DISSERTATION

# FIELD POPULATION GENETICS OF GLOBAL *PUCCINIA STRIIFORMIS* F. SP. *TRITICI* POPULATIONS IN WHEAT REVEAL A DYNAMIC MOLECULAR BATTLEFIELD

Submitted by

Rebecca Lyon

Department of Agricultural Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2021

Doctoral Committee:

Advisor: Pankaj Trivedi

Kirk Broders
Stephen Pearce
Ruth Hufbauer

# ABSTRACT

## FIELD POPULATION GENETICS OF GLOBAL *PUCCINIA STRIIFORMIS* F. SP. *TRITICI* POPULATIONS IN WHEAT REVEAL A DYNAMIC MOLECULAR BATTLEFIELD

Wheat stripe rust caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*) is one of the most devastating diseases of wheat. Globally, *Pst* is present in all major wheat-growing regions of the world and approximately 88% of the global wheat varieties are susceptible to *Pst* infection. Yield losses due to stripe rust can reach 100% in susceptible cultivars and mitigating the impact of wheat stripe rust costs over $1 billion annually. Identifying and deploying resistance genes (*R*-genes) into commercial wheat lines is currently the most effective defense against stripe rust. Unfortunately, the geographic footprint, frequency, and severity of stripe rust epidemics have increased globally due to climate change, and emerging new races are increasingly able to overcome resistant wheat varieties. Current surveillance methods include using a differential set of wheat cultivars to identify *Pst* races, however this method only focuses on the presence/absence of a few effectors and the differential set can take years to update. Surveillance methods for *Pst* infection need to be updated to efficiently detect and combat the spread of new isolates and improve preparedness for inevitable future epidemics. In addition, *Pst*-wheat interactions should be elucidated to understand the co-evolutionary arms race between the pathogen and host that regulate the disease outcome. Such information is vital for developing durable resistance strategies to combat *Pst* during active infection in susceptible wheat varieties. In this study, we used field patho-transcriptomics (the sequencing of transcriptomes from infected plant leaves gathered from fields) to (1) determine if modern North American *Pst*

populations represent introduction events or asexual changes within the resident population and to quantify the presence of effectors; (2) investigate the impact of *Pst* and wheat genetics on the transcriptome of each species; and (3) explore the interaction between *Pst* and wheat during active infection to reveal targets for the pathogen and insight into responses from the host. The samples used in this study were *Pst*-infected wheat leaves collected in fields from around the world. We gained insight into the evolution of *Pst* populations in North America and discovered worldwide conserved molecular interactions between host and pathogen.

The geographic footprint of *Pst* in the United States has expanded in the past two decades and novel races of *Pst* have emerged since 2010. Understanding the population structure of *Pst* and the origin of shifts in population structure can help with the evaluation of genes for the durable resistance against *Pst*. However, current methods of race-typing using a differential set of wheat cultivars is time-consuming and only gives limited information about the candidate effector diversity of *Pst* isolates in the field. Here we use field patho-transcriptomics to estimate pangenomic variability, assess population structure, and evaluate existing candidate effector diversity within *Pst*. Discriminant analysis of principal components (DAPC) using synonymous *Pst* single nucleotide polymorphisms (SNPs) identified eight distinct genetic clusters of the pathogen across the globe. The North American samples fell into three genetic clusters, with the largest cluster present throughout the continent and genetically distinct from other *Pst* populations around the world. Our results indicate most modern North American samples evolved asexually from within the endemic population (whether by novel mutations or somatic hybridizations) rather than from introductions of new isolates from outside the region. Results from co-transcriptomics analysis suggest that there are fewer candidate effectors in modern populations, which indicates that modern *Pst* races have evolved by shedding effectors to avoid

detection by the host immune system. Overall, our study provides information on recent evolutionary changes of *Pst* in North America. These findings have implications on wheat breeding and management.

Strategies for combating *Pst* when available resistance genes fail are necessary for developing long-term, durable resistance in novel varieties of cultivated wheat. In the final portion of this study, the co-transcriptomic data were used to evaluate both the host and pathogen exon expression and to explore potential intra- and inter-species molecular interactions during active infection. To start, the population structure of worldwide wheat was evaluated through use of biallelic synonymous SNPs in a DAPC analysis, and two genetic wheat clusters were detected. PERMANOVA analysis was used to clarify the effect of environment and genetic clusters on the transcriptomes. The results revealed that environmental and temporal variation accounted for a large portion of the variance seen in the exon expression of both organisms. Additionally, intra-specific genetic variation had a significant effect on the transcriptome structure within an organism, albeit a much lower effect within wheat. Interestingly, inter-specific genetic variation in *Pst* had a significant effect on the wheat transcriptome while wheat did not have a significant effect on the *Pst* transcriptome. These results imply that the extensive genetic variation within *Pst* is a major driving force in the epidemiology of global striped rust infections, while the major selective pressure of domestication and wheat breeding could be reducing the potential for the host to defend against *Pst* during active infection.

To further explore the mechanisms underlying wheat-*Pst* interaction, exon co-expression networks (ECNs) were created for each organism using Spearman correlations to identify functional exon networks among exons within each organism. Six highly connected *Pst* ECNs revealed exon networks associated with both core cellular functions and the infection process.

The same process identified 11 wheat ECNs having roles in photosynthesis, general plant defense, and transcriptional processing. These ECNs were condensed into single vectors using eigen-vector decomposition and a cross-kingdom dual interaction network was generated to explore both intra- and inter-species interactions among ECNs. This dual interaction network showed a large negative relationship between the host photosynthetic apparatus and *Pst* growth/development and virulence factors. This finding may be indicative of either a direct attack of *Pst* on wheat photosynthesis or a failure of chloroplasts to keep up with parasitic demand on photosynthate through source-sink relationships. Additionally, the dual interaction network revealed a wheat network associated with sulfur metabolism which is also transitively associated with *Pst* growth and virulence, thereby divulging *Pst*'s lack of endogenous sulfur metabolism genes and suggesting that *Pst* obtains needed sulfur from the host. Overall, both individual networks and the dual-transcriptomic network gave insight into biological mechanisms and interactions which occur during active infection.

Sequencing the co-transcriptomes of *Pst*-infected field samples of wheat allows for numerous opportunities to better understand the worldwide population structure and large-scale interactions within the *Pst*-wheat pathosystem. Chapter 2 explains how the global population structure of the pathogen was used to discover the origin of new races found within North America and the array of effectors within worldwide genetic groups. The first part of Chapter 3 presents how the worldwide population structure of the host was evaluated while analyzing the effect of genetics on transcriptional variation. The second part of Chapter 3 shows how these connections gave insight into biological mechanisms during active *Pst* infection and the potential novel breeding targets within the wheat host. This study furthers our understanding of pathogen migration and adaptation while serving as an initial step in probing for novel, durable resistance.

# ACKNOWLEDGEMENTS

I want to thank all of those who contributed to the success of these dissertation. Specifically, I thank both of my mentors, Dr. Pankaj Trivedi and Dr. Kirk Broders, for supporting and guiding me through this process. I also thank the rest of my doctoral committee members, Dr. Ruth Hufbauer and Dr. Stephen Pearce, for their feedback. I extend my gratitude to Diane Saunders and her team (Venessa Bueno, Dr. Guru Radhakrishnan, and Dr. Antoine Persoons) for sharing sequenced data and helping with the population study. I thank Dr. Jason Corwin for his extensive support – from bioinformatics to writing to inspiring the third chapter of this dissertation. I wish to acknowledge Dr. Gungxi Wu for giving me a solid start in collecting samples, working in the lab, and learning bioinformatics. I thank Karl VanGessel for his assistance with the Wheat GO analysis. I am grateful to Gloria Broders, Dr. Tessa Albrecht, and Dr. Stephen Wyka for helping me in the lab and discussing my project. I thank Dr. Xianming Chen for sample collection and support with inoculation and growth of *Pst* in growth chambers. I extend my gratitude for everyone who collected samples for this study, especially Bob Bowden. I thank my funding sources, the Colorado Wheat Administrative Committee and CSU's Department of Agricultural Biology. Finally, I thank my family and friends who supported me throughout this entire project, particularly my mother Dot Lyon who helped edit and format this dissertation.

TABLE OF CONTENTS

# Chapter 1: Impact of climate change and race evolution on the epidemiology and ecology of stripe rust in central and eastern United States and Canada

## Introduction

Stripe rust, caused by the fungal pathogen *Puccinia striiformis* f. sp. *tritici* (*Pst*), is one of the most devastating diseases of wheat and is present in all major wheat-growing regions of the world[1]. In the United States, stripe rust has traditionally been observed in cooler regions with higher moisture levels, such as the Pacific Northwest, but in the last 15 years its geographic footprint has expanded[2]. This expansion is associated with suspected incursions of novel *Pst* races that are better adapted to warmer environments, and have broader virulence profiles and are more aggressive than previously characterized races[3]. In North America, the frequency and severity of stripe rust epidemics have increased dramatically since the year 2000, with significant losses to stripe rust reported in seven of the last 15 years[4]. The worst epidemic to date was in 2015, when *Pst* infections reached epidemic levels throughout the United States, causing average yield losses of 11.2% and costing the wheat industry approximately $8.7 million[5]. The high level of residual inoculum at the end of the 2015 season resulted in significant infection rates in the autumn, which, in combination with consistent snow cover, led to the pathogen overwintering in areas historically deemed to have unfavorable conditions, including Colorado, Ohio, and South Dakota[4]. As a result, there was widespread disease in 2016 throughout the central United States and into Canada, as well as in the eastern United States. This lead to further inoculum build-up, increased autumn infection rates, and more northerly overwintering in the Puccinia Pathway, the path followed by overwintering urediniospores from the southern United States and Mexico

during the spring as they spread northward through the Great Plains (Figure 1.1). Urgent action is required to prepare for this severe and worsening threat to wheat production in the United States and Canada.

Furthermore, current climate predictions indicate that weather conditions, including milder winter temperatures, will become increasingly favorable for *Pst* overwintering at more northerly latitudes than observed in previous years, leading to more frequent and damaging epidemics in the medium to long term. While wheat stem rust (*Puccinia graminis* f. sp. *tritici*) has historically been considered a more economically important pathogen of wheat, the expanding geographic extent and increased production losses associated with stripe rust indicate that stripe rust is now the most damaging rust pathogen in North America and globally[6,7]. Therefore, the purpose of this perspective is to review the impact of emergence of new races and climate change on the ecology, evolution, and epidemiology of *Pst* in North America over the last decade. This will be followed by recommendations for moving forward, which aim to integrate the evolutionary biology of *Pst* with future climate change scenarios and deployment of host resistance to develop more resilient and durable stripe rust management strategies.

## Impact of *Pst* on North American wheat production

A recent study combining long-term crop loss data and climate models demonstrated the expansion of the geographic footprint of stripe rust due to climate change[2]. Using state-specific data compiled from the USDA's Cereal Disease Laboratory, Beddow et al. (2015) reinforced the notion that the spatial extent of losses from stripe rust has expanded substantially in recent decades compared with the historical prevalence of the disease[1]. For the entire period 1960–1999, just 11 of the 48 continental US states reported losses from stripe rust, whereas between 2000–2014, the disease was reported in a total of 26 states and three Canadian provinces[2]. Stripe rust

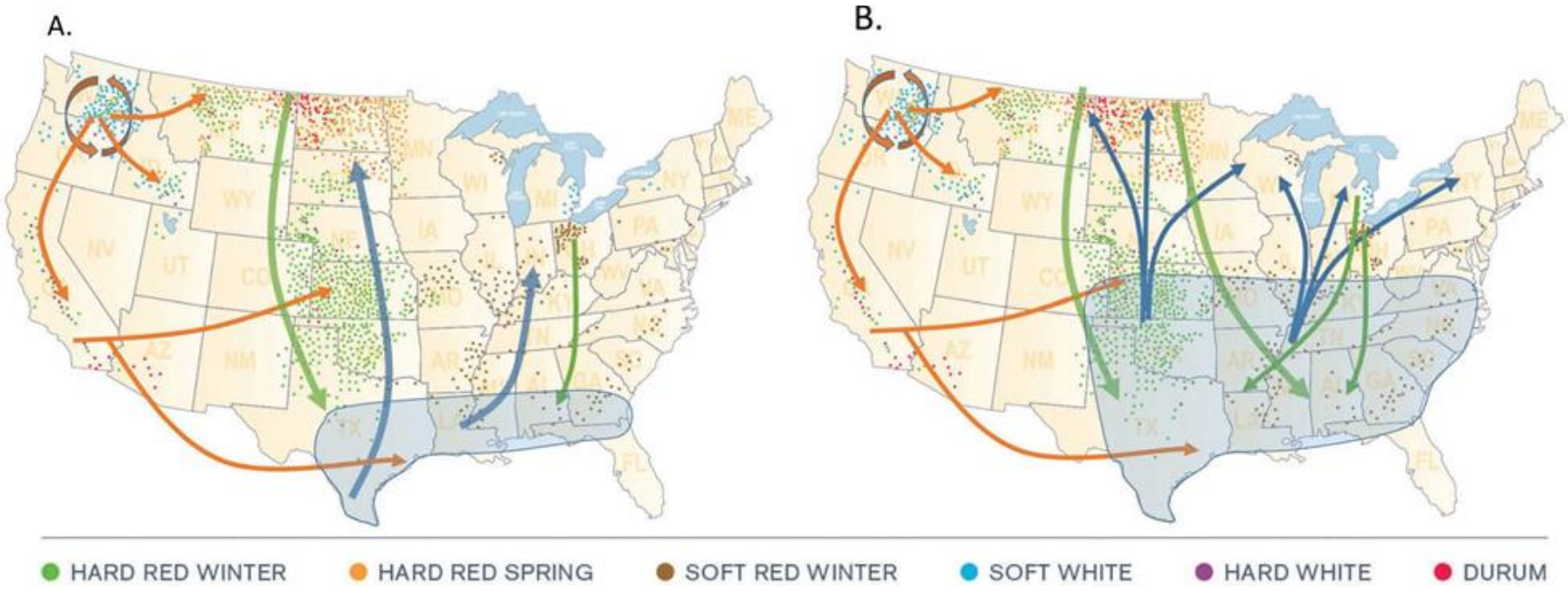


*Figure 1.1: Schematic diagrams of the annual movement of* P. striiformis f. sp. tritici *populations and overwintering range according to (A) our current understanding and (B) a conceptual model based on population genetic data*[8]*. Orange arrows represent movement of western* Pst *populations, blue and green arrows represent spring and autumn movement of eastern* Pst *populations, respectively. Blue ovals represent (A) current and (B) predicted overwintering range of* Pst *based on reports of 2015–2016 and 2016–2017 overwintering inoculum in the United States*[4,9]*. Wheat production map used with the permission of the National Association of Wheat growers.*

has expanded further since this report, with records of disease originating from 24 states and two Canadian provinces in 2015, and stripe rust was reported from 31 states and five Canadian provinces in 2016[4]. During the 2015 epidemic, many states in the central plains recorded wheat yield losses of 10–25%, representing one of the greatest yield losses due to stripe rust since the introduction of the new pathogen population in 2000[5]. The high level of inoculum at the end of the 2015 crop season resulted in significant autumn season infection rates, which, in combination with consistent snow cover, resulted in the pathogen overwintering in areas historically deemed to have unfavorable conditions. The significant inoculum source in combination with ample precipitation resulted in reports of *Pst* infection on wheat nearly 30–45 days ahead of the historical average along the Puccinia Pathway (Figure 1.2). While there were significant losses in the central plains during 2016 (1–15%), the biggest impact occurred in wheat growing regions in the eastern part of the United States. This was due to the early infections and the lack of resistant varieties in this region.

Prior to the year 2000, stripe rust was a rare problem in the Central Plains[10]. Since 2000, stripe rust has occurred more consistently and with greater severity[10]. This change in disease severity occurred at the same time as new pathogen populations emerged[10–12]. When isolates from the new races were tested, they had shorter latent periods and higher germination rates at warmer temperatures[13]. The post-2000 populations had increased virulence and were genetically distinct based on microsatellite markers[12,14]. However, it has been 17 years since this introduction and it is unclear to what extent mutation, recombination, or somatic hybridization has occurred among US populations of *Pst*. In addition, by using only microsatellite markers, we are only able to determine the evolution of these simple sequence repeats in the genome without gaining any further insight into regions of the genome that may be under selection for adaptation to the warmer and drier

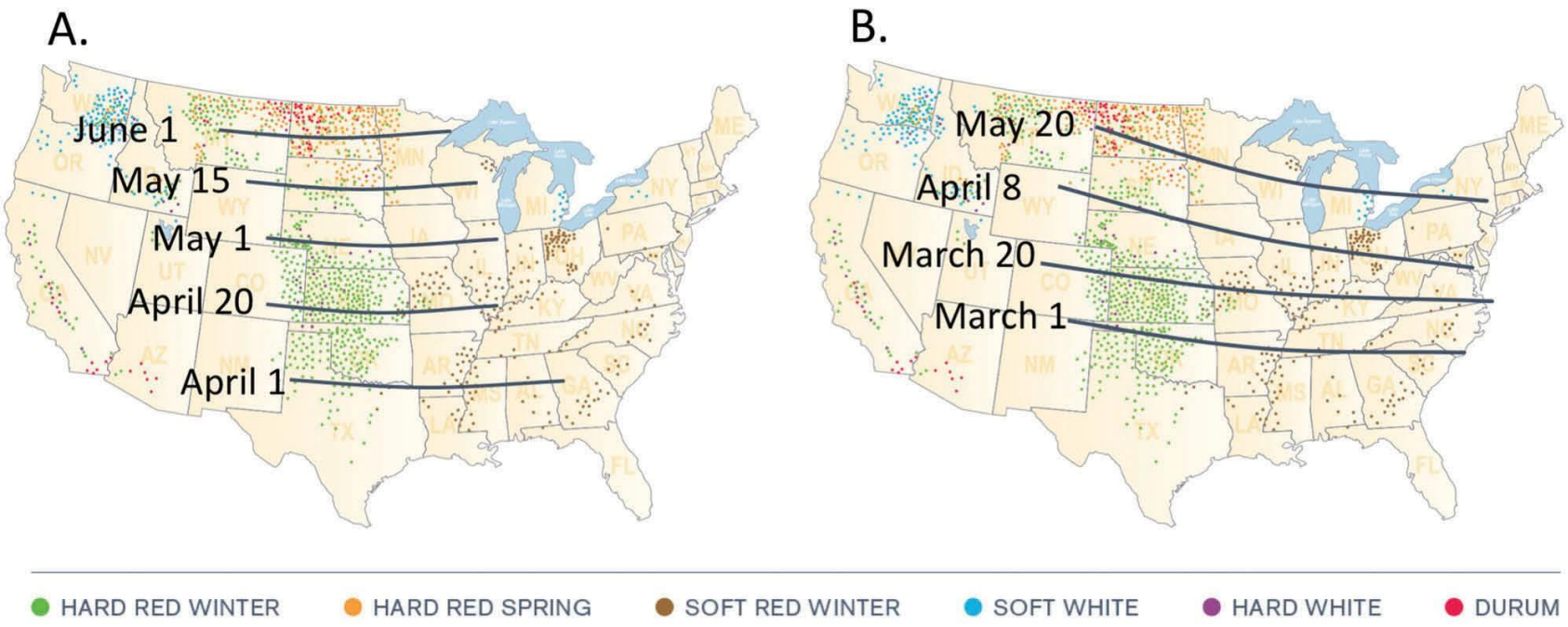


*Figure 1.2: Average first report of stripe rust (*P. striiformis f. sp. tritici*) along the Puccinia Pathway in the central and eastern United States during (A) 2015 and (B) 2016.*

climates of the central plains and warmer, more humid climates of the eastern United States. Future research will need to focus on understanding how these adaptive changes occur.

**Impact of new races on population dynamics of *Pst***

Even when novel pathogens are initially endemic to restricted geographic areas, they can rapidly emerge in new regions, provided they encounter a production system with susceptible hosts and favorable environmental conditions. In widely grown food crops, many pathogens were introduced long ago and are now geographically wide-spread; and therefore, they do not come to mind as invasive pathogens[15]. Although introductions could have occurred centuries ago, the evolution of such ubiquitous pathogens remains a significant cause of concern due to the risk of re-emergence caused by accidental dissemination of new, multi-virulence races, or new highly aggressive strains[16,17]. An accurate understanding of the origin, distribution, and diversity of inoculum reservoirs and past and recent migration routes of these pathogens is crucial to understand current epidemics, develop risk-assessment models, and alleviate the potential adverse effects of disease emergence[18,19]. This is particularly true for pathogens which are capable of long-distance migration, for which any newly advantageous mutant (increased virulence, aggressiveness, or resistance to fungicides) has the potential to spread over large geographic areas[20].

Stripe rust is an excellent example of a globally distributed pathogen that has major negative impacts on wheat production due to emergences and invasions of new races[6,14,17,21–23]. Despite the capacity for long-distance migration, the worldwide spread of *Pst* is relatively recent, with most emergences in new locations reported only within the last two decades. In addition to recently colonized areas, new races are known to periodically emerge through the acquisition of new virulences. This is what has likely occurred for the recent invasion of a new population of

*Pst* into North America, which is believed to have been introduced around the year 2000 and has since become the dominant population in the central plains and eastern United States[6,12,13]. The population structure of *Pst* is therefore likely to display the hallmarks of a complex mixture of re-emergence of diverse genotypes over continuous wheat-growing areas and rare founder events due to long-distance migration. Recent spread of the pathogen is likely to induce marked changes in patterns of population differentiation among geographic regions, potentially erasing the signature of more ancient colonization events.

A recent global analysis of the population diversity of *Pst* has found a strong population subdivision within *Pst*, with a clustering of isolates according to their geographic origin despite a capacity for long-distance migration[14,23]. This pattern stands in contrast with the previous understanding of the worldwide population structure of *Pst*, which considers the potential replacement of local populations by new invasions[20,22]. On the basis of survey data that monitors the occurrence of strains with newly acquired virulences that defeat recently deployed *Yr* (resistance) genes, the population structure of *Pst* was considered to be shaped by a continual replacement of pre-existing populations by the newly emerged races. This process is well known as the boom and bust cycle[24]. Such geographic migrations were documented in a recent global population analysis of *Pst*, but recently dispersed genotypes appear to coexist with older populations specific to the main geographic areas, suggesting that migrants do not replace local populations despite the capacity for recurrent long-distance dispersal[14]. In contrast, the invasion of a region by new genotypes in clonal populations would result in a population sweep and would replace the original population.

This was observed in the United States, where the post-2000 *Pst* population is dominated by aggressive races (such as PstS1 or its derivatives, which originated from the East African or

Middle East region), but several pre-2000 races continue to be observed in the United States each year at a low frequency[10,25,26]. The outcome of new race introductions may depend on the relative competitiveness of local populations in their region of origin, and the selective advantage of migrants (e.g. their virulence towards widely deployed resistance genes or an increased tolerance to abiotic stress, such as high temperature). Given the evolutionary potential of *Pst* and increased severity of recent epidemics, we believe the current *Pst* population has continued to adapt to the warmer climates of the eastern United States, leading to an increase in pathogen fitness and aggressiveness not associated with the presence or absence of specific virulence genes associated with host resistance genes. Future research is needed in order to test this theory.

**Impact of climate change on the ecology and epidemiology of *Pst***

The ability of stripe rust to infect wheat plants depends on many factors, including host resistance, time of infection, and environmental factors[27]. The three main climatic factors affecting disease severity are moisture, temperature, and wind. Bebber et al. (2013) documented a poleward shift of many fungal crop pathogens from southern latitudes moving northward due to increases in temperature and precipitation[2]. This may be occurring with *Pst* as temperature and precipitation affect the *Pst* life cycle, including spore germination, infection, and survival. It also affects latent period, sporulation, and host resistance[28,29]. The ideal temperature for many *Pst* isolates is between 7–12°C, so *Pst* prefers temperate regions or high-altitude areas within more tropical regions[1,30]. However, temperature range adaptation varies among isolates of *Pst*. Some isolates have a shorter latent period at 18°C than 12°C[31]. When creating prediction models, the pathogen population structure of the region should be taken into account. There are also limitations to prediction models, e.g. if a new race emerges with a different optimal temperature, then the prediction model may not be accurate.

As *Pst* is not able to survive the winter in all locations in North America, areas with high amounts of inoculum surviving over the winter and summer can be considered "hot spots" for future disease and development of new populations[32]. This is important as outbreak propagule pressure influences the landscape spread of wind-dispersed plant pathogens. Increased inoculum pressure measured in terms of urediniospore number is related to a higher probability of *Pst* persisting in the environment[30]. In fact, a study by Severns et al. (2014) examined how inoculum pressure influences the spread of *Pst*, and they found that a simple power function relates to outbreak disease levels for both magnitude of the final epidemic and the spatial expanse[28]. Therefore, urediniospores spread via wind allows small hot spots to have a big impact on a disease epidemic. The current understanding of the *Pst* population in the central and eastern United States and Canada postulates that inoculum primarily overwinters in southern Texas, northern Mexico, and along the Gulf Coast in Louisiana and Mississippi. The inoculum then moves northward through the central plains and eastern United States where races may be selected for based on their ability to infect wheat varieties with resistance genes. The population continues to migrate north into Canadian provinces, such as Saskatchewan, where the amount and virulence of the population depends on characteristics of the population in the central plains of the United States[33]. However, based on reports from the 2016 stripe rust epidemic and climate change projections, the overwintering range of *Pst* will likely move northward, creating a greater number of "hot spots" earlier in the growing season, which will result in greater levels of initial inoculum that will lead to greater secondary spore production and crops infected earlier in the season (Figure 1.2). This will inevitably lead to greater disease severity, more rapid expansion into northern latitudes, and greater yield losses throughout North America.

Over the summer months, temperature and humidity are key factors that affect *Pst* germination and infection. Constant high temperatures have a greater negative effect than temperatures that shift from high to low. Short exposure to higher temperatures does not kill the pathogen and so has less effect on germination and infection rates[31]. Very few recent studies have attempted to evaluate the importance of the over-summering range of *Pst* in the central and eastern United States; much of this is due to the fact that summer and autumn infection were rare in the past. As the rust population spreads north via the Puccinia Pathway, it first infects winter wheat in the southern great plains, which will be harvested in late June. *Pst* will continue to spread north into the Dakotas and Canada where spring wheat is planted in May and harvested in August. At this point, the *Pst* populations will begin the migration back south as winter wheat in South Dakota, Nebraska, and Colorado is planted in late August through September and then in Kansas and Oklahoma in September through October. It is this reverse Puccinia Pathway that may allow populations to complete their migration, but we know almost nothing about whether the predominant races moving northward in the spring are the predominant races moving southward in the autumn. In addition, wild grass species have been shown to harbor a more diverse population of *Pst* than is generally observed infecting wheat[34]. These two sources of inoculum have been largely overlooked in our assessment of *Pst* epidemiology. Much attention and many predictive models have focused on the movement of spores during the spring. Given recent developments, we need to begin investigating the annual meta-population movement of *Pst* on a continental scale and determine how this may change with climate variability in the future.

With regard to the ability of *Pst* to survive through the winter, it has a latent period that can last for several days, weeks, or months depending on weather conditions. During the

overwintering period, temperature is a key aspect of survival. The lowest monthly mean temperature under which *Pst* can continue to develop is believed to be -6°C to -7°C. If wheat plants are under snow, then *Pst* can survive to -10°C[31]. Given the importance of urediniospores as a primary inoculum source each spring, we predict that changing climate patterns will allow *Pst* to overwinter in regions previously thought unsuitable, leading to more rapid and northerly build-up of primary inoculum along the Puccinia Pathway (Figure 1.2).

**Adapting stripe rust management in the central United States and Canada to novel races and a changing climate**

The most effective defense against stripe rust is the identification and deployment of resistance (R) genes into wheat varieties through breeding programs. However, individual R-genes can quickly be overcome by virulent *Pst* races arising from novel genetic variation in pathogen secreted effector proteins. Effector proteins are important pathogenicity determinants that are recognized in certain host genotypes, where they induce a resistance response that prevents disease development[35]. Effectors identified by the host are under strong selective pressure to adapt to evade detection by the host resistance mechanisms[36]. To improve the speed and effectiveness of our responses to stripe rust epidemics, continued high-resolution surveillance of the genetic variability of *Pst* populations within the United States is required. These data can be applied to monitor the distribution of existing *Pst* races and to help predict future emergence and spread of novel *Pst* races.

Existing stripe rust surveillance mechanisms rely on race phenotyping, which involves inoculating each *Pst* isolate onto a series of plants carrying a known complement of R-genes. Infection profiles are scored to determine the presence of pathogen avirulence effector proteins carried by each isolate, allowing their classification into discrete races. This method is often combined with analysis of low-throughput genetic markers, such as microsatellites, to provide a

measure of genetic diversity within the population[37]. Aside from being low-throughput, cumbersome, and requiring the transport of live spores, this approach has the major limitation that novel effectors, effector diversity in the population, and effector loss in the population are not detected and described. One striking example is the finding that the most common *Pst* race in the United States, PSTv-37, is composed of 83 distinct multi-locus genotypes, including isolates belonging to genetic groups from both the eastern and western United States[8]. This means that within-race adaptation and evolution goes unnoticed, a critical shortcoming for the prediction of emerging epidemics[38].

There is an urgent need to update stripe rust surveillance mechanisms to efficiently detect the emergence and spread of new races of *Pst* within the United States to improve preparedness for inevitable future epidemics. Field pathogenomics is a novel approach to determine the genetic variability of *Pst* populations which improves upon many of the existing surveillance mechanisms[39]. Made possible by the recent and rapid advances in sequencing technologies, large volumes of transcriptome data are generated from infected wheat leaves, providing high resolution diagnostic information for both pathogen and host simultaneously[40]. This approach could be applied to monitoring populations of stripe rust in both agricultural systems and natural grass populations as a means for predicting new races of stripe rust and/or previously undetected races that may be important in future years. This method would also address questions regarding effector gene diversity and abundance in *Pst* populations. This has been a longstanding topic of interest to pathologists, as the abundance of a specific effector gene will rise and fall in the pathogen population depending on the prevalence of a corresponding resistance gene in the host population. It is unclear whether an effector associated with a new virulent race will decrease in abundance in the pathogen population when the corresponding resistance gene is

removed from production. This may be further complicated in *Pst*, as part of its success in the central United States and Canada appears to be an adaptation to warmer, drier climates and not just evolution of new virulent races. Therefore, epidemiological models will need to take into consideration all fitness attributes, including the effector composition of stripe rust populations, as well as adaptation of *Pst* populations to the climate of central and eastern United States and Canada.

Finally, as we attain a greater understanding of the evolution, ecology, and epidemiology of *Pst*, this information will need to be used to develop durable resistance strategies. With the expanding geographic footprint of stripe rust, plant breeders and pathologists need to work together in order to develop and deploy durable stripe rust resistance strategies[41]. In the case of *Pst*, this will include: (i) continuing to screen breeding material with a diverse population of *Pst* isolates; (ii) deployment of varieties with different resistance mechanisms (single-gene, quantitative, adult-plant, seedling stage); and (iii) development of varieties with multiple resistance genes or gene pyramids. While each of these strategies is theoretically determined to provide more durable resistance to *Pst*, it is unclear how durable quantitative resistance and gene pyramids will be in the field. It will be important to continue to test the longevity of these strategies, as well as their effect on the *Pst* population over time.

This perspective arose from a presentation at the Canadian Phytopathological Society annual meeting in a symposium entitled "Biovigilance: A framework for effective pest management." A title could not be more apropos, as rust never sleeps, so plant pathologists and breeders must remain ever vigilant against the development of novel races. But rather than waiting for the next new race to develop, we must change the paradigm and try to predict and be prepared for the emergence of the next new race. By utilizing new capabilities from low-cost, high-throughput sequencing to include a

more diverse sample of isolates, both historic and contemporary, we can gain greater insight into how *Pst* and other rust fungi evolve over time, and how populations evolve from year to year. Strategies for combating dynamic *Pst* populations, when available resistance genes fail, are essential for developing long-term, durable resistance in novel varieties of cultivated wheat. A dual-interaction network furthers our understanding of biological mechanisms during active *Pst* infection and probes for potential novel breeding targets within the wheat host for more durable resistance. These connections give insight into how the cross-kingdom interact during active infection and show conserved interactions from worldwide field samples. Using this knowledge coupled with further investigation into how resistance deployment impacts *Pst* populations, we can hopefully begin to develop long-term durable strategies for managing stripe rust.

