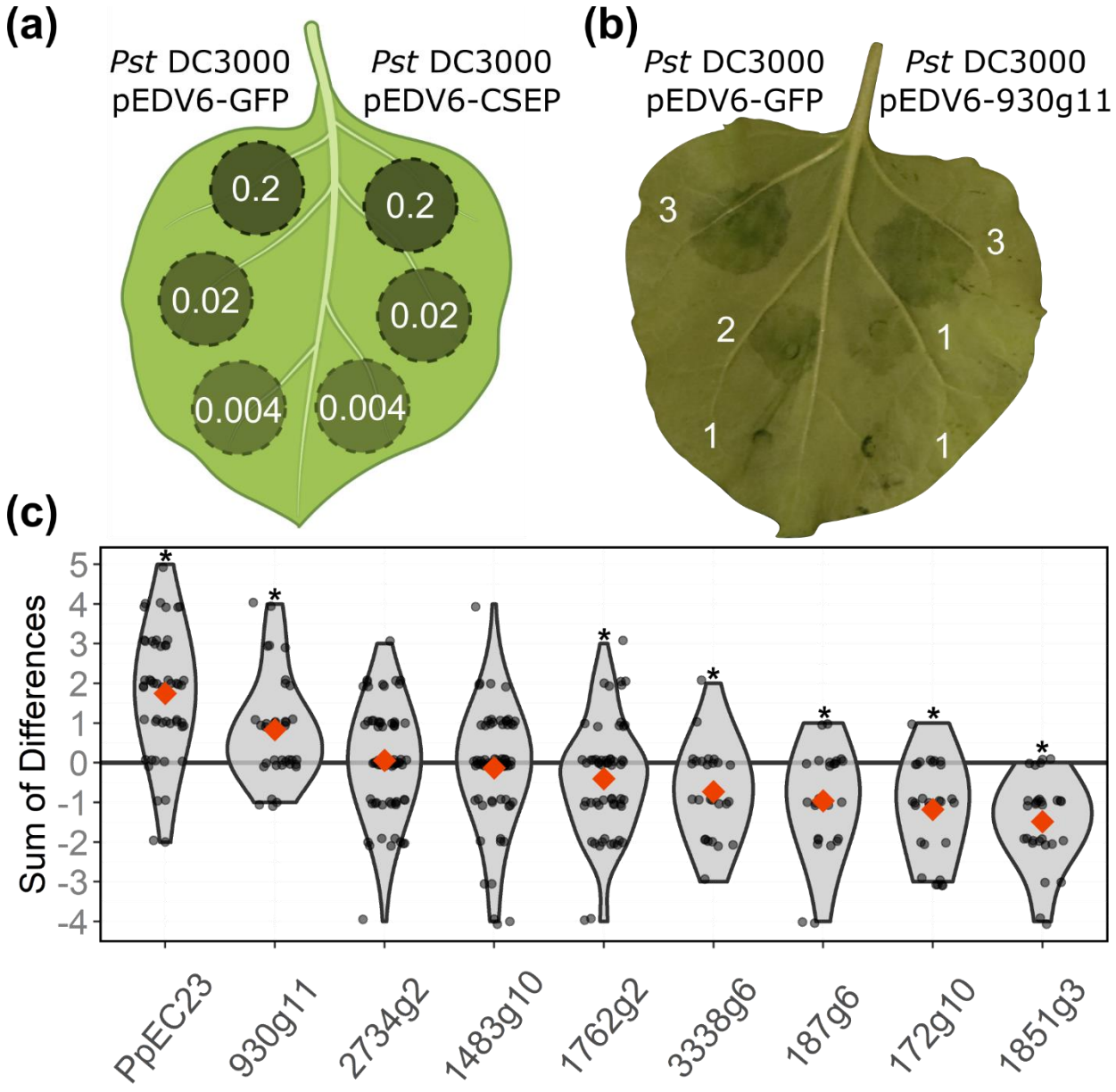


Figure 2. Quantification of hypersensitive response in *N. benthamiana* leaves after infiltration with *Pst* DC3000 carrying pEVD6-CSEP$_{ns}$ constructs. (a) Diagram showing the infiltration scheme for each leaf. The left leaf half was infiltrated with a negative GFP control and the right half was infiltrated with one of 9 CSEP$_{ns}$ constructs, which includes *Pp*EC23$_{ns}$ (a positive immune suppressor control) and the eight IA16 *Pp*EC23 homologs. The image was created with BioRender.com. (b) One example leaf photographed 22 HPI after infiltration with *Pst* DC3000 pEDV6-GFP or pEDV6-930g11$_{ns}$, with the scores for each infiltration region. (c) After subtracting the CSEP score from the GFP score, the differences were summed for each leaf. The violin plot represents the density of the sum of differences and the orange point represents the mean of each CSEP's sum of differences. The asterisks represent a significant difference of mean when compared to zero (two-sample t-tests for each CSEP, α=0.05).

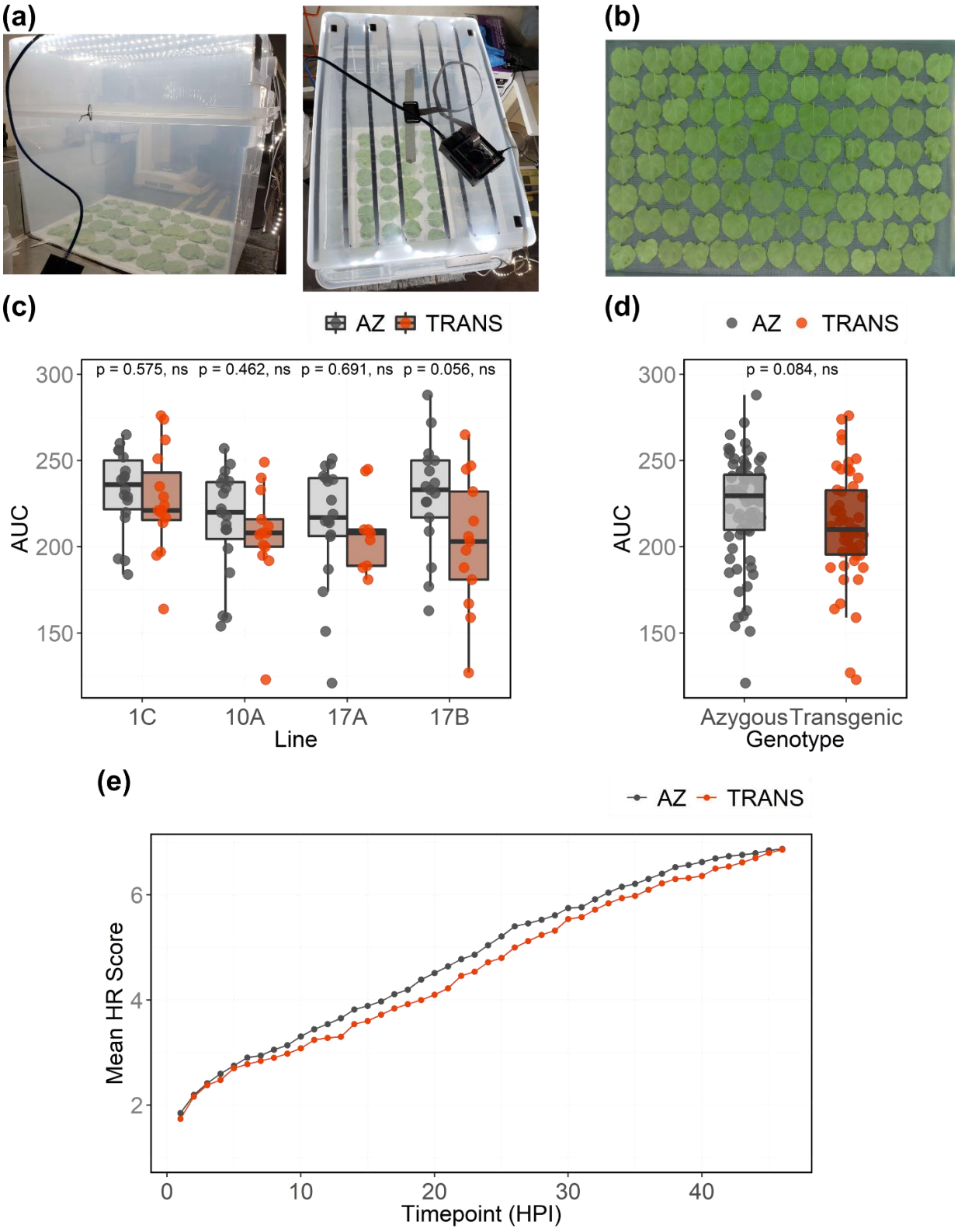


Figure 3. Hypersensitive response in transgenic (TRANS) and azygous (AZ) pBI121-3XFLAG-930g11$_{ns}$ *N. benthamiana* leaves after infiltration with *Pst* DC3000, as conducted in phenotyping

boxes. (a) The phenotyping boxes are composed of two plastic storage bins with a Raspberry Pi-operated camera lens mounted to the top. White LED strip lights cover the lid to supply consistent lighting during experiments. (b) An example of an image from a time lapse experiment, showing 96 leaves. (c) The area under the curve (AUC) was calculated for each leaf and plotted for each line or (d) each genotype. (e) The mean HR score at each timepoint was plotted according to genotype.

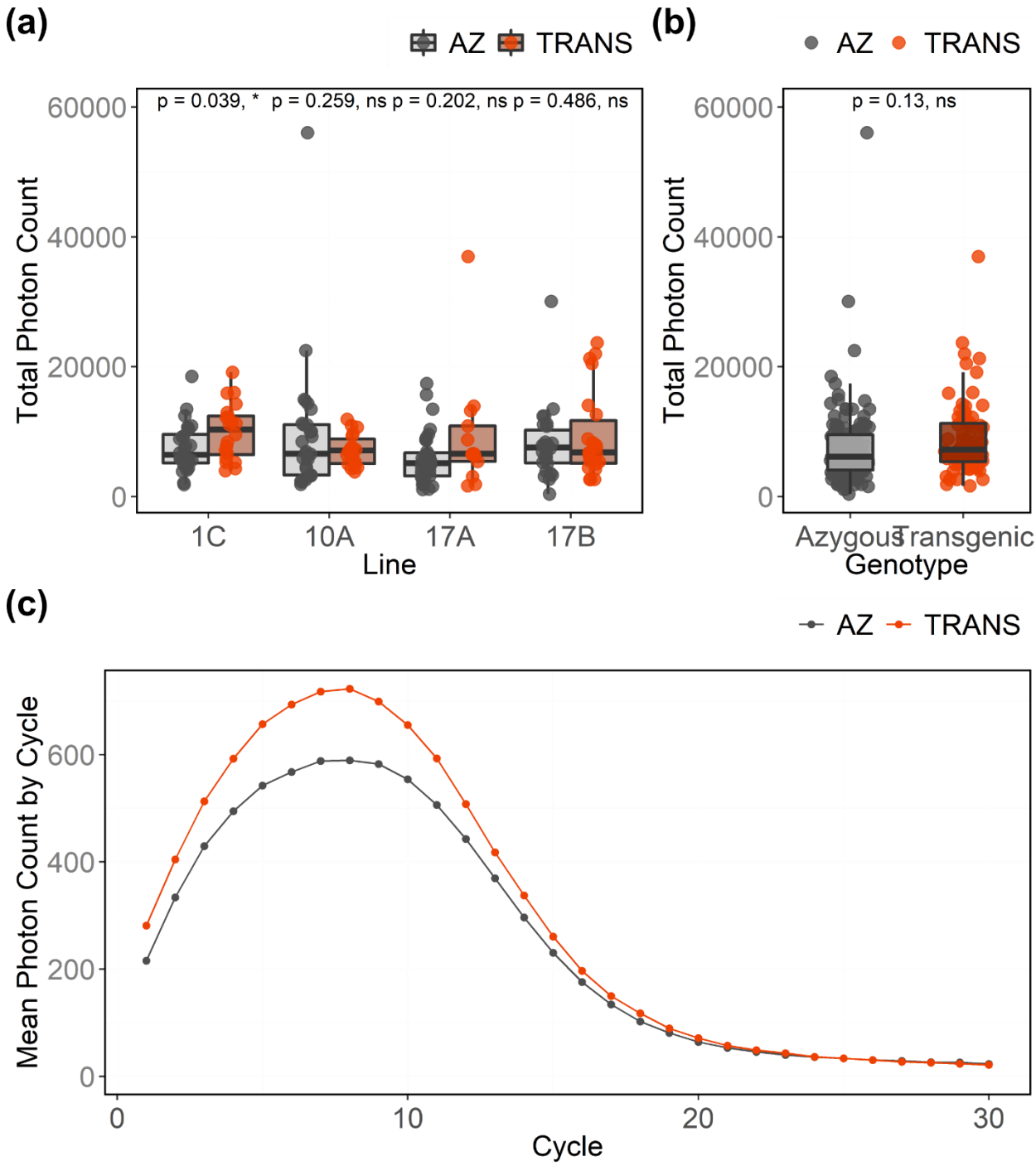


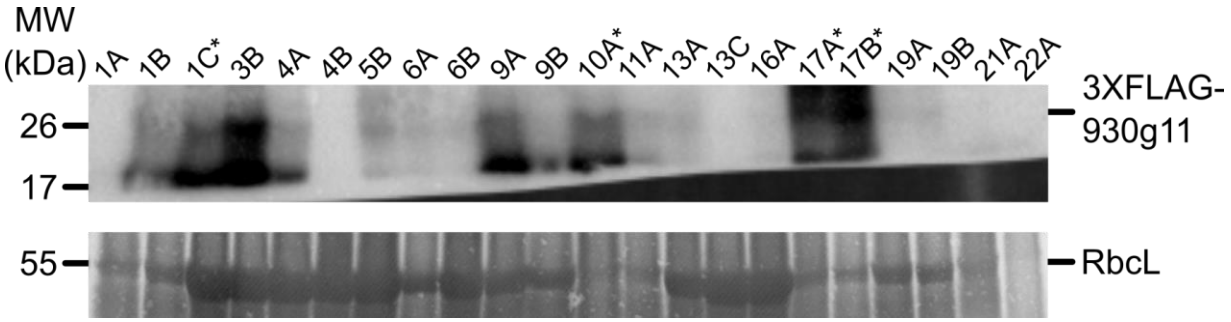
Figure 4. ROS burst assays were conducted on transgenic (TRANS) and azygous (AZ) pBI121-3xFLAG-930g11$_{ns}$ *N. benthamiana* leaves in the presence of flg22, and total photon count for each sample was summed and graphed according to (a) line or (b) genotype. (c) Mean photon count for all samples at each cycle was calculated and plotted according to genotype.

Tables

Table 1. A summary of the eleven identified *Pp*EC23 homologs in the published *P. sorghi* genome. Predicted % cysteine, predicted protein length, and predicted secretion signal are based on the published predicted CDS data. The cloned cysteines in motif and cloned motif length are based on the amplified CDS from the IA16 *P. sorghi* isolate.