## References


1. Chen, W., Wellings, C., Chen, X., Kang, Z. & Liu, T. Wheat stripe (yellow) rust caused by Puccinia striiformis f. sp. tritici. *Mol. Plant Pathol.* **15**, 433–446 (2014).
2. Beddow, J. M. *et al.* Research investment implications of shifts in the global geography of wheat stripe rust. *Nat. Plants* **1**, (2015).
3. Hovmøller, M. S., Walter, S. & Justesen, A. F. Escalating threat of wheat rusts. *Science* **329**, 369 (2010).
4. Hughes M. Cereal rust bulletins. (2016). Available at: https://www.ars.usda.gov/midwestarea/%0Ast-paul-mn/cereal-disease-lab/docs/cereal-rust-bulletins/cereal-rustbulletins/.
5. Hughes, M. Estimated small grain losses due to rust in 2015. (2015). Available at: https://www.ars.usda.gov/midwest-area/stpaul/cereal-disease-lab/docs/small-grain-losses-due-to-rust/small-grain-losses-due-to-rust/.
6. Milus, E. A., Kristensen, K. & Hovmøller, M. S. Evidence for increased aggressiveness in a recent widespread strain of Puccinia striiformis f. sp. tritici causing stripe rust of wheat. *Phytopathology* **99**, 89–94 (2009).
7. Wellings, C. R. Global status of stripe rust: A review of historical and current threats. *Euphytica* **179**, 129–141 (2011).
8. Xia, C. *et al.* Secreted protein gene derived-single nucleotide polymorphisms (SP-SNPs) reveal population diversity and differentiation of Puccinia striiformis f. sp. tritici in the United States. *Fungal Biol.* **120**, 729–744 (2016).
9. Liberatore, K. Cereal rust bulletins. (2016). Available at: https://www.ars.usda.gov/midwest-area/stpaul/cereal-disease-lab/docs/cereal-rust-bulletins/cereal-rust-bulletins/.
10. Liu, T., Wan, A., Liu, D. & Chen, X. Changes of races and virulence genes in Puccinia striiformis f. sp. tritici, the wheat stripe rust pathogen, in the United States from 1968 to 2009. *Plant Dis.* **101**, 1522–1532 (2017).
11. Chen, X., Moore, M., Milus, E., Long, D. & Line, R. Wheat stripe rust epidemics and races of Puccinia striiformis f. sp. tritici in the United States in 2000. *Plant Dis.* **86**, 39–46 (2002).
12. Markell, S. G. & Milus, E. A. Emergence of a Novel Population of Puccinia striiformis f . sp . tritici in Eastern United States. *Phytopathology* **98**, 632–639 (2008).
13. Milus, E. A., Seyran, E. & McNew, R. Aggressiveness of Puccinia striiformis f.sp. tritici isolates in the south-central States. *Plant Dis.* **90**, 847–852 (2006).
14. Ali, S. *et al.* Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia striiformis f.sp. tritici. *PLoS Pathog.* **10**, (2014).
15. Palm, M. E. *Systematics and the Impact of Invasive Fungi on Agriculture in the United States*. *BioScience* **51**, (2001).
16. Singh, R. P. *et al.* Current status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* **1**, (2006).
17. Hovmøller, M. S., Yahyaoui, A. H., Milus, E. A. & Justesen, A. F. Rapid global spread of two aggressive strains of a wheat rust fungus. *Mol. Ecol.* **17**, 3818–26 (2008).
18. Campbell, F. T. The science of risk assessment for phytosanitary regulation and the

impact of changing trade regulations. in *BioScience* **51**, 148–153 (Oxford Academic, 2001).
19. Perrings, C. *et al.* Biological invasion risks and the public good: An economic perspective. *Ecol. Soc.* **6**, (2002).
20. Brown, J. K. M. & Hovmøll, M. S. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537–541 (2002).
21. Wellings, C. R. Puccinia striiformis in Australia: A review of the incursion, evolution, and adaptation of stripe rust in the period 1979-2006. in *Australian Journal of Agricultural Research* **58**, 567–575 (CSIRO PUBLISHING, 2007).
22. Hovmøller, M. S. *et al.* Replacement of the European wheat yellow rust population by new races from the centre of diversity in the near-Himalayan region. *Plant Pathol.* **65**, 402–411 (2016).
23. Ali, S. *et al.* Yellow rust epidemics worldwide were caused by pathogen races from divergent genetic lineages. *Front. Plant Sci.* **8**, 1057 (2017).
24. de Vallavieille-Pope, C. *et al.* Virulence dynamics and regional structuring of Puccinia striiformis f. sp. tritici in France between 1984 and 2009. *Plant Dis.* **96**, 131–140 (2012).
25. Walter, S. *et al.* Molecular markers for tracking the origin and worldwide distribution of invasive strains of Puccinia striiformis. *Ecol. Evol.* **6**, 2790–2804 (2016).
26. Wan, A. M., Chen, X. M. & Yuen, J. Races of Puccinia striiformis f. sp. tritici in the United States in 2011 and 2012 and Comparison with Races in 2010. *Plant Dis.* **100**, 966–975 (2016).
27. Sharma-Poudyal, D. & Chen, X. M. Models for predicting potential yield loss of wheat caused by stripe rust in the U.S. Pacific Northwest. *Phytopathology* **101**, 544–554 (2011).
28. Severns, P. M., Estep, L. K., Sackett, K. E. & Mundt, C. C. Degree of host susceptibility in the initial disease outbreak influences subsequent epidemic spread. *J. Appl. Ecol.* **51**, 1622–1630 (2014).
29. Sha, Y., Phan, J. H. & Wang, M. D. Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data. in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* **2015**-**Novem**, 6461–6464 (2015).
30. Yan, J., Luo, Y., Chen, T., Huang, C. & Ma, Z. Field distribution of wheat stripe rust latent infection using real-time PCR. *Plant Dis.* **96**, 544–551 (2012).
31. Sharma-Poudyal, D., Chen, X. & Rupp, R. A. Potential oversummering and overwintering regions for the wheat stripe rust pathogen in the contiguous United States. *Int. J. Biometeorol.* **58**, 987–997 (2014).
32. Estep, L. K., Sackett, K. E. & Mundt, C. C. Influential disease foci in epidemics and underlying mechanisms: A field experiment and simulations. *Ecol. Appl.* **24**, 1854–1862 (2014).
33. Brar, G. S. & Kutcher, H. R. Race Characterization of Puccinia striiformis f. sp. tritici , the Cause of Wheat Stripe Rust, in Saskatchewan and Southern Alberta, Canada and Virulence Comparison with Races from the United States. *Plant Dis.* **100**, 1744–1753 (2016).
34. Cheng, P., Chen, X. & See, D. Grass Hosts Harbor More Diverse Isolates of Puccinia striiformis Than Cereal Crops. *Phytopathology* **106**, 362–371 (2016).
35. Cantu, D. *et al.* Genome analyses of the wheat yellow (stripe) rust pathogen Puccinia striiformis f. sp. tritici reveal polymorphic and haustorial expressed secreted proteins as

candidate effectors. *BMC Genomics* **14**, (2013).
36. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
37. Wan, A. & Chen, X. Virulence characterization of Puccinia striiformis f . sp . tritici using a new set of Yr single-gene line differentials in the United States in 2010. *Phytopathology* **98**, 1534–1542 (2014).
38. Naccache, S. N. *et al.* A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.* **24**, 1180–1192 (2014).
39. Hubbard, A. *et al.* Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biol.* **16**, 23 (2015).
40. Westermann, A. J., Gorski, S. A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nature Reviews Microbiology* **10**, 618–630 (2012).
41. Mundt, C. C. Durable resistance: A key to sustainable management of pathogens and pests. *Infection, Genetics and Evolution* **27**, 446–455 (2014).

## Chapter 2: Field patho-transcriptomics approach determines North American *Pst* population is distinct from global population due to asexual changes within the endemic population and newer *Pst* have less candidate effectors

### Introduction

Wheat is an important global staple crop and an economically important commodity for growers in North America. For example, the United States produced about 2.4 billion bushels of wheat valued at $9.1 billion in 2016[1]. Unfortunately, this crop is threatened by plant pathogens. Stripe rust, caused by the fungal pathogen *Puccinia striiformis* f. sp. *tritici* (*Pst*), is one of the most devastating diseases of wheat in the United States and globally. This obligate biotrophic basidiomycete fungi threatens wheat production by diminishing photosynthesis during the important grain-filling stage, leading to yield losses. *Pst* causes the highest yield losses when infection occurs after the flag leaf has emerged[2]. Stripe rust commonly causes between 10–70% crop loss, but can cause up to 100% yield loss in a heavily infected field[2].

*Pst* was first introduced to North America from northwestern Europe sometime early in the 20$^{th}$ century[3]. In the Pacific Northwest, stripe rust is observed each year due to favorable conditions for *Pst* growth and infection, including low night temperatures, dew formation, and year-round availability of host plants[2,4]. In contrast, there have been less frequent epidemics causing high yield losses east of the Rocky Mountains. For example, *Pst* was first reported in 1941 in Texas and a larger epidemic occurred in 1957, the same year *Pst* was first reported in Kansas[5,6].

A major shift in prevalent virulence factors was seen in North America in 2000. These newer US races had a broader virulence profile, were more aggressive at higher temperatures, and were found to be genetically distinct from previous races identified in the eastern United

States[7,8]. These novel races appeared to be from the introduction of races originating in the East African-Middle Eastern region[3,6]. The geographic footprint of *Pst* in the United States has expanded in the past two decades and is thought to be due to mutations or introductions of novel races[3,9,10]. In addition, the number of races in the United States has continued to increase since 2000, with hundreds of races identified by 2018[11]. Fungicides may be used to control this growing stripe rust problem. For example, the strobilurin class of fungicides are applied before *Pst* infects wheat or during early stages of *Pst* spore growth, and the triazole fungicides are applied after *Pst* infection has occurred[12,13]. Although stripe rust can be controlled with fungicides, a preferred approach is to breed wheat to have resistance genes against *Pst* infection. This approach is more economical and environmentally friendly than fungicide applications. Understanding the population structure of *Pst*, and the origin of shifts in population structure within North America and worldwide, help wheat breeders evaluate which resistance genes in wheat are most likely to be needed in future years.

Race-typing currently plays a major role in determining the population structure of *Pst*, which is vital in shaping which resistance genes to include in wheat-breeding programs. Historically, *Pst* has been classified into races based on its virulence phenotype, which is the compatibility of *Pst* with wheat plants possessing different resistance genes. Wheat plants with a major resistance gene (*Yr* gene) are resistant to pathogens with a corresponding avirulence (*Avr*) gene product. The race of a *Pst* isolate is determined by comparing the isolate against a differential set of wheat cultivars, each with a single major *Yr* gene. Specifically, the race of the *Pst* isolate is determined by evaluating which plant resistance genes the *Pst* isolate is able to infect or defeat.

Around the world, over 40 *Yr* resistance genes have been described, but the differential set of wheat cultivars used to determine stripe rust races represents only 18 major resistance genes, with an additional 9 major resistance genes in the supplemental set[14,15]. The differential set of wheat cultivars was last updated in 2010 and is usually modified when more *Yr* resistance genes are overcome in the field by *Pst* infection[15]. In the last five years (2015 to 2019), 125 *Pst* races have been identified on wheat plants in the United States (including 69 new races)[16]. Unfortunately, race-typing can be time-consuming, subjective, and labor intensive while only giving limited phenotypic information[17]. Since close monitoring of the *Pst* population for new incursions is essential for breeders and growers to react quickly to changing populations of *Pst*, a new approach using field patho-transcriptomics (also known as field pathogenomics) was developed to speed up the process of identifying strains and candidate effector diversity in the pathogen population, providing valuable information for breeders as they develop new wheat varieties[17,18].

Fundamental to race-typing are the candidate effector proteins which interact with the resistance genes of the host. Candidate effectors are proteins produced by pathogens that modify host-cell structure and function, such as enhance infection for the pathogen. To counteract pathogen infection, the plant develops mechanisms to recognize candidate effectors and inhibit pathogen infection, leading to a rapid evolution for both the pathogen candidate effectors and plant responses[19,20]. In filamentous plant pathogens like *Pst*, candidate effector genes are frequently under selection pressure to drive evolution of these resistance genes[21]. A number of *Puccinia* candidate effectors, including those in *Pst*, have been identified that increase host susceptibility to the pathogen; however, the diversity of these candidate effectors has not been studied on a global population scale[22,23]. While race-typing uses the differential set of wheat

cultivars to focus on the presence or absence of a few avirulence genes, field patho-transcriptomics evaluates all candidate effectors.

The fungal pathogen *Pst* is clonally propagated with little sexual reproduction and low genetic diversity throughout most of its global range[3]. Asexual reproduction can occur for this heteroecious fungus when wheat plants are present, while sexual reproduction can occur when the alternate host *Berberis spp.* (barberry) is present[24]. The Himalayan region has *Berberis spp.* and a higher diversity of *Pst,* indicating sexual reproduction occurs in the Himalayan region while *Pst* in the rest of the globe only undergoes asexual reproduction[3,25]. There is no evidence of sexual reproduction occurring in North America. Although *Berberis vulgaris* (barberry) is present in the Pacific Northwest, it does not act as an alternate host for *Pst* due to *Pst* teliospore degradation and barberry phenology[26].

Although asexual reproduction limits genetic diversity, changes in the *Pst* population can occur in several ways. The primary drivers for increased variation are either introductions of new isolates from outside the region or asexual genetic changes[26]. New races from around the world can be introduced into the United States through the transport of plant material or through the movement of spores by winds like the westerlies or trade winds. Once in North America, the wind corridor known as the Puccinia Pathway could provide a route for races to move within North America because spores overwintering in the southern United States and northern Mexico can be carried by springtime winds up north through the United States and into Canada. Asexual genetic changes in the *Pst* population can occur when races evolve through mutations that change the protein structure of virulence factors or delete virulence factors, thereby limiting the effectiveness of host resistance genes[6]. In a co-evolutionary arms race, *Pst* populations mutate to accumulate or lose specific *Avr* genes, resulting in a small number of individuals being able to

avoid recognition by the predominant resistance genes present in the host population. These novel *Pst* races will then quickly become the most abundant races in the population, resulting in widespread infection and yield loss. The goal of breeders is to rapidly identify resistant germplasm that can be used to develop new wheat varieties that are able to resist infection by these new races of *Pst*. Unfortunately for breeders, current *Pst* races can rapidly transform into new *Pst* races through a change in a single avirulence protein or through loss of virulence factors[10].

It is important to note that processes such as sexual, parasexual, and somatic recombination can be important drivers of genotypic diversity[27]. These asexual changes could be in the form of single-step mutation (which has been the origin of new avirulence and virulence combinations in *Pst*) or somatic hybridization (which occurred with a different *Puccinia* species when a new, virulent lineage of *Puccinia graminis* f. sp. *tritici* named Ug99 arose from somatic hybridization[27,28]).

When comparing *Pst* virulence profiles, international collections are similar to US collections. This similarity may be a reflection of the same resistance genes being deployed over multiple continents rather than a reflection of similar genetic background[29]. For example, despite high genetic diversity, there was a limited number of pathotypes in Nepal due to a deployment of a limited number of resistance genes in the wheat population which lead to a convergence of pathotypes[30]. In addition, while virulence profiles were similar between China and the United States, isolates from those two countries appeared to generally evolve independently[31].

In Europe in 2011, there was an introduction of a new *Pst* race, possibly from the near-Himalayan region to Europe, known as the "Warrior race." This new race caused disease damage on adult wheat plants which had previously demonstrated effective resistance to *Pst*.

In the first year after detection, the Warrior race was seen in high frequencies in multiple countries around Europe[32]. To date, it is unknown whether this Warrior race has been introduced into North America.

The ever-changing *Pst* population presents a challenge for researchers using race-typing to keep up with the population. For example, novel races have been observed in the central United States that can defeat the *Yr17* gene and the TAM111QTL; however, these races cannot be distinguished from race 37 using the standard differential set of major resistance genes. In contrast, the field patho-transcriptomics approach has found mutations, defined *Pst* strains, assigned strains to genetic lineages that correlated with their virulence profiles, and demonstrated a shift within a United Kingdom population of *Pst*[17,18].

In the last two decades, *Pst* incidence and disease loss have increased significantly in North America, particularly east of the Rocky Mountains. Additionally, novel races of *Pst* have emerged since 2010, but our current method of race-typing using the differential set of wheat cultivars is time-consuming and only gives limited information about the candidate effector diversity of *Pst* isolates in the field. Based on these problems, this study uses a field patho-transcriptomics approach to determine if novel races of *Pst* represent introduction events to North America or asexual changes within the resident population. This approach also allows us to quantify the presence/absence of specific candidate effectors in the isolates.

**Methods**

*Sample Collection, RNA Extraction, and Sequencing*

Sixteen wheat leaves infected with *Puccinia striiformis* f. sp. *tritici* were collected in 2017 from wheat fields in 8 states in the United States and placed in tubes containing RNAlater. In addition, one isolate (PST-100) originally collected in 2001 from Kansas (as part of the race

group introduced to North America in 2000) was inoculated on a susceptible cultivar (Ripper) and placed in a dew chamber in darkness and 100% humidity at 10°C for 36 hours. The plants were then grown in a growth chamber set at 18°C for the highest temperature and 4°C for the lowest temperature. The *Pst*-infected wheat leaf was collected in RNAlater when the seedling was at the 4-leaf stage.

Total RNA from the *Pst*-infected wheat leaves stored in RNAlater were extracted using a Qiagin RNeasy kit. A NanoDrop spectrophotometer was used to test the quality before sequencing using Illuminia HiSeq 4000 through BGI Americas Corporation. There were between 24 and 90 million total reads for each of the 17 *Pst*-infected wheat leaves.

*Sequence Acquisition and Alignment*

A total of 313 *Pst*-infected wheat transcriptomes and *Pst* genomes from around the world were included in this study. There were 59 samples from North America, 5 from South America, 11 from Africa, 39 from Asia, 195 from Europe, and 4 from Oceania. Of these samples, 17 were new transcriptomes generated by this study while 296 samples were obtained from previously published papers through either the European Nucleotide Archive or shared by authors[17,18,33]. The raw reads from these 17 collected samples and from the 296 transcriptomes were trimmed using fastp with default parameters[34]. The trimmed reads were aligned to the reference genome PST-78, which was originally collected in the Great Plains of the United States in 2000 and obtained through Ensemble[35]. This reference genome was indexed, and the first pass alignments were performed using STAR Version 2.7.5a with a max number of loci anchors of 200 and split reads no longer than 30. A second index was created from the initial alignments which annotated all splice junctions with the length of the genomic sequence around the junction set to 99. The second pass was performed using the same parameters as the first pass, but with the second

indexed genome[36]. The resulting bam files were sorted, indexed, and the pileup format generated using SAMtools Version 1.9[37].

The percentage of reads mapped to the *Pst* genome ranged from 1–86% and the percentage of reads aligned to the wheat genome ranged from 1–69%. The allele frequency revealed that each of the samples consisted of a single isolate.

*SNP Genotyping and Filtering*

Single nucleotide polymorphisms (SNPs) were identified using the pipeline and scripts from previously published work[38]. Only SNPs with at least 20x coverage were included in the final analysis. The frequency of the alleles was graphed to determine if there were one or more isolates present per sample. Synonymous and nonsynonymous SNPs were identified using snpEff [39]. Samples with less than 50% coverage were removed. The synonymous SNPs were further filtered; only SNPs that were present in 80% of samples and had a minimum allele frequency of 0.2 were used for further analysis.

In total, there were 313 samples with 143,183 SNPs having at least 20x coverage. By setting a minimum allele frequency of 0.2, the number of SNPs was decreased to 91,717. By stating that SNPs needed to be present in at least 80% of samples, the number of SNPs decreased to 12,359. Samples with less than 50% of their SNPs present in the data were removed, which decreased the number of samples to 295.

After filtering, there were 57 samples from North America collected in the years 2015–2017 and 1 sample originally collected in North America in 2001. There were 10 samples from Africa collected between 2014 and 2016, 39 samples from Asia collected between 2014 and 2017, 179 samples from Europe collected between 1978 and 2017, 4 samples from Oceania collected between 2006 and 2012, and 5 samples from Chile collected in 2014 (Appendix 1). All

*Pst* samples were collected on wheat except for 11 samples from Europe, where one sample was collected on durum wheat, one sample was collected on perennial grass, one sample was collected from rye, and 8 samples were collected from triticale.

*Population Genetics*

A multivariate discriminant analyses of principal components (DAPC) was performed using the adegenet package in R environment[40–43]. The number of genetic clusters was determined by comparing the value of Bayesian Information Criterion (BIC) to the value of the K-means clustering algorithm. Individuals were assigned to clusters based on membership probability. To establish the appropriate number of principal components, the data were cross-validated using poppr Version 2.8.6[44,45]. A phylogenetic analysis was performed using FastTree Version 2.1.11 SSE3 and the generalized time-reversible model[46]. The input file was created by including genes with at least 2x coverage, then only genes with at least 50% of the gene covered in 50% of the isolate were included to create the file for the phylogenetic analysis[38]. The tree was visualized using iTOL[47].

*Candidate Effector Analysis*

To determine candidate effectors, SignalP was used in order to establish which genes had a signal peptide[48]. Non-synonymous SNPs with at least 20x coverage were used to calculate how many SNPs per candidate effector. A binary matrix indicating presence and absence of candidate effectors was created. If nucleotides were present in 80% or more of the candidate effector, the candidate effector was considered present and given a "1;" otherwise, the value was a zero. There were 80 uninformative candidate effectors that were removed, 54 candidate effectors were present in every isolate, and 26 candidate effectors were present in no isolates. This process left

2,216 candidate effectors for subsequent analysis. A boxplot was created and a pairwise comparison using Wilcoxon rank sum test was calculated using RStudio Version 4.0.1.

**Results**

The results of the DAPC analysis identified 8 distinct genetic clusters (Figure 2.1) among the global population of *Pst* included in this study. Multiple genetic groups of *Pst* were found in Europe (5 clusters), North America (3 clusters), and Asia (2 clusters). Genetic Cluster 1 consisted of samples from Europe collected between 2013 and 2015 (Germany, France, United Kingdom, Italy, and Poland), samples from South Africa collected in 2016, samples from New Zealand collected in 2011–2012, and the historical *Pst* isolate PST-100 originally collected in 2001 in Kansas, US (part of the race group introduced to North America in 2000). Genetic Cluster 2 was composed entirely of samples from North America collected between 2015 and 2017 (United States and Canada). The final cluster with North American samples was Genetic Cluster 7, which included samples from the United States collected between 2015 and 2016 (specifically Oregon and Washington) and samples from China collected between 2015 and 2017. While all the samples in Genetic Cluster 2 came from locations distributed throughout North America between 2015 and 2017, the North American samples in Genetic Cluster 7 only came from the Pacific Northwest between 2015 and 2016, and the only North American sample in Genetic Cluster 1 came from Kansas in 2001 (Figure 2.2).

The remaining clusters contained no samples from North America. Specifically, Genetic Cluster 3 consisted entirely of samples from Europe collected between 2011 and 2014 (Belarus, Germany, France, United Kingdom, and Poland). The four samples known to be the Warrior race clustered together in Genetic Cluster 4. All the other samples in Genetic Cluster 4 were originally collected in Europe between 2011 and 2014 (Austria, Belarus, Bulgaria,

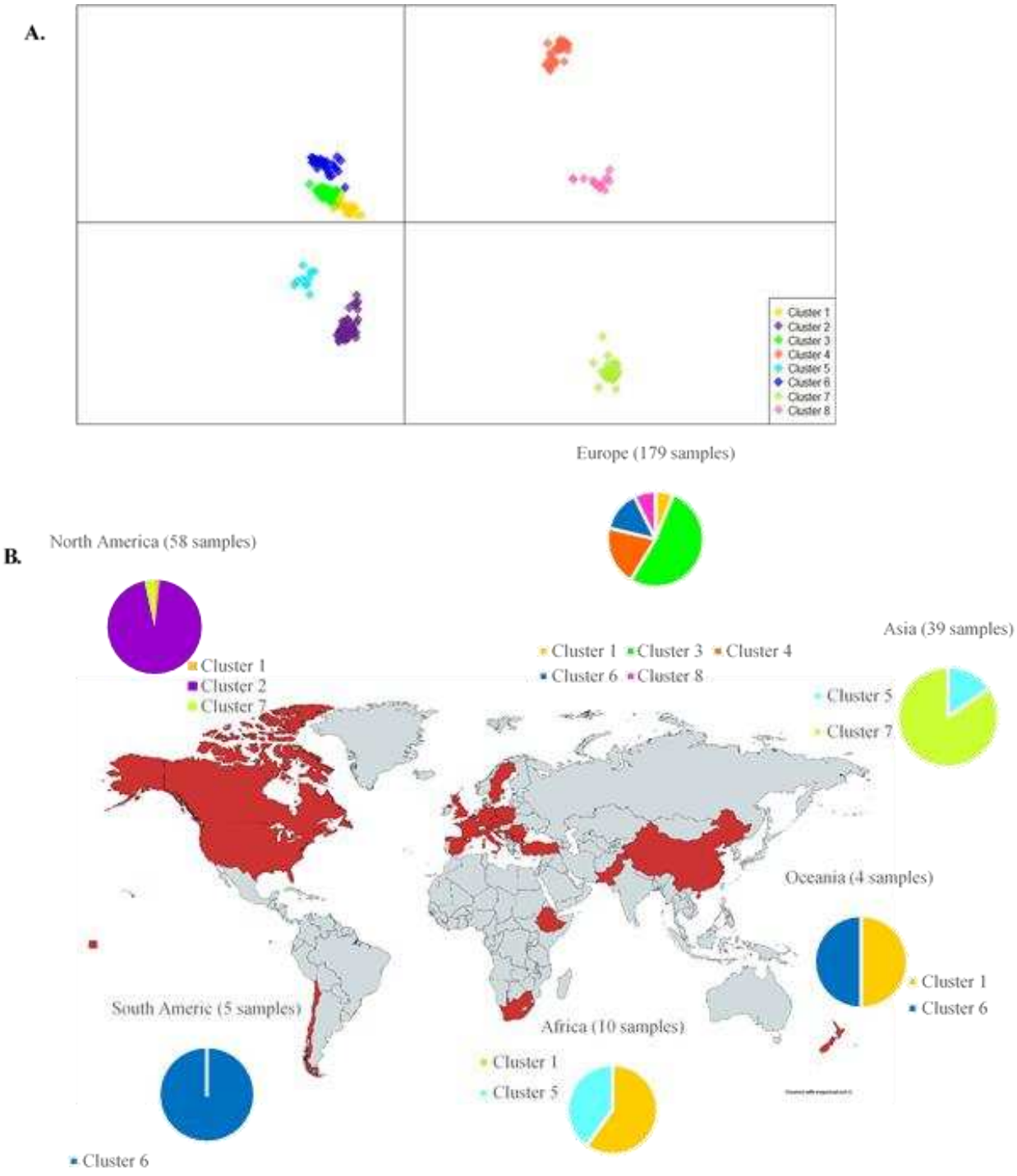


*Figure 2.1A: DAPC analysis showing the data forming into 8 clusters.*
*B: Countries where samples were taken are in red and pie charts by each continent show the percentage of isolates in each cluster.*

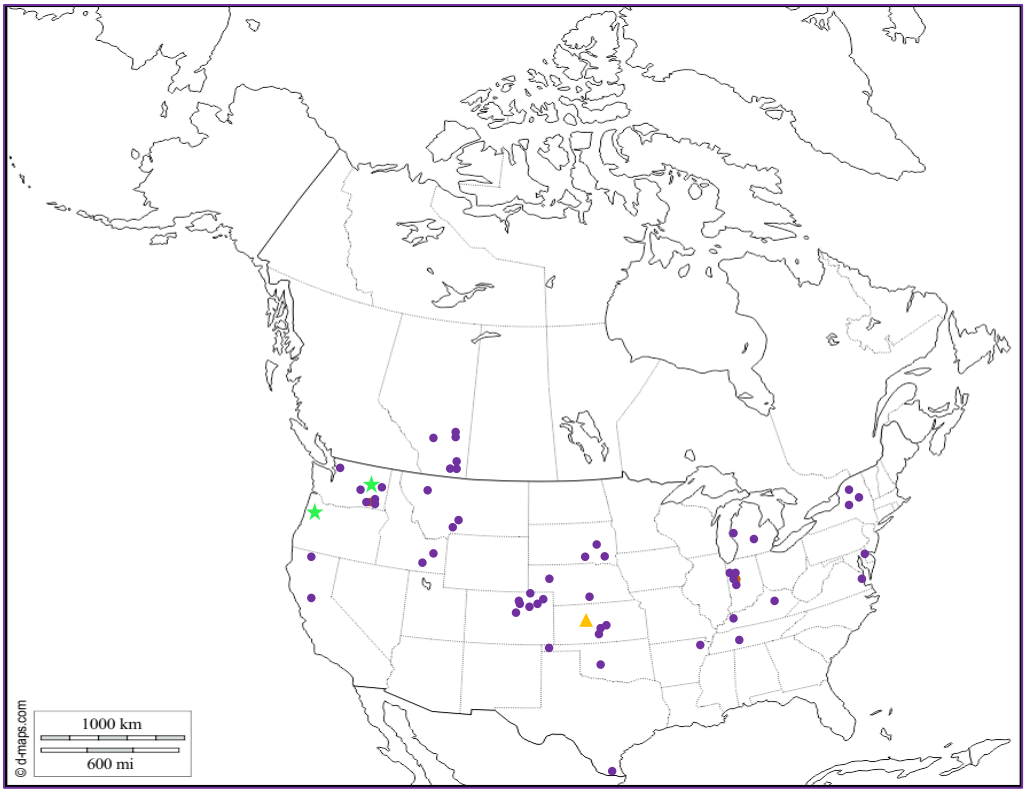


*Figure 2.2: Samples collected in the United States colored by genetic cluster. The green stars represent Genetic Cluster 7. The purple circles represents Genetic Cluster 2, and the orange triangle represents Genetic Cluster 1.*

Czech Republic, Spain, France, United Kingdom, Croatia, Italy, Poland, Romania, Serbia, Slovokia, and Turkey). Genetic Cluster 5 was comprised of samples from Ethiopia and Pakistan collected between 2014 and 2016. Genetic Cluster 6 was made up of samples from Chile, Germany, France, United Kingdom, and New Zealand collected between 1978 and 2014. Lastly, Genetic Cluster 8 consisted of samples collected between 2014 and 2017 in the United Kingdom, Poland, and Sweden.

After filtering, 10,984 genes were included to create the phylogenetic tree, which further confirmed population subdivisions with strong geographic connection (Figure 2.3). In fact, each genetic cluster from the DAPC analysis grouped together in the phylogeny with one exception. Originally identified as part of Genetic Cluster 6 (blue color), a single United Kingdom sample

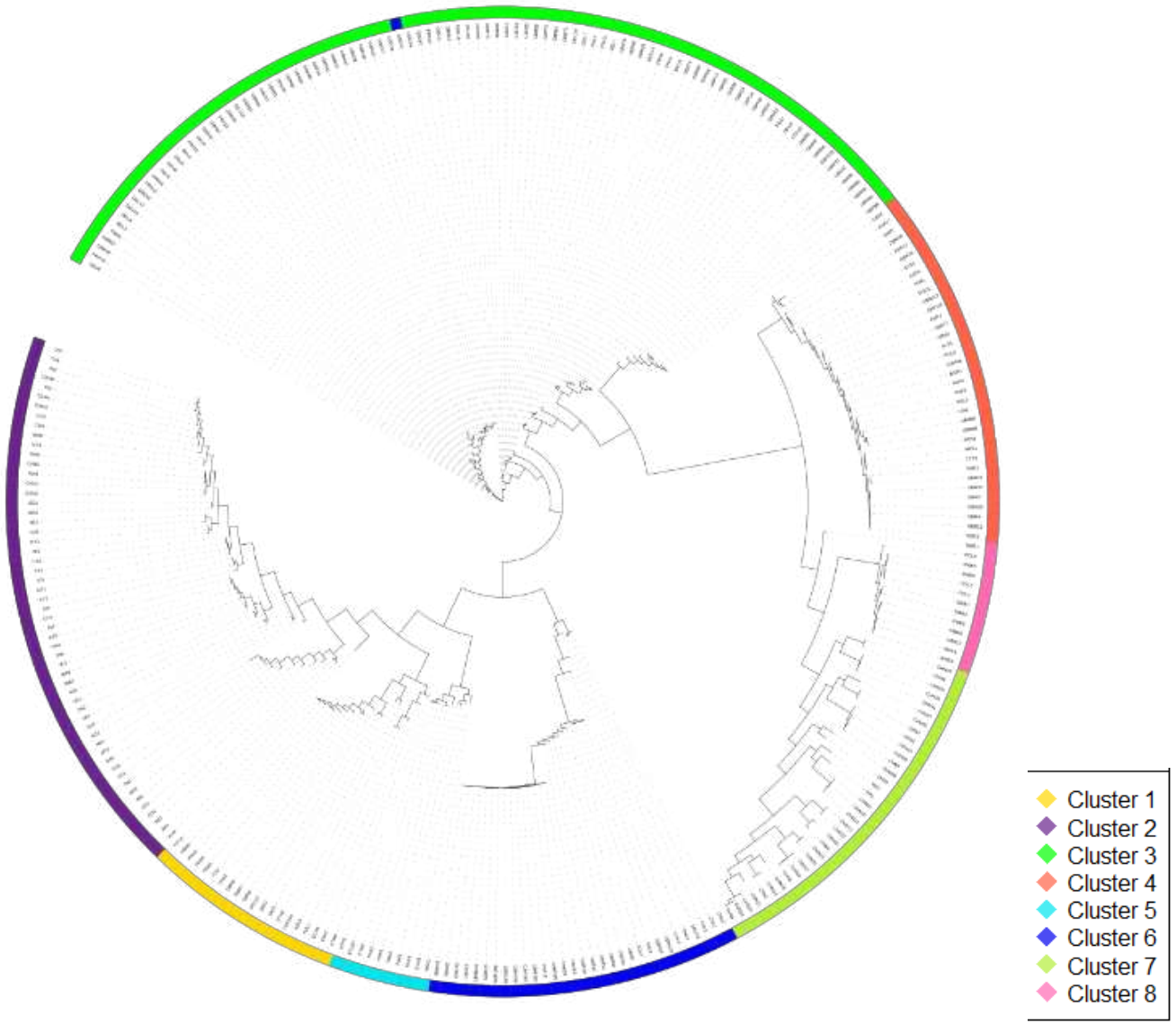


*Figure 2.3: Phylogeny of* Pst *isolates by FastTree with the ring of colors being the same as the clusters from the DAPC analysis.*

collected in 2013 was now among the Genetic Cluster 7 (green), which had samples from China and the Pacific Northwest of the United States.

As part of the candidate effector analysis, 2,301 candidate effectors were identified through SignalP. After removing 54 candidate effectors present in every isolate and removing 26 candidate effectors present in no isolates, 2,216 candidate effectors were left. A boxplot was created for the presence/absence matrix that organized the results by the 8 genetic clusters determined by the DAPC analysis (Figure 2.4).

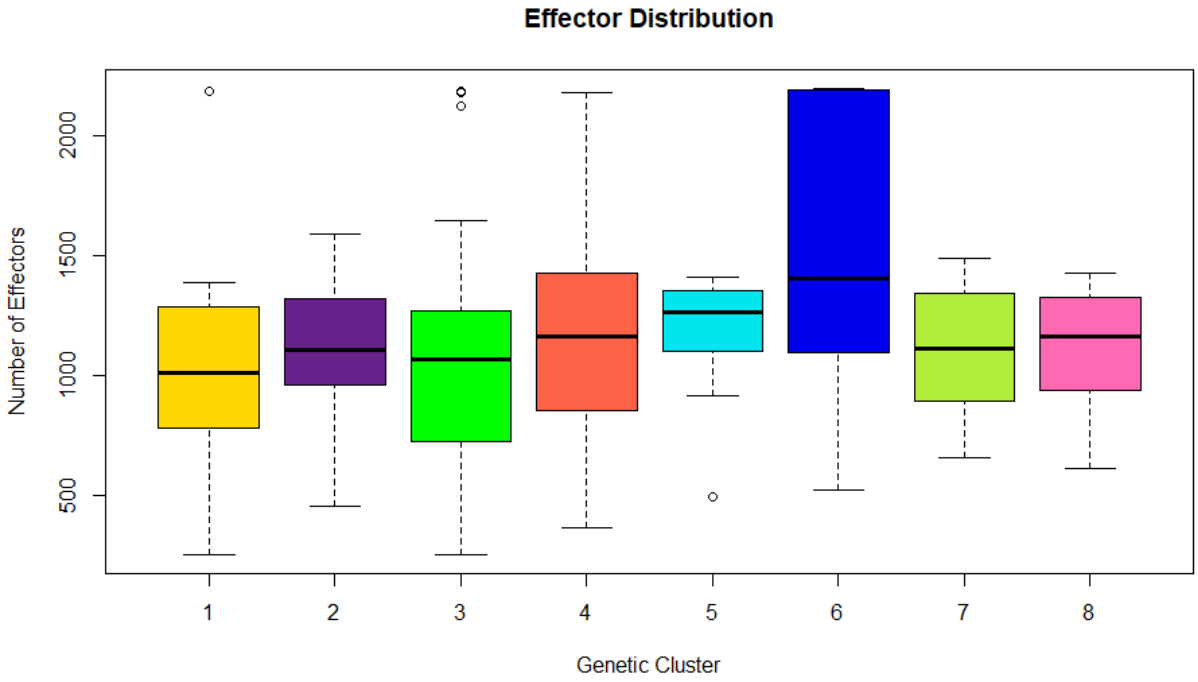


*Figure 2.4: Boxplot of presence/absence matrix with samples organized by genetic cluster.*

To determine p-values, pairwise comparisons of presence/absence data by genetic cluster were done using Wilcoxon rank sum test with continuity correction (Table 2.1). Genetic Cluster 6 showed a significant difference in the number of candidate effectors present, as

p-values were consistently below 0.05 when Genetic Cluster 6 was compared to each of the other clusters. No other genetic cluster showed significant differences with other genetic clusters.

*Table 2.1: Pairwise comparisons of presence/absence data by genetic cluster using Wilcoxon rank sum test with continuity correction. The p-values are shown when comparing genetic clusters.*

| | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **Cluster 5** | **Cluster 6** | **Cluster 7** |
|---|---|---|---|---|---|---|---|
| **Cluster 2** | 0.32294 | - | - | - | - | - | - |
| **Cluster 3** | 0.85809 | 0.24031 | - | - | - | - | - |
| **Cluster 4** | 0.29296 | 0.62627 | 0.18546 | - | - | - | - |
| **Cluster 5** | 0.18288 | 0.28748 | 0.16982 | 0.63166 | - | - | - |
| **Cluster 6** | 0.00096 | 0.00057 | 3.3e-05 | 0.01041 | 0.05300 | - | - |
| **Cluster 7** | 0.26918 | 0.97689 | 0.23438 | 0.58486 | 0.42657 | 0.00134 | - |
| **Cluster 8** | 0.41251 | 0.97512 | 0.48936 | 0.72556 | 0.52238 | 0.01460 | 0.93528 |

**Discussion**

While *Pst* is known to sexually reproduce when the alternate host *Berberis spp.* (barberry) is present, sexual reproduction of *Pst* does not occur with the barberry population of the Pacific Northwest of the United States due to teliospore degradation and barberry phenology[24,26]. Therefore, the population structure of *Pst* in North America is driven by clonal lineages with diversity created by either introductions from outside the continent or from asexual changes within the endemic population. In 2017 in the United States, there were 64 *Pst* races identified[11]. The DAPC analysis in this study determined that North American samples fell into only three genetic clusters, which reinforces the concept that within North America, there are a limited number of clonally propagated genotypes despite the large number of pathotypes identified by race-typing.