Gene ID	Annotation	Predicted % Cysteine	Predicted Protein Length	Predicted Secretion Signal	Cloned Cysteines in Motif	Cloned Motif Length
VP01_134g9*	Hypothetical protein	5.69	123	Yes	-	-
VP01_172g10	Uncharacterized protein	6.12	196	Yes	10	86
VP01_187g6	Putative signal peptide protein	6.92	130	Yes	10	92
VP01_930g11	Hypothetical protein	3.94	254	Yes	10	86
VP01_1483g10	Hypothetical protein	4.63	216	Yes	10	86
VP01_1638g2*	Hypothetical protein	6.22	209	No	-	-
VP01_1762g2	Hypothetical protein	10.31	97	Yes	10	86
VP01_1851g3	Hypothetical protein	4.37	252	Yes	10	89
VP01_1930g3*	Hypothetical protein	4.97	302	No	-	-
VP01_2734g2	Hypothetical protein	8.27	133	Yes	10	85
VP01_3338g6	Hypothetical protein	8.7	138	Yes	10	91
*No amplification from 7 DAI cDNA.						

Supplemental Figures



Supplemental Figure 1. A western blot showing the production of 3XFLAG-930g11$_{ns}$ (26.71 kDa) in transgenic *N. benthamiana* lines. Total protein was extracted from T0 plants, separated by SDS-PAGE, and transferred to nylon membrane. The membrane was blocked with 5% skim milk powder, washed, and then incubated with an anti-FLAG-HRP. Asterisks indicate lines used in subsequent experiments. The upper image shows anti-FLAG, and the lower image shows the Coomassie-stained Rubisco large subunit (RbcL).

Supplemental Tables

Supplemental Table 1. Oligonucleotide primers used in this study.

ID	Notes	Direction	Primer Sequence
134g9	Amplification primer, no secretion signal, not amplified	F	ACCCCTCCTTATCCACCGC
134g9	Amplification primer, no secretion signal, not amplified	R	CTCACCCTTAGCGGCAGC
172g10	Amplification primer, no secretion signal	F	AGTCCGCACTCCAGTCGTA
172g10	Amplification primer, no secretion signal	R	AACGAACAAAGCGAGAAGGC
187g6	Amplification primer, no secretion signal	F	TATAACTCTTACGTCGCGACCAC
187g6	Amplification primer, no secretion signal	R	TCGCTGAAAACAGTTACCACAAACC
930g11	Amplification primer, no secretion signal	F	GAGACATGGAGCTGCACTCA
930g11	Amplification primer, no secretion signal	R	CAAAAGGGTAACGCCCGC
1483g10	Amplification primer, no secretion signal	F	GATCACGAAGACCACGACCACG
1483g10	Amplification primer, no secretion signal	R	GAGGAAAGTGGCGGATACGA
1638g2	Amplification primer, no secretion signal, not amplified	F	ATGTCGAGAACCATCTGGTCCAG
1638g2	Amplification primer, no secretion signal, not amplified	R	GTCATAGGAGAGCCACGTGGAC
1762g2	Amplification primer, no secretion signal	F	TCTACAAAATTCATACCACAAGCACA
1762g2	Amplification primer, no secretion signal	R	AGTGCTTTTGAATCGGCAACC
1851g3	Amplification primer, no secretion signal	F	GATGCTGGAGACGGGCAA
1851g3	Amplification primer, no secretion signal	R	AGCAAACATAAAGGACATGAAGGT
1930g3	Amplification primer, no secretion signal, not amplified	F	ATGGAGAAAGTGCATGAAGTGAC
1930g3	Amplification primer, no secretion signal, not amplified	R	TTAGTTAGGCTTGAGTAATGTGGAGA
2734g2	Amplification primer, no secretion signal	F	ATCGGTGAGACGATGACCTG
2734g2	Amplification primer, no secretion signal	R	TGAAGCCCAAGGGATTTTGTTG

Supplemental Table 1 Continued			
ID	**Notes**	**Direction**	**Primer Sequence**
3338g6	Amplification primer, no secretion signal	F	GCCAGCATCGATTCACCTAC
3338g6	Amplification primer, no secretion signal	R	ACTCAGAGGCACGAGATCTG
pBI121-930	*N. benthamaiana* transformation plasmid	F	CGCGGATCCATGGACTACAAAGACCAT GACGGTGATTATAAAGATCATGACATC GACTACAAGGATGACGATGACAAGAT CACAAGTTTGTAC
pBI121-930	*N. benthamaiana* transformation plasmid	R	GACGAGCTCTCACAAAAGGGTAACGC CCGCG

Supplemental Files

Supplemental File 1. Clustal Omega 1.2.4 alignment output for the RO10H11247 predicted *P. sorghi Pp*EC23 homologs CDSs (RO) and the cloned IA16 CDSs (IA16). Initial missing matches for the IA16 CDSs are indicative of the predicted secretion signals that were not cloned.

172g10

```
RO_172g10         ATGAGTTGTACAACGATTTGCGTCCTCGTATTAGGTCTGGTCACAACATCCCTCTCCAGT   60
IA16_172g10       ---------------------------------------------------------AGT   3
                                                                           ***

RO_172g10         CCGCACTCCAGTCGTAGCAGCAGTAGTAGTGGCAGCACTAGTGTTCCCCCCACGTCCGGC   120
IA16_172g10       CCGCACTCCAGTCGTAGCAGCAGTAGTAGTGGCAGCACTAGTGTTCCCCCCACGTCCGGC   63
                  ************************************************************

RO_172g10         GGCGGAAACACCCTCAAGTGCACACGCTATTCCGGCATCAAAGGTCCTAACCCTACTTGT   180
IA16_172g10       GGCGGAAACACCCTCAAGTGCACACGCTATTCCGGCATCAAAGGTCCTAACCCTACTTGT   123
                  ************************************************************

RO_172g10         AACGATGACCGCAAACAAGTGTGCTCAGGAGGCTGTACAGGAGCAATCGTGGCAAGCCAA   240
IA16_172g10       AACGATGACCGCAAACAAGTGTGCTCAGGAGGCTGTACAGGAGCAATCGTGGCAAGCCAA   183
                  ************************************************************

RO_172g10         TGCAAGAAAGAGGGCGACCCTTCGACTCAAGCAGTCCAGCTGATGTGCACGATCAGCTAC   300
IA16_172g10       TGCAAGAAAGAGGGCGACCCTTCGACTCAAGCAGTCCAGCTGATGTGCACGATCAGCTAC   243
                  ************************************************************

RO_172g10         AGTCAGCCCAGTCTGGATCGCGGAGTCTGCGTCAATGAACAAGGCACCTTTGTTTGCCAA   360
IA16_172g10       AGTCAGCCCAGTCTGGATCGCGGAGTCTGCGTCAATGAACAAGGCACCTTTGTTTGCCAA   303
                  ************************************************************

RO_172g10         GGGATCCCGGAGGGCGAAGCCACTTGCACAGGGTGCGCGCTCATCGTCGGTGACCCGAGC   420
IA16_172g10       GGGATCCCGGAGGGCGAAGCCACTTGCACAGGGTGCGCGCTCATCGTCGGTGACCCGAGC   363
                  ************************************************************

RO_172g10         ATCGTCAACAGCTCGTCGCCCGATCCTCCCACTGCGCCCGGAATGAAAACCCCCTCCCCC   480
IA16_172g10       ATCGCCAACAGCTCGTCGCCCGATCCTCCCACTGCGCCCGGAATGAAAACCCCCTCCCCC   423
                  **** *******************************************************

RO_172g10         TCCTCCTCAACATCTCCCGCAACTCCGCTCCCTAGAAAACTCGCTCTCTTTCTCGTCGCG   540
IA16_172g10       TCCTCCTCAACATCTCCCGCAACTCCGCTCCCTAGAAAACTCGCTCTCTTTCTCGTCGCG   483
                  ************************************************************

RO_172g10         TCCAAGCTCCTCGGGCTGTTGGGCCTTCTCGCTTTGTTCGTTTAA      585
IA16_172g10       TCCAAGCTCCTCGGGCTGTTGGGCCTTCTCGCTTTGTTCGTT---      525
                  ******************************************
```

186g6

```
RO_187g6        ATGCCTCCGTTCAATCTTTCGTCTACCTGCCTCTTGTTCGTCCTGGCCTTGTCGGCTGCT    60
IA16_187g6      ------------------------------------------------------------    0


RO_187g6        TCGCTGGTTTCTGCTTATAACTCTTACGTCGCGACCACTTGCAGCAGCTTCAATGTGAAT    120
IA16_187g6      ---------------TATAACTCTTACGTCGCGACCACTTGCAGCAGCTTCAATGTGAAT    45
                               *******************************************


RO_187g6        CCTCAAGGGTCCGAATATATAGCAACATGCAACGGAGACCTGGATTGCAGCAAAGATTGC    180
IA16_187g6      CCTCAAGGGTCCGAATATATAGCAACATGCAACGGAGACCTGGATTGCAGCAAAGATTGC    105
                ************************************************************


RO_187g6        GTCAGTGAGATCGTCGGATCGGACTGTCAGGGTCCGGCCGACAGCGTCACGGTCACTACT    240
IA16_187g6      GTCAGTGAGATCGTCGGATCGGACTGTCAGGGTCCGGCCGACAGCGTCACGGTCACTACT    165
                ************************************************************


RO_187g6        AGTGCATCCCAGAAGTGTCTTACTGGATTCCAAGTCCTCAATCACTCTTCGAAACTTGCT    300
IA16_187g6      AGTGCATCCCAGAAGTGTCTTACTGGATTCCAAGTCCTCAATCACTCTTCGAAACTTGCT    225
                ************************************************************


RO_187g6        TATCTCTGTAAAAACGAATCAGGAAGTTTTACTTGTCAGACAAGAACAGGTGGTGATGCG    360
IA16_187g6      TATCTCTGTAAAAACGAATCAGGAAGTTTTACTTGTCAGACAAGAACAGGTGGTGATGCG    285
                ************************************************************


RO_187g6        GACCGTTGGACATCACAAATGCATATGTAA 390
IA16_187g6      GTTTGTGGTAACTGTTTTCAGCGA------ 309
                *:  ** * *..* : ::.:**.:
```

930g11

```
RO_930g11         ATGGCCACGCTCGCGGTTGCATCCATCATTCCCGCGCTCTCCGCCGGCCTGACCACCGGC   60
IA16_930g11       ------------------------------------------------------------   0


RO_930g11         GAGACATGGAGCTGCACTCAATATGCGCAAGCGAATACCCAATCTGCCATCTGCGGCGAT   120
IA16_930g11       GAGACATGGAGCTGCACTCAATATGCGCAAGCGAATACCCAATCTGCCATCTGCGGCGAT   60
                  ************************************************************


RO_930g11         CGACCAAATGCCGTGTGCTCTCAAAAATGTGTCGGCGGTGTTGTTGTTGATGGCTGCTCG   180
IA16_930g11       CGACCAAATGCCGTGTGCTCTCAAAAATGTGTCGGCGGTGTTGTTGTTGATGGCTGCTCG   120
                  ************************************************************


RO_930g11         TCAGCCTCAAATGCATCGGGCCCCTTGACACGTCAAACCTGCACCATCAGCTTTAGCCAC   240
IA16_930g11       TCAGCCTCAAATGCATCGGGCCCCTTGACACGTCAAACCTGCACCATCAGCTTTAGCCAC   180
                  ************************************************************


RO_930g11         ATCTCACCTACCAGTAGCAATTGTGTCAATGGCCAAGGTTCATTCATGTGTTCCGGTGTC   300
IA16_930g11       ATCTCACCTACCAGTAGCAATTGTGTCAATGGCCAAGGTTCATTCATGTGTTCCGGTGTC   240
                  ************************************************************


RO_930g11         CCCACTGGACAAGCCACGTGCTCCGGATGTGTTGATTCAACGCTGCACCTAGACCCAGCC   360
IA16_930g11       CCCACTGGACAAGCCACGTGCTCCGGATGTGTTGATTCAACGCTGCACCTAGACCCAGCC   300
                  ************************************************************


RO_930g11         CTGTCCCAAACTGTGCCACCATTGGCAGCCCCCGCGATTGGTGCCCCGGCTCCTGTTGTC   420
IA16_930g11       CTGTCCCAAACTGTGCCACCATTGGCAGCCCCCGCGATTGGTGCCCCGGCTCCTGTTGTC   360
                  ************************************************************


RO_930g11         GCTGCCAGCTCTTCCCCAGCCGCCTCAACTGTTGCCCCACCCCTTGCAGTTAAGGCGACC   480
IA16_930g11       GCTGCCAGCTCTTCCCCAGCCGCCTCAACTGTTGCCCCACCCCTTGCAGTTAAGGCGACC   420
                  ************************************************************


RO_930g11         GCTCCCTCCGTCTCTTCTGCGGCGCCTGCGGCGCCTGCTGCCCCCGCCCCAGCTGTCACG   540
IA16_930g11       GCTCCCTCCGTCTCTTCTGCGGC---------GCCTGCTGCCCCCGCCCCAGCTGTCACG   471
                  **********************         ****************************


RO_930g11         GCGGCACCTGCTATCTCCACTCAGGTCGTCACCGTATATAGTCAACCCACGACATCCACC   600
IA16_930g11       GCGGCACCTGCTATCTCCACTCAGGTCGTCACCGTATATAGTCAACCCACGACATCCACC   531
                  ************************************************************


RO_930g11         AGGACCTCTGCCCAAGACGAAGATACCCCCCAAAGTGACTCCTCCTCCTCGAACAGTAAC   660
IA16_930g11       AGGACCTCTGCCCAAGACGAAGATACCCCCCAAAGTGACTCCTCCTCCTCGAACAGTAAC   591
                  ************************************************************


RO_930g11         AGTACTACATCGATTGCTTCCATCACTTACACCAATCCCGCGACGCCGTTTGTGGTAACC   720
IA16_930g11       AGTACTACATCGATTGCTTCCATCACTTACACCAATCCCGCGACGCCGTTTGTGGTAACC   651
                  ************************************************************


RO_930g11         CTTGGGCTCTTGGCCGCGGGCGTTACCCTTTTGTGA 756
IA16_930g11       CTTGGGCTCTTGGCCGCGGGCGTTACCCTTTTG--- 684
                  *********************************
```