This study describes the population structure of the *P. striiformis* population in North America compared to the population in the rest of the world. Of the three genetic clusters found

in North America (Figure 2.2), Genetic Cluster 1 had a single sample from the United States that was originally collected in 2001 in Kansas. The Kansas sample clustered with samples collected from Europe between the years 2013 and 2015, Oceania collected between the years 2011 and 2012, and Africa collected between the years 2014 and 2016. The Kansas sample was collected in 2001 after a major shift in prevalent virulence factors was seen in North America in 2000 due to the introduction of novel races from outside North America, which changed the population structure in the United States[3,6]. The post-2000 race group in North America was shown to have increased aggressiveness at higher temperatures in a study analyzing Amplified Fragment Length Polymorphism (AFLP) markers, latent period, length and area of lesion, and number of spores[7]. Another study used virulence diversity and AFLP markers to show that the post-2000 population replaced the older population as a result of an introduction[8]. In addition, a study using the differential set of wheat cultivars also gave evidence supporting an introduction, while stating that other races later evolved from the endemic population in the United States[6]. Later using SSRs, the post-2000 *Pst* population in North and South America was shown to be an introduction from East African-Middle Eastern region[3]. These studies support the findings of the DAPC analysis in this study, which determined that Genetic Cluster 1 was associated with the post-2000 global spread of *Pst* isolates that were more adapted to warmer, drier climates such as those in the central plains of the United States. More modern North American samples (collected between 2015 and 2017) were not part of Genetic Cluster 1, nor were older *Pst* isolates such as the genomes collected in the 1970s and 1980s in the United Kingdom that were instead found in Genetic Cluster 6.

Of the genotypes tested from North America, 94.8% clustered into Genetic Cluster 2. This genetic cluster was found throughout the sampled areas in North America and consisted of

isolates collected between 2015 and 2017. No samples from other countries clustered with these genotypes, indicating there had been no introduction event but rather the occurrence of asexual changes such as mutations or somatic hybridization within the endemic *Pst* population. This conclusion is supported by previous research using the differential set of wheat cultivars, which determined most of the new races identified in the United States in 2011 and 2012 evolved from previously existing races through a change in a single virulence factor from the endemic population, suggesting mutations were the source of the new races[10]. In addition, somatic hybridization may have occurred with co-infection on wheat plants, similar to how stem rust shows evidence of somatic hybridization on wheat[28]. In laboratory settings, it has been demonstrated that *Puccinia striiformis* can undergo somatic hybridization[49]. Whether by mutations or somatic hybridization, the largest population of *Pst* in the United States between 2015 and 2017 was both genetically distinct from the post-2000 racial group (Genetic Cluster 1) and genetically distinct from *Pst* populations around the world.

The final genetic cluster that included North American samples was Genetic Cluster 7, which consisted of samples collected in China between 2015 and 2017 and two samples from the United States (one sample collected from the state of Oregon in 2015 and the other collected from the state of Washington in 2016). The composition of this genetic cluster supports a previous study where a phylogeny based on polymorphic SNPs was performed and samples from China and the United States fell into the same clade[17]. An introduction event between long distances such as across the Pacific Ocean is most likely due to human activity[3]. The lack of samples from China in other North American clusters indicates there has been a limited amount of genetic exchange between Asia and North America. While regional spread within North

America may occur due to winds, this introduction event into the Pacific Northwest did not spread beyond that region[3].

This study established that the Warrior race samples clustered in Genetic Cluster 4 along with samples from multiple countries in Europe. No samples in this study from outside the European continent clustered with the Warrior race. Therefore, it appears that the Warrior race had not left the European continent as of 2017. Furthermore, no known Warrior race isolates were identified in the genetic clusters found in North America. This finding indicates that, as of 2017, the Warrior race had not been introduced into North America. In addition, this finding further supports the conclusion that race shifts in the United States during 2013 and 2015 were not associated with an introduction event, but rather adaptation within the endemic population in the United States.

A few additional observations can be made about this study's genetic clusters. The DAPC analysis in this study determined that the South American samples all clustered into Genetic Cluster 6. This phenomena may be due to the limited number of samples collected from South America (only 5 samples) that were all collected in the year 2014. Conversely, while *Pst* samples from six continents were evaluated in this study, samples from northern Europe were overrepresented due to this region's extensive surveys[32]. This overrepresentation may explain why *Pst* samples from Europe clustered into five genetic clusters (specifically, Genetic Clusters 1, 3, 4, 6, and 8). Lastly, Genetic Cluster 5 was made up of samples from Pakistan and Ethiopia that were all collected between 2014 and 2016. Samples from these two countries also clustered together when SNPs were used in a DAPC analysis in a previous study[17].

As part of the candidate effector analysis, the number of present candidate effectors versus the number of absent candidate effectors were analyzed in terms of the 8 genetic clusters. Only

Genetic Cluster 6 was significantly enriched in the number of candidate effectors when compared to other genetic clusters. Specifically, Genetic Cluster 6 showed a significant difference in the number of candidate effectors present, as p-values were consistently below 0.05 when Genetic Cluster 6 was compared to each of the other genetic clusters (Table 2.1). It is important to note that the oldest samples with genomes dating back to 1978 are included in this genetic cluster. For example, Genetic Cluster 6 included samples from Europe collected between 1978 and 2014, samples from New Zealand collected in 2006 and 2012, and the five samples from Chile collected in 2014. On average, these samples had a larger number of candidate effectors present than the rest of the genetic clusters (Figure 2.4). Isolates collected after 2000 have fewer candidate effectors expressed from this list than the genetic and transcriptomic samples in Genetic Cluster 6. This loss of candidate effectors could partially explain the shift in races, as possible avirulence genes were no longer expressed and the host resistance genes were not able to recognize the pathogen, thus leading the newer races to infect the wheat hosts.

Since the newer *Pst* isolates had less candidate effector genes than older isolates, this phenomenon suggests there is some fitness cost to having more candidate effector genes than what is needed to overcome the resistance in the host population. Several studies support this hypothesis. The predominant races that caused severe *Pst* epidemics in 2010 and 2011 had a reduced number of virulence genes than the older race they appeared to evolve from in the western United States[15,50]. This situation is similar to *Puccinia coronata*, the causal agent of oat crown rust, where a fitness cost was associated with strains that had more virulence genes (specifically, more virulence genes were associated with longer latent periods and smaller pustules)[51]. Similarly, researchers investigating *P. infestans* demonstrated that, in the absence of resistance genes in the host population, fitness was negatively correlated with virulence

complexity in genetically related isolates[52]. In addition, there was a negative correlation in wheat leaf rust (*Puccinia triticina*) between two traits that conferred a positive effect (specifically, latent period and fecundity)[53]. However, virulence genes do not always confer a fitness cost. In *Pst*, there did not appear to be a fitness cost for resistance to the host gene *Yr2*, based on the fitness parameters tested[54].

Regardless of the possibility of fitness costs for resistance genes, evidence shows that the host population has a strong evolutionary effect on the pathogen population. Research with *Pst* determined a high frequency of virulence to deployed resistance genes, but an absence of virulence genes for resistance genes that were not deployed in the host population[55]. Both in *Pst* and *P. coronata,* the host population still drove the evolution of virulence levels in local populations despite the trade-off with some life-history traits and, even in the presence of racial diversity, the dominant lineage was driven by host resistance genes[51,55]. In addition, this study found fewer candidate effectors in more modern populations, which indicates the likelihood that the evolution of *Pst* candidate effectors is affected by the genetics of the host population.

In *Puccinia triticina*, the selection for quantitative traits can happen at the isolate level independent of the pathotype, indicating more diversity present than the pathotype[53]. This is similar to *Pst*, where the differential set of wheat cultivars for America was last updated in 2010 and may not reflect the latest changes in the population if changes occur outside the tested avirulence genes[15]. Field patho-transcriptomics, such as the techniques used in this study, are a useful tool in rapidly evaluating shifting populations of *Pst* as they adapt to the local wheat population[18]. Sequencing the transcriptome of *Pst* field samples can be used to both evaluate the synonymous SNPs and look directly at all the expressed candidate effectors.

## References


1. USDA. *Crop values 2016 Summary*. *National Agricultural Statistics Service* **2**, (2017).
2. Chen, X. M. Epidemiology and control of stripe rust [Puccinia striiformis f. sp. tritici] on wheat. *Can. J. Plant Pathol.* **27**, 314–337 (2005).
3. Ali, S. *et al.* Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia striiformis f.sp. tritici. *PLoS Pathog.* **10**, (2014).
4. Cheng, P. & Chen, X. M. Virulence and Molecular Analyses Support Asexual Reproduction of Puccinia striiformis f. sp. tritici in the U.S. Pacific Northwest. *Phytopathology* **104**, 1208–1220 (2014).
5. Milus, E. A., Seyran, E. & McNew, R. Aggressiveness of Puccinia striiformis f.sp. tritici isolates in the south-central States. *Plant Dis.* **90**, 847–852 (2006).
6. Chen, X., Penman, L., Anmin, W. & Cheng, P. Virulence races of Puccinia striiformis f. sp. tritici in 2006 and 2007 and development of wheat stripe rust and distributions, dynamics, and evolutionary relationships of races from 2000 to 2007 in the United States. *Can. J. Plant Pathol.* **32**, 315–333 (2010).
7. Milus, E. A., Kristensen, K. & Hovmøller, M. S. Evidence for increased aggressiveness in a recent widespread strain of Puccinia striiformis f. sp. tritici causing stripe rust of wheat. *Phytopathology* **99**, 89–94 (2009).
8. Markell, S. G. & Milus, E. A. Emergence of a Novel Population of Puccinia striiformis f . sp . tritici in Eastern United States. *Phytopathology* **98**, 632–639 (2008).
9. Beddow, J. M. *et al.* Research investment implications of shifts in the global geography of wheat stripe rust. *Nat. Plants* **1**, (2015).
10. Wan, A. M., Chen, X. M. & Yuen, J. Races of Puccinia striiformis f. sp. tritici in the United States in 2011 and 2012 and Comparison with Races in 2010. *Plant Dis.* **100**, 966–975 (2016).
11. WSU Extension Resource. Annual Stripe Rust Race Data Reports | Stripe Rust | Washington State University. (2020). Available at: https://striperust.wsu.edu/races/data/. (Accessed: 25th October 2020)
12. Bartlett, D. W. *et al.* The strobilurin fungicides. *Pest Management Science* **58**, 649–662 (2002).
13. De Wolf, E. Kansas State University Agricultural Experiment Station and Cooperative Extension Service. (2010).
14. Chen, W. Q. *et al.* Race Dynamics, Diversity, and Virulence Evolution in Puccinia striiformis f. sp tritici, the Causal Agent of Wheat Stripe Rust in China from 2003 to 2007. *Plant Dis.* **93**, 1093–1101 (2009).
15. Wan, A. & Chen, X. Virulence characterization of Puccinia striiformis f . sp . tritici using a new set of Yr single-gene line differentials in the United States in 2010. *Phytopathology* **98**, 1534–1542 (2014).
16. Chen, X. Annual Stripe Rust Race Data Reports | Washington State University. Available at: http://striperust.wsu.edu/races/data/. (Accessed: 26th March 2018)
17. Radhakrishnan, G. V. *et al.* MARPLE, a point-of-care, strain-level disease diagnostics and surveillance tool for complex fungal pathogens. *BMC Biol.* **17**, 65 (2019).
18. Hubbard, A. *et al.* Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biol.* **16**, 23 (2015).
19. Win, J. *et al.* Effector biology of plant-associated organisms: Concepts and perspectives. *Cold Spring Harb. Symp. Quant. Biol.* **77**, 235–247 (2012).

20. Hogenhout, S. A., Van der Hoorn, R. A. L., Terauchi, R. & Kamoun, S. Emerging Concepts in Effector Biology of Plant-Associated Organisms. *Mol. Plant-Microbe Interact.* **22**, 115–122 (2009).
21. De Jonge, R., Bolton, M. D. & Thomma, B. P. How filamentous pathogens co-opt plants: The ins and outs of fungal effectors. *Current Opinion in Plant Biology* **14**, 400–406 (2011).
22. Ramachandran, S. R. *et al.* Effectors from Wheat Rust Fungi Suppress Multiple Plant Defense Responses. *Phytopathology* **107**, 75–83 (2017).
23. Liu, C. *et al.* The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases. *New Phytologist* (2016). doi:10.1111/nph.14034
24. Jin, Y., Szabo, L. J. & Carson, M. Century-old mystery of Puccinia striiformis life history solved with the identification of Berberis as an alternate host. *Phytopathology* **100**, 432–435 (2010).
25. Duan, X. *et al.* Puccinia striiformis f.sp. tritici presents high diversity and recombination in the over-summering zone of Gansu, China. *Mycologia* **102**, 44–53 (2010).
26. Wang, M. N. & Chen, X. M. Barberry does not function as an alternate host for Puccinia striiformis f. sp. tritici in the US Pacific Northwest due to teliospore degradation and barberry phenology. *Plant Dis.* **99**, 1500–1506 (2015).
27. Hovmøller, M. S., Sørensen, C. K., Walter, S. & Justesen, A. F. Diversity of Puccinia striiformis on Cereals and Grasses. *Annu. Rev. Phytopathol.* **49**, 197–217 (2011).
28. Li, F. *et al.* Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. doi:10.1101/692640
29. Sharma-Poudyal, D. *et al.* Virulence Characterization of International Collections of the Wheat Stripe Rust Pathogen, Puccinia striiformis f. sp. tritici. *Plant Dis.* **97**, 379–386 (2013).
30. Ali, S. *et al.* Low pathotype diversity in a recombinant Puccinia striiformis population through convergent selection at the eastern Himalayan centre of diversity (Nepal). *Plant Pathol.* **67**, 810–820 (2018).
31. Zhan, G. *et al.* Comparative virulence phenotypes and molecular genotypes of Puccinia striiformis f. sp. tritici, the wheat stripe rust pathogen in China and the United States. *Fungal Biol.* **116**, 643–653 (2012).
32. Hovmøller, M. S. *et al.* Replacement of the European wheat yellow rust population by new races from the centre of diversity in the near-Himalayan region. *Plant Pathol.* **65**, 402–411 (2016).
33. Bueno-Sancho, V. *et al.* Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects Seasonal Variation and Host Specificity. *Genome Biol. Evol.* **9**, 3282–3296 (2017).
34. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* **34**, i884–i890 (2018).
35. Cuomo, C. A. *et al.* Comparative Analysis Highlights Variable Genome Content of Wheat Rusts and Divergence of the Mating Loci. *G3 Genes, Genomes, Genet.* **7**, 361–376 (2017).
36. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
37. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

38. Periyannan, S. *Wheat Rust Diseases: Methods and Protocols*. **1659**, (Humana Press, 2017).
39. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* **6**, 80–92 (2012).
40. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
41. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
42. Jombart, T. & Ahmed, I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
43. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
44. Kamvar, Z. N., Tabima, J. F. & Gr̈unwald, N. J. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, 1–14 (2014).
45. Kamvar, Z. N., Brooks, J. C. & Grünwald, N. J. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* **6**, 208 (2015).
46. Price, M. N., Dehal, P. S. & Arkin, A. P. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
47. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, (2019).
48. Petersen, T. N., Brunak, S., Von Heijne, G. & Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786 (2011).
49. Lei, Y. *et al.* Virulence and molecular characterization of experimental isolates of the stripe rust pathogen (Puccinia striiformis) indicate somatic recombination. *Phytopathology* **107**, 329–344 (2017).
50. Wan, A. & Chen, X. Virulence, frequency, and distribution of races of puccinia striiformis f. sp. tritici and P. striiformis f. sp. hordei identified in the United States in 2008 and 2009. *Plant Dis.* **96**, 67–74 (2012).
51. Bruns, E., Carson, M. L. & May, G. The jack of all trades is master of none: A pathogen's ability to infect a greater number of host genotypes comes at a cost of delayed reproduction. *Evolution (N. Y).* **68**, 2453–2466 (2014).
52. Montarry, J., Hamelin, F. M., Glais, I., Corbi, R. & Andrivon, D. Fitness costs associated with unnecessary virulence factors and life history traits: Evolutionary insights from the potato late blight pathogen Phytophthora infestans. *BMC Evol. Biol.* **10**, (2010).
53. Pariaud, B. *et al.* Shared influence of pathogen and host genetics on a trade-off between latent period and spore production capacity in the wheat pathogen, Puccinia triticina. *Evol. Appl.* **6**, 303–312 (2013).
54. Sørensen, C. K., Justesen, A. F. & Hovmøller, M. S. Spontaneous loss of Yr2 avirulence in two lineages of Puccinia striiformis did not affect pathogen fitness. *Plant Pathol.* **62**, 19–27 (2013).
55. Ali, S. *et al.* Yellow rust epidemics worldwide were caused by pathogen races from divergent genetic lineages. *Front. Plant Sci.* **8**, 1057 (2017).

## Chapter 3: Evaluating the role of genetics on pathogen and host transcription and creating a dual-interaction network

### Introduction

Wheat (*Triticum aestivum*) is the second largest cereal crop produced in the world, with an estimated 766 million tons of wheat produced worldwide on 216 million hectares in 2019[1]. One major impediment for wheat growers throughout the world is stripe rust disease, which is caused by the pathogen *Puccinia striiformis* f. sp. *tritici* (*Pst*). *Pst* is a particularly detrimental fungal pathogen belonging to the genus *Puccinia spp*., which is considered one of the top ten fungal pathogens in the field of molecular plant pathology based on scientific and economic importance[2]. Late-stage symptoms of *Pst* infection include elongated lesions of yellow-orange spores on the host's leaves which tax photosynthesis and ultimately lead to a dramatic reduction in grain development. Stripe rust can be a serious threat to regional and global wheat production during epidemic years, particularly with the growing geographic footprint of the disease. Controlling wheat stripe rust globally costs upwards of $1 billion per annum and local epidemics can result in complete yield loss for individual growers if left unchecked[3,4]. Currently, the primary method for controlling *Pst* infection is breeding of *Pst*-resistant wheat varieties through the introgression of single resistance genes (R-genes) that convey complete resistance to the fungus[4]. If these single resistance genes fail to convey resistance to local genotypes of *Pst* during the growing season, broad-spectrum fungicides can be used to control acute infections[5,6]. However, use of these broad-spectrum fungicides can have several off-target problems, including yield loss due to application and the inability to use them in organic management practices. Beyond breeding for single gene resistance and broad-spectrum fungicides, tools do not yet exist for reducing yield loss during active *Pst* infection (infection after wheat resistance fails).

Strategies for combating *Pst* when available resistance genes fail are necessary for developing long-term, durable resistance in novel varieties of cultivated wheat.

As mentioned above, breeding of *Pst*-resistant varieties of wheat consists primarily of selective genetic introgression for single, major-effect R-genes into the wheat host that convey resistance to a small subset of pathogenic *Pst* races. These R-genes primarily consist of plasma membrane localized receptor-like kinases (RLKs) or cytoplasmic NBS-binding leucine-rich repeat proteins (NBS-LRR proteins)[7]. These single gene receptors recognize either race-specific pathogen-associated molecular patterns (PAMPs) or pathogen effectors, and are responsible for triggering downstream PAMP-triggered immunity (PTI) or a robust effector triggered immunity (ETI), respectively, within the wheat host[8]. In an evolutionary context, this single gene warfare between host and pathogen can be described as an arms battle and is commonly known as the zig-zag model plant pathogen evolution[8]. According to the zig-zag model, plants recognize individual pathogens by endogenous PAMPs produced by a pathogen, such as chitin or flagellin, using RLKs to initiate PTI responses in the host. Successful pathogens evolve and produce effectors that interfere with RLK signaling and suppress PTI response from the host[8]. Lastly, the plant host can evolve novel NBS-LRR proteins designed to detect the presence of pathogen effectors and initiate the stronger, more robust ETI responses in the plant host to combat the pathogen[8]. A pathogen can avoid ETI by shedding/diversifying effectors or acquiring novel effectors that suppress NBS-LRR detection[8]. Host disease susceptibility is a consequence of the pathogen evading detection from the host immune system during this evolutionary arms race between plants and pathogens. Wheat cultivars capable of resisting *Pst* infection through single R-genes create a strong selection pressure on the pathogen population, which can rapidly evolve into novel races capable of evading detection of host R-genes in the standing wheat field and

therefore drive massive epidemics[9]. One contemporary example of this evolutionary arms race in wheat is the evolution and introduction of a novel *Pst* race in the 2010s called the *Pst* Warrior race group, which overcame existing R-genes used in wheat breeding, out competed the old *Pst* population in Europe, and caused more disease and dramatic crop yield loss on adult plants carrying long-term resistance[10]. Therefore, it is important to understand the role of non-R-gene mediated resistance mechanisms, such as high-temperature adult-plant (HTAP) resistance, so that they may be breed into elite breeding lines of wheat to reduce crop loss during epidemics of novel pathogenic races of *Pst* [9,11]. HTAP resistance to *Pst* is a temperature-dependent type of genetic resistance in mature wheat plants that provides partial resistance against all *Pst* races and has been shown to be robust for over 60 years[11,12]. However, these resistance mechanisms are poorly understood from both a breeding and molecular perspective and often have some trade-offs. Exploring the interaction between *Pst* and wheat during active infection can reveal targets for the pathogen and insight into responses from the host.

Field patho-transcriptomics (sequencing the co-transcriptome of infected plant tissue from field samples) has previously been used to evaluate the population structure of *Pst*[13,14]. Additionally, *in vivo* transcriptomic analysis has also been useful to evaluate transcriptional responses in wheat infected with *Pst* (such as HTAP resistance or the resistance of specific wheat cultivars) as well as to evaluate *Pst* haustorial development and function with a primary focus on identification of *Pst* effectors[15–17]. While the data generated during these transcriptomic studies can enable *in planta*-measurement of pathogen transcriptome and subsequent potential molecular-plant pathogen interactions, these previous studies have focused solely on a single organism in the interaction and/or were conducted under controlled laboratory settings. To analyze both host and pathogen simultaneously, co-transcriptomic approaches can be used to

reveal potential cross-kingdom interaction and communication. Specifically, the gene co-expression network approach has emerged as a powerful systems biology approach whereby transcripts of highly co-expressed or connected genes can be identified and provide a meaningful way to examine the correlations in gene expression generated from complex RNA-seq datasets. Use of such approaches within the Arabidopsis-*B. cinerea* pathosystem was able to identify novel virulence factors in the pathogen, including a novel biochemical pathway for an uncharacterized cyclic peptide, as well as potential host targets for the resulting toxin[18].

Currently, there is only one other study using a co-transcriptomic approach within the wheat-*Pst* pathosystem, which showed the expression of defense mechanisms in both compatible and incompatible interactions[19]. The study showed temporally coordinated activation and suppression of gene expression such as pathways related to plant stress hormones, membrane proteins, and vesicle trafficking in the host while the pathogen showed increased gene expression during early infection related to germination, appressorium formation, and pathogenicity[19]. However, Dobin et. al. focused solely on early infection of *Pst* in wheat and used only a single isolate of *Pst* within a greenhouse setting at the wheat's three-leaf stage, which likely does not encapsulate the diversity of *Pst* infection strategies driving yield loss on mature plants in the field[19]. At writing, there is no existing study examining the wheat-*Pst* co-transcriptome under field conditions, which limits our ability to understand the interaction within wild and commercial settings.

Our study is the first to evaluate the co-transcriptomic interactions in *Pst*-infected wheat samples under field conditions from the six wheat-growing continents of the world. *Pst*-infected wheat leaves were collected in fields worldwide during the active stage of infection and the transcriptomes of both species were simultaneously sequenced in 246 samples. Our results show

the impact of *Pst* and wheat genetics on the transcriptome of each species. In the present analysis, six exon co-expression networks (ECNs) for the pathogen were identified and are involved with growth/development and virulence mechanisms. For the wheat host, eleven ECNs were identified that encode genes related to photosynthesis (including light and dark reactions, pigments, and REDOX), translation, sulfur metabolism, and isoprenoid biosynthesis. A dual-transcriptomic network was created by eigen-vector decomposition of individual networks and by combining both the *Pst* and wheat eigen-vector ECNs into a single network that illustrates both the connections within the host and pathogen individually, as well as cross-kingdom interactions. These connections give insight into how the host and pathogen interact during active infection and show conserved interactions from worldwide field samples.

## Methods

### *Transcriptomes and Filtering*

Sequenced and trimmed reads from the field patho-transcriptomics evaluation discussed in Chapter 2 were used to analyze *Pst*-wheat interaction[13,14,20]. For the patho-transcriptomic evaluation, transcriptomes were sequenced from 279 *Pst*-infected wheat, triticale, and durum wheat samples that were collected in fields around the world and combined with transcriptomic data from 17 additional samples (see Chapter 2 for details). Samples from plants other than common wheat were removed from the analysis to focus specifically on the interaction between *Pst* and wheat, leaving 280 *Pst*-infected wheat transcriptomes. Raw reads were trimmed using fastp with default parameters and were subsequently mapped to the indexed International Wheat Genome Sequencing Consortium (IWGSC) high-quality bread wheat reference genome (Chinese Spring wheat variety) Version 1.0 to obtain counts of wheat genes[21]. These reads were aligned to the reference using STAR Version 2.7.5a with a maximum of six mismatches and a maximum

intron length of 10,000[22]. The resulting bam files were sorted and indexed, and the pileup format was generated using SAMtools Version 1.10[23]. Counts of individual wheat exons per sample were obtained using HTSeq-count with default settings. Counts were filtered by removing exons with fewer than 200 counts as well as exons with counts in less than 50% of samples. Similarly, *Pst* counts data were filtered by removing exons with less than 100 counts over the entire experiment and/or removing exons present in less than 50% of samples. All counts were TMM normalized for each organism using the 'edgeR' package within R[24].

*Wheat Population Structure*

Wheat population structure was determined to group the wheat samples into genetic clusters that would then be used to evaluate the effect of wheat genetics on gene expression levels. To determine wheat population structure, a variant call format (vcf) file was created from the wheat bam files by using the samtools 'mpileup' function to get single nucleotide polymorphisms (SNPs)[23]. Synonymous SNPs were annotated using snpEff, and a custom python script was used to prepare and filter files in the following way; 1.) only positions with at least one sample present were included in the analysis, 2.) only biallelic SNPs were included in the analysis, and 3.) a minor allele frequency (MAF) of at least 20% were included in the analysis. Additionally, an occupancy filter was used where the SNP position needed to be confidently called in at least half of the samples to be included in subsequent analysis[25]. Samples with less than 20% of the relevant informative SNPs for our collection were removed, therefore reducing the total number of samples from 280 to 246 *Pst*-infected wheat transcriptomes.

*Statistical Analysis*

These high-quality SNPs were utilized for the multivariate discriminant analyses of principal components (DAPC) using the adegenet package in R environment[26–29]. The number of

wheat genetic clusters was determined by comparing the value of Bayesian Information Criterion (BIC) to the value of the K-means clustering algorithm using the lowest BIC at the 'elbow' of the curve. Individual wheat samples were assigned to clusters based on fuzzy k-means clustering, which indicated the likelihood that an individual was part of a genetic cluster. To establish the appropriate number of principal components for clustering, the data were cross-validated using poppr Version 2.8.6[30,31]. The final 246 samples were used for all subsequent analysis and are marked with an asterisk in Appendix 1.

To determine what metadata factors within our dataset have a significant effect on the transcriptome of *Pst*, a non-parametric multivariate statistical test called Permutational Multivariate Analysis of Variance (PERMANOVA) was run using the following model:

$$Y_{ctpw} = C_c + T_t + P_p + W_w + P{:}W_{pw} + \varepsilon_{ctpw}$$

where C is the country of collection, T is the year of collection, P is *Pst* population cluster based on the *Pst* genetic cluster defined in Chapter 2, and W is the host genetic cluster as determined by the wheat DAPC analysis[32]. The PERMANOVA analysis is based on TMM-normalized counts in a square matrix with distance testing performed using 999 permutations[33]. To visualize the relative relatedness of samples, a Nonmetric Multidimensional Scaling (NMDS) ordination was performed with Bray Curtis distances within the 'vegan' (v. 2.5-7) package[32].

*Primary Network Analysis: Individual Exon Co-Expression Networks for Pst and Wheat*

Individual exon co-expression networks (ECNs) were created for both *Pst* and wheat to determine subnetworks of exons that were co-expressed within each organism during host infection. Correlations were determined using Spearman's rank coefficients for each exon pair from the normalized *Pst* or wheat pseudo-counts using 'rcorr' in the Hmisc package[34]. The exon correlations in both *Pst* and wheat networks were filtered based on Spearman's R and

correlations R>0.92 were kept for network construction. Networks with ten or more exons were considered major ECNs. Gene ontology (GO) analysis for wheat was completed for the genes in each of the major ECNs using the IWGSC wheat annotation and the 'topGO' package in R to determine the general putative function of these co-expression ECNs[35]. Due to a lack of GO annotation for *Pst* that prevented calculation of functional enrichment, ECN putative functions were qualitatively determined using JGI MycoCosm GO terms (when available) in combination with BLAST annotation for individual genes using NCBI's database[36]. Additionally, SignalP was used to establish which genes had a putative secretion signal peptide, and to predict secreted gene products within each Pst_ECN[37]. Each major ECN was plotted using the 'igraph' package in Rstudio[38].

*Dual Interaction Network Analysis*

A dual interaction network was constructed between *Pst* and the wheat host to estimate direct and indirect interaction of intra-organismal co-expression networks. To create this network, 'vegan' package was used to condense the pseudo-counts of exon members from a single ECN into an eigengene vector (one-dimensional NMDS) for all major pathogen and host-plant ECNs[32]. The projected position of samples along these resulting eigengene vectors were used to calculate Spearman's rank correlation coefficients between major ECNs and were filtered by keeping correlations above an R of 0.39. The resulting dual interaction network was visualized using Cytoscape[39].

**Results**

*Population Structure of* Pst *and Wheat Populations*

To evaluate the interaction between *Pst* and wheat at the transcriptomic level, the transcriptomes of 246 *Pst*-infected wheat leaves from fields representing 23 countries across

9 different years (Appendix 1) were investigated. The trimmed reads of those samples were aligned to both PST-78, a *Pst* reference genome, and the IWGSC's wheat reference genome Version 1.0. Samples with less than 100,000 total wheat counts and then 20% or fewer SNPs were removed. The percentage of reads aligned to the *Pst* genome averaged between 43.79% ± 1.12 s.e. per sample while the percentage of reads aligned to the wheat genome average 27.54% ± 1.07 s.e. per sample. After filtering for depth and occupancy, 35,152 *Pst* exons (in 2,809 genes) representing 40.69% of the total potential exome remained and were used for subsequent analysis. For the wheat transcriptome, we identified a total of 96,388 wheat exons representing 12.65% of the total potential wheat exome after filtering for depth and sample occupancy. Data from these worldwide field samples of *Pst*-infected wheat leaves allowed us to study the effects of pathogen and host genetics on exon expression levels within each organism during active infection, as well as allowed us to explore how pathogen and host interact with each other.

To assess the role intra-specific genetic variation plays on driving gene expression within each organism as well as across organisms, we used a DAPC population structure analysis using non-synonymous SNPs derived from the transcriptomic mapping. The population analysis in Chapter 2 revealed 8 *Pst* genetic clusters, which were used to define our *Pst* groups for statistical analysis. Similarly, a DAPC analysis was used to evaluate the wheat genetic population structure of the samples. In this analysis, 5,109 total biallelic SNPs were found with an MAF > 20%. The DAPC results identified two genetic clusters for wheat. Wheat Genetic Cluster 1 included samples collected from 2014 to 2017 while Wheat Genetic Cluster 2 included samples collected from a longer time period (1978 to 2017). Many countries had samples in both wheat genetic clusters, specifically Canada, Chile, China, France, Poland, United Kingdom, and the United

States. However, all samples collected in Croatia, Italy, Pakistan, Romania, and Serbia, as well as Ethiopia and South Africa, were in Wheat Genetic Cluster 1. Wheat Genetic Cluster 2 included all samples collected in Austria, Belgium, Czech Republic, Germany, New Zealand, Slovakia, Sweden, and Turkey. These results showed that there is likely a significant amount of admixture in the wheat population of this study's samples, likely due to contemporary wheat breeding.

*Assessing Impact of Pathogen and Host Genetics on Exon Expression*

To assess the role of intra-specific genetics on transcriptional variation within the *Pst* and wheat exomes, PERMANOVA and two-dimensional NMDS were performed. The *Pst* PERMANOVA results indicated that country of collection had the largest effect ($R^2$ = 0.2759; p-value = 0.001***) followed by year of sample ($R^2$ = 0.0914; p-value = 0.001***) as presented in Table 3.1. Collectively, these spatiotemporal terms of country and year variables explained 36.73% of the total variance within the *Pst* exome and represented, in part, the environmental factors related to location and plant physiology. The genetics of the pathogen explained a significant amount of the variance within the *Pst* exome, with the *Pst* genetic cluster accounting for 4.20% (p-value = 0.001***) of variance. By comparison, the effect of wheat genetics on the *Pst* transcriptome during active infection was low ($R^2$ = 0.0043; p-value = 0.119), indicating wheat genetics had a much lower impact on the *Pst* transcriptome during active infection. Additionally, the interaction between *Pst* genetic clusters and wheat genetic clusters accounted for a non-significant 1.38% of total variance (p-value = 0.186) within the *Pst* transcriptome, suggesting that the relationship between *Pst* and wheat genetics on the *Pst* transcriptome is largely additive in nature.

*Table 3.1:* Pst *PERMANOVA Results*

| | Df | SumOfSqs | $R^2$ | F | Pr(>F) | Significant Code |
|---|---|---|---|---|---|---|
| **Country** | 23 | 5.4413 | 0.27590 | 4.2522 | 0.001 | *** |
| **Year** | 7 | 1.8017 | 0.09135 | 4.6261 | 0.001 | *** |
| ***Pst* Cluster** | 7 | 0.8283 | 0.04200 | 2.1269 | 0.001 | *** |
| **Wheat Cluster** | 1 | 0.0847 | 0.00430 | 1.5227 | 0.119 | |
| ***Pst* Cluster by Wheat Cluster** | 4 | 0.2719 | 0.01379 | 1.2218 | 0.186 | |
| **Residual** | 203 | 11.2942 | 0.57267 | | | |
| **Total** | 245 | 19.7221 | 1.00000 | | | |

*** p-value of essentially 0.

A two-dimensional NMDS of the *Pst* transcriptome (Figure 3.1) helped demonstrate the variation across the dataset (stress = 0.1638). The impact of country and time of collection can clearly be observed shaping the *Pst* transcriptome, but drastically obfuscated the weaker impacts of both *Pst* and wheat genetics within the ordination space. Countries such as China, the United Kingdom, and the United States were more widely dispersed, while countries that were clustered more tightly together in the NMDS results included Belgium and France. The years that were more dispersed in the NMDS results include 2014, 2015, and 2016 while the years 2013 and 2017 were more tightly clustered together, implying a canalization in the structure of the *Pst* transcriptome in recent years. The NMDS results showed that the *Pst* Genetic Cluster 5 was more tightly clustered together while the remaining genetic clusters were widely dispersed. The NMDS results for both wheat genetic clusters were even more diffused. Taken together, these results show that the *Pst* transcriptome is largely driven by environmental factors and *Pst* genetics rather than wheat genetics.