1483g10

```
RO_1483g10        ATGGTTTTCCGGCCTATAACTTTGAGATCAATGCTCTATTCCGCCGTGTTTATAATTGGC  60
IA16_1483g10      ------------------------------------------------------------  0


RO_1483g10        ATTGGAACGGTTGTTCTTGGAGATCACGAAGACCACGACCACGACCAGCATGACCACGAC  120
IA16_1483g10      ----------------------GATCACGAAGACCACGACCACGACCAGCATGACCACGAC  39
                                        ************************************


RO_1483g10        CAGCATGACCACCCCCAGCATGACCACGACCAGGACTGTCACACATATACCGATGCGCGG  180
IA16_1483g10      CAGCATGACCACCCCCAGCATGACCACGACCAGGACTGTCACACATATACCGATGCGCGG  99
                  ************************************************************


RO_1483g10        TCGCCTACTGCTACCTGCAACGAAATTTATAAGTGCACCGGAGGCTGCGCCGGCTATGTC  240
IA16_1483g10      TCGCCTACTGCTACCTGCAACGAAATTTATAAGTGCACCGGAGGCTGCGCCGGCTATGTC  159
                  ************************************************************


RO_1483g10        ACCGCCACTCAGTGCACGCGCAATCCTGGAAACGACGTGAAAGCGTCGAAGACCACGGAG  300
IA16_1483g10      ACCGCCACTCAGTGCACGCGCAATCCTGGAAACGACGTGAAAGCGTCGAAGACCACGGAG  219
                  ************************************************************


RO_1483g10        AAATGCACGCTTGGATATGGAAAGTCGTCGGCAGCCATGACCATTTGCATCAACGAGAAT  360
IA16_1483g10      AAATGCACGCTTGGATATGGAAAGTCGTCGGCAGCCATGACCATTTGCATCAACGAGAAT  279
                  ************************************************************


RO_1483g10        GGCAGTTTCAATTGTTTTGGCCCGACGGATGGGAAAGCCGAGTGCAAAGGCTGCGTTATG  420
IA16_1483g10      GGCAGTTTCAATTGTTTTGGCCCGACGGATGGGAAAGCCGAGTGCAAAGGCTGCGTTATG  339
                  ************************************************************


RO_1483g10        GACGCAAGTCGCCACGTTTCGAACACCACCACCACCTCCTCCAGCTCTCCTCAACCTGGC  480
IA16_1483g10      GACGCAAGTCGCCACGTTTCGAACACCACCACCACCTCCTCCAGCTCTCCTCAACCTGGC  399
                  ************************************************************


RO_1483g10        GATAACTCCAATGCAGGCTCTTCCAACCAACCTGGCTCTAACACCGTTCCCATAGTCGAT  540
IA16_1483g10      GATAACTCCAATGCAGGCTCTTCCAACCAACCTGGCTCTAACACCGTTCCCATAGTCGAT  459
                  ************************************************************


RO_1483g10        ACCAAAGCCCCCGAGAAACCTTCTAACTCTTCCGCCTCTAGCTTCAGTGTTAACATCATC  600
IA16_1483g10      ACCAAAGCCCCCGAGAAACCTTCTAACTCTTCCGCCTCTAGCTTCAGTGTTAACATCATC  519
                  ************************************************************


RO_1483g10        TTCCTGTCGATTGGCGCTCTCGTATCCGCCACTTTCCTCTGA       642
IA16_1483g10      TTCCTGTCGATTGGCGCTCTCGTATCCGCCACTTTCCTC---       558
                  ************************************
```

1762g2

```
RO_1762g2        ATGGTGGCGCTCGCTCTATTGGTCACGAATTGGCTTATTTGTGCTGCACTGGGCTCTACA   60
IA16_1762g2      ------------------------------------------------------TCTACA   6
                                                                       ******

RO_1762g2        AAATTCATACCACAAGCACAAAATTGCCACACATACACCAATGCAAACACGAATCATGCC   120
IA16_1762g2      AAATTCATACCACAAGCACAAAATTGCCACACATACACCAATGCAAACACGAATCATGCC   66
                 ************************************************************

RO_1762g2        ACTTGTGGAAATTTTATCTGTCCCTATGGCTGCTCTTGGCCTTTCGTCACCGCGGAGAAC   180
IA16_1762g2      ACTTGTGGAAATTTTATCTGTCCCTATGGCTGCTCTTGGCCTTTCGTCACCGCGGAGAAC   126
                 ************************************************************

RO_1762g2        TGTGTCCCAGCGGCAGCGGCAG---------------------------------------   202
IA16_1762g2      TGTGTCCCAGCGGCAGCGGCAGGTGCGCACGCCAATGCCACCTCACAAATATGCCACGTC   186
                 **********************

RO_1762g2        --------------------------------CGTGTATTACGAAGCTTGGTACGTACAAA   231
IA16_1762g2      GGGTTCTGGAAAGAAGACGATCAAGTTTCTTCGTGTATTACGAAGCTTGGTACGTACAAA   246
                                                 ****************************

RO_1762g2        TGCACTGGAGGAATCAGCGGTTTTGCTGCTTGCACGGGTTGCCGATTCAAAAGCACTTGA   291
IA16_1762g2      TGCACTGGAGGAATCAGCGGTTTTGCTGCTTGCACGGGTTGCCGATTCAAAAGCACT---   303
                 ************************************************************
```

1851g3

```
RO_1851g3       ATGCCTATCGCACTCCTCTTTGTCACTGCCTTGTGCACCTCCGTCGTGCTCACCCAAGCC   60
IA16_1851g3     ------------------------------------------------------------   0


RO_1851g3       GATGCTGGAGACGGGCAACCTATGGATCTGGACTGTACCACATAC---------------   105
IA16_1851g3     GAAGCTGGAGACGGGCAACCTATGGATCTGGACTGTACCACATACGTCAACATAAACACC   60
                **:*****************************************


RO_1851g3       ----------TGTGTAACAATGTTCCCAACAGGGTGTGCTCTGGTGGCTGCACCGGCTAC   155
IA16_1851g3     ACTTCTGCCTTGTGTAACAATGTTCCCAACAGGGTGTGCTCTGGTGGCTGCACCGGCTAC   120
                          **************************************************


RO_1851g3       GCCAGTAAGTAGTAGTTCATCCTCAAGAATTCCTTCCTCCCCTATCTGCCATTTCAAAAA   215
IA16_1851g3     GCCA------------------------------------------------------   124
                ****


RO_1851g3       TTCAACGAAGCATATGGTCACTTCAGCCGCCACCAAGTGCAGTACCAGCCAGGTCTTTGG   275
IA16_1851g3     --------------------------CCGCCACCAAGTGCAGTACCAGCCAGGTCTTTGG   158
                                         **********************************


RO_1851g3       GGACCAAACAGGTCCCCTCACTACTGAAAAATGCACCGTCTCGTTTGGGCCCACGTCTGC   335
IA16_1851g3     GGACCAAACAGGTCCCCTCACTACTGAAAAATGCACCGTCTCGTTTGGGCCCACGTCTGC   218
                ************************************************************


RO_1851g3       CACCGCCACAATCTGCATCAACGAGAAGAGGTCCTTTACCTGCTATGGCCCCGTAACCGG   395
IA16_1851g3     CACCGCCACAATCTGCATCAACGAGAAGAGGTCCTTTACCTGCTATGGCCCCGTAACCGG   278
                ************************************************************


RO_1851g3       CACAGCTAACTGCAAGGGATGCACACAATCCTCCGGCTCTACCGACAAGCCACCAAACAA   455
IA16_1851g3     CACAGCTAACTGCAAGGGATGCACACAATCCTCCGGCTCTACCGACAAGCCACCAAACAA   338
                ************************************************************


RO_1851g3       TCCCCCCGTCGGCGGCTCCTCAGGAAACAACTCTACCAGCACTTCCGGAAACAATTCCAC   515
IA16_1851g3     TCCCCCCGTCGGCGGCTCCTCAGGAAACAACTCTACCAGCACTTCCGGAAACAATTCCAC   398
                ************************************************************


RO_1851g3       CAACACTCCCGGAAGCACTTTCGGTGACGCTTCCGGCAACACCTCAGTAAGCACATCTGG   575
IA16_1851g3     CAACACTCCCGGAAGCACTTTCGGTGACGCTTCCGGCAACACCTCAGTAAGCACATCTGG   458
                ************************************************************


RO_1851g3       TAACACTTCCGGTGGCACCTCCGACAAACCTGCCGGCGGCGCAGGAACCGCATCTCCCCC   635
IA16_1851g3     TAACACTTCCGGTGGCACCTCCGACAAACCTGCCGGCGGCGCAGGAACCGCATCTCCCCC   518
                ************************************************************


RO_1851g3       TTCCACTGGTAGCACCGGCTCCTCTCAGAGCGGGGACAACTCCGCTGGTGCCGCCTTAGG   695
IA16_1851g3     TTCCACTGGTAGCACCGGCTCCTCTCAGAGCGGGGACAACTCCGCTGGTGCCGCCTTAGG   578
                ************************************************************


RO_1851g3       TCTCAACGGCGTATCCTTGGCCCTTGCAACCTTCATGTCCTTTATGTTTGCTTAA  750
IA16_1851g3     TCTCAACGGCGTATCCTTGGCCCTTGCAACCTTCATGTCCTTTATGTTTGCT---  630
                **************************************************
```

2734g2

```
RO_2734g2        ATGAACGCATTCTTTTACGCCCTGATGTCACTTGCTATTGCTGCAACCAATGTCAGCGCT   60
IA16_2734g2      ------------------------------------------------------------   0


RO_2734g2        ATCGGTGAGACGATGACCTGCACTACCTATTCCGCATCGCCGAAAGGCTTTGCCTGCAAC   120
IA16_2734g2      ATCGGTGAGACGATGACCTGCACTACCTATTCCGCATCGCCGAAAGGCTTTGCCTGCAAC   60
                 ************************************************************


RO_2734g2        GACAGGCCCGACATTGTATGCACGGAGGGATGCAAGACATTCGTCACCAGCAGCCAATGC   180
IA16_2734g2      GACAGGCCCGACATTGTATGCACGGAGGGATGCAAGACATTCGTCACCAGCAGCCAATGC   120
                 ************************************************************


RO_2734g2        AAATTTGACAAGTACCCCAAGAAGCCGACGACTTCCGAGCTTTGCACCGTTGGGTTTGGA   240
IA16_2734g2      AAATTTGACAAGTACCCCAAGAAGCCGACGACTTCCGAGCTTTGCACCGTTGGGTTTGGA   180
                 ************************************************************


RO_2734g2        TCTACCAGTGCTACTACCAAGAGTAACCTTTTCCTTCCTCATTTTGCTGTTTGCCTTTGC   300
IA16_2734g2      TCTACCAGTGCTACTACCAAGAT---------------------------------TTGC   207
                 **********************                                 ****


RO_2734g2        ATTACTGGTCAGGGCTCCTTCAGCTGCACCGGCAAGTCCACCGGGTCCGCAAAGTGCTAC   360
IA16_2734g2      ATTACTGGTCAGGGCTCCTTCAGCTGCACCGGCAAGTCCACCGGGTCTGCAAAGTGCTAC   267
                 ******************************************** ************


RO_2734g2        GGTTGTGTGGCCTACAACAAAATCCCTTGGGCTTCATAA    399
IA16_2734g2      GGCTGTGTGGCCTACAACAAAATCCCTTGGGCTTCA---    303
                 ** ********************************
```

3338g6

```
RO_3338g6        ATGAGACCTGCCACTTTTCTGAGAGTGTTTGCCACATGGACCAGTCTCGCTCTTTCGACT   60
IA16_3338g6      ------------------------------------------------------------   0


RO_3338g6        TGCGCGGCCAGCATCGATTCACCTACCGTTGAACATTGCCCCTTCTACACCGGCGCCAAT   120
IA16_3338g6      ------GCCAGCATCGATTCACCTACCGTTGAACATTGCCCCTTCTACACCGGCGCCAAT   54
                       ******************************************************


RO_3338g6        ACTAACTCCGCGACCTGCACCACACGCTTCGAGTTCCGATGCGAAGGAGGATGCCAGAAG   180
IA16_3338g6      ACTAACTCCGCGACCTGCACCACACGCTTCGAGTTCCGATGCGAAGGAGGATGCCAGAAG   114
                 ************************************************************


RO_3338g6        AGCTTTGTCGAGGCTCAAGGATGTCTACCGGCCGATAATACCGGCGGCAATTTAACTGCG   240
IA16_3338g6      AGCTTTGTCGAGGCTCAAGGATGTCTACCGGCCGATAATACCGGCGGCAATTTAACTGCG   174
                 ************************************************************


RO_3338g6        CCAACGACTCGAACCTGTGATGTGGAATTTAGTATGGTCACGCCCCGCATCAACTCATGC   300
IA16_3338g6      CCAACGACTCGAACCTGTGATGTGGAATTTAGTATGGTCACGCCCCGCATCAACTCATGC   234
                 ************************************************************


RO_3338g6        GTCAGCAGTACTTATGGCAAGTTCTCATGCAAAGGACCCTCGTCAGGATTTGCGACTTGC   360
IA16_3338g6      GTCAGCAGTACTTATGGCAAGTTCTCATGCAAAGGACCCTCGTCAGGATTTGCGACTTGC   294
                 ************************************************************


RO_3338g6        TCTCAGTGCATCAGGGCTCTGC------ATGGATGCTCAGATCTCGTGCCTCTGAGTTAA   414
IA16_3338g6      TCTCAGTGCATCAGGGCTCGTGAGTCGCGTGAAATCCCTGGACATATGACCTTGAAAGCG   354
                 ******************         .**.*: * *:*.:*: .**.*  ***.: ..


RO_3338g6        ------------------------------------------------------------   414
IA16_3338g6      ATCCTTAGGCCTCACGCGTGGTGATTGCGGTCTAAAGACTCAATTCATCCGCGATCCATT   414


RO_3338g6        ----------------------------------------------------    414
IA16_3338g6      CCCACTTCTTTTCCGTCCAGTGCATGGACGCTCAGATCTCGTGCCTCTGAGT    466
```

CHAPTER 4. APPLICATION OF A U-NET NEURAL NETWORK TO THE *PUCCINIA SORGHI*-MAIZE PATHOSYSTEM

Katerina L. Holan[1], Charles H. White[2], and Steven A. Whitham[1]

[1]Department of Plant Pathology, Entomology, and Microbiology, Iowa State University, Ames, IA

[2]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO

Modified from a manuscript to be submitted to *Phytopathology*

Abstract

Phenotyping in phytopathology has become a critical area of research, particularly as pathogen distribution changes and the ability to overcome both innate resistance and control methods develop among pathogen populations. Computer vision approaches to analyze plant disease data can be both faster and more reliable than traditional, manual approaches. However, the requirement of manually annotating training data for the majority of machine learning applications can present a challenge for pipeline development. Here, we describe a machine learning approach to quantify *Puccinia sorghi* incidence on maize leaves utilizing U-Net convolutional neural network models. We analyze several U-Net models with increasing amounts of training image data, either randomly chosen from a large data pool, or randomly chosen from a subset of disease time course data. As training dataset size increases, models perform better, but the rate of performance decreases. Additionally, the use of a diverse training dataset can improve model performance and reduce the amount of annotated training data required for satisfactory performance. Models with as few as 48 training images are able to replicate the ground truth results within our testing dataset. The final model utilizing our entire

training dataset performs similarly to our ground truth data, with an intersection over union value of 0.5002 and an F1 score of 0.6669. This work illustrates the capacity of U-Nets to accurately answer real world plant pathology questions related to quantification and estimation of plant disease symptoms.

Introduction

Methods for automated disease phenotyping in phytopathology have become a significant area of research as more image and data acquisition technologies are developed and their respective costs of implementation decrease (Mutka and Bart 2015; Simko et al. 2017; Tanner et al. 2022). Phenotyping systems in plant pathology are often employed for identification of disease or the quantification of disease symptoms. An ability to accurately quantify disease can be useful for discovering small, additive effects of pest and pathogen control when testing pesticides or polygenic resistance traits, and it can inform disease mitigation timing and practices (Riaz et al. 2016; Riaz and Hickey 2017).

The enormous scale of the outputs of many plant phenotyping platforms means data analysis usually requires the assistance of computer vision or machine learning (ML) (Saleem et al. 2019). As a result, many computer vision approaches have been applied to various rust pathogens of plants, belonging to the order Pucciniales. Several members of Pucciniales are noteworthy for their significant impact on crops species, including many that infect grass species such as *Puccinia graminis* (wheat stem rust), *Puccinia striiformis*, (wheat stripe rust), and *Puccinia sorghi* (common rust of maize) (Figueroa et al. 2023). Many researchers still rely on standard area diagrams or scoring charts to estimate disease severity (Bade and Carmona 2011; Peterson et al. 1948). However, these charts are inherently qualitative in nature, and variation among their interpretations can lead to inaccurate conclusions on when and how often control

treatments are deployed (Bock et al. 2021). This is further exacerbated by a lack of adequate control methods for many rust pathogens, such as Asian soybean rust (*Phakopsora pachyrhizi*), where control is reliant on repeated applications of expensive fungicides to curb infections that would otherwise raze the crop (Nascimento et al. 2021). Additionally, climate change alters environmental pressures and management practices, which can change disease distribution and severity. For example, there have been several recent instances of *P. sorghi* and related species in previously unrecorded regions and new strains able to overcome currently utilized maize resistance (R) genes (Check et al. 2022; Halvorson et al. 2021; Quade et al. 2021; Ren et al. 2021). The rise of resistance in pathogen populations to various control methods makes the need for finer-tuned phenotypic assessments of these diseases imperative.