The wheat transcriptome showed similar environmental and genetic impacts on exon expression. Specifically, the PERMANOVA results showed country of collection explained the largest proportion of variation ($R^2$ = 0.3105; p-value = 0.001***) followed by year of collection

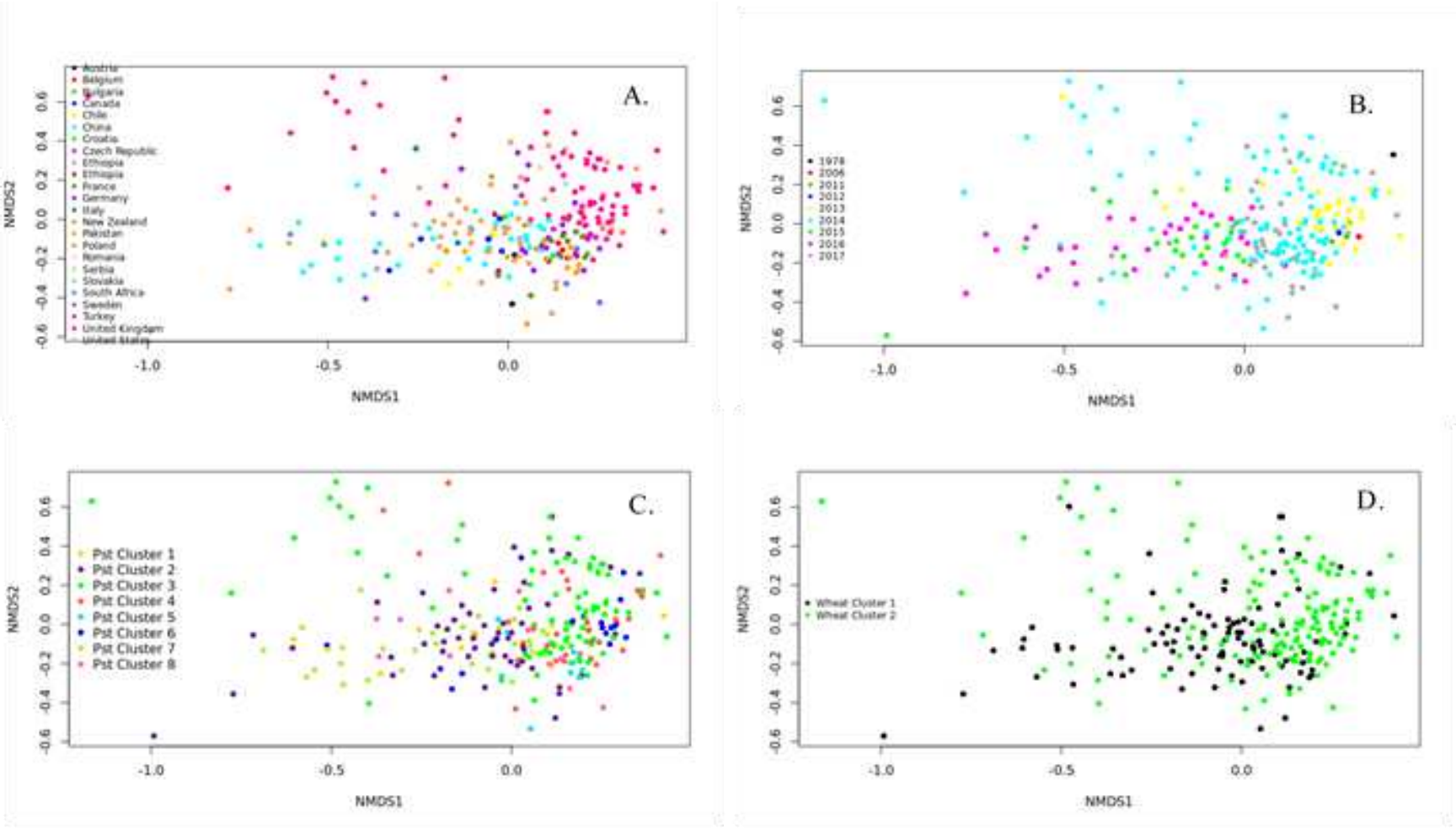


*Figure 3.1: Two-dimensional NMDS results for* Pst *colored according to A. country, B. year, C.* Pst *genetic cluster, and D. wheat genetic cluster.*

($R^2$ = 0.0588; p-value = 0.001***) representing a total of 36.9% of the variation in the wheat transcriptome (Table 3.2). Interestingly, while wheat genetic clusters had a significant impact on the wheat transcriptome ($R^2$ = 0.0056; p-value = 0.033*) during active infection, the *Pst* genetic clusters also had a significant impact, explaining nearly 6 times more variation than the wheat genetic clusters ($R^2$ = 0.0320; p-value = 0.001**). Lastly, the non-significant interaction between wheat and *Pst* genetic clusters (p-value = 0.156) further suggests an additive impact between wheat and *Pst* genetics while explaining a larger impact than wheat genetics alone ($R^2$ = 0.0137). These differences in the genetic effect between the two transcriptomes could be due to either the fewer number of genetic clusters found within the wheat germplasm, or due to less genetic diversity within domesticated wheat relative to the wild pathogen. An NMDS of the wheat transcriptome (Figure 3.2) also showed a good fit within two dimensions (stress = 0.1569). Similar to the *Pst* analysis, the impact of country and time of collection determined in the wheat analysis complicated the visualization showing the clustering of both the *Pst* and wheat genetics.

*Table 3.2: Wheat PERMANOVA Results*

| | **Df** | **SumOfSqs** | **$R^2$** | **F** | **Pr(>F)** | **Significant Code** |
|---|---|---|---|---|---|---|
| **Country** | 23 | 9.8027 | 0.31049 | 4.7301 | 0.001 | *** |
| **Year** | 7 | 1.8571 | 0.05882 | 2.9444 | 0.001 | *** |
| **Wheat Cluster** | 1 | 0.1763 | 0.00558 | 1.9569 | 0.033 | * |
| ***Pst* Cluster** | 7 | 1.0131 | 0.03209 | 1.6062 | 0.001 | ** |
| **Wheat Cluster by *Pst* Cluster** | 4 | 0.4311 | 0.01366 | 1.1962 | 0.156 | |
| **Residual** | 203 | 18.2915 | 0.57936 | | | |
| **Total** | 245 | 31.5719 | 1.00000 | | | |

*** p-value of essentially 0.
** p-value of approximately 0.001.
* p-value of approximately 0.01.

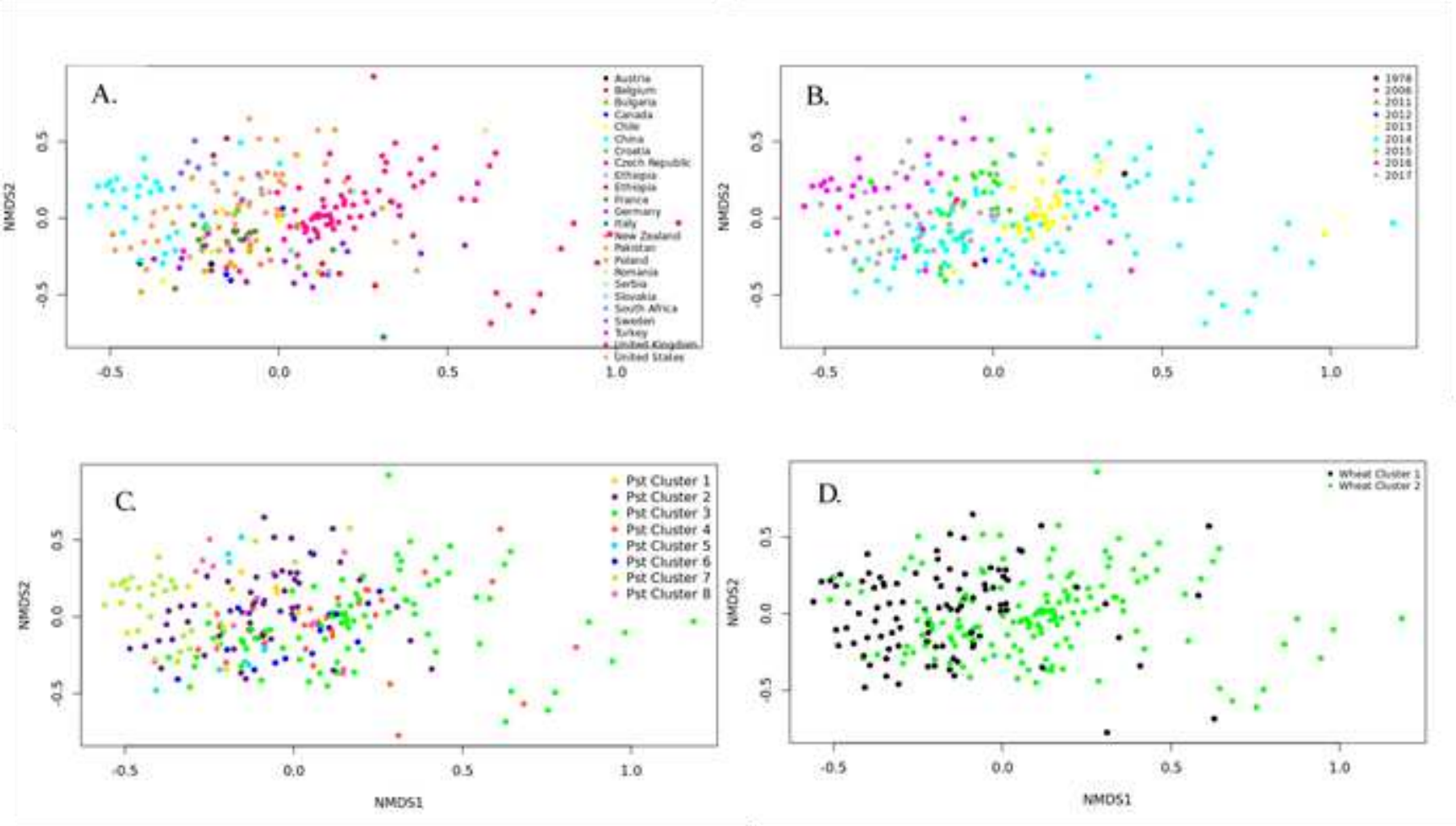


*Figure 3.2: Two-dimensional NMDS for wheat colored by A. country, B. year, C.* Pst *genetic cluster, and D. wheat genetic cluster.*

Within this ordination space, however, the NMDS results showed that samples from United Kingdom and the United States were the most widely dispersed. Samples from countries such as Canada, China, Ethiopia, France, and Poland were more tightly clustered together. Samples collected in 2014, 2016, and 2017 were more widely dispersed according to the NMDS results, while samples such as those collected in 2013 and 2015 were clustered closer together. Both wheat genetic clusters in the NMDS results demonstrated about the same amount of dispersion while the impacts of the *Pst* clusters varied. The wheat transcriptomes infected with *Pst* Genetic Clusters 1, 2, 5, 7, and 8 formed tight, independent clusters within the two-dimensional NMDS, while *Pst* Genetic Clusters 3 and 4 were more loosely aggregated. Taken together, this analysis suggests that spatiotemporal effects have the largest effect on the infected wheat transcriptome while intra-specific genetic variation within *Pst* has a larger effect on the wheat transcriptome during active infections than wheat genetics.

*Individual Exon Co-Expression Networks*

To identify functional gene networks among the mapped exons within *Pst,* exon-level co-expression networks were created. There were 53.36% of significant interactions among all exons identified (FDR P-val < 0.05), with an average Spearman correlations (*R*) of 0.02072 ± 0.21279. Spearman's rank coefficients for each *Pst* exon pair were calculated and filtered for positive correlations of *R* above 0.92 to optimize the number of large networks (membership > 10 nodes). These filtered gene pairs were used to construct ECNs. The *Pst* exon co-expression network revealed six major ECNs (networks with > 10 nodes) and spanned a variety of cellular functions, including translation/protein stability, maintenance of REDOX potential, and various virulence factors (Figure 3.3). Pst_ECN_I is the largest ECN (203 exons, 99 unique genes)

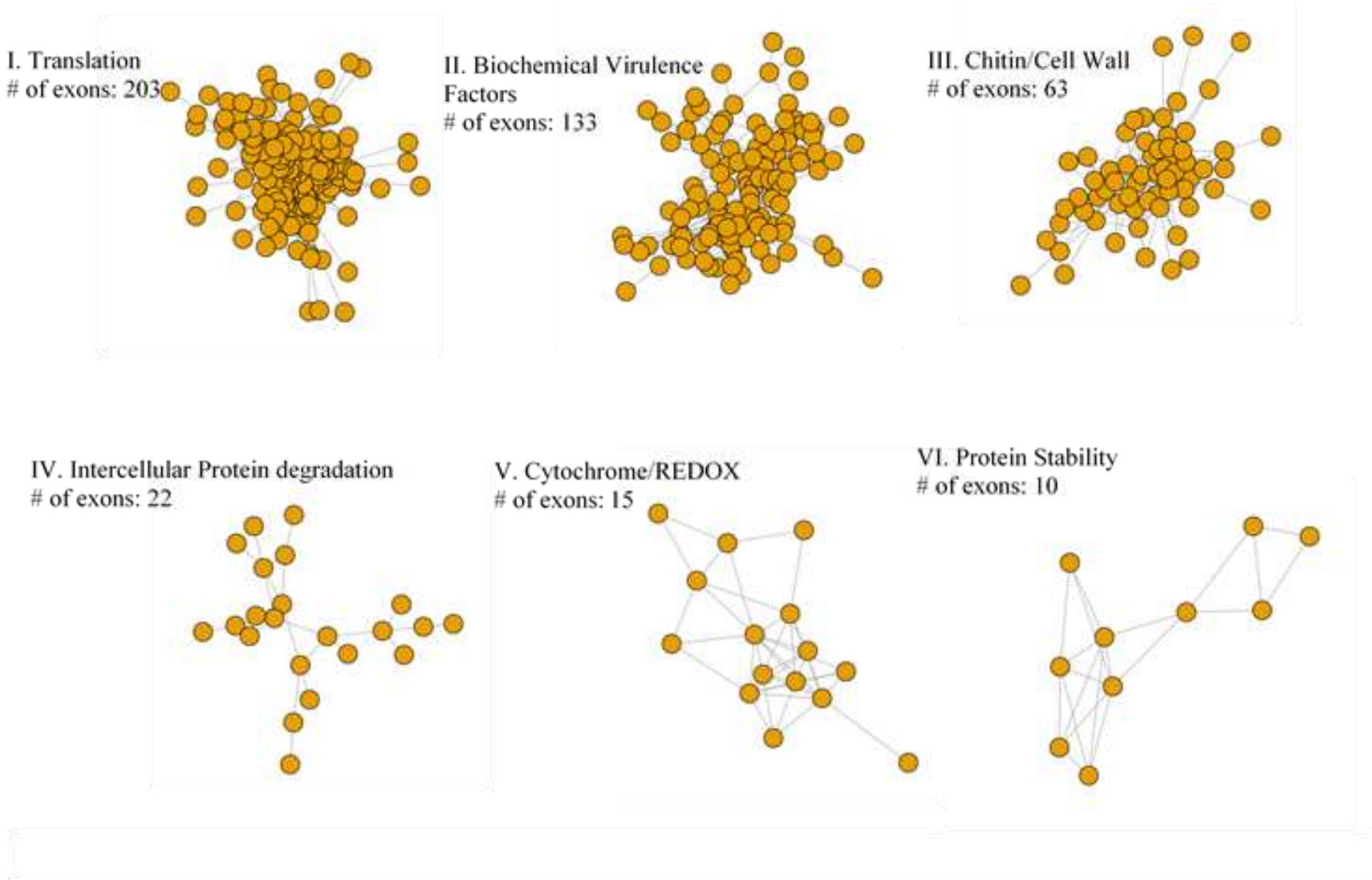


*Figure 3.3: Exon co-expression networks for* Pst. *The ECN number, main biological function, and number of exons are shown beside each ECN.*

consisting largely of 60s ribosomal proteins, 40s ribosomal proteins, and nucleotide binding proteins, which implicates this network's role in translation (Appendix 11). The second largest ECN is Pst_ECN_II (133 exons, 79 unique genes) and consists of secreted and non-secreted proteins and mainly hypothetical proteins with little to no GO annotation. Of the annotated genes within the network, several are homologous with enzymes such as protein kinase, tripeptidyl peptidase, chitin deacetylase, and lysophospholipase, implicating this ECN with biochemical virulence factors. Pst_ECN_III (63 exons, 60 unique genes) is associated with both secreted and non-secreted genes involved with membrane transport, chitin biosynthesis, and pectin degrading enzymes and subsequently is annotated as involved with chitin/cell wall formation. Pst_ECN_IV (22 exons, 20 unique genes) is associated with intercellular protein degradation and has genes with GO terms involved with transporter activity, protease or protein dimerization activity, and both secreted and non-secreted genes. The ECN associated with the function of cytochrome/REDOX is Pst_ECN_V (15 exons, 13 unique genes) and has non-secreted genes involved with NADH, ATP synthase, and cytochrome c oxidase. Finally, Pst_ECN_VI (10 exons, 2 unique genes) has non-secreted genes involved with several heat shock proteins and is putatively involved with protein stability. Overall, these six major ECNs appear to be primarily associated with gene translation, growth/development, and several general virulence factors.

Co-expression networks were also generated for the wheat exons to identify gene networks commonly up-regulated or down-regulated within the host during active infection. As with *Pst,* Spearman's rank coefficients for each wheat exon pair were calculated and positive correlations above 0.92 were kept to construct the ECNs. There were 11 major wheat ECNs

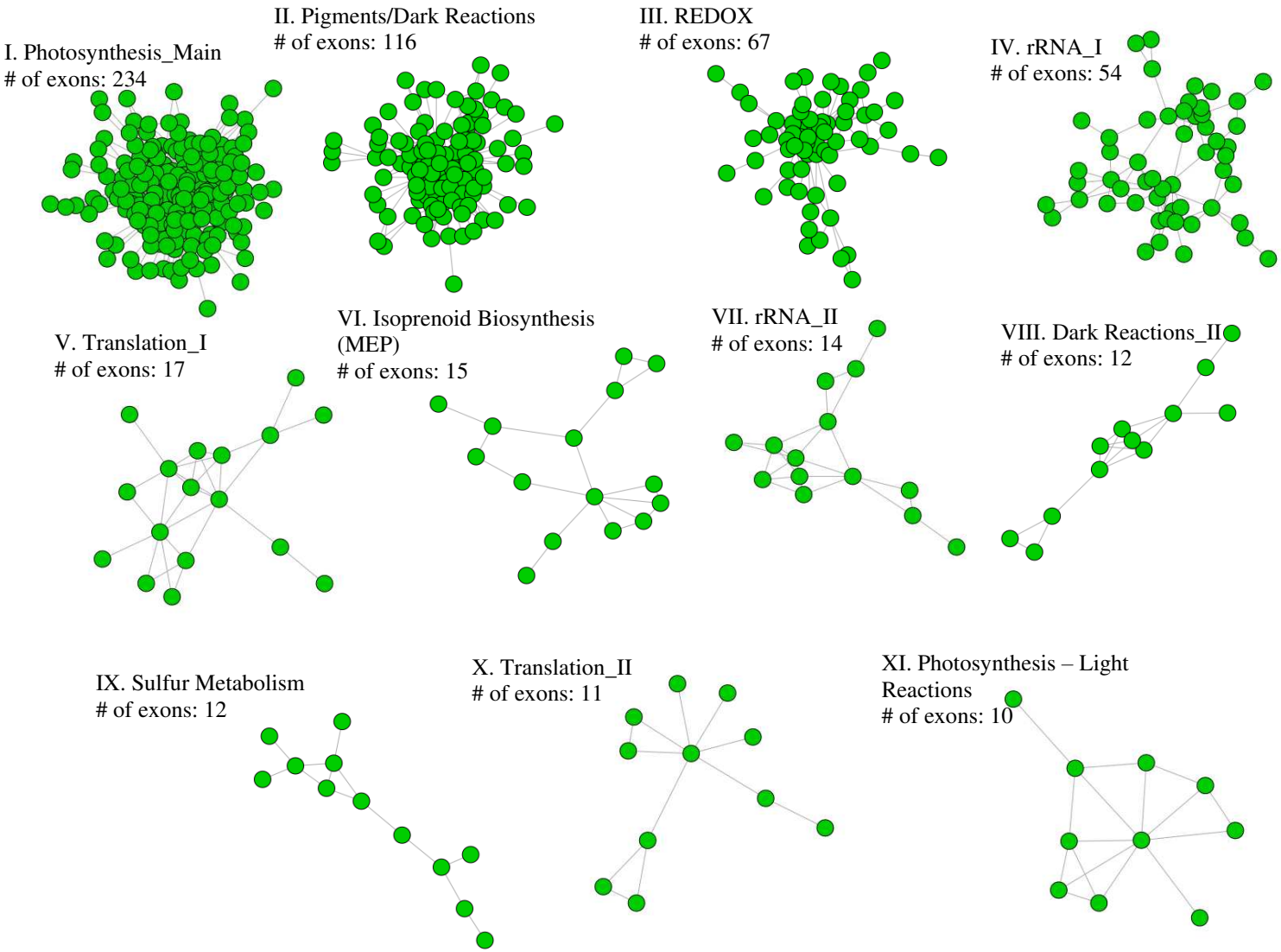


*Figure 3.4: Exon co-expression networks for wheat. The ECN number, main biological function, and number of exons are shown beside each ECN.*

consisting of 10 nodes or greater (Figure 3.4, Appendix 12). Five ECNs are involved with different sub-processes of light and dark reactions of photosynthesis (specifically, Ta_ECN_I, Ta_ECN_II, Ta_ECN_III, Ta_ECN_VIII, and Ta_ECN_XI). Among these five ECNs, Ta_ECN_I is by far the largest of this group (234 exons, 194 unique genes) and contains genes involved with both the light and dark reactions of photosynthesis, as well as accessory pigments (Appendix 2 through Appendix 10). The second largest wheat ECN (Ta_ECN_II, 116 exons, 106 unique genes) is primarily involved with accessory pigments and dark reactions, and is enriched for GO:0015979 (photosynthesis), GO:000534 (chloroplast thylakoid), and GO:0031409 (pigment binding). Ta_ECN_III (67 exons, 55 unique genes) is involved with light reactions and REDOX. Similar to Ta_ECN_II, Ta_ECN_VIII (12 exons, 12 unique genes) is involved with the dark reactions of photosynthesis and enriched for GO:0019685 (photosynthesis, dark reaction) and GO:0016829 (lyase activity). The smallest of the wheat ECNs involved in processes of photosynthesis (Ta_ECN_XI, 10 exons, 10 unique genes) contains a collection of light reactions and is enriched for GO:0015979 (photosynthesis), GO:0009535 (chloroplast thylakoid membrane), and GO:0009055 (electron transfer activity). Alternatively to photosynthesis, the remaining wheat ECNs encompass a variety of other cellular processes. Ta_ECN_IV (54 exons, 54 unique genes or non-coding ribosomal RNA) appears to be involved with protein translation and ribosome function. Two additional wheat ECNs involved with translation were Ta_ECN_X (11 exons, 11 unique genes) and Ta_ECN_VII (14 exons, 14 unique genes, 12 of which are non-coding and the rest are involved with rRNA). The next wheat ECN (Ta_ECN_V, 17 exons, 17 unique genes) is more generally involved with translation beyond rRNA. Several components of the chloroplast-specific isoprenoid pathway (MEP pathway) are contained within Ta_ECN_VI (15 exons, 15 unique genes). Ta_ECN_IX

(12 exons, 12 unique genes) is connected with SAMe generation and is involved with regulation of sulfur metabolism within the plant. Overall, the wheat ECNs are associated with different parts of photosynthesis, protein production, isoprenoid production, and sulfur metabolism.

*Interactions among Intra- and Inter-Species Exon Co-Expression Networks*

While simple pairwise correlation networks at the exon or even gene level can provide a glimpse into cellular processes involved with disease resistance, these simple pairwise correlation networks would be difficult to interpret at lower edge thresholds or if we included negative correlations as well. To assess the interactions among these *Pst* and wheat co-expression ECNs during infection, a dual interaction network was created where we explored the potential for direct/indirect interactions in both positive and negative relationships (Figure 3.5). To accomplish this, we decomposed the gene expression within these correlated ECNs into a single eigen-vector per network that described the total variation across the network and plotted the sample position along that vector. Sample positions along all eigen-vectors were correlated to look for positive or negative relationships both within each organism and across organisms. This one-dimensional decomposition for each ECN worked extremely well and each eigen-vector described the variation among genes within the network extremely well (stress = 0.09874 ± 0.01398), illustrating that these eigen-vectors captured the vast majority of variation contained within these networks. Of the 77.9% significant interactions (FDR P-val < 0.05), the Spearman correlations had an average R of 0.42679 ± 0.01961. A composite dual interaction network describing these correlations was generated to assess the relationships among the primary ECNs, both within and across ECNs. Spearman correlations with $R^2$ over 0.15 were kept. This analysis illustrated that *Pst* largely demonstrated a coordinated response between its biochemical

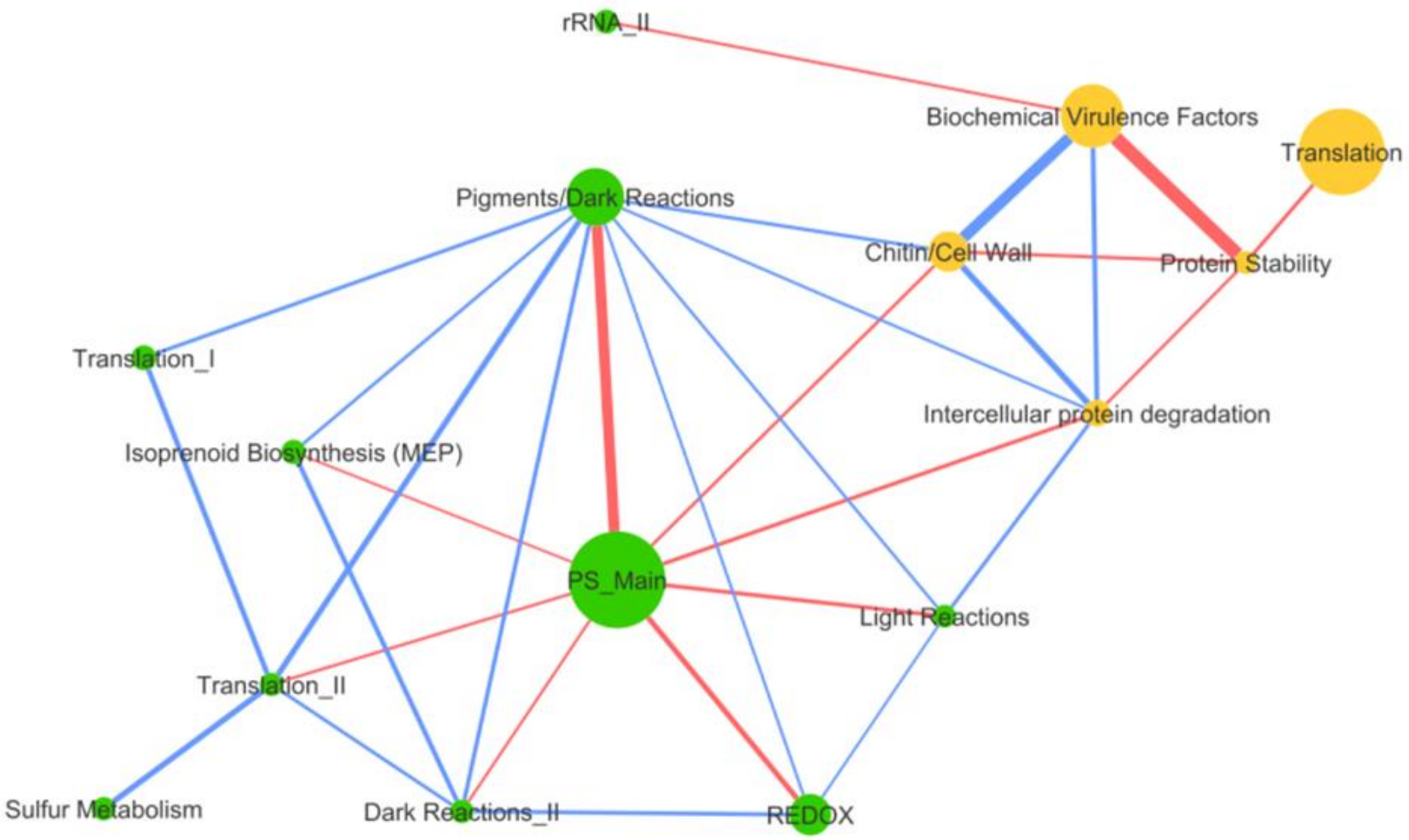


*Figure 3.5: Cross-kingdom co-expression network. Green circles represent major wheat ECNs while orange circles represent major* Pst *ECNs. The size of the circles show the number of exons in each ECN. Blue and red lines represent the positive and negative Spearman's rank correlation coefficients, respectively, while the thickness of the line represents the strength of the correlation.*

virulence factors, chitin formation/plant cell wall degradation, and inter-cellular protein degradation ECNs. Additionally, each of these ECNs were negatively correlated with protein stability ECN, suggesting an inverse relationship between growth, virulence, and fungal response to stress. Lastly, the largest *Pst* ECN (Pst_ECN_I Translation) only showed one negative correlation and that was with general protein stability within *Pst*, which suggested that there was a transitively positive relationship between protein translation and the virulence ECNs. The coordinated response between many of the major *Pst* ECNs suggested coordination between both cellular growth and virulence during active infection in the host.

The relationship among the wheat ECNs within the larger dual interaction network was equally interesting and may illustrate the abundance of photosynthetic ECNs. The primarily photosynthetic network (Ta_ECN_I) was negatively correlated with all other photosynthetic networks as well as isoprenoid biosynthesis, showing that these cellular processes were out of phase from each other. The second largest network (Ta_ECN_II), which contained genes involved with accessory pigments and dark reactions, was positively correlated with all other wheat ECNs except sulfur metabolism. All other ECNs were negatively correlated with the primary photosynthetic network (Ta_ECN_I), and positively correlated with the pigments/dark reactions ECN (Ta_ECN_II). The only exception was the sulfur metabolism ECN (which was connected with Translation_II), implying its role in supporting cysteine and methionine production for protein synthesis. This dual interaction network illustrated that the largest primarily photosynthetic network was out of phase with all other photosynthetic networks, implying that the main network was either down-regulated or susceptible while the other ECNs likely had a compensatory effect during active infection.

Across the organisms, there were several links in the dual interaction network that showed both positive and negative correlations with each other. The chitin/cell wall and intercellular protein degradation ECNs were both negatively correlated with the primary photosynthetic network in the host, implying that the host primary photosynthetic network was either a direct or indirect target of the *Pst*-derived virulence factors. Additionally, the compensatory accessory pigments/dark reactions and light reaction ECNs were positively associating with the virulence factors, indicating they were upregulated more severely when active infection was more severe. Lastly, the small rRNA_II network in the wheat host appeared to be negatively correlated with the biochemical/virulence factors in *Pst*, again implying that these virulence factors were directly or indirectly targeting host protein translation. These results showed that the host photosynthetic apparatus was under severe strain during active infection and compensatory photosynthetic genes were unable to take up the slack.

**Discussion**

While well-characterized R-genes in the host are the primary mode of plant resistance to obligate biotrophic pathogens, long-term durable resistance to these pathogens must include strong basal resistance when available R-genes fail to provide resistance to novel pathogen races. This study aims to identify the targets of *Pst* virulence factors and potential modes of host resistance during active infection. By studying the quantitative resistance of the dual host-pathogen transcriptome during late active infection, we can identify potential breeding targets for more durable basal resistance and move away from the gene-for-gene arms race where *Pst* rapidly evolves to overcome plant defenses.

Our global dataset pulled from a wide variety of environments and growing conditions, wheat breeding stock, and endemic *Pst* races, making it one of the largest and diverse public

datasets of its kind to explore wheat-*Pst* interactions. Importantly, we were able to account for large portions of environmental and temporal variation by controlling for country and year of collection, therefore revealing the effect of genetic variance on the transcriptomes of both the host and pathogen. The worldwide population structure of *Pst* revealed 8 distinctive genetic clusters, the evolution of *Pst* within the endemic population of North America, and the loss of *Pst* effectors in modern isolates as a way of overcoming wheat defenses. When evaluating the population structure of the wheat samples, only two genetic clusters were identified. This result recapitulates previous work on the genetic diversity within wheat-breeding populations, which showed that the vast majority of modern wheat cultivars can be grouped into 3 or 4 genetic clusters when including land races and modern wheat varieties[40]. Importantly, our study focused solely on common wheat, also known as bread wheat (*Triticum aestivum*), and did not include other commonly cultivated wheat species, such as the hexaploid spelt wheat (*Triticum spelta*) or tetraploid durum wheat (*Triticum durum*). The wheat population structure in our study agreed with the relatively narrow population structure of wheat, which could be a function of the low number of identified SNPs (5,109 biallelic SNPs) as well as the genetic bottleneck of wheat due to domestication. Thus, our dataset was likely representative of common wheat as well as the wild *Pst* germplasm throughout the globe.

Similar to most quantitative traits, variation in the *Pst* and wheat transcriptomes during active infection can be attributed to several variables[18,41]. Environmental factors had the largest effect on transcription within each organism. Inclusion of these variables (country and year of collection) within the PERMANOVA models facilitated the control of a significant portion of variation due to environmental factors. The intra-specific genetic variation of each organism had a significant effect on the structure of its own transcriptome during active infection, with wheat

having a far more canalized transcriptional response during active infection than *Pst*. Interestingly, a significant impact of *Pst* genetic variation on the wheat transcriptome could be seen when testing for cross-kingdom effects on the wheat transcriptome. Further, the *Pst* genetic clusters have a larger effect on the structure of the wheat transcriptome than wheat genetics. While the impact of wheat genetics on the *Pst* transcriptome did not share this reciprocal effect, these results suggest that the extensive genetic variation within *Pst* was a major driving force in the epidemiology of striped rust infections globally. Further, these results imply that the major selective pressure of domestication and breeding on wheat could be reducing the potential for the host to defend themselves from *Pst* during to active infection. In contrast, the *Pst* pangenome was under massive fluctuating selection. Grass hosts harbor more diverse isolates of *P. striiformis* (including isolates that infect both grasses and wheat) than the number of isolates found on wheat[42]. This pathogen also has the alternate host of barberry and needs to infect that host in order to sexually reproduce, which could further increase the pathogen's virulence during active infection[43]. Thus, it may be favorable to increase the level of genetic diversity within wheat breeding stocks to improve basal resistance and combat novel races of *Pst*.

While much of the current research on the virulence of *Pst* is focused on avirulence factors promoting infection, little is known about the virulence factors during active infection that negatively impact wheat seed set and yield. During active infection, *Pst* is largely undetected by the host as the hyphae grow and develop within plant tissues. In this co-expression analysis, the largest ECN for *Pst* (Pst_ECN_I; 203 exons) showed that the major ECNs are primarily associated with protein translation, which is a core cellular function for growth/development. Due to protein turnover, proteins must be made quickly during the infection process to combat host defenses, support mycelial growth, and funnel resources to developing spores. With both the

host and pathogen translating mRNA into proteins for growth and enhancing fitness, the efficiency of translation is important for determining the outcome between pathogen infection or host resistance. Translational efficiency can depend on a number of factors, including the availability of ribosomes[44]. The results of this study showed that translation in *Pst* is negatively correlated with Pst_ECN_VI (involved with protein stability), implying that growing fungus must rely on the generation of new virulence proteins during infection rather than protecting and supporting previously translated structural protein and enzymes. Further, this large network has transitively positive correlation with known/suspected virulence factors in growth and virulence (Pst_ECN_II, Pst_ECN_III, and Pst_ECN_IV) within the composite network (Figure 3.5). One of these transitively correlated ECNs is Pst_ECN_III, whose genes are related to chitin and cell walls (Figure 3.3). During active infection, *Pst* requires an increasing number of proteins related to chitin formation and plant cell wall degradation to grow and invade the plant host. Additionally, *Pst* uses a number of virulence factors during active infection to aid with persistence within the host. During successful establishment within the host, the pathogen releases effector molecules into the apoplast of the host that suppresses plant defense, leading to an active infection[8]. These effector molecules may be diverse, given that *Pst* is under massive fluctuating selection related to its wide host range including wheat, wild grasses, and the alternate barberry hosts[42,43]. Coupled with *Pst*'s high mutation rate and ability to sexually reproduce, this implies that *Pst* has the potential to evolve rapidly and overcome current and future resistant wheat varieties[45,46]. Lastly, the second largest ECN for *Pst* (Pst_ECN_II) has genes generally associated with secreted virulence proteins (Figure 3.3). This ECN is positively correlated with ECNs associated with fungal growth (Pst_ECN_I and Pst_ECN_III), which reflects the synergistic importance of virulence factors in this biotrophic pathogen's growth and

development. Pst_ECN_II is also positively correlated with the intercellular protein degradation ECN (Pst_ECN_IV) associated with enzymes like lysophospholipase activity and chitin deacetylase activity, which implicates their combined role in enhanced virulence (Figure 3.5). Together, genes associated with core cellular functions and genes associated with the infection process make up the highly correlated ECNs utilized during active infection.

Alternatively, this study identified a number of gene networks in the host that are co-regulated during active infection. Specifically, five out of eleven gene networks are associated with various aspects of photosynthesis, including light and dark reactions, REDOX potential, and pigments. The largest wheat ECN (Ta_ECN_I; 234 exons) is associated with multiple components of photosynthesis and is negatively correlated with the remaining four wheat photosynthetic ECNs. These associations suggest that, while the primary components of photosynthesis are down regulated during active infection, other components of photosynthesis attempt to compensate. Several studies have shown that photosynthesis varies locally within the leaf and lesion while the photosynthetic rate progressively declines through the infection process for many biotrophic pathogens, including *Pst*. This decline may be due to increased demand on the photosynthetic apparatus or a direct attack on the photosynthesis from the pathogen[47,48]. Early in fungal infections with rusts, the photosynthetic rate (measured by the uptake of $CO_2$ using Li-COR and chlorophyll fluorescence) is either stable or increases until further along in the infection when the photosynthetic rate decreases[47,49]. Sucrose availability for *Pst* increases during this early infection time period, indicating that *Pst* growth and development is a new and aggressive sink for photosynthate[49]. Later in the infection process during the sporulation stage, overall photosynthetic rate decreases but is most severe in actively infected wheat cells. This observation supports the idea that at least some portion of the negative impacts on the

photosynthetic apparatus are due to an increased demand on photosynthate from the fungus. Additionally, cellular regions that remain uninfected but are surrounded by actively infected cells (known as green islands) often demonstrate higher local photosynthetic rates, suggesting that they continue to act as a carbon source for the actively growing fungal mycelia[50,51]. Further, Berger et. al. used chlorophyll fluorescence to show that photosynthesis was inhibited within infection sites, while surrounding areas showed increased photosynthesis in tomato infection with *Botrytis cinerea* [52]. That study revealed an inverse relationship between a primary photosynthesis network and other individual components of the photosynthetic apparatus, reflecting the decrease in photosynthetic rate and nearby photosynthetic compensation typical in pathogen infections.

In addition to photosynthesis, several exon networks in wheat represent molecular systems that have potential dual roles in both photosynthesis and plant defense against pathogens. Two of the wheat ECNs negatively correlated with the main photosynthetic ECN are associated with the dark reactions of photosynthesis (Ta_ECN_II and Ta_ECN_VIII) and/or accessory pigments (Ta_ECN_II). While these biochemical pathways help fix carbon in the CBB cycle or produce compounds (e.g. carotenoids, etc.) that help protect against excess light for photosynthesis, they also can help deal with excess reactive oxygen species (ROS) production or produce defense compounds during infection[53]. Additionally, multiple defense-related pathways go through chloroplasts, including the production of ROS and biosynthesis of salicylic acid (SA), which can alter the rate of photorespiration and impact ROS accumulation in the peroxisome, chloroplast, and mitochondrion[54]. An ECN with genes directly involved in redox reactions (Ta_ECN_III) is positively correlated with the dark reactions subsystem (Ta_ECN_VIII), which is unsurprising given the role of peroxisomes scavenging

normal ROS generated during photosynthesis. ROS in the form of hydrogen peroxide is a common bi-product of the oxygenation reaction of Rubisco, and plants see a large spike in a variety of ROS during active infection[55,56]. This process begins with excess excitation energy that can occur when photosynthetic metabolism is limited. ROS accumulation can be damaging to the plant and must be prevented via several quenching processes[57]. While the overall decline of a host's photosynthetic rate is reflected in lower gene expression in many photosynthesis-related genes, other genes associated with defense and compensating for lower photosynthetic rates are upregulated. Wheat compensates for *Pst* disruption of the normal physiological processes by regulating genes involved with defense and damage mitigation.

While the assessment of individual gene networks within an organism is important for predicting the molecular process impacted during active infection, a cross-kingdom analysis of the networks can reveal the direct and indirect interaction of molecular components between the organisms. The cross-kingdom dual interaction network indicated that the relationship between *Pst* infection and photosynthesis in wheat is more complex than a simple negative correlation between fungal growth and wheat photosynthesis. For example, the main wheat photosynthetic Ta_ECN_I (Figure 3.4) was negatively correlated with the pathogen network Pst_ECN_III associated with chitin formation (Figure 3.5), suggesting that photosynthesis in wheat is inversely related to fungal growth. It has been previous shown that downregulation of photosynthesis gene expression occurs at a slower pace than the reduction of the rate of photosynthesis, which suggests that the downregulation of the genes in the main photosynthetic ECN in this study lags the reduction of the photosynthetic rate[48]. Our results suggest that later stages of active infection result in a decrease in the expression of photosynthetic-related genes in opposition to increased fungal growth. This decrease could be due to damage to the thylakoid

membrane ultrastructure during *Pst* infection in susceptible wheat plants, resulting in lower PSII activity[58]. Another possibility is the host plant could reduce photosynthesis in response to fungal infection in order to favor respiration and ROS production as part of the plant's defense, as shown in poplar plants infected by the fungal hemibiotroph *Marssonina brunnea*[50]. Regardless, host genes associated with photosynthesis are differently correlated with fungal growth. For example, Ta_ECN_II is directly and positively correlated with Pst_ECN_III and expresses genes involved with wheat pigments and dark reactions. Indirectly, all photosynthetic ECNs except the largest photosynthetic ECN are positively correlated with fungal growth. This correlation suggests that the leaf transitions from normal developmental and physiologically defined source-sink relationships to a source-sink host-to-pathogen relationship as photosynthesis decreases while pathogen demand for photosynthetic products increases[48]. Since *Pst* is a biotrophic pathogen that requires a living host in order to grow and reproduce, it uses sucrose from the plant and transforms the infected leaf into a sink[49]. The wheat plant both decreases the expression of photosynthetic genes while also increasing expression in others to compensate for the lower photosynthetic rate at this stage of infection. In other words, during *Pst* infection, the wheat leaf transitions from a source of sugar to a sink to compensate for the pathogen's demand for photosynthetic products.

Lastly, our network analysis identified an interesting role for sulfur metabolism in the interaction between wheat and *Pst*. One of the smaller wheat ECNs identified (Ta_ECN_IX) was associated with the role of sulfur metabolism in the interaction. While Ta_ECN_IX is not directly correlated with any *Pst* ECNs, it is transitively correlated with *Pst* ECNs involved with growth and infection in a positive manner. This correlation is interesting as the genomes of other obligate biotrophic pathogens, such as *Melampsora larici-populina* (causal agent of poplar leaf

rust) and *Puccinia graminis* f. sp. *tritici* (causal agent of wheat stem rust) lack the key genes required for sulfate assimilation pathways, thereby suggesting that these rust pathogens obtain needed sulfur from the host[59]. The wheat plant could increase the transcription of genes related to sulfur metabolism as the pathogen grows and acts as a sink for reduced sulfur in a similar fashion to photosynthate. This phenomenon suggests that targeted breeding for low sulfur tolerate wheat plants may help improve the basal resistance of wheat germplasm to *Pst* infection.

**Conclusions**

This study sequenced transcriptomes of *Pst*-infected wheat leaves from field samples representing worldwide, modern common wheat and wild *Pst* isolates. Environmental factors have the largest effect on transcriptional variation for both host and pathogen; however, there are also significant effects from both wheat and *Pst* genetic variation. *Pst* genetic variation is a major driving force in the global stripe rust epidemiology, while the selective pressure from wheat domestication and breeding narrow the wheat transcriptional response. Through network analysis, we revealed major interconnected *Pst* networks involved with growth, development, and virulence of the fungus. During *Pst* infection, the wheat leaf transitions from a source of sugars to a sink due to the needs of the pathogen. Despite a lower photosynthetic rate, reflected in an inverse relationship between *Pst* ECNs related to growth and the primary wheat photosynthetic network, other photosynthesis-related networks compensate and are positively correlated with *Pst* growth. This hypothesis-generating study has revealed both the importance of wheat diversity and highlighted some potential breeding targets for more durable basal resistance to move away from R-gene-based resistance.

## References


1. FAOSTAT. Available at: http://www.fao.org/faostat/en/#data/QC. (Accessed: 19th May 2021)
2. Dean, R. *et al.* The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* **13**, 414–430 (2012).
3. Chen, X. Pathogens which threaten food security: Puccinia striiformis, the wheat stripe rust pathogen. *Food Security* **12**, 239–251 (2020).
4. Chen, X. M. Epidemiology and control of stripe rust [Puccinia striiformis f. sp. tritici] on wheat. *Can. J. Plant Pathol.* **27**, 314–337 (2005).
5. Bartlett, D. W. *et al.* The strobilurin fungicides. *Pest Management Science* **58**, 649–662 (2002).
6. De Wolf, E. Kansas State University Agricultural Experiment Station and Cooperative Extension Service. (2010).
7. Tao, F. *et al.* Revealing Differentially Expressed Genes and Identifying Effector Proteins of Puccinia striiformis f. sp. tritici in Response to High-Temperature Seedling Plant Resistance of Wheat Based on Transcriptome Sequencing. *mSphere* **5**, (2020).
8. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
9. Wellings, C. R. Global status of stripe rust: A review of historical and current threats. *Euphytica* **179**, 129–141 (2011).
10. Hovmøller, M. S. *et al.* Replacement of the European wheat yellow rust population by new races from the centre of diversity in the near-Himalayan region. *Plant Pathol.* **65**, 402–411 (2016).
11. Chen, X. Review Article: High-Temperature Adult-Plant Resistance, Key for Sustainable Control of Stripe Rust. *Am. J. Plant Sci.* **04**, 608–627 (2013).
12. Chen, X., Coram, T., Huang, X., Wang, M. & Dolezal, A. Understanding Molecular Mechanisms of Durable and Non-durable Resistance to Stripe Rust in Wheat Using a Transcriptomics Approach. *Curr. Genomics* **14**, 111–126 (2013).
13. Hubbard, A. *et al.* Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biol.* **16**, 23 (2015).
14. Radhakrishnan, G. V. *et al.* MARPLE, a point-of-care, strain-level disease diagnostics and surveillance tool for complex fungal pathogens. *BMC Biol.* **17**, 65 (2019).
15. Coram, T. E., Settles, M. L. & Chen, X. Transcriptome analysis of high-temperature adult-plant resistance conditioned by Yr39 during the wheat-Puccinia striiformis f. sp. tritici interaction. *Mol. Plant Pathol.* **9**, 479–493 (2008).
16. Wang, Y. *et al.* Transcriptome analysis provides insights into the mechanisms underlying wheat cultivar Shumai126 responding to stripe rust. *Gene* **768**, 145290 (2021).
17. Cantu, D. *et al.* Genome analyses of the wheat yellow (stripe) rust pathogen Puccinia striiformis f. sp. tritici reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* **14**, (2013).
18. Zhang, W. *et al.* Plant–necrotroph co-transcriptome networks illuminate a metabolic battlefield. *Elife* **8**, (2019).
19. Dobon, A., Bunting, D. C. E., Cabrera-Quio, L. E., Uauy, C. & Saunders, D. G. O. The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. *BMC Genomics* **17**, 380 (2016).
20. Bueno-Sancho, V. *et al.* Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects

Seasonal Variation and Host Specificity. *Genome Biol. Evol.* **9**, 3282–3296 (2017).
21. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* **34**, i884–i890 (2018).
22. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
23. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
24. Robinson, M., Mccarthy, D., Chen, Y. & Smyth, G. K. edgeR : differential expression analysis of digital gene expression data. *October* **23**, 2887–2887 (2011).
25. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* **6**, 80–92 (2012).
26. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
27. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
28. Jombart, T. & Ahmed, I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
29. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
30. Kamvar, Z. N., Tabima, J. F. & Gr̈unwald, N. J. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, 1–14 (2014).
31. Kamvar, Z. N., Brooks, J. C. & Grünwald, N. J. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* **6**, 208 (2015).
32. Oksanen, J. *et al.* vegan: Community Ecology Package. (2020).
33. Kelly, B. J. *et al.* Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* **31**, 2461–2468 (2015).
34. Harrell Jr, F. E., with contributions from Charles Dupont & many others. Hmisc: Harrell Miscellaneous. (2020).
35. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology. (2019).
36. Cuomo, C. A. *et al.* Comparative Analysis Highlights Variable Genome Content of Wheat Rusts and Divergence of the Mating Loci. *G3 Genes, Genomes, Genet.* **7**, 361–376 (2017).
37. Petersen, T. N., Brunak, S., Von Heijne, G. & Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786 (2011).
38. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Sy**, 1695 (2006).
39. Paul Shannon, 1 *et al.* Cytoscape: A Software Environment for Integrated Models. *Genome Res.* **13**, 426 (1971).
40. Mahboubi, M., Mehrabi, R., Naji, A. M. & Talebi, R. Whole-genome diversity, population structure and linkage disequilibrium analysis of globally diverse wheat genotypes using genotyping-by-sequencing DArTseq platform. *3 Biotech* **10**, 48 (2020).
41. Poland, J. & Rutkoski, J. Advances and Challenges in Genomic Selection for Disease Resistance. *Annu. Rev. Phytopathol.* **54**, 79–98 (2016).
42. Cheng, P., Chen, X. & See, D. Grass Hosts Harbor More Diverse Isolates of Puccinia striiformis Than Cereal Crops. *Phytopathology* **106**, 362–371 (2016).

43. Jin, Y. Role of Berberis spp. as alternate hosts in generating new races of Puccinia graminis and P. striiformis. *Euphytica* **179**, 105–108 (2011).
44. Rodnina, M. V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Science* 1390–1406 (2016). doi:10.1002/pro.2950
45. Xia, C. *et al.* Genomic insights into host adaptation between the wheat stripe rust pathogen (Puccinia striiformis f. sp. tritici) and the barley stripe rust pathogen (Puccinia striiformis f. sp. hordei). *BMC Genomics* **19**, (2018).
46. Jin, Y., Szabo, L. J. & Carson, M. Century-old mystery of Puccinia striiformis life history solved with the identification of Berberis as an alternate host. *Phytopathology* **100**, 432–435 (2010).
47. Scholes, J. D. & Rolfe, S. A. Photosynthesis in localised regions of oat leaves infected with crown rust (Puccinia coronata): Quantitative imaging of chlorophyll fluorescence. *Planta* **199**, 573–582 (1996).
48. Berger, S., Sinha, A. K. & Roitsch, T. Plant physiology meets phytopathology: Plant primary metabolism and plant-pathogen interactions. *Journal of Experimental Botany* **58**, 4019–4026 (2007).
49. Chang, Q. *et al.* The effect of Puccinia striiformis f. sp. tritici on the levels of water-soluble carbohydrates and the photosynthetic rate in wheat leaves. *Physiol. Mol. Plant Pathol.* **84**, 131–137 (2013).
50. Chen, S. *et al.* Key genes and genetic interactions of plant-pathogen functional modules in poplar infected by marssonina brunnea. *Mol. Plant-Microbe Interact.* **33**, 1080–1090 (2020).
51. Camp, R. R. & Whittingham, W. F. Fine Structure of Chloroplasts in 'Green Islands' and in Surrounding Chlorotic Areas of Barley Leaves Infected by Powdery Mildew. *Am. J. Bot.* **62**, 403 (1975).
52. Berger, S., Papadopoulos, M., Schreiber, U., Kaiser, W. & Roitsch, T. Complex regulation of gene expression, photosynthesis and sugar levels by pathogen infection in tomato. *Physiol. Plant.* **122**, 419–428 (2004).
53. Foyer, C. H. Reactive oxygen species, oxidative signaling and the regulation of photosynthesis. *Environmental and Experimental Botany* **154**, 134–142 (2018).
54. Kangasjärvi, S., Neukermans, J., Li, S., Aro, E. M. & Noctor, G. Photosynthesis, photorespiration, and light signalling in defence responses. *Journal of Experimental Botany* **63**, 1619–1636 (2012).
55. Liu, Y. & He, C. A review of redox signaling and the control of MAP kinase pathway in plants. *Redox Biology* **11**, 192–204 (2017).
56. Buchanan BB, G. W. & (eds.), J. R. L. *Biochemistry and Molecular Biology of Plants (second edition). Wiley, Blackwell (Chichester, UK)* (2015).
57. Bechtold, U., Karpinski, S. & Mullineaux, P. M. The influence of the light environment and photosynthesis on oxidative signalling responses in plant-biotrophic pathogen interactions. *Plant, Cell and Environment* **28**, 1046–1055 (2005).
58. Chen, Y. E. *et al.* Influence of stripe rust infection on the photosynthetic characteristics and antioxidant system of susceptible and resistant wheat cultivars at the adult plant stage. *Front. Plant Sci.* **6**, 779 (2015).
59. Hamelin, R. C. *et al.* Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci.* **108**, 9166–9171 (2011).

## Appendix 1: Sample Collection Information

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| AR1* | Cash, Arkansas | USA | North America | 4/30/2015 | Wheat | LCS 2141 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| AUT1* | Groß-Enzersdorf | Austria | Europe | 5/19/2014 | Wheat | Sax | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327125 |
| AUT2* | Gießhübl | Austria | Europe | 5/22/2014 | Wheat | Sax | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327126 |
| BEL1* | Koksijde | Belgium | Europe | 6/2/2014 | Wheat | JB Asano | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327175 |
| BEL2* | Mignault | Belgium | Europe | 6/2/2014 | Wheat | Torch | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327178 |
| BEL3* | Zwevegem | Belgium | Europe | 6/2/2014 | Wheat | JB Asano | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327176 |
| BEL4* | Tiegem | Belgium | Europe | 6/2/2014 | Wheat | JB Asano | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327177 |
| BGR1* | Novi Han | Bulgaria | Europe | 6/11/2014 | Wheat | Enola | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327157 |
| CA1* | Siskiyou, California | USA | North America | 6/3/2015 | Wheat | YS 221 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CA2* | Woodland, California | USA | North America | 4/4/2016 | Wheat | breeding line | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CAN1* | Lethbridge, Alberta | Canada | North America | 7/21/2015 | Wheat | GYT3-190 DN | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CAN2* | Lethbridge, Alberta | Canada | North America | 7/21/2015 | Wheat | Morocco FF-B | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CAN3* | Lethbridge, Alberta | Canada | North America | 7/21/2015 | Wheat | Penhold | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CAN4* | Lacombe, Alberta | Canada | North America | 7/21/2015 | Wheat | HY2047 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CAN5* | St. Albert, Edmonton | Canada | North America | 7/21/2015 | Wheat | HY2042 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CAN6* | Lacombe, Alberta | Canada | North America | 7/21/2015 | Wheat | Yr31 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHL1* | StaRosa | Chile | South America | 11/24/2014 | Wheat | Fito1401-87/ cv. Pantera | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327104 |
| CHL2* | Carillana | Chile | South America | 11/24/2014 | Wheat | Onix | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327102 |
| CHL3* | Cajon | Chile | South America | 12/6/2014 | Wheat | Taita | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327105 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| CHL4* | Los Tilos | Chile | South America | 12/7/2014 | Wheat | Fito 1404-15 | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327106 |
| CHL5* | StaRosa | Chile | South America | 11/24/2014 | Wheat | Onix | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327103 |
| CHN1 | Nanhu, Zhuanglang | China | Asia | 11/26/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN10* | unknown | China | Asia | July 2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN11 | unknown | China | Asia | July 2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN12* | unknown | China | Asia | July 2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN13* | Nanhu, Zhuanglang | China | Asia | 11/26/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN14* | Nanhu, Zhuanglang | China | Asia | 11/26/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN15* | Pingnan, Quizhou | China | Asia | 11/27/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN16* | Pingnan, Quizhou | China | Asia | 11/27/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN17* | Pingnan, Quizhou | China | Asia | 11/27/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN18* | Pingnan, Quizhou | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN19* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN2* | Nanhu, Zhuanglang | China | Asia | 11/26/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN20* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN21* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN22* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN23* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN24* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN25* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN26* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN27* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| CHN28* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN29* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN3* | Pingnan, Quizhou | China | Asia | 11/27/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN30* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN31* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN32* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN33* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN4 | Pingnan, Quizhou | China | Asia | 11/27/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN5* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN6* | Jinshan, Gangu | China | Asia | 11/28/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN7* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN8* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CHN9* | unknown | China | Asia | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| CO1* | Akron, CO | USA | North America | 6/16/2017 | Wheat | Ruth | This study | RNA-seq | |
| CO2* | Colorado | USA | North America | 6/16/2017 | Wheat | CO12M0367 | This study | RNA-seq | |
| CO3* | Colorado | USA | North America | 6/16/2017 | Wheat | CO13003C | This study | RNA-seq | |
| CO4* | Orchard, CO | USA | North America | 6/16/2017 | Wheat | Brawl CL Plus | This study | RNA-seq | |
| CO5* | Orchard, CO | USA | North America | 6/16/2017 | Wheat | CO14A065 | This study | RNA-seq | |
| CO6* | Roggen, CO | USA | North America | 6/12/2017 | Wheat | unknown | This study | RNA-seq | |
| CO7* | site 56, CO | USA | North America | 6/6/2017 | Wheat | unknown | This study | RNA-seq | |
| CZE1* | Rokytnice | Czech Republic | Europe | 6/12/2014 | Wheat | Tilman | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327120 |
| CZE2 | Olomouc | Czech Republic | Europe | 6/19/2014 | Wheat | NIC08 -108 | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327154 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| CZE3 | Olomouc | Czech Republic | Europe | 6/19/2014 | Perennial grass | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327153 |
| CZE4 | Olomouc | Czech Republic | Europe | 6/19/2014 | Wheat | Turandot/Baryton cross | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327155 |
| DE1* | Middletown, DE | USA | North America | 6/8/2017 | Wheat | Shirley | This study | RNA-seq | |
| DEU1* | Moselürsch | Germany | Europe | 6/10/2014 | Wheat | Aktuer | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327142 |
| DEU12* | Dunau | Germany | Europe | 6/17/2014 | Wheat | Loft | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327148 |
| DEU14* | Gommershoven | Germany | Europe | 5/13/2014 | Wheat | JB Asano | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327163 |
| DEU15* | Groitzsch | Germany | Europe | 5/28/2014 | Wheat | Kometus | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327164 |
| DEU16* | Groitzsch | Germany | Europe | 6/3/2014 | Wheat | Akteur | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327165 |
| DEU17* | Weißensee | Germany | Europe | 6/3/2014 | Wheat | Ritmo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327166 |
| DEU18* | Weißensee | Germany | Europe | 6/3/2014 | Wheat | JB Asano | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327167 |
| DEU19* | Tachenhausen | Germany | Europe | 6/4/2014 | Wheat | Kometus | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327169 |
| DEU2* | Hingste Inokluiert | Germany | Europe | 6/16/2014 | Wheat | Contur | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327147 |
| DEU20* | Hüttendorf | Germany | Europe | 6/5/2014 | Wheat | Kometus | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327171 |
| DEU22 | Werbig | Germany | Europe | July 2014 | Rye | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327185 |
| DEU23* | Ruchheim | Germany | Europe | 6/4/2014 | Wheat | Victo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327161 |
| DEU3* | Grucking | Germany | Europe | 6/23/2014 | Wheat | JB Asano | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327149 |
| DEU4* | Böhl Bauch 10 | Germany | Europe | 6/17/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327162 |
| DEU5* | Ruchheim | Germany | Europe | 6/17/2014 | Wheat | Victo | Bueno-Sancho et al., 2017 | RNA-seq | To be shared |
| DEU6* | Weißensee | Germany | Europe | 6/3/2014 | Wheat | Akteur | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327168 |
| DEU7* | Hüttendorf | Germany | Europe | 6/5/2014 | Wheat | Hermann | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327170 |
| DEU8* | Etzdorf | Germany | Europe | 6/13/2014 | Wheat | JB Asano | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327143 |
| DEU9* | Etzdorf | Germany | Europe | 6/13/2014 | Wheat | Kometus | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327144 |
| ESP1 | Garínoain | Spain | Europe | 7/3/2014 | Wheat | Camargo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327158 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| ETH1* | Kulumsa, Oromia | Ethiopia | Africa | 5/16/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327111 |
| ETH10* | Sinana | Ethiopia | Africa | 12/5/2016 | Wheat | Digalu | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| ETH3* | Sinana | Ethiopia | Africa | 12/5/2016 | Wheat | Millenium | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| ETH4* | Sinana | Ethiopia | Africa | 12/5/2016 | Wheat | SPPSN#15 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| ETH5 | Sinana | Ethiopia | Africa | 12/5/2016 | Wheat | Hoggana | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| ETH6* | Sinana RS | Ethiopia | Africa | 12/1/2014 | Wheat | Hidasse | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327115 |
| ETH7* | Sinana RS | Ethiopia | Africa | 12/1/2014 | Wheat | Honqolo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327116 |
| ETH8* | Sinana RS | Ethiopia | Africa | 12/1/2014 | Wheat | Sanate | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327117 |
| FRA1 | Castanet | France | Europe | 6/4/2015 | Triticale | Orleac | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| FRA10* | Marchélepot | France | Europe | 6/16/2014 | Wheat | Trapez | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327122 |
| FRA11* | Marchélepot | France | Europe | 6/16/2014 | Wheat | Bergamo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327123 |
| FRA12* | Pontfaverger | France | Europe | 6/10/2014 | Wheat | Trapez | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327124 |
| FRA13* | Guégon | France | Europe | 6/4/2014 | Wheat | Expert | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327130 |
| FRA14* | Bucy Saint Liphard | France | Europe | 6/25/2014 | Wheat | Hysun | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327151 |
| FRA15* | Bucy Saint Liphard | France | Europe | 6/25/2014 | Wheat | Karur | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327152 |
| FRA16* | unknown | France | Europe | 6/11/2014 | Durum wheat | Luminur | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327183 |
| FRA17 | - | France | Europe | - | Wheat | unknown | Hubbard et al., 2015 | RNA-seq | ERS1327095 |
| FRA18 | - | France | Europe | 2000 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX671609 |
| FRA19 | - | France | Europe | 2002 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX671644 |
| FRA2 | Castanet | France | Europe | 6/4/2015 | Triticale | Tarzan | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| FRA20 | - | France | Europe | 2001 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX672139 |
| FRA3* | Gourhel | France | Europe | 6/2/2014 | Wheat | Altigo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327129 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| FRA4* | Loyat | France | Europe | 6/4/2014 | Wheat | Alixan | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327131 |
| FRA6* | Nangis, Seine-et-Marne | France | Europe | 6/3/2014 | Wheat | Rosario | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327174 |
| FRA7* | Marchélepot | France | Europe | 6/16/2014 | Wheat | Glasgow | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327121 |
| FRA8 | Castanet | France | Europe | 6/4/2015 | Triticale | Ragtac | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| FRA9 | Castanet | France | Europe | 6/4/2015 | Triticale | Fido | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| GBR1* | Cirencester, Gloucestershire | UK | Europe | 6/12/2013 | Wheat | Oakley | Hubbard et al., 2015 | RNA-seq | SRX684252 |
| GBR10* | Caythorpe, Lincolnshire | UK | Europe | 5/29/2013 | Wheat | Oakley | Hubbard et al., 2015 | RNA-seq | SRX684852 |
| GBR100 | Cascob, Powys (Wales) | UK | Europe | 8/13/2014 | Wheat | AC Barrie | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327232 |
| GBR103* | - | UK | Europe | 1978 | Wheat | unknown | Hubbard et al., 2015 | RNA-seq | SRX708529 |
| GBR104 | - | UK | Europe | 1987 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX667853 |
| GBR105 | - | UK | Europe | 1987 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX667852 |
| GBR106 | - | UK | Europe | 1988 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX667891 |
| GBR107 | - | UK | Europe | 1988 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX667866 |
| GBR108 | - | UK | Europe | 1988 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX667861 |
| GBR109 | - | UK | Europe | 1988 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX667861 |
| GBR11* | Boothby Graffoe, Lincolnshire | UK | Europe | 6/13/2013 | Wheat | KWS Santiago | Hubbard et al., 2015 | RNA-seq | SRX664976 |
| GBR110 | - | UK | Europe | 1998 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX667790 |
| GBR111* | Cambridgeshire | UK | Europe | 2014 | Wheat | Oakley | Hubbard et al., 2015 | RNA-seq | SRX702551 |
| GBR112* | Cambridgeshire | UK | Europe | 2014 | Wheat | Torch | Hubbard et al., 2015 | RNA-seq | SRX702556 |
| GBR12* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | Recital | Hubbard et al., 2015 | RNA-seq | SRX690356 |
| GBR13* | Mendlesham, Suffolk | UK | Europe | 6/5/2013 | Wheat | KWS 196 | Hubbard et al., 2015 | RNA-seq | SRX701325 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| GBR14* | Boothby Graffoe, Lincolnshire | UK | Europe | 6/13/2013 | Wheat | Solstice | Hubbard et al., 2015 | RNA-seq | SRX684439 |
| GBR15* | Caythorpe, Lincolnshire | UK | Europe | 6/5/2013 | Wheat | RW41088 | Hubbard et al., 2015 | RNA-seq | SRX684851 |
| GBR16* | Boothby Graffoe, Lincolnshire | UK | Europe | 6/13/2013 | Wheat | KWS Dali | Hubbard et al., 2015 | RNA-seq | SRX685991 |
| GBR17* | unknown | UK | Europe | 2017 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| GBR18* | Humbie, East Lothian | UK | Europe | 6/23/2014 | Wheat | KWS Kielder | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327190 |
| GBR19 | - | UK | Europe | 2001 | Wheat | unknown | Hubbard et al., 2015 | Genomic | SRX667770 |
| GBR2* | Cambridgeshire | UK | Europe | 6/18/2013 | Wheat | Claire | Hubbard et al., 2015 | RNA-seq | SRX684850 |
| GBR20 | - | UK | Europe | 2003 | Wheat | Brock | Hubbard et al., 2015 | Genomic | SRX667387 |
| GBR21 | - | UK | Europe | 2008 | Wheat | Solstice | Hubbard et al., 2015 | Genomic | SRX667083 |
| GBR22 | - | UK | Europe | 2008 | Wheat | Timber | Hubbard et al., 2015 | Genomic | SRX667388 |
| GBR23 | Lincolnshire | UK | Europe | 2011 | Wheat | Warrior | Hubbard et al., 2015 | Genomic | SRX703657 |
| GBR24* | - | UK | Europe | 2011 | Wheat | unknown | Hubbard et al., 2015 | RNA-seq | SRS707593 |
| GBR25* | - | UK | Europe | 2011 | Wheat | unknown | Hubbard et al., 2015 | RNA-seq | SRX708523 |
| GBR26 | Bedfordshire | UK | Europe | 2011 | Wheat | KWS Sterling | Hubbard et al., 2015 | Genomic | SRX667377 |
| GBR27* | Caythorpe, Lincolnshire | UK | Europe | 7/3/2013 | Wheat | Ambition | Hubbard et al., 2015 | RNA-seq | SRX702534 |
| GBR28* | Boston, Lincolnshire | UK | Europe | 7/3/2013 | Wheat | Ambition | Hubbard et al., 2015 | RNA-seq | SRX702401 |
| GBR29* | Boston, Lincolnshire | UK | Europe | 6/5/2013 | Wheat | Oakley | Hubbard et al., 2015 | RNA-seq | SRX684393 |
| GBR3* | Mendlesham, Suffolk | UK | Europe | 6/5/2013 | Wheat | Torch | Hubbard et al., 2015 | RNA-seq | SRX685990 |
| GBR30* | Unknown | UK | Europe | 2013 | Wheat | Crusoe | Hubbard et al., 2015 | RNA-seq | SRX702535 |
| GBR31* | Sutton Scotney, Hampshire | UK | Europe | 6/6/2013 | Wheat | KWS Rowan | Hubbard et al., 2015 | RNA-seq | SRX690338 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| GBR32* | Boothby Graffoe, Lincolnshire | UK | Europe | 6/13/2013 | Wheat | KWS Kielder | Hubbard et al., 2015 | RNA-seq | SRX690311 |
| GBR33* | Cirencester, Gloucestershire | UK | Europe | 6/12/2013 | Wheat | Horatio | Hubbard et al., 2015 | RNA-seq | SRX686006 |
| GBR34* | Cirencester, Gloucestershire | UK | Europe | 6/12/2013 | Wheat | Solstice | Hubbard et al., 2015 | RNA-seq | SRX690352 |
| GBR35* | Oxfordshire | UK | Europe | 6/14/2013 | Wheat | Solstice | Hubbard et al., 2015 | RNA-seq | SRX685190 |
| GBR36* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | ReR22 | Hubbard et al., 2015 | RNA-seq | SRX690343 |
| GBR37* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | ReR04 | Hubbard et al., 2015 | RNA-seq | SRX690312 |
| GBR38* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | ReR32 | Hubbard et al., 2015 | RNA-seq | SRX686004 |
| GBR39* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | ReR64 | Hubbard et al., 2015 | RNA-seq | SRX690541 |
| GBR4* | Mendlesham, Suffolk | UK | Europe | 6/5/2013 | Wheat | LGW56 (Panacea) | Hubbard et al., 2015 | RNA-seq | SRX686005 |
| GBR40* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | Allez-Y | Hubbard et al., 2015 | RNA-seq | SRX690378 |
| GBR41* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | ReR110 | Hubbard et al., 2015 | RNA-seq | SRX701334 |
| GBR42* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | ReR138 | Hubbard et al., 2015 | RNA-seq | SRX690348 |
| GBR43* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | Toisander | Hubbard et al., 2015 | RNA-seq | SRX690379 |
| GBR44* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | Fairplay | Hubbard et al., 2015 | RNA-seq | SRX690528 |
| GBR45* | Welbourne, Lincolnshire | UK | Europe | 6/14/2013 | Wheat | Audace | Hubbard et al., 2015 | RNA-seq | SRX690484 |
| GBR46* | NIAB HQ, Cambridgeshire | UK | Europe | 7/7/2013 | Wheat | Warrior | Hubbard et al., 2015 | RNA-seq | SRX701336 |
| GBR47* | Headley Hall, North Yorkshire | UK | Europe | 6/19/2013 | Wheat | Laurier | Hubbard et al., 2015 | RNA-seq | SRX690483 |
| GBR48* | Sutton Scotney, Hampshire | UK | Europe | 2/21/2014 | Wheat | Victo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327132 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| GBR49* | Sutton Scotney, Hampshire | UK | Europe | 2/21/2014 | Wheat | Oakley | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327133 |
| GBR5 | Rockland St Mary, Norfolk | UK | Europe | 6/6/2013 | Triticale | unknown | Hubbard et al., 2015 | RNA-seq | SRX690542 |
| GBR50* | unknown | UK | Europe | 3/7/2014 | Wheat | Panorama | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327134 |
| GBR51* | unknown | UK | Europe | 3/7/2014 | Wheat | Target | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327135 |
| GBR52* | unknown | UK | Europe | 3/18/2014 | Wheat | spreader | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327136 |
| GBR53 | unknown | UK | Europe | 2014 | Wheat | Cocoon | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327137 |
| GBR54* | Morley, Norfolk | UK | Europe | 6/13/2014 | Wheat | Kranich | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327182 |
| GBR55* | Bridgend, Mid Glamorgan | UK | Europe | June 2014 | Wheat | Vuka | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327186 |
| GBR56* | unknown | UK | Europe | July 2014 | Wheat | Vuka | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327187 |
| GBR57* | Bishop Burton, East Yorkshire | UK | Europe | August 2014 | Wheat | Vuka | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327188 |
| GBR59* | Wolferton, Norfolk | UK | Europe | 6/23/2014 | Wheat | Jorvik | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327191 |
| GBR6 | Headley Hall, North Yorkshire | UK | Europe | 6/26/2013 | Triticale | Phidahlia | Hubbard et al., 2015 | RNA-seq | SRX702404 |
| GBR60* | Humbie, East Lothian | UK | Europe | 6/23/2014 | Wheat | LGW65 (Energise) | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327192 |
| GBR62 | Humbie, East Lothian | UK | Europe | 6/23/2014 | Wheat | JB Diego | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327194 |
| GBR63* | York, North Yorkshire | UK | Europe | 4/28/2014 | Wheat | Solstice | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327195 |
| GBR64* | York, North Yorkshire | UK | Europe | 4/28/2014 | Wheat | KWS Kielder | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327196 |
| GBR65* | Morley, Norfolk | UK | Europe | 4/1/2014 | Wheat | Victo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327197 |
| GBR66 | Caythorpe, Lincolnshire | UK | Europe | 1/23/2014 | Wheat | Oakley | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327198 |
| GBR67 | Caythorpe, Lincolnshire | UK | Europe | 1/23/2014 | Wheat | Victo | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327199 |
| GBR68* | Titchmarsh, Northamptonshire | UK | Europe | 1/13/2014 | Wheat | Claire | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327200 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| GBR69* | Wallingford, Oxfordshire | UK | Europe | December 2013 | Wheat | KWS Kielder | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327201 |
| GBR7 | Wareham, Dorset | UK | Europe | 7/5/2013 | Triticale | Amarillo | Hubbard et al., 2015 | RNA-seq | SRX702423 |
| GBR71 | Caythorpe, Lincolnshire | UK | Europe | 1/23/2014 | Wheat | Claire | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327203 |
| GBR72 | Caythorpe, Lincolnshire | UK | Europe | 1/23/2014 | Wheat | Robigus | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327204 |
| GBR73* | Canterbury, Kent | UK | Europe | 5/23/2014 | Wheat | Cordiale | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327205 |
| GBR74* | Canterbury, Kent | UK | Europe | 5/23/2014 | Wheat | Solstice | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327206 |
| GBR75 | Caythorpe, Lincolnshire | UK | Europe | 5/6/2014 | Wheat | Solstice | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327207 |
| GBR76* | Caythorpe, Lincolnshire | UK | Europe | 5/6/2014 | Wheat | Robigus | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327208 |
| GBR77* | unknown | UK | Europe | 5/16/2014 | Wheat | Ambition | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327209 |
| GBR79* | Cambridgeshire | UK | Europe | 5/8/2014 | Wheat | Kranich | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327211 |
| GBR8 | Wool, Dorset | UK | Europe | 7/5/2013 | Triticale | Bennette | Hubbard et al., 2015 | RNA-seq | SRX702533 |
| GBR80* | Sturminster, Dorset | UK | Europe | 5/16/2014 | Wheat | Solstice | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327212 |
| GBR81* | Canterbury, Kent | UK | Europe | 5/23/2014 | Wheat | Claire | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327213 |
| GBR82 | Keisby, Lincolnshire | UK | Europe | 5/16/2014 | Wheat | Belvoir | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327214 |
| GBR83* | Canterbury, Kent | UK | Europe | 5/23/2014 | Wheat | KWS Sterling | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327215 |
| GBR84* | Canterbury, Kent | UK | Europe | 5/23/2014 | Wheat | Robigus | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327216 |
| GBR85* | Headley Hall, North Yorkshire | UK | Europe | 5/6/2014 | Wheat | Oakley | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327217 |
| GBR86* | Royston, Cambridgeshire | UK | Europe | 5/23/2014 | Wheat | Oakley | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327218 |
| GBR87* | Canterbury, Kent | UK | Europe | 5/23/2014 | Wheat | Oakley | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327219 |
| GBR88* | unknown | UK | Europe | 5/16/2014 | Wheat | Kranich | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327220 |
| GBR89* | Canterbury, Kent | UK | Europe | 5/23/2014 | Wheat | Warrior | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327221 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| GBR9* | Terrington St Clement, Norfolk | UK | Europe | 2013 | Wheat | Solstice | Hubbard et al., 2015 | RNA-seq | SRX684276 |
| GBR90* | Caythorpe, Lincolnshire | UK | Europe | 5/6/2014 | Wheat | Oakley | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327222 |
| GBR91* | Canterbury, Kent | UK | Europe | 5/23/2014 | Wheat | Relay | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327223 |
| GBR92* | Fulbourn, Cambridgeshire | UK | Europe | May 2014 | Wheat | Gallant | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327224 |
| GBR93 | Sutton Scotney, Hampshire | UK | Europe | 6/2/2014 | Wheat | KWS Sterling | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327225 |
| GBR94* | Morley, Norfolk | UK | Europe | May 2014 | Wheat | Solstice | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327226 |
| GBR95 | Headley Hall, North Yorkshire | UK | Europe | 6/2/2014 | Wheat | Relay | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327227 |
| GBR97 | Morley, Norfolk | UK | Europe | May 2014 | Wheat | Oakley | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327229 |
| GBR98* | Morley, Norfolk | UK | Europe | 2014 | Wheat | Cordiale | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327230 |
| GBR99* | Cascob, Powys (Wales) | UK | Europe | 8/13/2014 | Wheat | AC Barrie | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327231 |
| HRV1* | Osisek | Croatia | Europe | 5/21/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327119 |
| HRV2* | unknown | Croatia | Europe | 5/13/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327128 |
| ID1* | Ririe, Idaho | USA | North America | 6/9/2015 | Wheat | Deloris | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| ID2* | Rockland, Idaho | USA | North America | 4/26/2016 | Wheat | Bonneville | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| IN1* | West Lafeyette, Indiana | USA | North America | 5/23/2017 | Wheat | Ag1331 | CSU collection | RNA-seq | |
| IN2* | Lafayette, Indiana | USA | North America | 5/17/2016 | Wheat | DH13SRW02 4-182 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| IN3* | Layfayette, Indiana | USA | North America | 5/25/2017 | Wheat | NC13142-33 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| IN4* | Layfayette, Indiana | USA | North America | 5/25/2017 | Wheat | NC13142-35 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| IN5* | Layfayette, Indiana | USA | North America | 5/25/2017 | Wheat | NC13142-36 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| ITA1* | Ozzano, Bologna | Italy | Europe | 5/21/2014 | Wheat | Massari | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327159 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| ITA2* | Conselice | Italy | Europe | 5/21/2014 | Wheat | Irnerio | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327160 |
| KS1* | Wichita, Kansas | USA | North America | 4/25/2017 | Wheat | 2141 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| KS2* | Wichita, Kansas | USA | North America | 4/27/2017 | Wheat | LCS Compass | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| KS3* | Wichita, Kansas | USA | North America | 4/22/2015 | Wheat | Everest | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| KY1* | Lexington, KY | USA | North America | 5/17/2017 | Wheat | unkown | This study | RNA-seq | |
| KY2* | Princeton, KY | USA | North America | 6/2/2017 | Wheat | unkown | This study | RNA-seq | |
| MI1* | Scottville, MI | USA | North America | 6/2/2017 | Wheat | unkown | This study | RNA-seq | |
| MI2* | Lansing, Michigan | USA | North America | 5/23/2016 | Wheat | K505hw14 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| MT1* | Bozeman, Montana | USA | North America | 6/1/2015 | Wheat | Yellowstone | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| MT2* | Montana | USA | North America | 6/2/2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| MT3* | Kalispell, Montana | USA | North America | 4/26/2016 | Wheat | Decade | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| NE1* | Blue Hills, NE | USA | North America | 5/24/2017 | Wheat | unknown | This study | RNA-seq | |
| NE2* | Box Butte, Nebraska | USA | North America | 7/2/2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| NY1* | Penn Yan, NY | USA | North America | 5/26/2017 | Wheat | Erie (soft red winter) | This study | RNA-seq | |
| NY2* | Savannah, NY | USA | North America | 6/7/2017 | Wheat | unkown | This study | RNA-seq | |
| NY3* | Weedsport, New York | USA | North America | 5/27/2016 | Wheat | Seedway | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| NZL1 | Darfield | New Zealand | Oceania | November 2011 | Wheat | Morph | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327108 |
| NZL2* | - | New Zealand | Oceania | 2012 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327109 |
| NZL3 | Lincoln | New Zealand | Oceania | November 2012 | Wheat | Sy Epson | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327107 |
| NZL4* | - | New Zealand | Oceania | 2006 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327110 |
| OK1* | Chickasha, Oklahoma | USA | North America | 4/7/2017 | Wheat | Ruby Lee | Radhakrishnan et al., 2019 | RNA-seq | To be shared |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| OK2* | Altus, Oklahoma | USA | North America | 4/11/2017 | Wheat | Snowmass | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| OK3* | Cimarron, Oklahoma | USA | North America | 5/26/2015 | Wheat | Byrd | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| OR1* | Corvallis, Oregon | USA | North America | 5/27/2015 | Wheat | ORCF 101R | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| PAK1* | Attarshisha | Pakistan | Asia | 13/05/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327099 |
| PAK2* | Abbottabad | Pakistan | Asia | 16/05/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327096 |
| PAK3* | Qalandar Abad | Pakistan | Asia | 16/05/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327098 |
| PAK4* | Attarshisha | Pakistan | Asia | 13/05/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327100 |
| PAK5* | Attarshisha | Pakistan | Asia | 13/05/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327101 |
| PAK6* | Abbottabad | Pakistan | Asia | 16/05/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327097 |
| POL1* | Kwidzyn | Poland | Europe | 6/5/2014 | Wheat | Arkadia | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327139 |
| POL2* | Środa Wielko | Poland | Europe | 6/6/2014 | Wheat | Princeps | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327141 |
| POL3* | Choryn | Poland | Europe | 5/22/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327127 |
| POL4* | Kwidzyn | Poland | Europe | 6/5/2014 | Wheat | Kobra | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327138 |
| POL5* | Środa Wielko | Poland | Europe | 6/6/2014 | Wheat | Legenda | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327140 |
| POL6* | Jarosławiec | Poland | Europe | 6/15/2015 | Wheat | KWS Magic | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| PST100* | | USA | North America | 4/12/2017 | Wheat | Ripper | This study | RNA-seq | |
| ROU1* | unknown | Romania | Europe | 5/24/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327179 |
| SD1* | Gregory, South Dakota | USA | North America | 6/10/2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SD2* | Tripp, South Dakota | USA | North America | 6/29/2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SD3* | Clarkton, South Dakota | USA | North America | 7/2/2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SRB1* | unknown | Serbia | Europe | 5/24/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327181 |
| SVK1* | Detva | Slovakia | Europe | 5/26/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327172 |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| SVK2* | Detva | Slovakia | Europe | 5/26/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327173 |
| SWE1* | Österlen | Sweden | Europe | 5/7/2015 | Wheat | Primus | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SWE2* | Sanby | Sweden | Europe | 5/30/2017 | Wheat | Linz | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SWE3* | Svalöv | Sweden | Europe | 6/15/2017 | Wheat | Kranich | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SWE4 | Svalöv | Sweden | Europe | 6/15/2017 | Wheat | 555012 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SWE5* | Svalōv | Sweden | Europe | 5/26/2016 | Wheat | Hereford | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SWE6* | Svalōv | Sweden | Europe | 5/26/2016 | Wheat | Linus | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SWE7* | Sanby | Sweden | Europe | 5/30/2017 | Wheat | Linz | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SWE8* | Svalöv | Sweden | Europe | 6/2/2017 | Wheat | Kranich | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| SWE9* | Svalöv | Sweden | Europe | 6/2/2017 | Wheat | Norin | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| TN1* | Tennessee | USA | North America | 4/21/2017 | Wheat | Beck's 128 | This study | RNA-seq | |
| TUR1* | unknown | Turkey | Europe | 5/24/2014 | Wheat | unknown | Bueno-Sancho et al., 2017 | RNA-seq | ERS1327180 |
| TX1* | Hidalgo, Texas | USA | North America | 3/9/2015 | Wheat | Sisson | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| VA1* | Warsaw, Virginia | USA | North America | 4/21/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| WA1* | Walla Walla, Washington | USA | North America | 6/24/2015 | Wheat | Express | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| WA2* | Walla Walla, Washington | USA | North America | 6/13/2017 | Wheat | Otto | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| WA3* | Walla Walla, Washington | USA | North America | 6/13/2017 | Wheat | Artdeco | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| WA4* | Edicott, Washington | USA | North America | 6/11/2015 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| WA6* | Mt. Vernon, Washington | USA | North America | 4/10/2015 | Wheat | PS 279 | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| WA7* | Rearden, Washington | USA | North America | 5/13/2016 | Wheat | Ovation | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| WA8* | Othello, Washington | USA | North America | 4/20/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |

| ID | Location | Country | Continent | Date Collected | Host | Host Variety | Source | Data Type | Accession Number |
|---|---|---|---|---|---|---|---|---|---|
| WA9* | Walla Walla, Washington | USA | North America | 5/12/2016 | Wheat | LCS Jet | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| ZAF1* | Napier | South Africa | Africa | 10/18/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |
| ZAF2* | Napier | South Africa | Africa | 10/18/2016 | Wheat | unknown | Radhakrishnan et al., 2019 | RNA-seq | To be shared |

Legend:
* Sample used in Chapter 3 after all filtering steps.

## Appendix 2: Wheat GO Analysis – Ta_ECN_I

| Biological Process – Ta_ECN_I | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0015979 | Photosynthesis | 813 | 76 | 2.04 | < 1e-30 |
| 2 | GO:0006091 | Generation of precursor metabolites and energy | 1415 | 62 | 3.55 | < 1e-30 |
| 3 | GO:0019684 | Photosynthesis, light reaction | 511 | 45 | 1.28 | < 1e-30 |
| 4 | GO:0009768 | Photosynthesis, light harvesting in photosystem I | 64 | 19 | 0.16 | < 1e-30 |
| 5 | GO:0009765 | Photosynthesis, light harvesting | 105 | 21 | 0.26 | < 1e-30 |
| 6 | GO:0019253 | Reductive pentose-phosphate cycle | 41 | 14 | 0.1 | 5.30E-27 |
| 7 | GO:0019685 | Photosynthesis, dark reaction | 42 | 14 | 0.11 | 7.90E-27 |
| 8 | GO:0018298 | Protein-chromophore linkage | 164 | 19 | 0.41 | 1.80E-26 |
| 9 | GO:0009416 | Response to light stimulus | 878 | 30 | 2.2 | 1.10E-25 |
| 10 | GO:0009314 | Response to radiation | 988 | 30 | 2.48 | 3.20E-24 |
| 11 | GO:0015977 | Carbon fixation | 61 | 14 | 0.15 | 3.20E-24 |
| 12 | GO:0010207 | Photosystem II assembly | 168 | 17 | 0.42 | 9.70E-23 |
| 13 | GO:0009628 | Response to abiotic stimulus | 1867 | 33 | 4.68 | 1.90E-19 |
| 14 | GO:0051186 | Cofactor metabolic process | 1891 | 33 | 4.74 | 2.80E-19 |
| 15 | GO:0035304 | Regulation of protein dephosphorylation | 131 | 13 | 0.33 | 1.60E-17 |
| 16 | GO:0006090 | Pyruvate metabolic process | 576 | 20 | 1.44 | 2.30E-17 |
| 17 | GO:0016053 | Organic acid biosynthetic process | 1597 | 29 | 4 | 2.30E-17 |
| 18 | GO:0046394 | Carboxylic acid biosynthetic process | 1597 | 29 | 4 | 2.30E-17 |
| 19 | GO:0035303 | Regulation of dephosphorylation | 137 | 13 | 0.34 | 3.00E-17 |
| 20 | GO:0009773 | Photosynthetic electron transport in photosystem I | 73 | 11 | 0.18 | 4.50E-17 |

| Cellular Component – Ta_ECN_I | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0009507 | Chloroplast | 3737 | 96 | 9.89 | < 1e-30 |
| 2 | GO:0009579 | Thylakoid | 1067 | 67 | 2.82 | < 1e-30 |
| 3 | GO:0009534 | Chloroplast thylakoid | 833 | 56 | 2.21 | < 1e-30 |
| 4 | GO:0031976 | Plastid thylakoid | 835 | 56 | 2.21 | < 1e-30 |
| 5 | GO:0044436 | Thylakoid part | 846 | 56 | 2.24 | < 1e-30 |
| 6 | GO:0044434 | Chloroplast part | 1977 | 69 | 5.23 | < 1e-30 |
| 7 | GO:0034357 | Photosynthetic membrane | 810 | 54 | 2.14 | < 1e-30 |
| 8 | GO:0044435 | Plastid part | 2020 | 69 | 5.35 | < 1e-30 |
| 9 | GO:0042651 | Thylakoid membrane | 768 | 52 | 2.03 | < 1e-30 |
| 10 | GO:0009536 | Plastid | 8609 | 104 | 22.79 | < 1e-30 |
| 11 | GO:0009535 | Chloroplast thylakoid membrane | 697 | 50 | 1.84 | < 1e-30 |
| 12 | GO:0055035 | Plastid thylakoid membrane | 699 | 50 | 1.85 | < 1e-30 |
| 13 | GO:0009521 | Photosystem | 325 | 35 | 0.86 | < 1e-30 |
| 14 | GO:0009522 | Photosystem I | 151 | 29 | 0.4 | < 1e-30 |
| 15 | GO:0031984 | Organelle subcompartment | 2011 | 56 | 5.32 | < 1e-30 |
| 16 | GO:0009526 | Plastid envelope | 883 | 43 | 2.34 | < 1e-30 |
| 17 | GO:0009941 | Chloroplast envelope | 842 | 39 | 2.23 | < 1e-30 |
| 18 | GO:0010287 | Plastoglobule | 147 | 22 | 0.39 | < 1e-30 |
| 19 | GO:0044444 | Cytoplasmic part | 19525 | 111 | 51.68 | < 1e-30 |
| 20 | GO:0098796 | Membrane protein complex | 1280 | 40 | 3.39 | < 1e-30 |

| Molecular Function – Ta_ECN_I | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0031409 | Pigment binding | 64 | 19 | 0.13 | < 1e-30 |
| 2 | GO:0016168 | Chlorophyll binding | 160 | 23 | 0.31 | < 1e-30 |
| 3 | GO:0048037 | Cofactor binding | 3182 | 44 | 6.26 | 8.20E-27 |
| 4 | GO:0046906 | Tetrapyrrole binding | 1489 | 24 | 2.93 | 1.10E-15 |
| 5 | GO:0004332 | Fructose-bisphosphate aldolase activity | 18 | 6 | 0.04 | 9.00E-13 |
| 6 | GO:0016491 | Oxidoreductase activity | 5041 | 35 | 9.92 | 8.90E-12 |
| 7 | GO:0042132 | Fructose 1,6-bisphosphate 1-phosphatase activity | 14 | 5 | 0.03 | 5.20E-11 |
| 8 | GO:0016830 | Carbon-carbon lyase activity | 353 | 11 | 0.69 | 1.20E-10 |
| 9 | GO:0016832 | Aldehyde-lyase activity | 49 | 6 | 0.1 | 6.50E-10 |
| 10 | GO:0050308 | Sugar-phosphatase activity | 25 | 5 | 0.05 | 1.40E-09 |
| 11 | GO:0052852 | Very-long-chain-(S)-2-hydroxy-acid oxidase activity | 10 | 4 | 0.02 | 2.90E-09 |
| 12 | GO:0052853 | Long-chain-(S)-2-hydroxy-long-chain-acid oxidase activity | 10 | 4 | 0.02 | 2.90E-09 |
| 13 | GO:0052854 | Medium-chain-(S)-2-hydroxy-acid oxidase activity | 10 | 4 | 0.02 | 2.90E-09 |
| 14 | GO:0008891 | Glycolate oxidase activity | 14 | 4 | 0.03 | 1.40E-08 |
| 15 | GO:0003973 | (S)-2-hydroxy-acid oxidase activity | 16 | 4 | 0.03 | 2.50E-08 |
| 16 | GO:0019203 | Carbohydrate phosphatase activity | 52 | 5 | 0.1 | 6.40E-08 |
| 17 | GO:0043891 | Glyceraldehyde-3-phosphate dehydrogenase (NAD(P)+) (phosphorylating) activity | 21 | 4 | 0.04 | 8.20E-08 |
| 18 | GO:0009055 | Electron transfer activity | 665 | 11 | 1.31 | 8.60E-08 |
| 19 | GO:0016628 | Oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor | 115 | 6 | 0.23 | 1.20E-07 |
| 20 | GO:0051537 | 2 iron, 2 sulfur cluster binding | 120 | 6 | 0.24 | 1.50E-07 |

## Appendix 3: Wheat GO Analysis – Ta_ECN_II

| Biological Process – Ta_ECN_II | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0015979 | Photosynthesis | 813 | 42 | 1.24 | < 1e-30 |
| 2 | GO:0019684 | Photosynthesis, light reaction | 511 | 19 | 0.78 | 1.60E-21 |
| 3 | GO:0006091 | Generation of precursor metabolites and energy | 1415 | 24 | 2.15 | 2.70E-19 |
| 4 | GO:0009765 | Photosynthesis, light harvesting | 105 | 9 | 0.16 | 6.60E-14 |
| 5 | GO:0009069 | Serine family amino acid metabolic process | 253 | 11 | 0.39 | 1.70E-13 |
| 6 | GO:0015995 | Chlorophyll biosynthetic process | 150 | 9 | 0.23 | 1.70E-12 |
| 7 | GO:0019344 | Cysteine biosynthetic process | 153 | 9 | 0.23 | 2.10E-12 |
| 8 | GO:0009416 | Response to light stimulus | 878 | 15 | 1.34 | 3.50E-12 |
| 9 | GO:0006534 | Cysteine metabolic process | 174 | 9 | 0.26 | 6.60E-12 |
| 10 | GO:0009768 | Photosynthesis, light harvesting in photosystem I | 64 | 7 | 0.1 | 7.90E-12 |
| 11 | GO:0009314 | Response to radiation | 988 | 15 | 1.5 | 1.90E-11 |
| 12 | GO:0009070 | Serine family amino acid biosynthetic process | 197 | 9 | 0.3 | 2.00E-11 |
| 13 | GO:0006779 | Porphyrin-containing compound biosynthetic process | 201 | 9 | 0.31 | 2.40E-11 |
| 14 | GO:0035304 | Regulation of protein dephosphorylation | 131 | 8 | 0.2 | 2.80E-11 |
| 15 | GO:0015994 | Chlorophyll metabolic process | 204 | 9 | 0.31 | 2.80E-11 |
| 16 | GO:0035303 | Regulation of dephosphorylation | 137 | 8 | 0.21 | 3.90E-11 |
| 17 | GO:0019253 | Reductive pentose-phosphate cycle | 41 | 6 | 0.06 | 4.20E-11 |
| 18 | GO:0019685 | Photosynthesis, dark reaction | 42 | 6 | 0.06 | 4.90E-11 |
| 19 | GO:0033014 | Tetrapyrrole biosynthetic process | 226 | 9 | 0.34 | 6.90E-11 |
| 20 | GO:0046148 | Pigment biosynthetic process | 327 | 10 | 0.5 | 7.50E-11 |

| Cellular Component – Ta_ECN_II | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0009579 | Thylakoid | 1067 | 44 | 1.71 | < 1e-30 |
| 2 | GO:0009534 | Chloroplast thylakoid | 833 | 38 | 1.34 | < 1e-30 |
| 3 | GO:0031976 | Plastid thylakoid | 835 | 38 | 1.34 | < 1e-30 |
| 4 | GO:0044436 | Thylakoid part | 846 | 38 | 1.36 | < 1e-30 |
| 5 | GO:0034357 | Photosynthetic membrane | 810 | 37 | 1.3 | < 1e-30 |
| 6 | GO:0044434 | Chloroplast part | 1977 | 46 | 3.18 | < 1e-30 |
| 7 | GO:0044435 | Plastid part | 2020 | 46 | 3.24 | < 1e-30 |
| 8 | GO:0009536 | Plastid | 8609 | 67 | 13.83 | < 1e-30 |
| 9 | GO:0009507 | Chloroplast | 3737 | 53 | 6 | < 1e-30 |
| 10 | GO:0042651 | Thylakoid membrane | 768 | 33 | 1.23 | < 1e-30 |
| 11 | GO:0009535 | Chloroplast thylakoid membrane | 697 | 31 | 1.12 | < 1e-30 |
| 12 | GO:0055035 | Plastid thylakoid membrane | 699 | 31 | 1.12 | < 1e-30 |
| 13 | GO:0031984 | Organelle subcompartment | 2011 | 39 | 3.23 | < 1e-30 |
| 14 | GO:0009521 | Photosystem | 325 | 20 | 0.52 | 1.90E-26 |
| 15 | GO:0044444 | Cytoplasmic part | 19525 | 71 | 31.36 | 6.00E-26 |
| 16 | GO:0005737 | Cytoplasm | 22168 | 71 | 35.61 | 5.00E-22 |
| 17 | GO:0009526 | Plastid envelope | 883 | 22 | 1.42 | 1.60E-20 |
| 18 | GO:0009522 | Photosystem I | 151 | 13 | 0.24 | 2.50E-19 |
| 19 | GO:0043231 | Intracellular membrane-bounded organelle | 24272 | 71 | 38.99 | 3.10E-19 |
| 20 | GO:0043227 | Membrane-bounded organelle | 24447 | 71 | 39.27 | 5.20E-19 |

| Molecular Function – Ta_ECN_II | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0031409 | Pigment binding | 64 | 7 | 0.07 | 6.10E-13 |
| 2 | GO:0048037 | Cofactor binding | 3182 | 20 | 3.39 | 3.00E-11 |
| 3 | GO:0016491 | Oxidoreductase activity | 5041 | 23 | 5.37 | 3.60E-10 |
| 4 | GO:0016168 | Chlorophyll binding | 160 | 7 | 0.17 | 4.20E-10 |
| 5 | GO:0043891 | Glyceraldehyde-3-phosphate dehydrogenase (NAD(P)+) (phosphorylating) activity | 21 | 4 | 0.02 | 6.80E-09 |
| 6 | GO:0016628 | Oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor | 115 | 5 | 0.12 | 1.60E-07 |
| 7 | GO:0042132 | Fructose 1,6-bisphosphate 1-phosphatase activity | 14 | 3 | 0.01 | 4.10E-07 |
| 8 | GO:0050308 | Sugar-phosphatase activity | 25 | 3 | 0.03 | 2.60E-06 |
| 9 | GO:0016627 | Oxidoreductase activity, acting on the CH-CH group of donors | 214 | 5 | 0.23 | 3.40E-06 |
| 10 | GO:0042578 | Phosphoric ester hydrolase activity | 917 | 8 | 0.98 | 5.50E-06 |
| 11 | GO:0045550 | Geranylgeranyl reductase activity | 5 | 2 | 0.01 | 1.10E-05 |
| 12 | GO:0019203 | Carbohydrate phosphatase activity | 52 | 3 | 0.06 | 2.40E-05 |
| 13 | GO:0046906 | Tetrapyrrole binding | 1489 | 9 | 1.59 | 2.50E-05 |
| 14 | GO:0016620 | Oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor | 162 | 4 | 0.17 | 2.80E-05 |
| 15 | GO:0004324 | Ferredoxin-NADP+ reductase activity | 9 | 2 | 0.01 | 4.00E-05 |
| 16 | GO:0052852 | Very-long-chain-(S)-2-hydroxy-acid oxidase activity | 10 | 2 | 0.01 | 5.00E-05 |
| 17 | GO:0052853 | Long-chain-(S)-2-hydroxy-long-chain-acid oxidase activity | 10 | 2 | 0.01 | 5.00E-05 |
| 18 | GO:0052854 | Medium-chain-(S)-2-hydroxy-acid oxidase activity | 10 | 2 | 0.01 | 5.00E-05 |
| 19 | GO:0051287 | NAD binding | 190 | 4 | 0.2 | 5.20E-05 |
| 20 | GO:0004096 | Catalase activity | 11 | 2 | 0.01 | 6.10E-05 |

## Appendix 4: Wheat GO Analysis – Ta_ECN_III

| Biological Process – Ta_ECN_III | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0015979 | Photosynthesis | 813 | 10 | 0.48 | 1.50E-11 |
| 2 | GO:0010207 | Photosystem II assembly | 168 | 6 | 0.1 | 5.60E-10 |
| 3 | GO:0006091 | Generation of precursor metabolites and energy | 1415 | 10 | 0.83 | 3.20E-09 |
| 4 | GO:0035304 | Regulation of protein dephosphorylation | 131 | 5 | 0.08 | 1.30E-08 |
| 5 | GO:0019684 | Photosynthesis, light reaction | 511 | 7 | 0.3 | 1.30E-08 |
| 6 | GO:0035303 | Regulation of dephosphorylation | 137 | 5 | 0.08 | 1.60E-08 |
| 7 | GO:0000023 | Maltose metabolic process | 118 | 4 | 0.07 | 6.70E-07 |
| 8 | GO:0019220 | Regulation of phosphate metabolic process | 300 | 5 | 0.18 | 7.80E-07 |
| 9 | GO:0051174 | Regulation of phosphorus metabolic process | 300 | 5 | 0.18 | 7.80E-07 |
| 10 | GO:0031399 | Regulation of protein modification process | 307 | 5 | 0.18 | 8.80E-07 |
| 11 | GO:0034622 | Cellular protein-containing complex assembly | 1531 | 8 | 0.9 | 1.70E-06 |
| 12 | GO:0065003 | Protein-containing complex assembly | 1591 | 8 | 0.93 | 2.20E-06 |
| 13 | GO:0043933 | Protein-containing complex subunit organization | 1686 | 8 | 0.99 | 3.40E-06 |
| 14 | GO:0009902 | Chloroplast relocation | 182 | 4 | 0.11 | 3.80E-06 |
| 15 | GO:0019750 | Chloroplast localization | 182 | 4 | 0.11 | 3.80E-06 |
| 16 | GO:0051644 | Plastid localization | 182 | 4 | 0.11 | 3.80E-06 |
| 17 | GO:0051667 | Establishment of plastid localization | 182 | 4 | 0.11 | 3.80E-06 |
| 18 | GO:0019252 | Starch biosynthetic process | 190 | 4 | 0.11 | 4.50E-06 |
| 19 | GO:0032790 | Ribosome disassembly | 6 | 2 | 0 | 4.90E-06 |
| 20 | GO:0016051 | Carbohydrate biosynthetic process | 828 | 6 | 0.48 | 6.70E-06 |

| Cellular Component – Ta_ECN_III | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0044434 | Chloroplast part | 1977 | 22 | 1.25 | 5.30E-25 |
| 2 | GO:0044435 | Plastid part | 2020 | 22 | 1.28 | 8.50E-25 |
| 3 | GO:0009579 | Thylakoid | 1067 | 19 | 0.68 | 9.00E-25 |
| 4 | GO:0009507 | Chloroplast | 3737 | 22 | 2.37 | 5.40E-19 |
| 5 | GO:0009535 | Chloroplast thylakoid membrane | 697 | 14 | 0.44 | 1.70E-18 |
| 6 | GO:0055035 | Plastid thylakoid membrane | 699 | 14 | 0.44 | 1.80E-18 |
| 7 | GO:0042651 | Thylakoid membrane | 768 | 14 | 0.49 | 6.50E-18 |
| 8 | GO:0034357 | Photosynthetic membrane | 810 | 14 | 0.51 | 1.40E-17 |
| 9 | GO:0009534 | Chloroplast thylakoid | 833 | 14 | 0.53 | 2.00E-17 |
| 10 | GO:0031976 | Plastid thylakoid | 835 | 14 | 0.53 | 2.10E-17 |
| 11 | GO:0044436 | Thylakoid part | 846 | 14 | 0.54 | 2.50E-17 |
| 12 | GO:0009536 | Plastid | 8609 | 26 | 5.45 | 8.20E-17 |
| 13 | GO:0044446 | Intracellular organelle part | 8349 | 24 | 5.29 | 3.80E-14 |
| 14 | GO:0044422 | Organelle part | 8378 | 24 | 5.31 | 4.10E-14 |
| 15 | GO:0031984 | Organelle subcompartment | 2011 | 14 | 1.27 | 3.40E-12 |
| 16 | GO:0044444 | Cytoplasmic part | 19525 | 28 | 12.37 | 1.10E-10 |
| 17 | GO:0005737 | Cytoplasm | 22168 | 28 | 14.04 | 4.00E-09 |
| 18 | GO:0009941 | Chloroplast envelope | 842 | 8 | 0.53 | 3.70E-08 |
| 19 | GO:0043231 | Intracellular membrane-bounded organelle | 24272 | 28 | 15.38 | 5.10E-08 |
| 20 | GO:0009526 | Plastid envelope | 883 | 8 | 0.56 | 5.40E-08 |

| Molecular Function – Ta_ECN_III | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0004324 | Ferredoxin-NADP+ reductase activity | 9 | 2 | 0 | 5.70E-06 |
| 2 | GO:0016731 | Oxidoreductase activity, acting on iron-sulfur proteins as donors, NAD or NADP as acceptor | 11 | 2 | 0 | 8.80E-06 |
| 3 | GO:0004849 | Uridine kinase activity | 18 | 2 | 0.01 | 2.40E-05 |
| 4 | GO:0016730 | Oxidoreductase activity, acting on iron-sulfur proteins as donors | 18 | 2 | 0.01 | 2.40E-05 |
| 5 | GO:0019206 | Nucleoside kinase activity | 27 | 2 | 0.01 | 5.60E-05 |
| 6 | GO:0016655 | Oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor | 201 | 3 | 0.08 | 7.50E-05 |
| 7 | GO:0016651 | Oxidoreductase activity, acting on NAD(P)H | 342 | 3 | 0.14 | 0.00036 |
| 8 | GO:0019205 | Nucleobase-containing compound kinase activity | 102 | 2 | 0.04 | 0.0008 |
| 9 | GO:0048038 | Quinone binding | 102 | 2 | 0.04 | 0.0008 |
| 10 | GO:0008483 | Transaminase activity | 144 | 2 | 0.06 | 0.00159 |
| 11 | GO:0016769 | Transferase activity, transferring nitrogenous groups | 144 | 2 | 0.06 | 0.00159 |
| 12 | GO:0003746 | Translation elongation factor activity | 150 | 2 | 0.06 | 0.00172 |
| 13 | GO:0051741 | 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase activity | 7 | 1 | 0 | 0.00287 |
| 14 | GO:0016491 | Oxidoreductase activity | 5041 | 7 | 2.07 | 0.00288 |
| 15 | GO:0015114 | Phosphate ion transmembrane transporter activity | 8 | 1 | 0 | 0.00328 |
| 16 | GO:0030170 | Pyridoxal phosphate binding | 283 | 2 | 0.12 | 0.00595 |
| 17 | GO:0070279 | Vitamin B6 binding | 283 | 2 | 0.12 | 0.00595 |
| 18 | GO:0004619 | Phosphoglycerate mutase activity | 15 | 1 | 0.01 | 0.00613 |
| 19 | GO:0048037 | Cofactor binding | 3182 | 5 | 1.3 | 0.00799 |
| 20 | GO:0045158 | Electron transporter, transferring electrons within cytochrome b6/f complex of photosystem II activity | 22 | 1 | 0.01 | 0.00898 |