New phenotyping strategies for rusts have employed computer vision-based image processing to replace or complement use of standard area diagrams. For example, many rust disease symptoms are a contrasting color to surrounding leaf tissue, making it relatively easy to extract disease features based on thresholding of color ranges (Cui et al. 2010; Patil and Bodhe 2011; Ganthaler et al. 2018). Similar strategies have been employed for Poaceae-infecting rusts (Agarwal and Samantaray 2016; Yadav and Dutta 2018). However, thresholding in this manner is largely dependent on consistency in lighting, disease symptoms, and plant coloration, and may not be flexible enough for some datasets. Discrepancies in these features can result in a dramatic decrease in performance for a simple image processing pipeline.

As an alternative, ML or deep learning has been applied to rust pathosystems as well, as the variation between images can be larger as long as similar images are included in training datasets (Mochida et al. 2018; Xu et al. 2023), with several studies focusing on rust disease quantification of field-plot or aerial images (DeSalvio et al. 2022; Gao et al. 2020; Heineck et al.

2019; Mochida et al. 2018). Many ML methods for maize disease classification, including *Puccinia* species, have also been implemented (Mafukidze et al. 2022; Paliwal and Joshi 2022; Ullah et al. 2021). Regardless of approach, the proper use of computer vision techniques can lead to the development of a consistent and objective phenotyping pipeline that does not rely on prior knowledge or expertise. Although computer vision has its own biases and limitations, ML biases should be consistent across all images, as opposed to human biases, which can vary between scores, as well as inter- and intra-personally (Habib et al. 2022). Additionally, ML methods can easily be adjusted and reapplied to an entire experiment or existing datasets can be expanded to include new use cases.

There are limitations to using ML approaches, however. Most applications are restricted to narrow use-cases, meaning techniques or models developed for field data are likely not suited for individual leaf data. Another prominent hurdle to developing ML pipelines for disease identification and quantification is the reliance of most ML approaches on annotated training data, which can be tedious and time-consuming to generate and often requires significant domain expertise. This can be especially challenging for rust species, due to the small and numerous nature of rust pustules on plant tissue. To better understand the requirements for a robust ML model for quantifying rust diseases at a greenhouse scale, we utilized the *P. sorghi*-maize pathosystem, which develops numerous small, round to oval, brick-red to dark brown pustules called uredinia on maize leaf surfaces.

Very few studies have utilized ML for surface area quantification of specific diseases on maize leaves, and even fewer are at the single-plant scale. One recent example utilizes convolutional neural network (CNN) models to classify maize foliar diseases, including common rust, followed by quantification (Mafukidze et al. 2022). However, of the four diseases analyzed,

only quantification results for Northern corn leaf blight were shown. Another example is the PlantCV v2 naive Bayesian machine learning algorithm that was applied as a proof of concept for wheat stem rust (*Puccinia graminis* f. sp. *tritici*) (Gehan et al. 2017). Two other studies test the applicability of a Mask R-CNN algorithm to quantifying common rust symptoms on maize leaves (Gerber et al. 2021; Pillay et al. 2021).

In this study, we aimed to further investigate the applicability of deep learning to *P. sorghi*-maize image data to determine if the developed model could generate informative, quantitative data accurate enough to replace labor-intensive manual quantification. We focused on quantification of plant disease symptoms at a single leaf scale for use in greenhouse applications, as experiments related to plant-pathogen interactions or effector biology are often conducted at this scale. Because manually generated training annotations are quite labor-intensive, we also wanted to better understand the amount of annotated training data required to corroborate ground truth data. In line with this goal, we trained several U-Net CNN models with supervised learning using increasing amounts of training data. U-Nets consist of an encoder-decoder architecture with convolutional layers that operate at different spatial scales, and skip-connections allowing for the preservation of fine-scale detail in their output (Ronneberger et al. 2015; Zhou et al. 2018). To test the models, we generated two datasets from commonly conducted plant pathology experiments to serve as biologically meaningful data. Finally, we aimed to develop a pipeline to acquire, process, and quantitatively phenotype future datasets. Overall, we found that a relatively small amount of training data is required to obtain biologically meaningful results. However, it is imperative that training and validation datasets are diverse, especially with respect to small training datasets, which greatly improves model performance.

Methods

Differential resistance experiment

The maize genotype H95 *Rp1-D* was planted in peat-based substrate and grown in two separate growth chambers under 16 hour days and 8 hour nights at 25/21˚C. When plants were approximately two weeks old, they were inoculated with urediniospores of *P. sorghi* isolate IA16 or IN2 by dusting the spores manually onto seedlings. H95 *Rp1-D* is susceptible to IA16 and resistant to IN2. Plants were then misted with water and covered with a tall plastic cover and left in the dark for approximately 24 hours, after which the cover was removed, and the plants were returned to their previous day-night cycle. Nine days after the initial inoculation, leaf three from each plant was scanned. This experiment was repeated three times for each *P. sorghi* isolate. A total of 78 images were analyzed by the NN models.

Fungicide gradient experiment

Sweet corn (cultivar Golden Bantam), susceptible to both the IA16 and IN2 isolates, was planted in peat-based substrate and grown in a greenhouse with supplemental light at 16-hour days and 8-hour nights. Approximately three weeks after planting, the fungicide Tilt® (Syngenta, active ingredient propiconazole) was applied to the seedlings at the equivalent rates of 0, 0.5, 2.0, and 4.0 fluid ounces per acre. Five days after fungicide application, seedlings were spray-inoculated with urediniospores of the IA16 *P. sorghi* isolate. This time point was chosen to ensure pustules are still able to develop at the higher rates of fungicide, as fungicide efficacy is lowered (Mueller et al. 2004). Spores were collected from previously inoculated sweet corn seedlings and mixed with mineral oil. Using an airbrush sprayer connected to a compressor set to 30 PSI, seedlings were spray-inoculated evenly across each experiment. After spraying, the oil was allowed to dry, and the seedlings were misted with water and placed in a dark chamber for

24 hours before returning to the greenhouse. Nine days after rust inoculation, leaf three from each plant was scanned in three of the experimental replicates and leaf four was scanned in one of the replicates. The imaged leaf number for each experimental replicate was chosen so that the scanned leaf was both sprayed with fungicide and was still alive at the time of scanning. A total of 720 images from these experiments were analyzed by the NN models.

Leaf scanning protocol

Leaves used in the testing and training datasets were scanned using a Canon CanoScan LiDE220 flatbed scanner. Images were taken at 1,200 DPI, with varying numbers of leaves per scan, depending on leaf size. Young maize leaves at various time points during disease development were scanned to comprise the training dataset. Sweet corn variety Golden Bantam was the most common maize genotype imaged, but leaves from the inbred line H95 are also represented. Only leaves inoculated with *P. sorghi* isolate IA16 were included in the training dataset, but pustules between the IA16 and IN2 rust isolates have indistinguishable phenotypes. The flatbed scanner setup consisted of a blue cardstock (Astrobrights® 65 lb in Lunar Blue™) background and a label at the top of each scan. Leaves were cut close to the main stem and laid across the scanner glass face down so the adaxial surface was imaged. If the leaf was longer than the width of the scanner bed, the proximal end of the leaf was allowed to hang over the side so that the distal end, where the majority of pustules develop, was imaged. To enable automatic division of each scan into individual leaf images, leaves were laid out on the scanner bed so that a rectangular box could be drawn around the entire leaf, limiting overlap with other leaves as much as possible. Leaves were carefully laid out to minimize wrinkling. To keep them flat, another sheet of blue cardstock was used to slowly cover the leaves from one end to the other to prevent curling or twisting. Although this is not represented in the training or testing dataset,

current methods substitute the extra sheet of cardstock with several blue rubber bands similar in color to the background cardstock that wrap around the scanner bed to hold leaves down while setting up the rest of the leaves. This helps minimize wrinkling of the leaves and reduces the setup time. The scanner glass was cleaned with 75% ethanol and paper towels between each scan to reduce noise from loose spores or debris. Our training dataset pool was comprised of 510 images. These images contain leaves from multiple experiments and dates and have a wide variety of pustule and leaf phenotypes.

Scanned image processing

The scanned images were preprocessed to segment leaves into individual images. A Python script was used to threshold the leaf regions and crop out individual leaves. First, the edge pixels and labels are removed, and an Otsu threshold is applied to the blue channel of each scan to identify all leaf regions. Based on the leaf region, a box is drawn around each leaf and cropped, resulting in individual leaf images. During the leaf thresholding, the total leaf area in pixels is saved for each leaf. Occasionally, the bounding boxes of leaves overlap, resulting in images with additional partial or full leaves. Such images were treated as “one” leaf, and the total leaf area was saved for each image. These images were annotated and processed the same as truly individual leaf images.

Annotation of datasets

All ground truth pustule annotations were marked using the ellipse tool in ImageJ Fiji and then concatenated into one CSV file for each image in the training and testing dataset (Schindelin et al. 2012). In the 510 images of the training dataset, there were 53,714 total annotations, which include 22,037,886 positive pustule class pixels. There were a total of 12,156

annotations in the 720 images of the testing dataset, which includes 4,002,338 positive class pixels.

Training of U-Net models

All U-Net models were trained with a batch size of 24 256x256 pixel images. The specific 256x256 segments of each image were chosen so that every pixel in the training dataset was represented at least once. Training data was augmented with the Albumentations package (Buslaev et al. 2020) using the horizontal flips, vertical flips, random 90° rotations, transpose, and random sized crop functions. Each 256x256 pixel segment in each batch had a 50% chance of being horizontally flipped, a 50% chance of being vertically flipped, a 75% chance of a random 90° rotation, a 50% chance of transposition, and a 25% chance of a random size crop (minimum size of 128x128 pixels). We used the Adam optimizer with an initial learning rate of 5e-4 (Kingma and Ba 2015). Performance on a validation set was calculated every 250 iterations. The learning rate is reduced by a factor of 10 after 750 iterations of no improvement on the validation set to a minimum rate of 5e-6. Training was stopped after 1,500 iterations with no improvement. The models are trained with focal loss with α=0.25, and γ=2.0 (Lin et al. n.d.).

Relative to Zhou et al. (2018), we used one fewer downsampling operation in the backbone (denoted as Unet ++ L3), used 32 filters per layer with a 3x3 kernel, and increased the number of filters by 32 after each downsampling layer, relative to doubling in the original implementation. These changes were made to decrease the total number of trainable parameters and decrease computational expense, particularly for use on systems without GPUs. We found that these changes did not substantially change the performance of these models for our application.

Approximately 20% of each model's training set was randomly chosen as the validation set, with training images rounded up to the nearest integer and validation images rounded down to the nearest integer. A best threshold (BT) that optimized intersection over union (IoU) of the validation set and a 0.5 threshold were tested for each model. The neural network was developed in TensorFlow and Keras (Abadi et al. 2016; Harris et al. 2020), with additional processing performed using NumPy (Chollet, F. & others, 2015. Keras. Available at: https://github.com/fchollet/keras).

The name of an image not present in our 510-image training dataset was added when randomizing data for the UTC models. As a result, some models were trained on one fewer image. These models were: UTC1-12, UTC4-12, UTC5-24, UTC2-48, UTC3-48, UTC5-48, UTC1-96, UTC2-96, UTC3-96, UTC4-96, UTC5-96, UTC1-192, UTC2-192, UTC3-192, UTC4-192, and UTC5-192. However, based on the results of these models, it does not appear one fewer image had a measurable impact on their performance.

Metrics and performance analysis

Each model was tested on the entire 798 image testing dataset. The overall pixel IoU of the positive pustule class was used to estimate performance for each of the 70 models, which was calculated from the total number of true positive predicted pustule pixels over the intersection of true positive, false positive, and false negative pixels over the entire testing dataset. The F1 score was calculated from the harmonic mean of pixel precision and recall. The negative class, namely leaf and background, was not considered in performance metrics as it tends to be uninformative or misleading when it makes up a large portion of the dataset, as is the case in our data.

Results

NN training dataset

The training dataset consisted of 510 images of maize leaves that have been inoculated with *P. sorghi* urediniospores and their respective annotations. Leaves were imaged with a portable flatbed scanner at a 1,200 dots per inch (DPI) resolution, with time points ranging from four to fifteen days after rust inoculation (DAI). Since *P. sorghi* takes approximately seven days from urediniospore contact with a leaf to form a developed uredinium, these time points captured images of common rust of maize throughout the disease progression from the first appearance of pustules at four DAI to fully developed and spore-shedding pustules at seven to fifteen DAI.

NN testing dataset

The testing dataset consisted of 798 images taken at nine DAI gathered from two plant pathology experiments that simulate real-world applications for our models. The first experiment, with 78 images, is a differential disease response trial where the maize line H95 *Rp1-D* was inoculated with two *P. sorghi* isolates, IA16 and IN2. The second experiment, consisting of 720 images, is a fungicide gradient experiment in which the fungicide Tilt® (Syngenta, active ingredient propiconazole) was sprayed onto sweet corn seedlings at increasing concentrations before inoculation with the *P. sorghi* isolate IA16.

Annotation of datasets

Pustules on each whole-leaf image in the training and testing datasets were hand-annotated with the ellipse tool in Fiji/ImageJ to mark the positive pustule class. Coordinates for the individual pustule annotations were collated and saved to a single CSV file for each image. Any non-annotated pixels were deemed the non-pustule, negative class, which includes both non-pustule leaf tissue and background. Approximately 100 person-hours were required to annotate the 1,308 images of the training and testing datasets. A total of 53,714 regions were

annotated in the training dataset, and a total of 12,156 regions were annotated in the testing dataset.

NN architecture and training strategies

Using a U-Net (Ronneberger et al. 2015) convolutional neural network (Lecun et al. 2015) to perform pixel-wise segmentation of the training images, we trained a total of 70 models. Specifically, we used a variation of the original U-Net known as U-Net++, which adds nested convolution layers between encoder and decoder to allow for deeper representations at finer spatial scales (Zhou et al. 2018). During training, the Albumentations package was used to augment the image data with the horizontal flip, vertical flip, random 90˚ rotation, transpose, or random crop functions (Buslaev et al. 2020). The model predicts the probability that a particular pixel belongs to the pustule class, on a scale of 0 to 1. For each model, 20% of the training dataset was randomly chosen as the validation set and two thresholds were tested for each model. Thirty-five models, designated U-Net random (UR), were trained on random subsets from the training data pool. Models were further divided into model groups, with each group being trained on different amounts of training data, namely 6, 12, 24, 48, 96, 192, or 510 images. The training process was repeated five times for each model group. The other 35 models, named U-Net time course (UTC), were trained with data first picked from a two-week common rust time course experiment. These models were divided similarly into model groups. The time course data consists of 76 images taken from 4 DAI, when leaves are young and common rust symptoms first start appearing, until 15 DAI, where pustules are fully developed and maize leaves are dying or dead at their tips. To ensure as much diversity of training images was present for each model without being overly biased, leaves were randomly and evenly picked from each DAI. When a model group required more images than were available in the time course subset, namely 96 and

up, the rest of the training data was chosen randomly from the remainder of the 510 image pool. For example, for the models in group 6, we randomly picked one leaf each from either 4 or 5 DAI, 6 or 7 DAI, 8 or 9 DAI, 10 or 11 DAI, 12 or 13 DAI, and 14 or 15 DAI, to ensure the training data contained images across the entire time course. For models in groups 12 and up, an image was randomly picked from each DAI until the training dataset for a particular model was complete.

We tested two thresholding methods for each model to determine the cutoff point for classifying pixels as the pustule class, namely a set threshold of 0.5 and a model-specific BT, where the threshold that maximized IoU of the validation set was used. Results for each of the 798 testing images were generated for each model at each threshold, for a total of 140 results files. Approximately 48 hours were required to train the 70 models and generate the 140 evaluation files on the testing datasets in sequence on an Nvidia Quadro RTX 6000 GPU.

Number of training images affects model performance

As model group increases, model performance also increases, both for UR and UTC models (Figure 1a). When considering a single training strategy, we see this effect is more pronounced in the UR models than in UTC models (Figure 1a). The largest improvement in IoU occurred between model groups 6 and 12 and groups 12 and 24, particularly for the UR models (Figure 1a). After group 96, the mean IoU value remains fairly stagnant, with no or minimal improvement between model groups, and in the case of the UTC models, a slight reduction in improvement (Figure 1a). Variability in performance between models in the same group tends to decrease as the training dataset expands. Model group 6, followed closely by group 12, have the widest distribution of IoU values, and distributions get tighter as training images are added until model group 96, at which point variability increases somewhat, but not to the same extent as in

the lowest two groups (Figure 1a). These trends are consistent at an individual leaf image level, where we tend to see an improvement in IoU as training dataset size increases, and as more pustules are correctly identified (Figure 1b). Overall, models tend to perform better, both in terms of higher IoU and reduced variability, when more training data is used, but the added benefit of additional data greatly decreases when the number of images in the training datasets exceeds 96.

Training strategy affects model performance

Although similar trends in performance were observed within each training strategy, the UTC training strategy consistently performed better than the UR strategy (Figure 1a). This advantage is most obvious in model groups 6 and 12 and is seen to a lesser extent in groups 24, 48, and 96, as the size of the advantage decreases as overall performance gained from additional training data also decreases (Figure 1a). Moreover, the worst performing UTC model, with a BT IoU of 0.303, performs 26 times better than the worst performing UR model, which has a BT IoU of 0.0115. Six UR models, three each from model groups 6 and 12, perform worse than the lowest performing UTC model, UTC5-6 (Figure 1a). Again, although both training strategies tighten the distribution of model performance in each model group, the UTC model groups consistently are less variable in performance (Figure 1a). The UTC training strategy seems to offer a substantial advantage over a fully random UR strategy, particularly at lower model groups, giving a higher likelihood of better model performance and reduced variability between models in the same group.

Diversity of training images affects model performance

Models trained on even the smallest dataset have the potential to be high performing, but they also have a greater likelihood of poor performance, particularly for UR models (Figure 1a).

This disparity in performance is quite obvious in UR model group 6, which had a large range in IoU between the two best and the two worst performing models (Figure 1a). To investigate potential reasons for this result, we looked at the training image data for these four models, namely UR1-6, UR4-6, and UR2-6, UR5-6, which were the two best and two worst UR-6 models, respectively. The training images, and particularly the validation subset, for the best UR-6 models were drastically more diverse in both leaf and pustule phenotype than the worst UR-6 models (Figure 2).

Both of the worst performing models had a validation image with almost no pustules, as opposed to the best performing UR-6 models, of which each validation image had 175 or more annotated pustules (Figure 2, Table 1, Supplemental Table 1). The best UR-6 models training datasets also included images of yellowing leaves, a common occurrence in our datasets, of which none were present in the worst UR-6 models (Figure 2). In summary, the best performing UR-6 models had images that were much more representative of the training and testing datasets as a whole than the worst performing UR-6 models.

Threshold performance depends on training strategy and model group

The two thresholds, BT and 0.5, greatly affect performance metrics, with BT outperforming a 0.5 threshold in most models (Figure 3a,c). Generally, the BT is determined to be lower than 0.5, and as a result, the model outputs more false positives but fewer false negatives (Supplemental Table 2). In other words, fewer true positives are missed by the model when utilizing the BT. The BT, however, leads to poorer performance in UR models with fewer training images, namely groups 6, 12, and 24 (Figure 3a). One of these models, UR1-12, performs especially poorly when using the BT, dramatically reducing the IoU value (Figure 3b). When the differences in IoU between the 0.5 threshold and the BT were compared, we found

only two UTC models had a lower performance with the BT, as compared to 17 UR models (Figure 3c). In conjunction with previous results, it appears that a best threshold as determined by the validation set performs better than a set threshold of 0.5, given a diverse training and validation set.

All models can distinguish between binary positive and negative data

The inbred maize line H95 is generally susceptible to *P. sorghi* isolates, however, H95 lines carrying different *Rp* resistance genes are resistant to select rust isolates. An H95 line carrying the common rust resistance gene *Rp1-D* is resistant to *P. sorghi* isolate IN2 (Hulbert 2002), but it is susceptible to an Iowan isolate that we named IA16. These two isolates were used in a binary positive and negative test for the NN models, where maize plants were inoculated with one of these two *P. sorghi* isolates. Overall, no pustules developed on the H95 *Rp1-D* leaves inoculated with IN2, whereas leaves inoculated with IA16 consistently developed varying numbers of pustules (Figure 4a-c). When assessing model performance, we looked at total pustule predictions per image and the percent coverage of pustules on leaf tissue. Since U-Nets are unable to differentiate between individual pustules that overlap, a percent coverage could give a more accurate result over the number of pustule predictions, particularly for leaves with high amounts of pustules. In the ground truth results, we found the means for both total pustule annotations and percent coverage of the IN2 and IA16 inoculated leaves were significantly different for both annotations and pustule coverage (α=0.05, Figure 4a,b). When the same two-sample t-test was conducted for each model and threshold level, all models were able to correctly determine that the IA16 and IN2 inoculated leaves had statistically different pustule coverages (Table 2).

UTC models are most likely to corroborate fungicide gradient ground truth results

To generate images with a gradient of disease symptoms, maize seedlings were inoculated with *P. sorghi* isolate IA16 five days after being sprayed with fungicide at the equivalent rates of 0, 0.5, 2, and 4 fluid ounces per acre, which were expected to provide increasing levels of protection (Mueller and Buck 2004). In line with this expectation, we observed that as the rate of fungicide increased, the pustule coverage decreased (Figure 4d, 4e). Overall, in an ANOVA and follow-up Tukey HSD, we found that the 0 and 0.5 fungicide rates were not statistically different from each other, and the 2 and 4 rates were not statistically different from each other, with all other pairwise comparisons found to be significantly different for both total number of pustules and pustule coverage (α=0.05, Figure 4d, 4e, Supplemental Table 3). For this statistical test, fewer models were able to match the ground truth results. Models with smaller training datasets, UR models, and UR models using the BT were all less likely to corroborate ground truth results (Table 2). Overall, UTC models were the most likely to match ground truth results, with 33 out of 35 models being able to do so at both thresholds (Table 2). Based on the statistical testing, the UR models in model group 48 or higher and UTC models in model group 12 or higher give similar f-statistics and p-values to the ground truth data between fungicide treatment groups.

Identifying a true mean of zero is difficult for all tested models

The general trends within the ground truth data can be correctly identified by the majority of the models tested, but we further investigated the IN2 data. The leaves in these images never have any pustules, making this a challenging test for our models. Essentially, a perfect score would have no false positives, which is difficult to achieve. We tested our models against the null hypothesis that the total annotations and pustule coverage of IN2-inoculated leaves is zero,

which gives us a ground truth p-value of 0.162 for both. In other words, there is not enough evidence to support the IN2 true mean is not zero, given $\alpha=0.05$. None of the models were able to correctly predict zero pustules, and all had p-values under 0.05 for number of predicted pustules, pustule coverage, or both. However, most models using the best threshold have a predicted mean close to zero, especially the UTC models, of which all predicted pustule means are less than 20 pustules per image, with the majority being less than 10 (Figure 5a). The exception to this is a few of the UR models in smaller model groups, which greatly overestimate the amount of pustule coverage on the IN2-inoculated leaves (Figure 5). Although no models are able to statistically show that the IN2 mean equals zero, the majority of models, and in particular UTC models in bigger model groups, have a very low predicted mean. We expect that it is unlikely a researcher would come to an incorrect conclusion, namely H95 *Rp1-D* leaves are not resistant to isolate IN2, when interpreting the predicted results for those models.

Final model yields similar results to ground truth annotations

We chose our best performing 510 model to compare directly to ground truth annotations, namely UTC4-510 using the validation set-determined best threshold. When comparing UTC4-510's individual image results to the ground truth results, we see a similar trend for both the differential experiment and the fungicide gradient experiment (Figure 6a-d). The model tends to overestimate the number of pustules and pustule coverage but predicts a large number of pustules for IA16-inoculated leaves, few pustules on IN2-inoculated leaves, and a decreasing pustule coverage on leaves treated with increasing rates of the Tilt® fungicide (Figure 6a-d). This model is also capable of outperforming hand-annotated data in specific instances. For example, the model correctly identified a pustule that had been missed during manual annotation (Figure 6e). In contrast, for an image with 8 pustule annotations, nearly all models had an IoU of 0 for that

image. Upon closer inspection, none of the annotations were clearly discernible pustules, i.e., they were falsely marked (Figure 6f). UTC4-510 had a total of 4 predictions for this leaf, but the total area of pustules was very small, and was an improvement over the manual, ground truth annotations. In summary, although pustule coverage tends to be slightly overestimated by the model in the training dataset, UTC4-510's results remain relatively similar to the ground truth.

Discussion

A significant bottleneck to beginning a new phenotyping project that employs ML is the annotation of high-quality training data. For our *P. sorghi* system, we have a large collection of potential training images to choose from. However, we wanted to simulate different levels of training data to better inform the development of a ML pipeline in a new pathosystem. From our results, additional training data is a large contributor to model success, and the likelihood of a particularly poor performing model is greatly decreased as more training data is used. However, we consistently found that additional training data had diminishing returns in model performance. The 10-fold difference in manual annotation time between model group 48 and model group 510 when compared to the negligible return in model performance implies smaller amounts of training data may have been sufficient in this use-case.

Ensuring that the training data captures the phenotypic diversity of a given pathosystem helped to increase model success. Utilizing diverse disease time course data proved to be a big advantage for smaller training datasets, which informs the types of training data to preferentially acquire for future ML applications in phytopathology. The training data in our poorest performing models UR-6, which were randomly selected from our entire training pool, had much lower pustule counts and the leaf diversity was minimal. In contrast, the training data in our highest performing UR-6 models, had many more pustule annotations and included both mostly

green leaves and leaves with yellowing leaf tips. Inclusion of these yellowing leaves is especially important for experiments with additional treatments beyond *P. sorghi* inoculation, such as the fungicide gradient experiment. Scanned leaves from this experiment are older, as they needed an additional fungicide treatment before being inoculated with urediniospores. As leaf tips senesce, it becomes more difficult to distinguish pustules from dying leaf tissue, and inclusion of this phenotypic diversity in training data seems to be imperative for model success. Utilization of disease time course experiments to gather training data enables the collection of comprehensive diversity of phenotypes quickly and easily.

Interestingly, the diversity of the training dataset has a close relationship with which threshold technique performed best. For UR models that performed poorly, it was much more likely for a set threshold of 0.5 to give a better overall IoU value. When these models used the BT, they performed much worse (Figure 5). Selecting a threshold based on a validation set with low diversity likely overfits to that narrow phenotype, greatly impacting performance on the significantly more diverse testing dataset. Ensuring diverse training and validation sets enables the use of the BT strategy, increasing the likelihood of better model predictions and overall performance.

One specific advantage ML phenotyping can have over manual annotation or estimation is an equalization of biases across images. Visual and manual assessments are subject to various biases, extent of background knowledge, or even time allotted for phenotyping, all of which can affect the results of an experiment. Although ML has its own biases and errors, a given model will perform nearly identically every time on a given set of data. This can increase reproducibility both within and between experiments. In fact, although our final model was not perfect, it is able to out-perform manual annotation in some instances.

The final model from this study is likely specific to rust symptoms on Poaceae species on images from smaller scale greenhouse experiments. There are limitations on the use-cases for this model, as all leaves in this study were gathered by the same person using the same scanner and background color, and the plants were scanned when they were still relatively young. However, future datasets could be expanded to include data gathered from other research groups using different equipment, older leaves, or from field conditions to increase the robustness of the model. Additional testing would be required to determine if this model architecture works with different data, such as different background colors, a different scanner, or older maize leaves. On the other hand, although not included in our testing data, some training leaves had mosaic symptoms caused by foxtail mosaic virus, and this model may perform well on similarly infected leaves. Furthermore, although this model would likely have limited applicability to significantly different pathosystems, diseases with similar phenotypes, such as Southern rust of maize (*Puccinia polysora*) may be phenotyped with similar success.

Conclusions

As more methods to collect plant phenotypic data are developed, there is an increasing need for image processing pipelines. Even when utilizing the same data acquisition platform, different problems and datasets will have varying image processing requirements. Additionally, high-throughput, accurate, and quantitative phenotyping is needed for plant disease research to inform breeding decisions and aid in research involving plant immunity, resistance gene and effector biology, and polygenic pathogen resistance traits.

Starting a plant phenotyping platform from scratch can be intimidating for researchers primarily trained as biologists. To better understand the minimum requirements for a new phenotyping pipeline, we mimicked various levels of annotated training data and possible

outcomes at those levels to inform future pipeline development. Based on our results, a relatively small amount of data may be sufficient to begin development of accurate ML models. However, it is important to note that training data should be as diverse as possible and include a large number of annotations. Models trained on data with limited diversity and few annotations performed extremely poorly, whereas ensuring diversity by utilizing disease time course data helped to maximize model performance, especially for models trained on smaller training datasets.

Overall, we found that a U-Net neural network is sufficient to quantify common rust symptoms on scanned maize leaves. With at least a 48-image training dataset, we were able to corroborate ground truth results for both a differential experiment and a fungicide gradient experiment. Models trained with larger training datasets, and thus requiring more time devoted to annotation of training data, did not have an equal return in performance, but are likely more robust when analyzing future datasets. Our final model, UTC4-510, generates results similar to ground truth. Additionally, although false negatives and false positives can occur, model-predicted pustule regions are generally tighter than is feasible by manual annotation and models are able to outperform manual annotations in some instances. Our results show that ML models hold significant promise for quantifying plant disease in greenhouse experiments and relatively little annotated training data can be used to begin development of new pipelines. Reduced upfront costs in annotation time can hasten future disease quantification pipelines and model training, with the outcome being a more consistent and reliable tool for disease quantification.

Acknowledgements

This research was funded by the Agricultural and Food Research Initiative grant no. 2019-07318 from the USDA National Institute of Food and Agriculture, the Iowa State

University Predictive Plant Phenomics graduate training program funded by the National Science Foundation (DGE #1545453), and the Iowa State University Plant Sciences Institute. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders.

We would also like to thank John Shriver and Alison Robertson, for input into the design of the fungicide experiment, John Shriver for the application of the fungicide, Dave Berger for feedback on the testing experimental designs, Baskar Ganapathysubramanian, for input into the initial setup of leaf scanning, Zaki Jubery for assistance with the scanned image preprocessing workflow, and Peter Balint-Kurti and Saet-Byul Kim for providing the *P. sorghi* IN2 isolate.