## Appendix 5: Wheat GO Analysis – Ta_ECN_V

| Biological Process – Ta_ECN_V | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0006412 | Translation | 2294 | 11 | 0.59 | 1.10E-14 |
| 2 | GO:0043043 | Peptide biosynthetic process | 2305 | 11 | 0.59 | 1.10E-14 |
| 3 | GO:0043604 | Amide biosynthetic process | 2439 | 11 | 0.63 | 2.10E-14 |
| 4 | GO:0006518 | Peptide metabolic process | 2668 | 11 | 0.69 | 5.60E-14 |
| 5 | GO:0043603 | Cellular amide metabolic process | 2958 | 11 | 0.76 | 1.70E-13 |
| 6 | GO:1901566 | Organonitrogen compound biosynthetic process | 4688 | 11 | 1.21 | 2.80E-11 |
| 7 | GO:0034645 | Cellular macromolecule biosynthetic process | 9336 | 11 | 2.4 | 5.40E-08 |
| 8 | GO:0009059 | Macromolecule biosynthetic process | 9513 | 11 | 2.45 | 6.70E-08 |
| 9 | GO:0044271 | Cellular nitrogen compound biosynthetic process | 9573 | 11 | 2.47 | 7.20E-08 |
| 10 | GO:0010467 | Gene expression | 9704 | 11 | 2.5 | 8.30E-08 |
| 11 | GO:0002181 | Cytoplasmic translation | 222 | 4 | 0.06 | 2.30E-07 |
| 12 | GO:0044267 | Cellular protein metabolic process | 11875 | 11 | 3.06 | 7.70E-07 |
| 13 | GO:0042254 | Ribosome biogenesis | 851 | 5 | 0.22 | 1.30E-06 |
| 14 | GO:0044249 | Cellular biosynthetic process | 12683 | 11 | 3.27 | 1.60E-06 |
| 15 | GO:0019538 | Protein metabolic process | 13401 | 11 | 3.45 | 2.90E-06 |
| 16 | GO:1901576 | Organic substance biosynthetic process | 13435 | 11 | 3.46 | 3.00E-06 |
| 17 | GO:0022613 | Ribonucleoprotein complex biogenesis | 1091 | 5 | 0.28 | 4.40E-06 |
| 18 | GO:0009058 | Biosynthetic process | 14157 | 11 | 3.65 | 5.30E-06 |
| 19 | GO:0034641 | Cellular nitrogen compound metabolic process | 14791 | 11 | 3.81 | 8.60E-06 |
| 20 | GO:1901564 | Organonitrogen compound metabolic process | 16308 | 11 | 4.2 | 2.50E-05 |

| Cellular Component – Ta_ECN_V | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0022626 | Cytosolic ribosome | 627 | 11 | 0.16 | 4.30E-21 |
| 2 | GO:0044445 | Cytosolic part | 728 | 11 | 0.18 | 2.20E-20 |
| 3 | GO:0044391 | Ribosomal subunit | 791 | 11 | 0.2 | 5.60E-20 |
| 4 | GO:0005840 | Ribosome | 1241 | 11 | 0.31 | 8.20E-18 |
| 5 | GO:1990904 | Ribonucleoprotein complex | 2116 | 11 | 0.53 | 3.00E-15 |
| 6 | GO:0005829 | Cytosol | 2695 | 11 | 0.67 | 4.20E-14 |
| 7 | GO:0043228 | Non-membrane-bounded organelle | 3549 | 11 | 0.88 | 8.80E-13 |
| 8 | GO:0043232 | Intracellular non-membrane-bounded organelle | 3549 | 11 | 0.88 | 8.80E-13 |
| 9 | GO:0022625 | Cytosolic large ribosomal subunit | 387 | 6 | 0.1 | 1.90E-10 |
| 10 | GO:0032991 | Protein-containing complex | 6582 | 11 | 1.64 | 7.90E-10 |
| 11 | GO:0022627 | Cytosolic small ribosomal subunit | 201 | 5 | 0.05 | 8.40E-10 |
| 12 | GO:0015934 | Large ribosomal subunit | 496 | 6 | 0.12 | 8.50E-10 |
| 13 | GO:0015935 | Small ribosomal subunit | 293 | 5 | 0.07 | 5.50E-09 |
| 14 | GO:0044446 | Intracellular organelle part | 8349 | 11 | 2.08 | 1.10E-08 |
| 15 | GO:0044422 | Organelle part | 8378 | 11 | 2.09 | 1.10E-08 |
| 16 | GO:0044444 | Cytoplasmic part | 19525 | 11 | 4.86 | 0.00012 |
| 17 | GO:0005737 | Cytoplasm | 22168 | 11 | 5.52 | 0.0005 |
| 18 | GO:0043229 | Intracellular organelle | 25553 | 11 | 6.36 | 0.00241 |
| 19 | GO:0043226 | Organelle | 25579 | 11 | 6.37 | 0.00244 |
| 20 | GO:0044424 | Intracellular part | 28808 | 11 | 7.17 | 0.00901 |

| Molecular Function – Ta_ECN_V | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0003735 | Structural constituent of ribosome | 1223 | 11 | 0.28 | 2.40E-18 |
| 2 | GO:0005198 | Structural molecule activity | 1421 | 11 | 0.32 | 1.20E-17 |
| 3 | GO:0003723 | RNA binding | 2699 | 4 | 0.61 | 0.0023 |
| 4 | GO:0003676 | Nucleic acid binding | 11391 | 4 | 2.57 | 0.2424 |
| 5 | GO:1901363 | Heterocyclic compound binding | 23868 | 4 | 5.38 | 0.8723 |
| 6 | GO:0097159 | Organic cyclic compound binding | 23870 | 4 | 5.38 | 0.8724 |
| 7 | GO:0005488 | Binding | 35771 | 4 | 8.07 | 0.9981 |
| 8 | GO:0003674 | Molecular_function | 48782 | 11 | 11 | 1 |
| 9 | GO:0000014 | Single-stranded DNA endodeoxyribonuclease activity | 12 | 0 | 0 | 1 |
| 10 | GO:0000030 | Mannosyltransferase activity | 75 | 0 | 0.02 | 1 |
| 11 | GO:0000049 | tRNA binding | 60 | 0 | 0.01 | 1 |
| 12 | GO:0000062 | Fatty-acyl-CoA binding | 32 | 0 | 0.01 | 1 |
| 13 | GO:0000104 | Succinate dehydrogenase activity | 8 | 0 | 0 | 1 |
| 14 | GO:0000107 | Imidazoleglycerol-phosphate synthase activity | 5 | 0 | 0 | 1 |
| 15 | GO:0000149 | SNARE binding | 126 | 0 | 0.03 | 1 |
| 16 | GO:0000150 | Recombinase activity | 26 | 0 | 0.01 | 1 |
| 17 | GO:0000155 | Phosphorelay sensor kinase activity | 59 | 0 | 0.01 | 1 |
| 18 | GO:0000156 | Phosphorelay response regulator activity | 11 | 0 | 0 | 1 |
| 19 | GO:0000166 | Nucleotide binding | 12967 | 0 | 2.92 | 1 |
| 20 | GO:0000171 | Ribonuclease MRP activity | 6 | 0 | 0 | 1 |

## Appendix 6: Wheat GO Analysis – Ta_ECN_VI

| Biological Process – Ta_ECN_VI | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0019288 | Isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway | 264 | 4 | 0.03 | 7.10E-09 |
| 2 | GO:0009240 | Isopentenyl diphosphate biosynthetic process | 278 | 4 | 0.03 | 8.70E-09 |
| 3 | GO:0046490 | Isopentenyl diphosphate metabolic process | 278 | 4 | 0.03 | 8.70E-09 |
| 4 | GO:0019682 | Glyceraldehyde-3-phosphate metabolic process | 329 | 4 | 0.04 | 1.70E-08 |
| 5 | GO:0006081 | Cellular aldehyde metabolic process | 458 | 4 | 0.05 | 6.50E-08 |
| 6 | GO:0006090 | Pyruvate metabolic process | 576 | 4 | 0.07 | 1.60E-07 |
| 7 | GO:0008654 | Phospholipid biosynthetic process | 583 | 4 | 0.07 | 1.70E-07 |
| 8 | GO:0008299 | Isoprenoid biosynthetic process | 587 | 4 | 0.07 | 1.70E-07 |
| 9 | GO:0006720 | Isoprenoid metabolic process | 666 | 4 | 0.08 | 2.90E-07 |
| 10 | GO:0006644 | Phospholipid metabolic process | 798 | 4 | 0.09 | 6.00E-07 |
| 11 | GO:0008610 | Lipid biosynthetic process | 1504 | 4 | 0.18 | 7.40E-06 |
| 12 | GO:0090407 | Organophosphate biosynthetic process | 1507 | 4 | 0.18 | 7.50E-06 |
| 13 | GO:0044255 | Cellular lipid metabolic process | 1877 | 4 | 0.22 | 1.80E-05 |
| 14 | GO:0032787 | Monocarboxylic acid metabolic process | 1996 | 4 | 0.23 | 2.30E-05 |
| 15 | GO:0019637 | Organophosphate metabolic process | 2106 | 4 | 0.25 | 2.80E-05 |
| 16 | GO:0006412 | Translation | 2294 | 4 | 0.27 | 4.00E-05 |
| 17 | GO:0043043 | Peptide biosynthetic process | 2305 | 4 | 0.27 | 4.10E-05 |
| 18 | GO:1901135 | Carbohydrate derivative metabolic process | 2365 | 4 | 0.28 | 4.50E-05 |
| 19 | GO:0043604 | Amide biosynthetic process | 2439 | 4 | 0.29 | 5.10E-05 |
| 20 | GO:0006629 | Lipid metabolic process | 2471 | 4 | 0.29 | 5.30E-05 |

| Cellular Component – Ta_ECN_VI | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0009536 | Plastid | 8609 | 10 | 1.95 | 7.80E-08 |
| 2 | GO:0019013 | Viral nucleocapsid | 185 | 3 | 0.04 | 8.50E-06 |
| 3 | GO:0019028 | Viral capsid | 191 | 3 | 0.04 | 9.30E-06 |
| 4 | GO:0019012 | Virion | 207 | 3 | 0.05 | 1.20E-05 |
| 5 | GO:0044423 | Virion part | 207 | 3 | 0.05 | 1.20E-05 |
| 6 | GO:0009941 | Chloroplast envelope | 842 | 4 | 0.19 | 2.50E-05 |
| 7 | GO:0009526 | Plastid envelope | 883 | 4 | 0.2 | 3.00E-05 |
| 8 | GO:0005840 | Ribosome | 1241 | 4 | 0.28 | 0.00011 |
| 9 | GO:0044444 | Cytoplasmic part | 19525 | 10 | 4.42 | 0.00028 |
| 10 | GO:0031967 | Organelle envelope | 1598 | 4 | 0.36 | 0.0003 |
| 11 | GO:0031975 | Envelope | 1603 | 4 | 0.36 | 0.0003 |
| 12 | GO:0044434 | Chloroplast part | 1977 | 4 | 0.45 | 0.00067 |
| 13 | GO:0044435 | Plastid part | 2020 | 4 | 0.46 | 0.00073 |
| 14 | GO:0009507 | Chloroplast | 3737 | 5 | 0.85 | 0.00076 |
| 15 | GO:1990904 | Ribonucleoprotein complex | 2116 | 4 | 0.48 | 0.00087 |
| 16 | GO:0005737 | Cytoplasm | 22168 | 10 | 5.02 | 0.00101 |
| 17 | GO:0043231 | Intracellular membrane-bounded organelle | 24272 | 10 | 5.49 | 0.00249 |
| 18 | GO:0043227 | Membrane-bounded organelle | 24447 | 10 | 5.53 | 0.00268 |
| 19 | GO:0043229 | Intracellular organelle | 25553 | 10 | 5.78 | 0.00417 |
| 20 | GO:0043226 | Organelle | 25579 | 10 | 5.79 | 0.00421 |

| Molecular Function – Ta_ECN_VI | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0033926 | Glycopeptide alpha-N-acetylgalactosaminidase activity | 18 | 2 | 0 | 5.80E-06 |
| 2 | GO:0140103 | Catalytic activity, acting on a glycoprotein | 35 | 2 | 0.01 | 2.20E-05 |
| 3 | GO:0003735 | Structural constituent of ribosome | 1223 | 4 | 0.25 | 7.30E-05 |
| 4 | GO:0005198 | Structural molecule activity | 1421 | 4 | 0.29 | 0.00013 |
| 5 | GO:0019843 | rRNA binding | 293 | 2 | 0.06 | 0.00157 |
| 6 | GO:0051920 | Peroxiredoxin activity | 15 | 1 | 0 | 0.00307 |
| 7 | GO:0003676 | Nucleic acid binding | 11391 | 6 | 2.34 | 0.01404 |
| 8 | GO:0004553 | Hydrolase activity, hydrolyzing O-glycosyl compounds | 1145 | 2 | 0.23 | 0.02186 |
| 9 | GO:0016798 | Hydrolase activity, acting on glycosyl bonds | 1314 | 2 | 0.27 | 0.02826 |
| 10 | GO:0004601 | Peroxidase activity | 425 | 1 | 0.09 | 0.08379 |
| 11 | GO:0016684 | Oxidoreductase activity, acting on peroxide as acceptor | 432 | 1 | 0.09 | 0.08512 |
| 12 | GO:0016209 | Antioxidant activity | 483 | 1 | 0.1 | 0.09472 |
| 13 | GO:0003723 | RNA binding | 2699 | 2 | 0.55 | 0.1025 |
| 14 | GO:0003700 | DNA-binding transcription factor activity | 1801 | 1 | 0.37 | 0.31355 |
| 15 | GO:0140110 | Transcription regulator activity | 1948 | 1 | 0.4 | 0.33473 |
| 16 | GO:1901363 | Heterocyclic compound binding | 23868 | 6 | 4.89 | 0.35085 |
| 17 | GO:0097159 | Organic cyclic compound binding | 23870 | 6 | 4.89 | 0.35095 |
| 18 | GO:0000166 | Nucleotide binding | 12967 | 3 | 2.66 | 0.52134 |
| 19 | GO:1901265 | Nucleoside phosphate binding | 12967 | 3 | 2.66 | 0.52134 |
| 20 | GO:0036094 | Small molecule binding | 13505 | 3 | 2.77 | 0.55321 |

## Appendix 7: Wheat GO Analysis – Ta_ECN_VIII

| Biological Process – Ta_ECN_VIII | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0019253 | Reductive pentose-phosphate cycle | 41 | 3 | 0.01 | 2.90E-08 |
| 2 | GO:0019685 | Photosynthesis, dark reaction | 42 | 3 | 0.01 | 3.10E-08 |
| 3 | GO:0015977 | Carbon fixation | 61 | 3 | 0.01 | 9.70E-08 |
| 4 | GO:0019464 | Glycine decarboxylation via glycine cleavage system | 8 | 2 | 0 | 6.40E-07 |
| 5 | GO:0009853 | Photorespiration | 130 | 3 | 0.02 | 9.60E-07 |
| 6 | GO:0043094 | Cellular metabolic compound salvage | 232 | 3 | 0.04 | 5.50E-06 |
| 7 | GO:0015976 | Carbon utilization | 23 | 2 | 0 | 5.80E-06 |
| 8 | GO:0006546 | Glycine catabolic process | 36 | 2 | 0.01 | 1.40E-05 |
| 9 | GO:0042135 | Neurotransmitter catabolic process | 38 | 2 | 0.01 | 1.60E-05 |
| 10 | GO:0009071 | Serine family amino acid catabolic process | 41 | 2 | 0.01 | 1.90E-05 |
| 11 | GO:0006544 | Glycine metabolic process | 55 | 2 | 0.01 | 3.40E-05 |
| 12 | GO:0042133 | Neurotransmitter metabolic process | 75 | 2 | 0.01 | 6.40E-05 |
| 13 | GO:0001505 | Regulation of neurotransmitter levels | 77 | 2 | 0.01 | 6.70E-05 |
| 14 | GO:0015979 | Photosynthesis | 813 | 3 | 0.13 | 0.00023 |
| 15 | GO:0016051 | Carbohydrate biosynthetic process | 828 | 3 | 0.14 | 0.00024 |
| 16 | GO:1901606 | Alpha-amino acid catabolic process | 166 | 2 | 0.03 | 0.00031 |
| 17 | GO:0009063 | Cellular amino acid catabolic process | 176 | 2 | 0.03 | 0.00035 |
| 18 | GO:0009069 | Serine family amino acid metabolic process | 253 | 2 | 0.04 | 0.00072 |
| 19 | GO:0016054 | Organic acid catabolic process | 332 | 2 | 0.05 | 0.00123 |
| 20 | GO:0046395 | Carboxylic acid catabolic process | 332 | 2 | 0.05 | 0.00123 |

| Cellular Component – Ta_ECN_VIII | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0005960 | Glycine cleavage complex | 6 | 2 | 0 | 3.20E-07 |
| 2 | GO:0009507 | Chloroplast | 3737 | 5 | 0.59 | 7.80E-05 |
| 3 | GO:1990204 | Oxidoreductase complex | 258 | 2 | 0.04 | 0.0007 |
| 4 | GO:0044444 | Cytoplasmic part | 19525 | 7 | 3.09 | 0.0033 |
| 5 | GO:0009536 | Plastid | 8609 | 5 | 1.36 | 0.0041 |
| 6 | GO:0009570 | Chloroplast stroma | 838 | 2 | 0.13 | 0.0071 |
| 7 | GO:0009532 | Plastid stroma | 874 | 2 | 0.14 | 0.0077 |
| 8 | GO:0005737 | Cytoplasm | 22168 | 7 | 3.51 | 0.008 |
| 9 | GO:0043231 | Intracellular membrane-bounded organelle | 24272 | 7 | 3.84 | 0.0151 |
| 10 | GO:0043227 | Membrane-bounded organelle | 24447 | 7 | 3.87 | 0.0158 |
| 11 | GO:0043229 | Intracellular organelle | 25553 | 7 | 4.05 | 0.0216 |
| 12 | GO:0043226 | Organelle | 25579 | 7 | 4.05 | 0.0217 |
| 13 | GO:0044434 | Chloroplast part | 1977 | 2 | 0.31 | 0.0361 |
| 14 | GO:1902494 | Catalytic complex | 1980 | 2 | 0.31 | 0.0362 |
| 15 | GO:0044435 | Plastid part | 2020 | 2 | 0.32 | 0.0376 |
| 16 | GO:0044424 | Intracellular part | 28808 | 7 | 4.56 | 0.0499 |
| 17 | GO:0005622 | Intracellular | 29512 | 7 | 4.67 | 0.0591 |
| 18 | GO:0044464 | Cell part | 32576 | 7 | 5.16 | 0.1181 |
| 19 | GO:0005623 | Cell | 32702 | 7 | 5.18 | 0.1213 |
| 20 | GO:0032991 | Protein-containing complex | 6582 | 2 | 1.04 | 0.2804 |

| Molecular Function – Ta_ECN_VIII | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0016984 | Ribulose-bisphosphate carboxylase activity | 26 | 3 | 0 | 1.30E-09 |
| 2 | GO:0016829 | Lyase activity | 1039 | 5 | 0.11 | 4.30E-09 |
| 3 | GO:0016831 | Carboxy-lyase activity | 263 | 3 | 0.03 | 1.50E-06 |
| 4 | GO:0016830 | Carbon-carbon lyase activity | 353 | 3 | 0.04 | 3.70E-06 |
| 5 | GO:0004089 | Carbonate dehydratase activity | 40 | 2 | 0 | 6.50E-06 |
| 6 | GO:0004497 | Monooxygenase activity | 1110 | 3 | 0.11 | 0.00011 |
| 7 | GO:0016836 | Hydro-lyase activity | 228 | 2 | 0.02 | 0.00022 |
| 8 | GO:0016835 | Carbon-oxygen lyase activity | 390 | 2 | 0.04 | 0.00063 |
| 9 | GO:0016491 | Oxidoreductase activity | 5041 | 3 | 0.52 | 0.00939 |
| 10 | GO:0008270 | Zinc ion binding | 4250 | 2 | 0.44 | 0.06351 |
| 11 | GO:0003824 | Catalytic activity | 29558 | 5 | 3.03 | 0.08166 |
| 12 | GO:0046914 | Transition metal ion binding | 6136 | 2 | 0.63 | 0.12203 |
| 13 | GO:0046872 | Metal ion binding | 10274 | 2 | 1.05 | 0.28458 |
| 14 | GO:0043169 | Cation binding | 10327 | 2 | 1.06 | 0.28683 |
| 15 | GO:0043167 | Ion binding | 20930 | 2 | 2.15 | 0.71138 |
| 16 | GO:0005488 | Binding | 35771 | 2 | 3.67 | 0.9801 |
| 17 | GO:0003674 | Molecular_function | 48782 | 5 | 5 | 1 |
| 18 | GO:0000014 | Single-stranded DNA endodeoxyribonuclease activity | 12 | 0 | 0 | 1 |
| 19 | GO:0000030 | Mannosyltransferase activity | 75 | 0 | 0.01 | 1 |
| 20 | GO:0000049 | tRNA binding | 60 | 0 | 0.01 | 1 |

## Appendix 8: Wheat GO Analysis – Ta_ECN_IX

| Biological Process – Ta_ECN_IX | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0019344 | Cysteine biosynthetic process | 153 | 3 | 0.02 | 4.50E-07 |
| 2 | GO:0006534 | Cysteine metabolic process | 174 | 3 | 0.02 | 6.60E-07 |
| 3 | GO:0009070 | Serine family amino acid biosynthetic process | 197 | 3 | 0.02 | 9.60E-07 |
| 4 | GO:0000097 | Sulfur amino acid biosynthetic process | 246 | 3 | 0.03 | 1.90E-06 |
| 5 | GO:0009069 | Serine family amino acid metabolic process | 253 | 3 | 0.03 | 2.00E-06 |
| 6 | GO:0019288 | Isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway | 264 | 3 | 0.03 | 2.30E-06 |
| 7 | GO:0009240 | Isopentenyl diphosphate biosynthetic process | 278 | 3 | 0.03 | 2.70E-06 |
| 8 | GO:0046490 | Isopentenyl diphosphate metabolic process | 278 | 3 | 0.03 | 2.70E-06 |
| 9 | GO:0000096 | Sulfur amino acid metabolic process | 279 | 3 | 0.03 | 2.70E-06 |
| 10 | GO:0019682 | Glyceraldehyde-3-phosphate metabolic process | 329 | 3 | 0.04 | 4.50E-06 |
| 11 | GO:0044272 | Sulfur compound biosynthetic process | 432 | 3 | 0.05 | 1.00E-05 |
| 12 | GO:0006081 | Cellular aldehyde metabolic process | 458 | 3 | 0.05 | 1.20E-05 |
| 13 | GO:1901566 | Organonitrogen compound biosynthetic process | 4688 | 5 | 0.55 | 1.60E-05 |
| 14 | GO:1901607 | Alpha-amino acid biosynthetic process | 548 | 3 | 0.06 | 2.10E-05 |
| 15 | GO:0006090 | Pyruvate metabolic process | 576 | 3 | 0.07 | 2.40E-05 |
| 16 | GO:0008654 | Phospholipid biosynthetic process | 583 | 3 | 0.07 | 2.50E-05 |
| 17 | GO:0008299 | Isoprenoid biosynthetic process | 587 | 3 | 0.07 | 2.50E-05 |
| 18 | GO:0008652 | Cellular amino acid biosynthetic process | 623 | 3 | 0.07 | 3.00E-05 |
| 19 | GO:0006720 | Isoprenoid metabolic process | 666 | 3 | 0.08 | 3.70E-05 |
| 20 | GO:0006412 | Translation | 2294 | 4 | 0.27 | 4.00E-05 |

| Cellular Component – Ta_ECN_IX | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0000311 | Plastid large ribosomal subunit | 6 | 2 | 0 | 3.20E-07 |
| 2 | GO:0009547 | Plastid ribosome | 11 | 2 | 0 | 1.20E-06 |
| 3 | GO:0009536 | Plastid | 8609 | 7 | 1.36 | 1.10E-05 |
| 4 | GO:0005840 | Ribosome | 1241 | 4 | 0.2 | 2.00E-05 |
| 5 | GO:0015934 | Large ribosomal subunit | 496 | 3 | 0.08 | 4.80E-05 |
| 6 | GO:0000315 | Organellar large ribosomal subunit | 76 | 2 | 0.01 | 6.10E-05 |
| 7 | GO:0022626 | Cytosolic ribosome | 627 | 3 | 0.1 | 9.50E-05 |
| 8 | GO:0044445 | Cytosolic part | 728 | 3 | 0.12 | 0.00015 |
| 9 | GO:0000313 | Organellar ribosome | 119 | 2 | 0.02 | 0.00015 |
| 10 | GO:1990904 | Ribonucleoprotein complex | 2116 | 4 | 0.34 | 0.00016 |
| 11 | GO:0044391 | Ribosomal subunit | 791 | 3 | 0.13 | 0.00019 |
| 12 | GO:0009532 | Plastid stroma | 874 | 3 | 0.14 | 0.00025 |
| 13 | GO:0019013 | Viral nucleocapsid | 185 | 2 | 0.03 | 0.00036 |
| 14 | GO:0019028 | Viral capsid | 191 | 2 | 0.03 | 0.00038 |
| 15 | GO:0019012 | Virion | 207 | 2 | 0.03 | 0.00045 |
| 16 | GO:0044423 | Virion part | 207 | 2 | 0.03 | 0.00045 |
| 17 | GO:0043228 | Non-membrane-bounded organelle | 3549 | 4 | 0.56 | 0.00119 |
| 18 | GO:0043232 | Intracellular non-membrane-bounded organelle | 3549 | 4 | 0.56 | 0.00119 |
| 19 | GO:0044435 | Plastid part | 2020 | 3 | 0.32 | 0.0029 |
| 20 | GO:0044444 | Cytoplasmic part | 19525 | 7 | 3.09 | 0.00328 |

| Molecular Function – Ta_ECN_IX | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0003735 | Structural constituent of ribosome | 1223 | 4 | 0.18 | 1.30E-05 |
| 2 | GO:0005198 | Structural molecule activity | 1421 | 4 | 0.2 | 2.30E-05 |
| 3 | GO:0016851 | Magnesium chelatase activity | 7 | 1 | 0 | 0.001 |
| 4 | GO:0051002 | Ligase activity, forming nitrogen-metal bonds | 7 | 1 | 0 | 0.001 |
| 5 | GO:0051003 | Ligase activity, forming nitrogen-metal bonds, forming coordination complexes | 7 | 1 | 0 | 0.001 |
| 6 | GO:0003723 | RNA binding | 2699 | 3 | 0.39 | 0.005 |
| 7 | GO:0003676 | Nucleic acid binding | 11391 | 5 | 1.63 | 0.0095 |
| 8 | GO:0019843 | rRNA binding | 293 | 1 | 0.04 | 0.0413 |
| 9 | GO:1901363 | Heterocyclic compound binding | 23868 | 6 | 3.42 | 0.0557 |
| 10 | GO:0097159 | Organic cyclic compound binding | 23870 | 6 | 3.43 | 0.0558 |
| 11 | GO:0016874 | Ligase activity | 1407 | 1 | 0.2 | 0.1853 |
| 12 | GO:0000166 | Nucleotide binding | 12967 | 3 | 1.86 | 0.2771 |
| 13 | GO:1901265 | Nucleoside phosphate binding | 12967 | 3 | 1.86 | 0.2771 |
| 14 | GO:0036094 | Small molecule binding | 13505 | 3 | 1.94 | 0.3011 |
| 15 | GO:0005488 | Binding | 35771 | 6 | 5.13 | 0.4042 |
| 16 | GO:0005524 | ATP binding | 8472 | 1 | 1.22 | 0.737 |
| 17 | GO:0008144 | Drug binding | 8905 | 1 | 1.28 | 0.7561 |
| 18 | GO:0035639 | Purine ribonucleoside triphosphate binding | 9051 | 1 | 1.3 | 0.7623 |
| 19 | GO:0032559 | Adenyl ribonucleotide binding | 10516 | 1 | 1.51 | 0.8173 |
| 20 | GO:0030554 | Adenyl nucleotide binding | 10522 | 1 | 1.51 | 0.8175 |

## Appendix 9: Wheat GO Analysis – Ta_ECN_X

| Biological Process – Ta_ECN_X | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0006412 | Translation | 2294 | 3 | 0.16 | 0.00015 |
| 2 | GO:0043043 | Peptide biosynthetic process | 2305 | 3 | 0.16 | 0.00016 |
| 3 | GO:0043604 | Amide biosynthetic process | 2439 | 3 | 0.17 | 0.00019 |
| 4 | GO:0006518 | Peptide metabolic process | 2668 | 3 | 0.19 | 0.00024 |
| 5 | GO:0043603 | Cellular amide metabolic process | 2958 | 3 | 0.21 | 0.00033 |
| 6 | GO:1901566 | Organonitrogen compound biosynthetic process | 4688 | 3 | 0.33 | 0.00132 |
| 7 | GO:0034645 | Cellular macromolecule biosynthetic process | 9336 | 3 | 0.66 | 0.01045 |
| 8 | GO:0009059 | Macromolecule biosynthetic process | 9513 | 3 | 0.67 | 0.01105 |
| 9 | GO:0044271 | Cellular nitrogen compound biosynthetic process | 9573 | 3 | 0.67 | 0.01126 |
| 10 | GO:0010467 | Gene expression | 9704 | 3 | 0.68 | 0.01173 |
| 11 | GO:0044267 | Cellular protein metabolic process | 11875 | 3 | 0.83 | 0.0215 |
| 12 | GO:0044249 | Cellular biosynthetic process | 12683 | 3 | 0.89 | 0.0262 |
| 13 | GO:0019538 | Protein metabolic process | 13401 | 3 | 0.94 | 0.0309 |
| 14 | GO:1901576 | Organic substance biosynthetic process | 13435 | 3 | 0.94 | 0.03114 |
| 15 | GO:0009058 | Biosynthetic process | 14157 | 3 | 0.99 | 0.03643 |
| 16 | GO:0034641 | Cellular nitrogen compound metabolic process | 14791 | 3 | 1.04 | 0.04155 |
| 17 | GO:1901564 | Organonitrogen compound metabolic process | 16308 | 3 | 1.15 | 0.05569 |
| 18 | GO:0044260 | Cellular macromolecule metabolic process | 20000 | 3 | 1.41 | 0.10273 |
| 19 | GO:0043170 | Macromolecule metabolic process | 23321 | 3 | 1.64 | 0.16287 |
| 20 | GO:0006807 | Nitrogen compound metabolic process | 25126 | 3 | 1.77 | 0.20369 |

| Cellular Component – Ta_ECN_X | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0019013 | Viral nucleocapsid | 185 | 2 | 0.02 | 0.00017 |
| 2 | GO:0019028 | Viral capsid | 191 | 2 | 0.02 | 0.00018 |
| 3 | GO:0005840 | Ribosome | 1241 | 3 | 0.14 | 0.00021 |
| 4 | GO:0019012 | Virion | 207 | 2 | 0.02 | 0.00022 |
| 5 | GO:0044423 | Virion part | 207 | 2 | 0.02 | 0.00022 |
| 6 | GO:1990904 | Ribonucleoprotein complex | 2116 | 3 | 0.24 | 0.00102 |
| 7 | GO:0044391 | Ribosomal subunit | 791 | 2 | 0.09 | 0.00309 |
| 8 | GO:0043228 | Non-membrane-bounded organelle | 3549 | 3 | 0.4 | 0.00457 |
| 9 | GO:0043232 | Intracellular non-membrane-bounded organelle | 3549 | 3 | 0.4 | 0.00457 |
| 10 | GO:0044444 | Cytoplasmic part | 19525 | 5 | 2.21 | 0.01682 |
| 11 | GO:0032991 | Protein-containing complex | 6582 | 3 | 0.74 | 0.02608 |
| 12 | GO:0005737 | Cytoplasm | 22168 | 5 | 2.51 | 0.03173 |
| 13 | GO:0009536 | Plastid | 8609 | 3 | 0.97 | 0.05397 |
| 14 | GO:0043229 | Intracellular organelle | 25553 | 5 | 2.89 | 0.06457 |
| 15 | GO:0043226 | Organelle | 25579 | 5 | 2.89 | 0.0649 |
| 16 | GO:0044424 | Intracellular part | 28808 | 5 | 3.26 | 0.1176 |
| 17 | GO:0005622 | Intracellular | 29512 | 5 | 3.34 | 0.13269 |
| 18 | GO:0044464 | Cell part | 32576 | 5 | 3.69 | 0.21744 |
| 19 | GO:0005623 | Cell | 32702 | 5 | 3.7 | 0.22168 |
| 20 | GO:0044446 | Intracellular organelle part | 8349 | 2 | 0.94 | 0.24014 |

| Molecular Function – Ta_ECN_X | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0003735 | Structural constituent of ribosome | 1223 | 3 | 0.13 | 0.00015 |
| 2 | GO:0005198 | Structural molecule activity | 1421 | 3 | 0.15 | 0.00024 |
| 3 | GO:0019843 | rRNA binding | 293 | 1 | 0.03 | 0.02967 |
| 4 | GO:0003676 | Nucleic acid binding | 11391 | 3 | 1.17 | 0.08688 |
| 5 | GO:0003723 | RNA binding | 2699 | 1 | 0.28 | 0.24768 |
| 6 | GO:0000166 | Nucleotide binding | 12967 | 2 | 1.33 | 0.40052 |
| 7 | GO:1901265 | Nucleoside phosphate binding | 12967 | 2 | 1.33 | 0.40052 |
| 8 | GO:0036094 | Small molecule binding | 13505 | 2 | 1.38 | 0.42367 |
| 9 | GO:1901363 | Heterocyclic compound binding | 23868 | 3 | 2.45 | 0.4799 |
| 10 | GO:0097159 | Organic cyclic compound binding | 23870 | 3 | 2.45 | 0.47998 |
| 11 | GO:0005488 | Binding | 35771 | 3 | 3.67 | 0.87808 |
| 12 | GO:0003674 | Molecular_function | 48782 | 5 | 5 | 1 |
| 13 | GO:0000014 | Single-stranded DNA endodeoxyribonuclease activity | 12 | 0 | 0 | 1 |
| 14 | GO:0000030 | Mannosyltransferase activity | 75 | 0 | 0.01 | 1 |
| 15 | GO:0000049 | tRNA binding | 60 | 0 | 0.01 | 1 |
| 16 | GO:0000062 | Fatty-acyl-CoA binding | 32 | 0 | 0 | 1 |
| 17 | GO:0000104 | Succinate dehydrogenase activity | 8 | 0 | 0 | 1 |
| 18 | GO:0000107 | Imidazoleglycerol-phosphate synthase activity | 5 | 0 | 0 | 1 |
| 19 | GO:0000149 | SNARE binding | 126 | 0 | 0.01 | 1 |
| 20 | GO:0000150 | Recombinase activity | 26 | 0 | 0 | 1 |