Author Contributions

KH generated all leaf image data, designed the testing data experiments, and annotated the majority of the training dataset. KH also analyzed the neural network results, the ground truth results, and the ground truth and neural network comparisons and composed the manuscript. CW annotated the majority of the testing dataset, developed the machine learning approach, produced the U-Net pipeline, and developed the U-Net models and results files. SW was involved in the experimental design and writing and editing the manuscript.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. ArXiv.

Agarwal, C., and Samantaray, S. D. 2016. A Novel Image Processing based Approach for Identification of Yellow Rust in Wheat Plants. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 6:2277.

Bade, C. I. A., and Carmona, M. A. 2011. Comparison of methods to assess severity of common rust caused by *Puccinia sorghi* in maize. Trop. Plant Pathol. 36:264–266.

Bock, C. H., Chiang, K.-S., and Del Ponte, E. M. 2021. Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. Trop. Plant Pathol. 2021. :1–18.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. 2020. Albumentations: Fast and flexible image augmentations. Inf. 11:125.

Check, J. C., Aime, M. C., Byrne, J. M., and Chilvers, M. I. 2022. First Report of Southern Rust (*Puccinia polysora*) on Corn (*Zea mays*) in Michigan. Plant Dis. 106:2262.

Cui, D., Zhang, Q., Li, M., Hartman, G. L., and Zhao, Y. 2010. Image processing methods for quantitatively detecting soybean rust from multispectral images. Biosyst. Eng. 107:186–193.

DeSalvio, A. J., Adak, A., Murray, S. C., Wilde, S. C., and Isakeit, T. 2022. Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms. Sci. Rep. 12:1–14.

Figueroa, M., Dodds, P. N., Henningsen, E. C., and Sperschneider, J. 2023. Global Landscape of Rust Epidemics by *Puccinia* Species: Current and Future Perspectives. In Plant Relationships, The Mycota, Springer, Cham, p. 391–423..

Ganthaler, A., Losso, A., and Mayr, S. 2018. Using image analysis for quantitative assessment of needle bladder rust disease of Norway spruce. Plant Pathol. 67:1122–1130.

Gao, Z., Luo, Z., Zhang, W., Lv, Z., and Xu, Y. 2020. Deep Learning Application in Plant Stress Imaging: A Review. AgriEngineering. 2:430–446.

Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., et al. 2017. PlantCV v2: Image analysis software for high-throughput plant phenotyping. PeerJ. 2017:e4088.

Gerber, M., Pillay, N., Holan, K. L., Whitham, S. A., and Berger, D. K. 2021. Automated Hyper-Parameter Tuning of a Mask R-CNN for Quantifying Common Rust Severity in Maize. In *Proceedings of the International Joint Conference on Neural Networks*, Institute of Electrical and Electronics Engineers Inc.

Habib, A., Abdullah, A., and Puyam, A. 2022. Visual Estimation: A Classical Approach for Plant Disease Estimation. In Trends in Plant Disease Assessment, Springer, Singapore, p. 19–45..

Halvorson, J., Kim, Y., Gill, U., and Friskop, A. 2021. First report of the southern corn rust pathogen *Puccinia polysora* on *Zea mays* in North Dakota. Can. J. Plant Pathol. 43:S352–S357.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. 2020. Array programming with NumPy. Nature. 585:357–362.

Heineck, G. C., McNish, I. G., Jungers, J. M., Gilbert, E., and Watkins, E. 2019. Using R-Based Image Analysis to Quantify Rusts on Perennial Ryegrass. Plant Phenome J. 2:1–10.

Kingma, D. P., and Ba, J. L. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*, ICLR 2015 - Conference Track Proceedings.

Lecun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. Nature. 521:436–444.

Mafukidze, H. D., Owomugisha, G., Otim, D., Nechibvute, A., Nyamhere, C., and Mazunga, F. 2022. Adaptive Thresholding of CNN Features for Maize Leaf Disease Classification and Severity Estimation. Appl. Sci. 12:8412.

Mochida, K., Koda, S., Inoue, K., Hirayama, T., Tanaka, S., Nishii, R., et al. 2018. Computer vision-based phenotyping for improvement of plant productivity: A machine learning perspective. Gigascience. 8:1–12.

Mueller, D. S., Jeffers, S. N., and Buck, J. . W. 2004. Effect of timing of fungicide applications on development of rusts on daylily, geranium, and sunflower. Plant Dis. 88:657–661.

Mutka, A. M., and Bart, R. S. 2015. Image-based phenotyping of plant disease symptoms. Front. Plant Sci. 5.

Nascimento, R. S. M., Ferreira, L. R., Zambolim, L., Parreira, D. F., da Costa, Y. K. S., Damascena, J. F., et al. 2021. Spray mixture volume in the control of Asian soybean rust. Crop Prot. 146:105662.

Paliwal, J., and Joshi, S. 2022. An Overview of Deep Learning Models for Foliar Disease Detection in Maize Crop. J. Artif. Intell. Syst. 4:1–21.

Patil, S. B., and Bodhe, S. K. 2011. Leaf disease severity measurement using image processing. Int. J. Eng. Technol. 3:297–301.

Peterson, R. F., Campbell, A. B., and Hannah, A. E. 1948. A Diagrammatic Scale for Estimating Rust Intensity on Leaves and Stems of Cereals. Can. J. Res. 26c:496–500.

Pillay, N., Gerber, M., Holan, K. L., Whitham, S. A., and Berger, D. K. 2021. Quantifying the Severity of Common Rust in Maize Using Mask R-CNN. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, p. 202–213.

Quade, A., Ash, G. J., Park, R. F., and Stodart, B. 2021. Resistance in Maize (*Zea mays*) to Isolates of *Puccinia sorghi* from Eastern Australia. Phytopathology. 111:1751–1757.

Ren, J., Li, Z., Wu, P., Zhang, A., Liu, Y., Hu, G., et al. 2021. Genetic Dissection of Quantitative Resistance to Common Rust (*Puccinia sorghi*) in Tropical Maize (*Zea mays* L.) by Combined Genome-Wide Association Study, Linkage Mapping, and Genomic Prediction. Front. Plant Sci. 12.

Riaz, A., and Hickey, L. T. 2017. Rapid phenotyping adult plant resistance to stem rust in wheat grown under controlled conditions. In Methods in Molecular Biology, Humana Press, New York, NY, p. 183–196.

Riaz, A., Periyannan, S., Aitken, E., and Hickey, L. 2016. A rapid phenotyping method for adult plant resistance to leaf rust in wheat. Plant Methods. 12:1–10.

Ronneberger, O., Fischer, P., and Brox, T. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, p. 234–241.

Saleem, M. H., Potgieter, J., and Arif, K. M. 2019. Plant disease detection and classification by deep learning. Plants. 8:468.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. 2012. Fiji: An open-source platform for biological-image analysis. Nat. Methods. 9:676–682.

Simko, I., Jimenez-Berni, J. A., and Sirault, X. R. R. 2017. Phenomic approaches and tools for phytopathologists. Phytopathology. 107:6–17.

Tanner, F., Tonn, S., de Wit, J., Van den Ackerveken, G., Berger, B., and Plett, D. 2022. Sensor-based phenotyping of above-ground plant-pathogen interactions. Plant Methods. 18:1–18.

Ullah, K., Jan, M. A., and Sayyed, A. 2021. Automatic Diseases Detection and Classification in Maize Crop using Convolution Neural Network. Int. J. Adv. Trends Comput. Sci. Eng. 10:675–679.

Xu, L., Cao, B., Zhao, F., Ning, S., Xu, P., Zhang, W., et al. 2023. Wheat leaf disease identification based on deep learning algorithms. Physiol. Mol. Plant Pathol. 123:101940.

Yadav, A., and Dutta, M. K. 2018. An Automated Image Processing Method for Segmentation and Quantification of Rust Disease in Maize Leaves. Int. Conf. Computational Intell. Commun. Technol. CICT 2018.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. 2018. Unet++: A nested u-net architecture for medical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer, Cham., p. 3–11.

Appendix. Data Availability

The entirety of the training and testing datasets are available here: https://doi.org/10.25380/iastate.23264180. This data repository includes the raw TIF files for each cropped image and its corresponding CSV annotation file and all 140 non-summarized model analysis CSV files.

For more information on the code and techniques used in this publication, see the GitHub repository at https://github.com/katholan/Psorghi-UNet-Quantification. Additionally, the names of all images used to train each model (training_images.xlsx), the results summaries for each model (model_results_summaries.xlsx), the p-values (α=0.05) for all statistical tests for each model (statistical_test_results.xlsx), and the model summaries and statistical results for only the IN2 leaves (in2_summaries_analyses.xlsx).

Figures

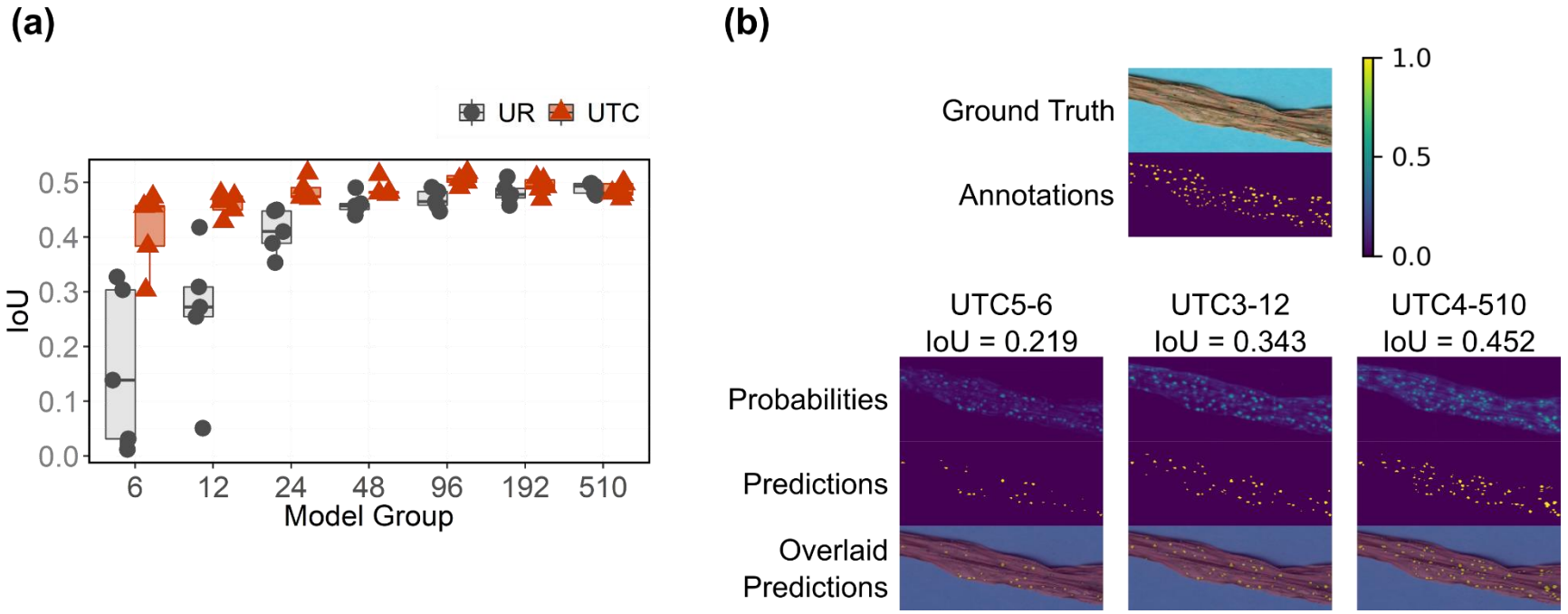


Figure 1. The effect of size of training data set on neural network (NN) model performance. Performance of each NN model as measured by the overall intersection over union (IoU) of the testing dataset for each model using the best threshold (BT) to determine predictions. Models are grouped by number of images used in the training set. "UR" denotes U-Net models generated with randomly selected images. "UTC" denotes U-Net models generated with images selected from a time course spanning common rust disease development. Whiskers represent 1.5 times the interquartile range, with points outside classified as outliers. (a) The performance according to IoU for all 70 models separated by training strategy. (b) A portion of a single ground truth image and its respective annotations. The pixel probabilities for the positive pustule class, their resulting predictions using the BT, and those predictions overlaid on the ground truth image are shown for three models using increasing amounts of training data (UTC5-6, UTC3-12, and UTC4-510).

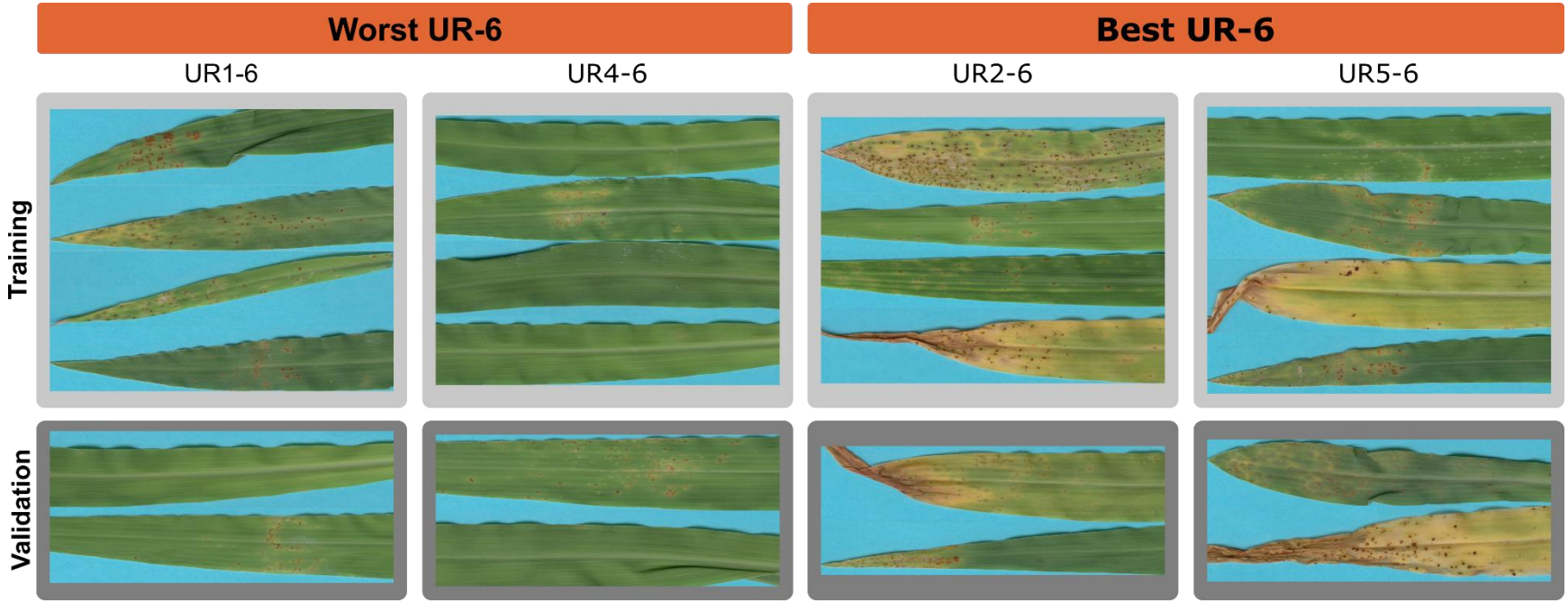


Figure 2. The effect of image diversity in small training data sets on neural network (NN) model performance. The images used in four of the UR-6 models that were trained using six randomly selected leaves: the two UR-6 models that produced the lowest IoU values (Worst UR-6), the two UR-6 models that produced the greatest IoU values (Best UR-6). Images are in scale with each other. Each image was cropped at a location that best represented the phenotypes within the image. The bottom two leaves of each model comprise the validation set for that model. The leaves used for UR1-6 and UR4-6 had minimal pustule and leaf diversity. The leaves used for UR2-6 and UR5-6 had large numbers of pustules and were more diverse in appearance.

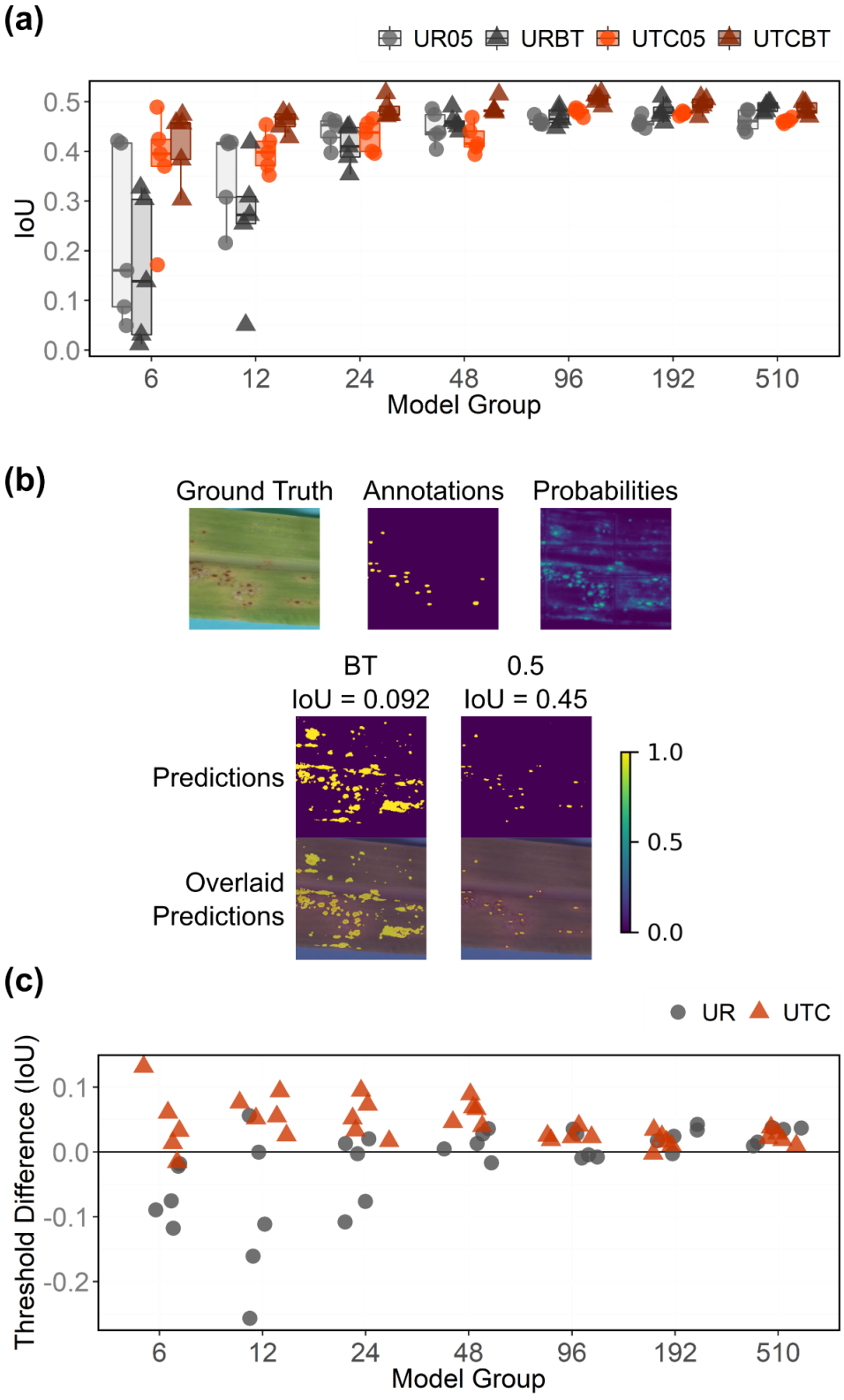


Figure 3. The effect of threshold on NN model performance as determined by the overall IoU of the testing dataset for each model. The models are separated by training strategy and threshold.

Models are grouped by the number of images used for training. Whiskers represent 1.5 times the interquartile range, with points outside classified as outliers. (a) IoU values for each of the model groups. UR05 denotes U-Net models trained with randomly selected images and a threshold of 0.5. URBT denotes U-Net models trained with randomly selected images and the BT. UTC05 denotes U-Net models generated with images selected from a time course spanning common rust disease development and threshold of 0.5. UTCBT denotes U-Net models generated with images selected from a time course spanning common rust disease development and the BT. (b) An example of thresholds on a portion of a single testing image analyzed by UR1-12. The original ground truth and its corresponding annotations are shown alongside UR1-12's pixel probabilities. Below are the predictions and predictions overlaid on the ground truth image for both thresholds. A BT in this case produces an IoU of 0.09185438 while a 0.5 threshold produces an IoU of 0.45368737. (c) The difference in IoU value between the BT and the 0.5 threshold for each model. Models are grouped by number of training images used in the training dataset. A positive value indicates the BT performs better and a negative value indicates that the 0.5 threshold performs better for a given model.

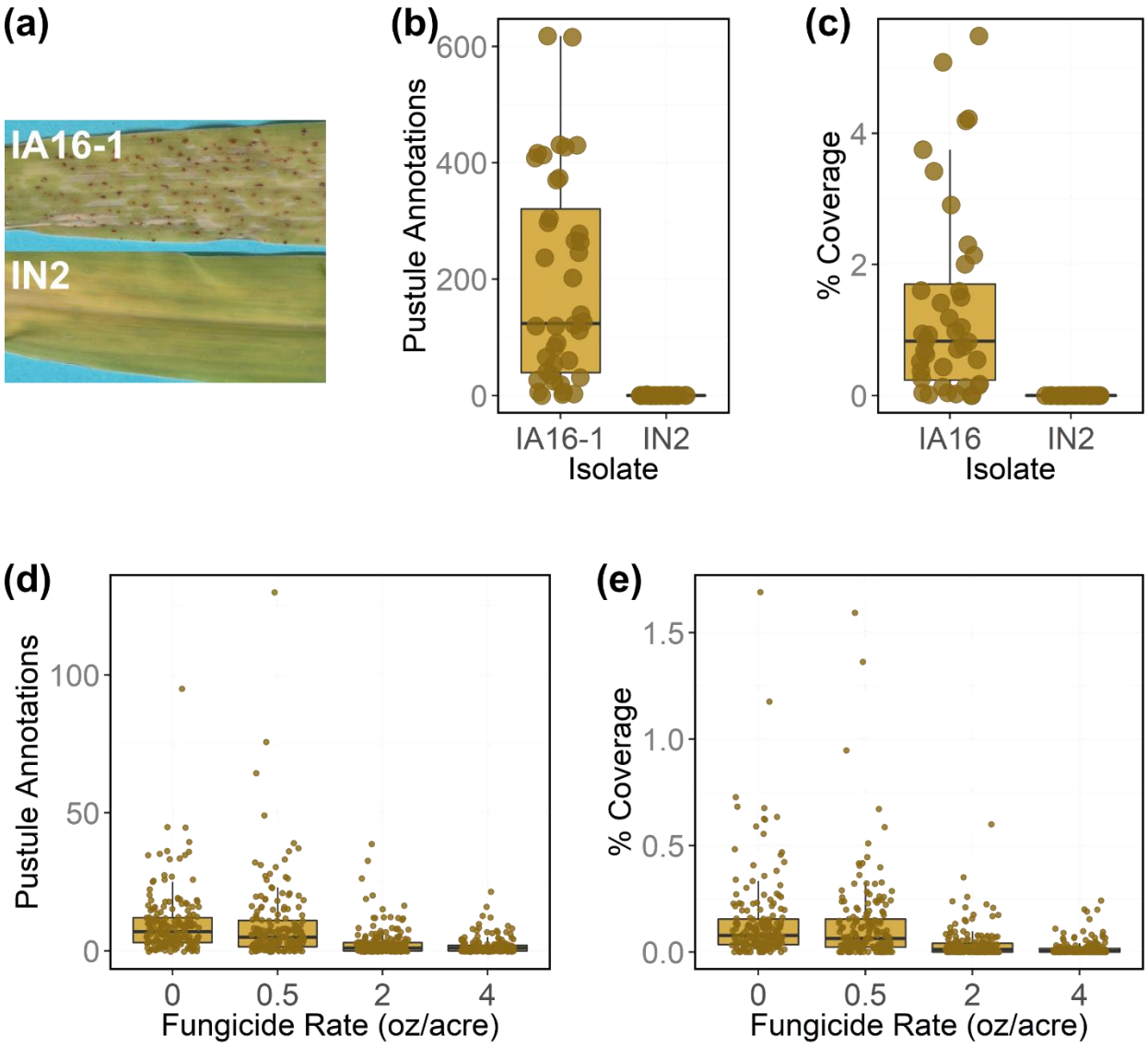


Figure 4. Ground truth results from the H95 *Rp1-D* differential experiment and the fungicide application experiment. (a) Representative leaves from the *Rp1-D* differential experiments. Leaf 3 of H95 *Rp1-D* plants inoculated with either the virulent IA16 isolate or the avirulent IN2 isolate at 9 days after inoculation (DAI). For each *P. sorghi* isolate, (b) the total number of manually annotated pustules per image, and (c) the % pustule coverage per image. For each fungicide application rate (fluid oz/acre), (d) the number of annotated pustules per image and (e) the % pustule coverage per image.

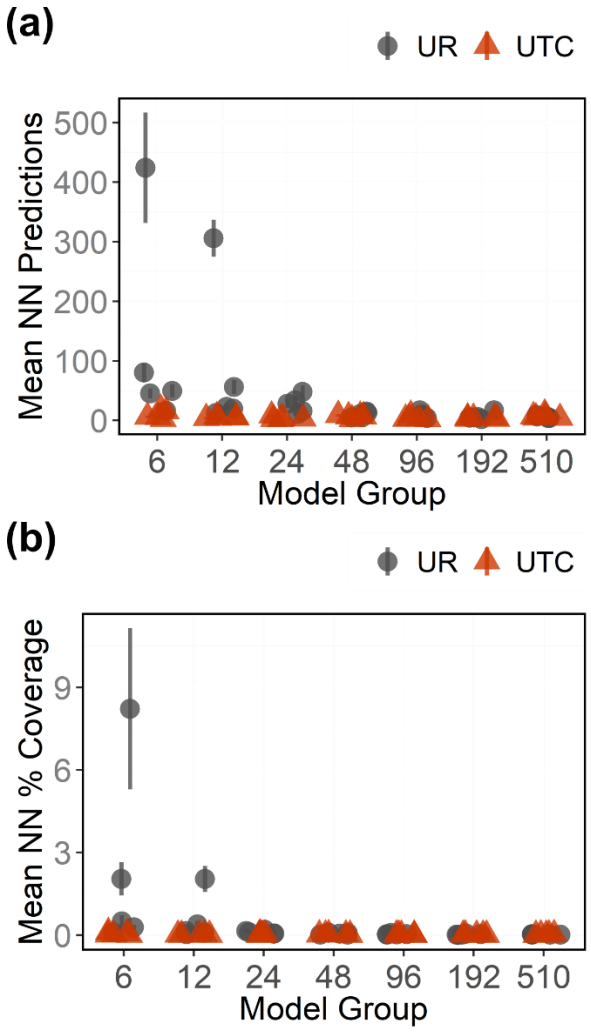


Figure 5. Performance of NN models on the IN2 images from the H95 *Rp1-D* differential experiment. The predicted means and 95% confidence intervals for (a) the average number of predictions per image and (b) the predicted % pustule coverage per image for each model as determined by the BT are shown.

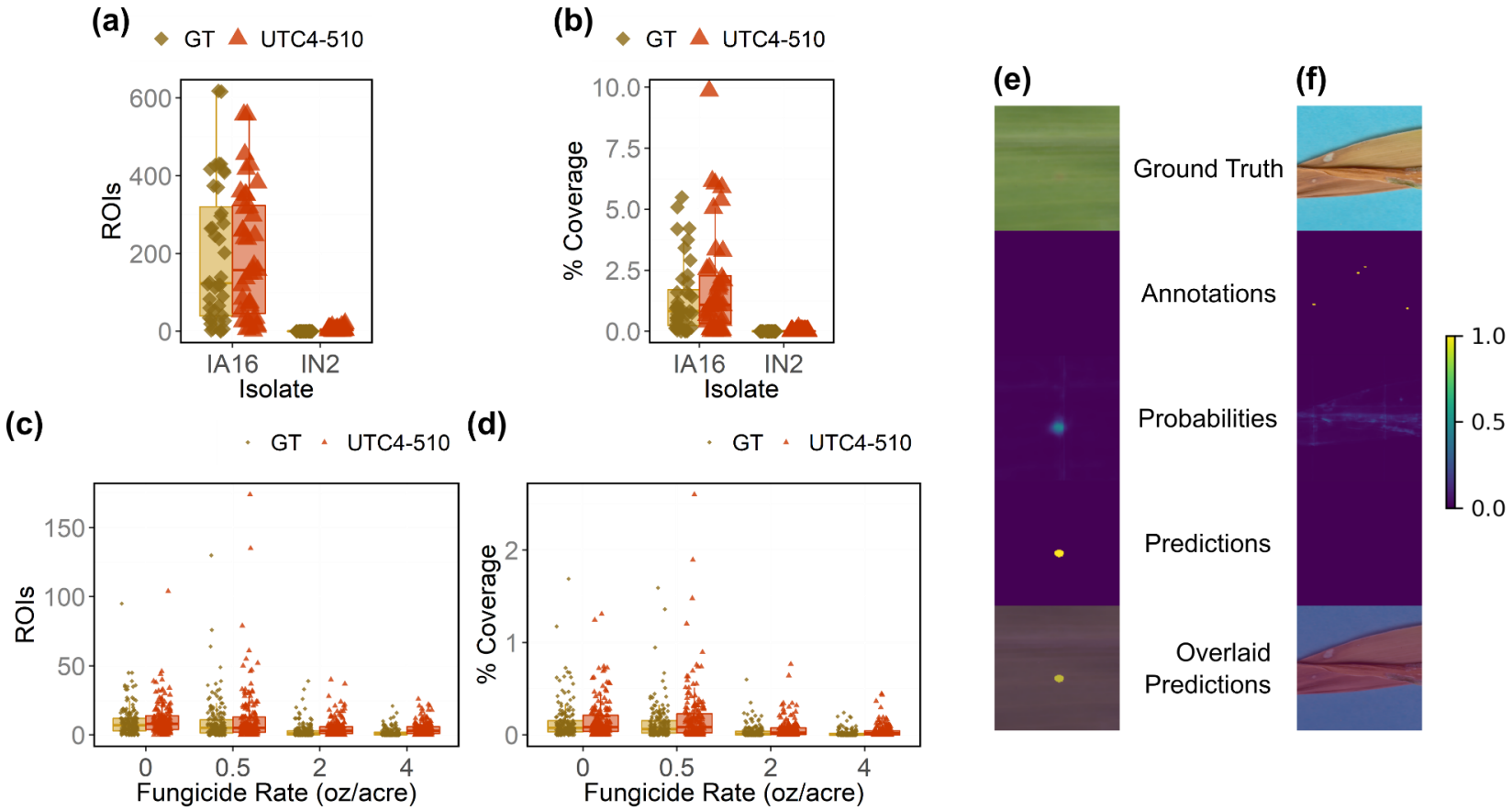


Figure 6. Comparison of the top performing neural network (NN) model trained on 510 images (UTC4-510) to ground truth (GT) data. The (a) number of regions of interest (ROIs) (either pustule annotations for GT data or pustule predictions for NN data) and (b) % pustule coverage for ground truth vs NN predictions for the H95 *Rp1-D* differential data. The (c) number of ROIs and (d) the % pustule coverage for ground truth vs NN predictions for the fungicide gradient data. An example of an instance where the NN model outperforms manual annotation, showing a crop of the ground truth image, its corresponding annotations, the model's pixel probabilities and subsequent predictions, and predictions overlaid on the ground truth image. (e) An example of a pustule that was originally missed during manual annotation, and (f) an example of not marking ground truth pustules that are not true pustules and were originally mis-annotated.

Tables

Table 1. The IoU and number of annotations in the training and validation subsets from the best two and worst two performing UR-6 models. Results were assessed at the BT.

Model	BT IoU	Training GT Annotations	Validation GT Annotations	Total GT Annotations
UR1_6	0.031	453	117	570
UR4_6	0.011	144	155	299
UR2_6	0.304	2282	406	2688
UR5_6	0.327	496	384	880

Table 2. The percentage of each model group for each threshold that is able to corroborate all ground truth results for the IA16 vs IN2 two-sample t-test of the differential data, the ANOVA and subsequent TukeyHSD groups for the fungicide gradient, and the IN2 mean equals zero one-sample t-test of the differential data. N for each percentage = 5.

Model Group	IA16 vs IN2, two-sample t-test				Fungicide Gradient, TukeyHSD				IN2 mu=0, one-sample t-test			
	UR		UTC		UR		UTC		UR		UTC	
	0.5	BT	0.5	BT	0.5	BT	0.5	BT	0.5	BT	0.5	BT
6	100%	100%	100%	100%	20%	20%	80%	80%	0%	0%	0%	0%
12	100%	100%	100%	100%	60%	20%	100%	100%	0%	0%	0%	0%
24	100%	100%	100%	100%	100%	80%	100%	100%	0%	0%	0%	0%
48	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%
96	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%
192	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%
510	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%

Supplemental Tables

Supplemental Table 1. Images used in the best two and worst two UR-6 models. "Performance" indicates whether it was high or low performing in the UR-6 model group. "# GT Annotations" is the total number of individual ground truth annotations labeled for each image.

Model	Performance	Leaf Name	Subset	# GT Annotations
UR1_6	Worst	20200912_0002_16	Training	202
UR1_6	Worst	20200912_0005_37	Training	88
UR1_6	Worst	B73_*Rp1-D*21_007	Training	65
UR1_6	Worst	spray_l3-08_203	Training	98
UR1_6	Worst	20210109_tilt0.5_71	Validation	7
UR1_6	Worst	20190501_0001_03	Validation	110
UR4_6	Worst	20190501_0006_39	Training	129
UR4_6	Worst	20210109_tilt0.0_28	Training	10
UR4_6	Worst	20210109_tilt0.5_85	Training	5
UR4_6	Worst	20210109_tilt1.0_129	Training	0
UR4_6	Worst	20210109_tilt0.5_61	Validation	1
UR4_6	Worst	20190501_0005_24	Validation	154
UR2_6	Best	cut_l2-02_047	Training	1951
UR2_6	Best	individual_leaves_2_week_day_12_13	Training	93
UR2_6	Best	7_day_day_7_34	Training	151
UR2_6	Best	20190501_0001_01	Training	87
UR2_6	Best	individual_leaves_2_week_day_06_59	Validation	175
UR2_6	Best	dust_l3-09_185	Validation	231
UR5_6	Best	20190501_0005_23	Training	87
UR5_6	Best	cut_l3-07_241	Training	167
UR5_6	Best	individual_leaves_2_week_day_11_12	Training	42
UR5_6	Best	2020_10_09_09	Training	200
UR5_6	Best	individual_leaves_2_week_day_12_17	Validation	191
UR5_6	Best	2020_09_30_05	Validation	193

Supplemental Table 2. The average performance metrics for each model group-training strategy at both thresholds (BT or 0.5). A darker color indicates better performance. Color scales are specific to each metric.

Training Strategy	Model Group	Threshold	IoU	Pustule F1	% TP Pixels	% FP Pixels	% FN Pixels
UR	6	0.5	0.106	0.192	10.62%	75.90%	13.48%
UR	6	BT	0.037	0.072	3.73%	95.02%	1.25%
UTC	6	0.5	0.375	0.545	37.48%	14.23%	48.29%
UTC	6	BT	0.408	0.580	40.80%	36.27%	22.93%
UR	12	0.5	0.360	0.529	35.97%	15.52%	48.51%
UR	12	BT	0.138	0.243	13.84%	80.51%	5.64%
UTC	12	0.5	0.400	0.571	39.99%	10.82%	49.20%
UTC	12	BT	0.459	0.629	45.93%	31.85%	22.22%
UR	24	0.5	0.442	0.613	44.17%	12.97%	42.86%
UR	24	BT	0.406	0.577	40.57%	47.28%	12.15%
UTC	24	0.5	0.432	0.603	43.17%	11.10%	45.74%
UTC	24	BT	0.485	0.653	48.48%	31.92%	19.60%
UR	48	0.5	0.448	0.618	44.77%	12.26%	42.97%
UR	48	BT	0.460	0.630	45.98%	39.82%	14.20%
UTC	48	0.5	0.426	0.598	42.61%	10.41%	46.99%
UTC	48	BT	0.488	0.656	48.76%	33.06%	18.17%
UR	96	0.5	0.461	0.631	46.07%	11.93%	42.00%
UR	96	BT	0.468	0.638	46.83%	39.37%	13.80%
UTC	96	0.5	0.480	0.648	47.96%	12.13%	39.91%
UTC	96	BT	0.505	0.671	50.55%	35.08%	14.38%
UR	192	0.5	0.458	0.629	45.84%	10.84%	43.32%
UR	192	BT	0.481	0.649	48.06%	38.26%	13.67%
UTC	192	0.5	0.476	0.645	47.64%	11.83%	40.53%
UTC	192	BT	0.492	0.660	49.23%	37.26%	13.51%
UR	510	0.5	0.463	0.633	46.28%	10.54%	43.18%
UR	510	BT	0.489	0.657	48.94%	37.18%	13.88%
UTC	510	0.5	0.462	0.632	46.19%	10.54%	43.26%
UTC	510	BT	0.485	0.653	48.51%	38.11%	13.39%

Supplemental Table 3. The statistical test results for the ground truth data of the fungicide gradient experiment for both the number of ROIs and the % coverage of ROIs on the leaves. The ANOVA tests for both metrics had f-statistics of <2e-16. A follow-up Tukey HSD test was conducted for each metric to obtain pairwise comparison between the fungicide concentration treatment groups (fluid oz/acre), with the resultant p-values reported. Asterisks indicate significant results (p-value <0.0005).

Comparison (fluid oz/acre)	ROIs	% Coverage
0.0 - 0.5	8.89e-01	6.79e-01
0.0 - 2.0	0.00e+00***	0.00e+00***
0.0 - 4.0	0.00e+00***	0.00e+00***
0.5 - 2.0	9.78e-09***	6.88e-08***
0.5 - 4.0	0.00e+00***	0.00e+00***
2.0 - 4.0	5.17e-01	6.51e-01

CHAPTER 5. GENERAL CONCLUSIONS

As the rust fungi world enters the era of long-read sequencing, genomic resources are expected to continue improving at a rapid rate. Paired with high-quality transcriptome data and bioinformatics tools, candidate effector proteins can be quickly identified and selected for further characterization. Large-scale characterization studies of candidate effectors are very popular, with several common assays used to identify functions like suppression of basal and effector-triggered immunity or avirulence. These studies have led to several interesting results, detailing host targets involved in transcript regulation or immune signaling (Qi et al. 2016; Liu et al. 2016). However, the success rate of a detectable phenotype in these studies can often be quite low (Lorrain et al. 2019). The development of computer vision-based phenotyping has not yet been extensively applied to effector studies in rust pathosystems but is sorely needed to increase throughput and consistency in results and their interpretation.

To improve and build upon the current genomic resources for *Puccinia sorghi*, Chapter 2 presents a long-read based genome assembly for Iowa isolate IA16. This assembly is expected to provide many advantages to the study of this pathogen, as it is highly contiguous with extensive resolved repeat regions. Interestingly, this assembly is nearly double the size of the previously reported Argentine RO10H11247 isolate (Rochi et al. 2018), which seems to be primarily the result of additional repeat sequences in the IA16 genome. This characteristic has been noted in other rust species, as well. However, it is important to note that the RO10H11247 assembly was generated from short-reads, and thus likely underestimates true genome size due to collapsed repeat regions. Deeper examination into the types and distribution of these transposable elements in the IA16 assembly are sure to provide insight into the diversification of the species. Pseudophasing of the genome was also performed, resulting in two haplotype sequences. The

majority of these sequences are expected to be collapsed, but there are likely differences between the two, and further study into those differences are bound to be enlightening as well. Within the haploid assembly, we predicted a total of 1,845 secreted proteins, of which 742 are also predicted to be effectors, providing additional candidates for screening. This list also provides an opportunity to compare the IA16 predicted effectors to those from the RO10H11247 assembly, undoubtedly providing insight into the differences between the two isolates.

As there is limited functional annotation for most predicted effectors, *in planta* characterization is necessary to elucidate function. From the hundreds of candidate effectors in *P. sorghi*, we cloned eight candidates homologous to *Pp*EC23, a *Phakopsora pachyrhizi* effector of the rust protein family cluster 112 shown to suppress plant immunity (Link et al. 2014; Qi et al. 2016). One *P. sorghi* candidate was able to suppress hypersensitive immune response in a heterologous system, albeit at a lower level than *Pp*EC23 (Qi et al. 2016). Unlike *Pp*EC23, we were unable to show that the *P. sorghi* homolog suppressed plant basal immunity, suggesting it may impact basal immunity at a different timepoint than those tested here or may affect immunity in a different manner than *Pp*EC23. Additional research into maize factors that interact with the *Pp*EC23 homologs may elucidate the differences between the candidates from *P. sorghi* and *Pp*EC23. A yeast two-hybrid library of *P. sorghi*-infected maize was generated for this purpose.

Another interesting note regards the *Pp*EC23 homologs that we were originally unable to amplify from cDNA. The PCR amplification primers were based on the predicted CDSs from the RO10H11247 assembly, and any significant differences at the 5’ or 3’ end of the CDSs between RO10H11247 and IA16 would mean no product was amplified, even if the gene was transcribed in both isolates. This turned out to be the case for two of the three unamplified candidates. These

two candidates were predicted in IA16, but with large variations of predicted protein sequence at the N-terminus when compared to the RO10H11247 predicted sequence. It will be interesting to delve further into differences between the two isolate's annotations, particularly for predicted effector genes.

Within Chapter 3, we also described the use of a phenotyping box that allowed us to generate time course image data in a hands-off manner for an immune response assay. The boxes, adapted from another publication (Barbacci et al. 2020), are very inexpensive and flexible and their use is expected to increase throughput and consistency between experiments. Although not applicable to our dataset, the initial publication regarding these boxes utilized automated phenotyping. A similar pipeline could easily be applied to hypersensitive response immune assays, especially considering images between timepoints and experiments are very similar.

We additionally expanded on the phenotyping methods available to rust fungi researchers in Chapter 4. Machine learning (ML) is a powerful tool, but many applications require extensive, manual annotations of relevant features. By utilizing our large amount of phenotyping data, we were able to assess model performance at varying amounts of training data. The probability that a particular model's performance can be increased if researchers utilize datasets with high phenotypic diversity, such as those generated by disease time courses. Beyond this, the final presented machine learning model performs very well when compared to ground truth data, yielding remarkably similar answers. This model can be applied to data from various experiments but will first be applied to virus-induced gene silencing (VIGS) (Lange et al. 2013; Beernink and Whitham 2023) and host-induced gene silencing (HIGS) (Zand Karimi and Innes 2022) studies. VIGS experiments involve the silencing of maize proteins suspected to be integral to disease pathways while HIGS targets *P. sorghi* proteins. Silencing of these factors is expected

to influence phenotype, and an automatic phenotyping system provides a reproducible way to quantify disease. Additionally, phenotypic changes may be small, thus undetectable by manual scoring with a standard area diagram. We expect that subtle changes in phenotype may be parsed with the U-Net model. Additionally, quantification with the ML model will enable consistent quantification of pustules across all treatments. If multiple researchers gather image data, the U-Net model ensures consistency between the results of each researcher's images as well.

Overall, this dissertation provides a comprehensive approach to the identification and characterization of candidate effector proteins of *P. sorghi*, from the development of genomic resources for effector mining in a new *P. sorghi* isolate, to the immune suppression characterization of existing predicted effector candidates, to the deployment of phenotyping platforms for additional characterization studies.

References

Barbacci, A., Navaud, O., Mbengue, M., Barascud, M., Godiard, L., Khafif, M., et al. 2020. Rapid identification of an Arabidopsis NLR gene as a candidate conferring susceptibility to *Sclerotinia sclerotiorum* using time-resolved automated phenotyping. Plant J. 103:903–917.

Beernink, B. M., and Whitham, S. A. 2023. Foxtail mosaic virus: A tool for gene function analysis in maize and other monocots. Mol. Plant Pathol..

Lange, M., Yellina, A. L., Orashakova, S., and Becker, A. 2013. Virus-induced gene silencing (VIGS) in plants: An overview of target species and the virus-derived vector systems. Methods Mol. Biol. 975:1–14.

Link, T. I., Lang, P., Scheffler, B. E., Duke, M. V., Graham, M. A., Cooper, B., et al. 2014. The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. Mol. Plant Pathol. 15:379–393.

Liu, C., Pedersen, C., Schultz-Larsen, T., Aguilar, G. B., Madriz-Ordeñana, K., Hovmøller, M. S., et al. 2016. The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases. New Phytol.

Lorrain, C., Gonçalves dos Santos, K. C., Germain, H., Hecker, A., and Duplessis, S. 2019. Advances in understanding obligate biotrophy in rust fungi. New Phytol. 222:1190–1206.

Qi, M., Link, T. I., Müller, M., Hirschburger, D., Pudake, R. N., Pedley, K. F., et al. 2016. A Small Cysteine-Rich Protein from the Asian Soybean Rust Fungus, *Phakopsora pachyrhizi*, Suppresses Plant Immunity ed. Peter N Dodds. PLoS Pathog. 12:e1005827.

Rochi, L., Diéguez, M. J., Burguener, G., Darino, M. A., Pergolesi, M. F., Ingala, L. R., et al. 2018. Characterization and comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize common rust. Fungal Genet. Biol. 112:31–39.

Zand Karimi, H., and Innes, R. W. 2022. Molecular mechanisms underlying host-induced gene silencing. Plant Cell. 34:3183–3199.