## Appendix 10: Wheat GO Analysis – Ta_ECN_XI

| Biological Process – Ta_ECN_XI | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0015979 | Photosynthesis | 813 | 3 | 0.08 | 2.70E-05 |
| 2 | GO:0018298 | Protein-chromophore linkage | 164 | 2 | 0.02 | 8.80E-05 |
| 3 | GO:0009767 | Photosynthetic electron transport chain | 191 | 2 | 0.02 | 0.00012 |
| 4 | GO:0019684 | Photosynthesis, light reaction | 511 | 2 | 0.05 | 0.00084 |
| 5 | GO:0022900 | Electron transport chain | 623 | 2 | 0.06 | 0.00125 |
| 6 | GO:0055114 | Oxidation-reduction process | 5037 | 3 | 0.47 | 0.00598 |
| 7 | GO:0006091 | Generation of precursor metabolites and energy | 1415 | 2 | 0.13 | 0.0063 |
| 8 | GO:0009772 | Photosynthetic electron transport in photosystem II | 71 | 1 | 0.01 | 0.00663 |
| 9 | GO:0044267 | Cellular protein metabolic process | 11875 | 3 | 1.11 | 0.06807 |
| 10 | GO:0019538 | Protein metabolic process | 13401 | 3 | 1.26 | 0.09452 |
| 11 | GO:1901564 | Organonitrogen compound metabolic process | 16308 | 3 | 1.53 | 0.15896 |
| 12 | GO:0006464 | Cellular protein modification process | 8493 | 2 | 0.8 | 0.17908 |
| 13 | GO:0036211 | Protein modification process | 8493 | 2 | 0.8 | 0.17908 |
| 14 | GO:0006412 | Translation | 2294 | 1 | 0.21 | 0.19818 |
| 15 | GO:0043043 | Peptide biosynthetic process | 2305 | 1 | 0.22 | 0.19906 |
| 16 | GO:0043412 | Macromolecule modification | 9166 | 2 | 0.86 | 0.20368 |
| 17 | GO:0043604 | Amide biosynthetic process | 2439 | 1 | 0.23 | 0.20963 |
| 18 | GO:0044237 | Cellular metabolic process | 29031 | 4 | 2.72 | 0.21359 |
| 19 | GO:0006518 | Peptide metabolic process | 2668 | 1 | 0.25 | 0.22746 |
| 20 | GO:0043603 | Cellular amide metabolic process | 2958 | 1 | 0.28 | 0.2496 |

| Cellular Component – Ta_ECN_XI | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0009535 | Chloroplast thylakoid membrane | 697 | 3 | 0.06 | 1.50E-05 |
| 2 | GO:0055035 | Plastid thylakoid membrane | 699 | 3 | 0.06 | 1.60E-05 |
| 3 | GO:0042651 | Thylakoid membrane | 768 | 3 | 0.07 | 2.10E-05 |
| 4 | GO:0034357 | Photosynthetic membrane | 810 | 3 | 0.07 | 2.40E-05 |
| 5 | GO:0009534 | Chloroplast thylakoid | 833 | 3 | 0.08 | 2.60E-05 |
| 6 | GO:0031976 | Plastid thylakoid | 835 | 3 | 0.08 | 2.60E-05 |
| 7 | GO:0044436 | Thylakoid part | 846 | 3 | 0.08 | 2.80E-05 |
| 8 | GO:0009507 | Chloroplast | 3737 | 4 | 0.34 | 5.10E-05 |
| 9 | GO:0009579 | Thylakoid | 1067 | 3 | 0.1 | 5.50E-05 |
| 10 | GO:0009521 | Photosystem | 325 | 2 | 0.03 | 0.00032 |
| 11 | GO:0044434 | Chloroplast part | 1977 | 3 | 0.18 | 0.00035 |
| 12 | GO:0031984 | Organelle subcompartment | 2011 | 3 | 0.18 | 0.00036 |
| 13 | GO:0044435 | Plastid part | 2020 | 3 | 0.18 | 0.00037 |
| 14 | GO:0031360 | Intrinsic component of thylakoid membrane | 11 | 1 | 0 | 0.001 |
| 15 | GO:0031361 | Integral component of thylakoid membrane | 11 | 1 | 0 | 0.001 |
| 16 | GO:0044446 | Intracellular organelle part | 8349 | 4 | 0.76 | 0.00127 |
| 17 | GO:0044422 | Organelle part | 8378 | 4 | 0.76 | 0.00129 |
| 18 | GO:0009536 | Plastid | 8609 | 4 | 0.78 | 0.00144 |
| 19 | GO:0098796 | Membrane protein complex | 1280 | 2 | 0.12 | 0.00484 |
| 20 | GO:0032991 | Protein-containing complex | 6582 | 3 | 0.6 | 0.01173 |

| Molecular Function – Ta_ECN_XI | | | | | | |
|---|---|---|---|---|---|---|
| | **GO.ID** | **Term** | **Annotated** | **Significant** | **Expected** | **Classic Fisher** |
| 1 | GO:0009055 | Electron transfer activity | 665 | 3 | 0.05 | 1.00E-05 |
| 2 | GO:0045156 | Electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity | 75 | 2 | 0.01 | 1.40E-05 |
| 3 | GO:0016168 | Chlorophyll binding | 160 | 2 | 0.01 | 6.40E-05 |
| 4 | GO:0045158 | Electron transporter, transferring electrons within cytochrome b6/f complex of photosystem II activity | 22 | 1 | 0 | 0.0018 |
| 5 | GO:0016491 | Oxidoreductase activity | 5041 | 3 | 0.41 | 0.0041 |
| 6 | GO:0046906 | Tetrapyrrole binding | 1489 | 2 | 0.12 | 0.0054 |
| 7 | GO:0051539 | 4 iron, 4 sulfur cluster binding | 176 | 1 | 0.01 | 0.0144 |
| 8 | GO:0048037 | Cofactor binding | 3182 | 2 | 0.26 | 0.0234 |
| 9 | GO:0019843 | rRNA binding | 293 | 1 | 0.02 | 0.0238 |
| 10 | GO:0051536 | Iron-sulfur cluster binding | 368 | 1 | 0.03 | 0.0298 |
| 11 | GO:0051540 | Metal cluster binding | 368 | 1 | 0.03 | 0.0298 |
| 12 | GO:0000287 | Magnesium ion binding | 511 | 1 | 0.04 | 0.0412 |
| 13 | GO:0003735 | Structural constituent of ribosome | 1223 | 1 | 0.1 | 0.0966 |
| 14 | GO:0005506 | Iron ion binding | 1272 | 1 | 0.1 | 0.1003 |
| 15 | GO:0020037 | Heme binding | 1331 | 1 | 0.11 | 0.1048 |
| 16 | GO:0005198 | Structural molecule activity | 1421 | 1 | 0.12 | 0.1115 |
| 17 | GO:0046872 | Metal ion binding | 10274 | 2 | 0.84 | 0.1973 |
| 18 | GO:0043169 | Cation binding | 10327 | 2 | 0.85 | 0.199 |
| 19 | GO:0003723 | RNA binding | 2699 | 1 | 0.22 | 0.2036 |
| 20 | GO:0043168 | Anion binding | 12382 | 2 | 1.02 | 0.2682 |

## Appendix 11: Pst ECN Membership

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| I | KNF06096-2 | PSTG_00607 | - | large subunit ribosomal protein L35e |
| I | KNF06096-3 | PSTG_00607 | - | large subunit ribosomal protein L35e |
| I | KNF06096-4 | PSTG_00607 | - | large subunit ribosomal protein L35e |
| I | KNF06142-1 | PSTG_00651 | - | 40S ribosomal protein S1 |
| I | KNF06142-4 | PSTG_00651 | - | 40S ribosomal protein S1 |
| I | KNF06142-5 | PSTG_00651 | - | 40S ribosomal protein S1 |
| I | KNF05893-5 | PSTG_00887 | - | GTP-binding protein YchF |
| I | KNF05916-3 | PSTG_00908 | - | hypothetical protein |
| I | KNF05916-4 | PSTG_00908 | - | hypothetical protein |
| I | KNF05916-5 | PSTG_00908 | - | hypothetical protein |
| I | KNF05961-4 | PSTG_00954 | - | 40S ribosomal protein S27 |
| I | KNF05587-2 | PSTG_01396 | - | hypothetical protein |
| I | KNF05587-3 | PSTG_01396 | - | hypothetical protein |
| I | KNF05587-4 | PSTG_01396 | - | hypothetical protein |
| I | KNF05432-5 | PSTG_01646 | - | hypothetical protein |
| I | KNF05116-2 | PSTG_01745 | - | 30S ribosomal protein S28e |
| I | KNF05116-3 | PSTG_01745 | - | 30S ribosomal protein S28e |
| I | KNF05189-4 | PSTG_01816 | - | 60S ribosomal protein L37-A |
| I | KNF05190-6 | PSTG_01817 | - | ubiquitin-40S ribosomal protein S27a-2 |
| I | KNF04673-5 | PSTG_02158 | - | hypothetical protein |
| I | KNF04685-1 | PSTG_02170 | - | 40S ribosomal protein S29 |
| I | KNF04686-3 | PSTG_02171 | - | 40S ribosomal protein S8 |
| I | KNF04686-2 | PSTG_02171 | - | 40S ribosomal protein S8 |
| I | KNF04686-4 | PSTG_02171 | - | 40S ribosomal protein S8 |
| I | KNF04770-1 | PSTG_02248 | - | hypothetical protein |
| I | KNF04770-5 | PSTG_02248 | - | hypothetical protein |
| I | KNF04270-2 | PSTG_02614 | - | hypothetical protein |
| I | KNF04270-5 | PSTG_02614 | - | hypothetical protein |
| I | KNF04306-5 | PSTG_02648 | - | 40S ribosomal protein S16 |
| I | KNF03999-2 | PSTG_02708 | - | hypothetical protein |
| I | KNF03999-4 | PSTG_02708 | - | hypothetical protein |
| I | KNF04041-2 | PSTG_02749 | - | 60S ribosomal protein L32 |
| I | KNF03864-3 | PSTG_02954 | - | 40S ribosomal protein S9-B |
| I | KNF03864-4 | PSTG_02954 | - | 40S ribosomal protein S9-B |
| I | KNF03865-1 | PSTG_02955 | - | 60S ribosomal protein L21-A |
| I | KNF03865-2 | PSTG_02955 | - | 60S ribosomal protein L21-A |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| I | KNF03865-3 | PSTG_02955 | - | 60S ribosomal protein L21-A |
| I | KNF03898-5 | PSTG_02985 | - | ribosomal protein L28e |
| I | KNF03578-6 | PSTG_03101 | - | hypothetical protein |
| I | KNF03578-7 | PSTG_03101 | - | hypothetical protein |
| I | KNF03579-2 | PSTG_03102 | - | 40S ribosomal protein S14a |
| I | KNF03590-1 | PSTG_03113 | - | ribonucleoprotein-associated protein |
| I | KNF03590-3 | PSTG_03113 | - | ribonucleoprotein-associated protein |
| I | KNF03043-4 | PSTG_03636 | - | 60S ribosomal protein L3 |
| I | KNF03043-6 | PSTG_03636 | - | 60S ribosomal protein L3 |
| I | KNF03043-7 | PSTG_03636 | - | 60S ribosomal protein L3 |
| I | KNF03043-5 | PSTG_03636 | - | 60S ribosomal protein L3 |
| I | KNF02956-1 | PSTG_03905 | - | 60S ribosomal protein L15-B |
| I | KNF02956-2 | PSTG_03905 | - | 60S ribosomal protein L15-B |
| I | KNF02956-4 | PSTG_03905 | - | 60S ribosomal protein L15-B |
| I | KNF02749-4 | PSTG_04034 | - | hypothetical protein |
| I | KNF02749-5 | PSTG_04034 | - | hypothetical protein |
| I | KNF02504-5 | PSTG_04409 | - | 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase |
| I | KNF02504-8 | PSTG_04409 | - | 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase |
| I | KNF02518-3 | PSTG_04424 | - | hypothetical protein |
| I | KNF02518-4 | PSTG_04424 | - | hypothetical protein |
| I | KNF02519-1 | PSTG_04425 | - | 40S ribosomal protein S21 |
| I | KNF02519-2 | PSTG_04425 | - | 40S ribosomal protein S21 |
| I | KNF02519-3 | PSTG_04425 | - | 40S ribosomal protein S21 |
| I | KNF01938-1 | PSTG_04763 | - | hypothetical protein |
| I | KNF01938-2 | PSTG_04763 | - | hypothetical protein |
| I | KNF01622-2 | PSTG_05054 | - | 40S ribosomal protein S13 |
| I | KNF01622-4 | PSTG_05054 | - | 40S ribosomal protein S13 |
| I | KNF01623-1 | PSTG_05055 | - | small subunit ribosomal protein S6e |
| I | KNF01623-3 | PSTG_05055 | - | small subunit ribosomal protein S6e |
| I | KNF01623-4 | PSTG_05055 | - | small subunit ribosomal protein S6e |
| I | KNF01711-3 | PSTG_05140 | - | 40S ribosomal protein S19-A |
| I | KNF01711-4 | PSTG_05140 | - | 40S ribosomal protein S19-A |
| I | KNF01429-1 | PSTG_05214 | - | hypothetical protein |
| I | KNF01429-2 | PSTG_05214 | - | hypothetical protein |
| I | KNF01429-3 | PSTG_05214 | - | hypothetical protein |
| I | KNF01514-4 | PSTG_05294 | - | hypothetical protein |
| I | KNF01514-5 | PSTG_05294 | - | hypothetical protein |
| I | KNF01284-5 | PSTG_05381 | - | ribosomal protein L37ae |
| I | KNF01009-2 | PSTG_05643 | - | 60S ribosomal protein L5 |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| I | KNF01009-3 | PSTG_05643 | - | 60S ribosomal protein L5 |
| I | KNF01105-5 | PSTG_05734 | - | 40S ribosomal protein S5 |
| I | KNF01105-3 | PSTG_05734 | - | 40S ribosomal protein S5 |
| I | KNF00913-4 | PSTG_05806 | - | 60S ribosomal protein L29 |
| I | KNF00749-2 | PSTG_05888 | - | 40S ribosomal protein S7 |
| I | KNF00749-4 | PSTG_05888 | - | 40S ribosomal protein S7 |
| I | KNF00749-3 | PSTG_05888 | - | 40S ribosomal protein S7 |
| I | KNF00604-3 | PSTG_06020 | - | 40S ribosomal protein S0 |
| I | KNF00604-4 | PSTG_06020 | - | 40S ribosomal protein S0 |
| I | KNF00621-8 | PSTG_06036 | - | glyceraldehyde-3-phosphate dehydrogenase |
| I | KNF00461-2 | PSTG_06392 | - | 60S ribosomal protein L4-A |
| I | KNF00461-3 | PSTG_06392 | - | 60S ribosomal protein L4-A |
| I | KNF00461-5 | PSTG_06392 | - | 60S ribosomal protein L4-A |
| I | KNF00280-5 | PSTG_06451 | - | hypothetical protein |
| I | KNF00280-8 | PSTG_06451 | - | hypothetical protein |
| I | KNF00322-1 | PSTG_06495 | - | hypothetical protein |
| I | KNF00322-3 | PSTG_06495 | - | hypothetical protein |
| I | KNF00322-4 | PSTG_06495 | - | hypothetical protein |
| I | KNF00322-2 | PSTG_06495 | - | hypothetical protein |
| I | KNF00337-4 | PSTG_06510 | - | 50S ribosomal protein L36e |
| I | KNF00338-2 | PSTG_06511 | - | hypothetical protein |
| I | KNF00338-5 | PSTG_06511 | - | hypothetical protein |
| I | KNF00124-5 | PSTG_06536 | - | hypothetical protein |
| I | KNF00124-7 | PSTG_06536 | - | hypothetical protein |
| I | KNE99890-1 | PSTG_06743 | - | large subunit ribosomal protein L31e |
| I | KNE99890-2 | PSTG_06743 | - | large subunit ribosomal protein L31e |
| I | KNE99656-4 | PSTG_07148 | - | hypothetical protein |
| I | KNE99656-7 | PSTG_07148 | - | hypothetical protein |
| I | KNE99656-5 | PSTG_07148 | - | hypothetical protein |
| I | KNE99490-3 | PSTG_07207 | - | 60S ribosomal protein L27-A |
| I | KNE99490-2 | PSTG_07207 | - | 60S ribosomal protein L27-A |
| I | KNE99146-2 | PSTG_07623 | - | large subunit ribosomal protein L31e |
| I | KNE99146-4 | PSTG_07623 | - | large subunit ribosomal protein L31e |
| I | KNE99083-4 | PSTG_07734 | - | hypothetical protein |
| I | KNE99083-5 | PSTG_07734 | - | hypothetical protein |
| I | KNE99084-3 | PSTG_07735 | - | 40S ribosomal protein S17 |
| I | KNE99084-4 | PSTG_07735 | - | 40S ribosomal protein S17 |
| I | KNE98972-1 | PSTG_07816 | - | 40S ribosomal protein S3 |
| I | KNE98972-2 | PSTG_07816 | - | 40S ribosomal protein S3 |
| I | KNE98972-3 | PSTG_07816 | - | 40S ribosomal protein S3 |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| I | KNE98984-5 | PSTG_07828 | - | 30S ribosomal protein S26e |
| I | KNE98598-10 | PSTG_08150 | - | T-complex protein 1 subunit epsilon |
| I | KNE98630-4 | PSTG_08181 | - | hypothetical protein |
| I | KNE98630-6 | PSTG_08181 | - | hypothetical protein |
| I | KNE98049-1 | PSTG_08724 | - | 50S ribosomal protein L22 |
| I | KNE98049-2 | PSTG_08724 | - | 50S ribosomal protein L22 |
| I | KNE98049-3 | PSTG_08724 | - | 50S ribosomal protein L22 |
| I | KNE98049-4 | PSTG_08724 | - | 50S ribosomal protein L22 |
| I | KNE98069-6 | PSTG_08743 | - | guanine nucleotide-binding protein subunit beta-like protein |
| I | KNE98069-8 | PSTG_08743 | - | guanine nucleotide-binding protein subunit beta-like protein |
| I | KNE98070-3 | PSTG_08744 | - | 60S ribosomal protein L12 |
| I | KNE98070-4 | PSTG_08744 | - | 60S ribosomal protein L12 |
| I | KNE97625-8 | PSTG_09030 | - | hypothetical protein |
| I | KNE97533-4 | PSTG_09227 | - | hypothetical protein |
| I | KNE97198-3 | PSTG_09460 | - | hypothetical protein |
| I | KNE97198-7 | PSTG_09460 | - | hypothetical protein |
| I | KNE96845-1 | PSTG_09828 | - | 60S ribosomal protein L33-B |
| I | KNE96845-2 | PSTG_09828 | - | 60S ribosomal protein L33-B |
| I | KNE96845-3 | PSTG_09828 | - | 60S ribosomal protein L33-B |
| I | KNE96810-1 | PSTG_09946 | - | 40S ribosomal protein S11 |
| I | KNE96810-2 | PSTG_09946 | - | 40S ribosomal protein S11 |
| I | KNE96810-4 | PSTG_09946 | - | 40S ribosomal protein S11 |
| I | KNE96707-6 | PSTG_09978 | - | 60S ribosomal protein L11 |
| I | KNE96707-4 | PSTG_09978 | - | 60S ribosomal protein L11 |
| I | KNE96635-1 | PSTG_10042 | - | 60S ribosomal protein L15-B |
| I | KNE96635-2 | PSTG_10042 | - | 60S ribosomal protein L15-B |
| I | KNE96635-4 | PSTG_10042 | - | 60S ribosomal protein L15-B |
| I | KNE96498-6 | PSTG_10205 | - | heat shock protein SSB |
| I | KNE96288-1 | PSTG_10407 | - | 60S ribosomal protein L28 |
| I | KNE96288-3 | PSTG_10407 | - | 60S ribosomal protein L28 |
| I | KNE95905-5 | PSTG_10822 | - | hypothetical protein |
| I | KNE95641-5 | PSTG_11006 | - | hypothetical protein |
| I | KNE95641-6 | PSTG_11006 | - | hypothetical protein |
| I | KNE95523-3 | PSTG_11130 | - | 60S ribosomal protein L1-B |
| I | KNE95523-5 | PSTG_11130 | - | 60S ribosomal protein L1-B |
| I | KNE95523-6 | PSTG_11130 | - | 60S ribosomal protein L1-B |
| I | KNE95024-6 | PSTG_11621 | - | ubiquitin-60S ribosomal protein L40 |
| I | KNE94999-2 | PSTG_11696 | - | hypothetical protein |
| I | KNE94999-4 | PSTG_11696 | - | hypothetical protein |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| I | KNE94391-3 | PSTG_12291 | - | translationally-controlled tumor protein |
| I | KNE94400-5 | PSTG_12300 | - | pyruvate dehydrogenase E1 component subunit beta |
| I | KNE93748-1 | PSTG_12851 | - | hypothetical protein |
| I | KNE93748-6 | PSTG_12851 | - | hypothetical protein |
| I | KNE93631-2 | PSTG_13007 | - | 50S ribosomal protein L19e |
| I | KNE93631-4 | PSTG_13007 | - | 50S ribosomal protein L19e |
| I | KNE93632-2 | PSTG_13008 | - | 40S ribosomal protein S18 |
| I | KNE93632-4 | PSTG_13008 | - | 40S ribosomal protein S18 |
| I | KNE93481-6 | PSTG_13111 | - | elongation factor 1-alpha |
| I | KNE93481-7 | PSTG_13111 | - | elongation factor 1-alpha |
| I | KNE93481-8 | PSTG_13111 | - | elongation factor 1-alpha |
| I | KNE93481-9 | PSTG_13111 | - | elongation factor 1-alpha |
| I | KNE93500-4 | PSTG_13128 | - | hypothetical protein |
| I | KNE93500-5 | PSTG_13128 | - | hypothetical protein |
| I | KNE93500-6 | PSTG_13128 | - | hypothetical protein |
| I | KNE93413-1 | PSTG_13237 | - | endoribonuclease L-PSP |
| I | KNE93198-2 | PSTG_13448 | - | 40S ribosomal protein S15 |
| I | KNE93198-4 | PSTG_13448 | - | 40S ribosomal protein S15 |
| I | KNE93199-1 | PSTG_13449 | - | hypothetical protein |
| I | KNE93199-3 | PSTG_13449 | - | hypothetical protein |
| I | KNE92706-10 | PSTG_13914 | - | pyruvate kinase |
| I | KNE92706-8 | PSTG_13914 | - | pyruvate kinase |
| I | KNE92706-9 | PSTG_13914 | - | pyruvate kinase |
| I | KNE92453-1 | PSTG_14174 | - | 50S ribosomal protein L34e |
| I | KNE92453-4 | PSTG_14174 | - | 50S ribosomal protein L34e |
| I | KNE92347-6 | PSTG_14247 | - | hypothetical protein |
| I | KNE91789-3 | PSTG_14806 | - | hypothetical protein |
| I | KNE91789-4 | PSTG_14806 | - | hypothetical protein |
| I | KNE91408-2 | PSTG_15155 | - | 60S ribosomal protein L8 |
| I | KNE91408-3 | PSTG_15155 | - | 60S ribosomal protein L8 |
| I | KNE91408-5 | PSTG_15155 | - | 60S ribosomal protein L8 |
| I | KNE91332-1 | PSTG_15262 | - | small subunit ribosomal protein S2e |
| I | KNE91332-3 | PSTG_15262 | - | small subunit ribosomal protein S2e |
| I | KNE91277-4 | PSTG_15301 | - | 60S ribosomal protein L23 |
| I | KNE90888-7 | PSTG_15678 | - | hypothetical protein |
| I | KNE90223-4 | PSTG_16330 | - | S25 ribosomal protein |
| I | KNE90223-2 | PSTG_16330 | - | S25 ribosomal protein |
| I | KNE90178-3 | PSTG_16362 | - | hypothetical protein |
| I | KNE90178-5 | PSTG_16362 | - | hypothetical protein |
| I | KNE90178-4 | PSTG_16362 | - | hypothetical protein |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| I | KNE89788-2 | PSTG_16747 | - | 60S ribosomal protein L20 |
| I | KNE89788-1 | PSTG_16747 | - | 60S ribosomal protein L20 |
| I | KNE89788-3 | PSTG_16747 | - | 60S ribosomal protein L20 |
| I | KNE89600-6 | PSTG_16938 | - | adenosylhomocysteinase |
| I | KNE89600-7 | PSTG_16938 | - | adenosylhomocysteinase |
| I | KNE89584-1 | PSTG_16947 | - | hypothetical protein |
| I | KNE89584-2 | PSTG_16947 | - | hypothetical protein |
| I | KNE89584-3 | PSTG_16947 | - | hypothetical protein |
| I | KNE89584-4 | PSTG_16947 | - | hypothetical protein |
| II | KNF06693-6 | PSTG_00009 | - | hypothetical protein |
| II | KNF06384-1 | PSTG_00266 | - | hypothetical protein |
| II | KNF06384-2 | PSTG_00266 | - | hypothetical protein |
| II | KNF06387-1 | PSTG_00269 | - | hypothetical protein |
| II | KNF06140-1 | PSTG_00649 | Yes | hypothetical protein |
| II | KNF06140-13 | PSTG_00649 | Yes | hypothetical protein |
| II | KNF06141-1 | PSTG_00650 | Yes | hypothetical protein |
| II | KNF05923-1 | PSTG_00916 | - | hypothetical protein |
| II | KNF05923-2 | PSTG_00916 | - | hypothetical protein |
| II | KNF06019-1 | PSTG_01012 | - | hypothetical protein |
| II | KNF06019-3 | PSTG_01012 | - | hypothetical protein |
| II | KNF05780-3 | PSTG_01177 | - | hypothetical protein |
| II | KNF05780-4 | PSTG_01177 | - | hypothetical protein |
| II | KNF05358-1 | PSTG_01573 | Yes | hypothetical protein |
| II | KNF05358-5 | PSTG_01573 | Yes | hypothetical protein |
| II | KNF04815-4 | PSTG_01871 | - | hypothetical protein |
| II | KNF04871-1 | PSTG_01925 | - | hypothetical protein |
| II | KNF04605-2 | PSTG_02092 | Yes | hypothetical protein |
| II | KNF04652-1 | PSTG_02137 | - | hypothetical protein |
| II | KNF04524-3 | PSTG_02434 | Yes | hypothetical protein |
| II | KNF04244-1 | PSTG_02592 | - | hypothetical protein |
| II | KNF04244-2 | PSTG_02592 | - | hypothetical protein |
| II | KNF04085-2 | PSTG_02790 | Yes | hypothetical protein |
| II | KNF04132-12 | PSTG_02836 | - | hypothetical protein |
| II | KNF04132-4 | PSTG_02836 | - | hypothetical protein |
| II | KNF03592-7 | PSTG_03115 | - | hypothetical protein |
| II | KNF03592-9 | PSTG_03115 | - | hypothetical protein |
| II | KNF03674-1 | PSTG_03194 | Yes | hypothetical protein |
| II | KNF03413-1 | PSTG_03353 | - | hypothetical protein |
| II | KNF03457-2 | PSTG_03398 | - | hypothetical protein |
| II | KNF03457-3 | PSTG_03398 | - | hypothetical protein |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| II | KNF03457-5 | PSTG_03398 | - | hypothetical protein |
| II | KNF03260-3 | PSTG_03525 | - | hypothetical protein |
| II | KNF03260-4 | PSTG_03525 | - | hypothetical protein |
| II | KNF03260-5 | PSTG_03525 | - | hypothetical protein |
| II | KNF02619-1 | PSTG_04216 | Yes | hypothetical protein |
| II | KNF02619-4 | PSTG_04216 | Yes | hypothetical protein |
| II | KNF02092-3 | PSTG_04590 | - | hypothetical protein |
| II | KNF02092-5 | PSTG_04590 | - | hypothetical protein |
| II | KNF02092-7 | PSTG_04590 | - | hypothetical protein |
| II | KNF02092-1 | PSTG_04590 | - | hypothetical protein |
| II | KNF02027-1 | PSTG_04848 | Yes | hypothetical protein |
| II | KNF02044-2 | PSTG_04865 | Yes | hypothetical protein |
| II | KNF02044-3 | PSTG_04865 | Yes | hypothetical protein |
| II | KNF02044-5 | PSTG_04865 | Yes | hypothetical protein |
| II | KNF01539-2 | PSTG_05319 | - | hypothetical protein |
| II | KNF01283-1 | PSTG_05380 | - | hypothetical protein |
| II | KNF00807-2 | PSTG_05945 | - | hypothetical protein |
| II | KNF00849-1 | PSTG_05983 | - | hypothetical protein |
| II | KNF00502-1 | PSTG_06194 | - | hypothetical protein |
| II | KNF00502-2 | PSTG_06194 | - | hypothetical protein |
| II | KNF00540-1 | PSTG_06232 | - | hypothetical protein |
| II | KNF00540-3 | PSTG_06232 | - | hypothetical protein |
| II | KNF00540-5 | PSTG_06232 | - | hypothetical protein |
| II | KNF00334-1 | PSTG_06507 | Yes | hypothetical protein |
| II | KNF00334-4 | PSTG_06507 | Yes | hypothetical protein |
| II | KNF00357-2 | PSTG_06528 | - | hypothetical protein |
| II | KNF00182-2 | PSTG_06592 | - | hypothetical protein |
| II | KNF00182-6 | PSTG_06592 | - | hypothetical protein |
| II | KNF00207-3 | PSTG_06620 | Yes | hypothetical protein |
| II | KNE99894-1 | PSTG_06747 | - | hypothetical protein |
| II | KNE99986-3 | PSTG_06838 | - | hypothetical protein |
| II | KNE99986-4 | PSTG_06838 | - | hypothetical protein |
| II | KNE99522-1 | PSTG_07035 | - | hypothetical protein |
| II | KNE99748-2 | PSTG_07035 | - | hypothetical protein |
| II | KNE99748-3 | PSTG_07035 | - | hypothetical protein |
| II | KNE99225-1 | PSTG_07533 | Yes | hypothetical protein |
| II | KNE99225-2 | PSTG_07533 | Yes | hypothetical protein |
| II | KNE99225-3 | PSTG_07533 | Yes | hypothetical protein |
| II | KNE99225-5 | PSTG_07533 | Yes | hypothetical protein |
| II | KNE98963-2 | PSTG_07807 | Yes | hypothetical protein |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| II | KNE98963-4 | PSTG_07807 | Yes | hypothetical protein |
| II | KNE98801-1 | PSTG_07989 | - | hypothetical protein |
| II | KNE98801-2 | PSTG_07989 | - | hypothetical protein |
| II | KNE98801-3 | PSTG_07989 | - | hypothetical protein |
| II | KNE98801-5 | PSTG_07989 | - | hypothetical protein |
| II | KNE98801-7 | PSTG_07989 | - | hypothetical protein |
| II | KNE98579-2 | PSTG_08131 | Yes | hypothetical protein |
| II | KNE98081-1 | PSTG_08755 | Yes | hypothetical protein |
| II | KNE98081-3 | PSTG_08755 | Yes | hypothetical protein |
| II | KNE98081-4 | PSTG_08755 | Yes | hypothetical protein |
| II | KNE97907-5 | PSTG_08780 | - | CMGC/MAPK |
| II | KNE97907-6 | PSTG_08780 | - | CMGC/MAPK |
| II | KNE97641-1 | PSTG_09047 | Yes | hypothetical protein |
| II | KNE97641-3 | PSTG_09047 | Yes | hypothetical protein |
| II | KNE97641-5 | PSTG_09047 | Yes | hypothetical protein |
| II | KNE97422-1 | PSTG_09255 | - | hypothetical protein |
| II | KNE97422-2 | PSTG_09255 | - | hypothetical protein |
| II | KNE97422-3 | PSTG_09255 | - | hypothetical protein |
| II | KNE96842-2 | PSTG_09825 | - | hypothetical protein |
| II | KNE96080-3 | PSTG_10652 | Yes | hypothetical protein |
| II | KNE96080-5 | PSTG_10652 | Yes | hypothetical protein |
| II | KNE95699-1 | PSTG_10918 | Yes | hypothetical protein |
| II | KNE95699-5 | PSTG_10918 | Yes | hypothetical protein |
| II | KNE95701-1 | PSTG_10920 | Yes | hypothetical protein |
| II | KNE95701-5 | PSTG_10920 | Yes | hypothetical protein |
| II | KNE95760-2 | PSTG_10976 | - | hypothetical protein |
| II | KNE95619-3 | PSTG_11109 | Yes | hypothetical protein |
| II | KNE95625-1 | PSTG_11115 | - | hypothetical protein |
| II | KNE95280-1 | PSTG_11366 | - | hypothetical protein |
| II | KNE95266-1 | PSTG_11438 | Yes | hypothetical protein |
| II | KNE95205-1 | PSTG_11470 | - | hypothetical protein |
| II | KNE95205-2 | PSTG_11470 | - | hypothetical protein |
| II | KNE95088-4 | PSTG_11565 | - | hypothetical protein |
| II | KNE94593-2 | PSTG_12057 | - | hypothetical protein |
| II | KNE94593-5 | PSTG_12057 | - | hypothetical protein |
| II | KNE94566-5 | PSTG_12118 | Yes | hypothetical protein |
| II | KNE94421-6 | PSTG_12216 | Yes | hypothetical protein |
| II | KNE94396-1 | PSTG_12296 | - | hypothetical protein |
| II | KNE94283-4 | PSTG_12309 | - | TKL/LISK/LISK-DD1 protein kinase |
| II | KNE94223-1 | PSTG_12451 | - | hypothetical protein |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| II | KNE94149-1 | PSTG_12479 | Yes | hypothetical protein |
| II | KNE94149-4 | PSTG_12479 | Yes | hypothetical protein |
| II | KNE94173-1 | PSTG_12504 | - | hypothetical protein |
| II | KNE93237-2 | PSTG_13411 | Yes | hypothetical protein |
| II | KNE92714-1 | PSTG_13922 | Yes | hypothetical protein |
| II | KNE92505-1 | PSTG_14106 | - | hypothetical protein |
| II | KNE91482-1 | PSTG_15112 | Yes | hypothetical protein |
| II | KNE91482-2 | PSTG_15112 | Yes | hypothetical protein |
| II | KNE91339-9 | PSTG_15268 | - | hypothetical protein |
| II | KNE91169-1 | PSTG_15425 | - | hypothetical protein |
| II | KNE91169-11 | PSTG_15425 | - | hypothetical protein |
| II | KNE90769-8 | PSTG_15777 | Yes | hypothetical protein |
| II | KNE90580-3 | PSTG_15973 | Yes | hypothetical protein |
| II | KNE90580-5 | PSTG_15973 | Yes | hypothetical protein |
| II | KNE90191-4 | PSTG_16373 | Yes | hypothetical protein |
| II | KNE90191-6 | PSTG_16373 | Yes | hypothetical protein |
| II | KNE89665-1 | PSTG_16862 | - | hypothetical protein |
| II | KNE89459-1 | PSTG_17090 | - | hypothetical protein |
| II | KNE89459-4 | PSTG_17090 | - | hypothetical protein |
| II | KNE89459-7 | PSTG_17090 | - | hypothetical protein |
| II | KNE89425-1 | PSTG_17118 | - | hypothetical protein |
| II | KNE87876-1 | PSTG_18734 | - | hypothetical protein |
| III | KNF06597-2 | PSTG_00471 | - | hypothetical protein |
| III | KNF06141-10 | PSTG_00650 | Yes | hypothetical protein |
| III | KNF05923-4 | PSTG_00916 | - | hypothetical protein |
| III | KNF05965-4 | PSTG_00958 | Yes | hypothetical protein |
| III | KNF05780-10 | PSTG_01177 | - | hypothetical protein |
| III | KNF05316-9 | PSTG_01530 | - | AGC/NDR protein kinase |
| III | KNF05358-6 | PSTG_01573 | Yes | hypothetical protein |
| III | KNF05079-6 | PSTG_01711 | Yes | hypothetical protein |
| III | KNF04871-3 | PSTG_01925 | - | hypothetical protein |
| III | KNF04871-6 | PSTG_01925 | - | hypothetical protein |
| III | KNF04605-6 | PSTG_02092 | Yes | hypothetical protein |
| III | KNF04652-2 | PSTG_02137 | - | hypothetical protein |
| III | KNF04524-5 | PSTG_02434 | Yes | hypothetical protein |
| III | KNF03674-7 | PSTG_03194 | Yes | hypothetical protein |
| III | KNF03413-6 | PSTG_03353 | - | hypothetical protein |
| III | KNF03457-6 | PSTG_03398 | - | hypothetical protein |
| III | KNF03060-1 | PSTG_03652 | Yes | hypothetical protein |
| III | KNF03060-3 | PSTG_03652 | Yes | hypothetical protein |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| III | KNF02751-5 | PSTG_04036 | - | hypothetical protein |
| III | KNF02619-12 | PSTG_04216 | Yes | hypothetical protein |
| III | KNF02092-8 | PSTG_04590 | - | hypothetical protein |
| III | KNF02044-6 | PSTG_04865 | Yes | hypothetical protein |
| III | KNF00808-8 | PSTG_05945 | - | hypothetical protein |
| III | KNF00502-3 | PSTG_06194 | - | hypothetical protein |
| III | KNE99894-9 | PSTG_06747 | - | hypothetical protein |
| III | KNE99946-3 | PSTG_06799 | - | hypothetical protein |
| III | KNE99747-6 | PSTG_07034 | - | hypothetical protein |
| III | KNE99522-2 | PSTG_07235 | - | hypothetical protein |
| III | KNE99225-9 | PSTG_07533 | Yes | hypothetical protein |
| III | KNE98939-1 | PSTG_07784 | - | hypothetical protein |
| III | KNE98767-2 | PSTG_07954 | - | hypothetical protein |
| III | KNE97908-7 | PSTG_08780 | - | CMGC/MAPK protein kinase%2C variant |
| III | KNE97637-6 | PSTG_09044 | Yes | hypothetical protein |
| III | KNE97641-6 | PSTG_09047 | Yes | hypothetical protein |
| III | KNE97422-4 | PSTG_09255 | - | hypothetical protein |
| III | KNE96842-4 | PSTG_09825 | - | hypothetical protein |
| III | KNE96132-2 | PSTG_10552 | - | hypothetical protein |
| III | KNE96062-12 | PSTG_10638 | - | hypothetical protein |
| III | KNE96080-7 | PSTG_10652 | Yes | hypothetical protein |
| III | KNE95625-2 | PSTG_11115 | - | hypothetical protein |
| III | KNE95280-4 | PSTG_11366 | - | hypothetical protein |
| III | KNE95266-5 | PSTG_11438 | Yes | hypothetical protein |
| III | KNE95088-7 | PSTG_11565 | - | hypothetical protein |
| III | KNE94396-2 | PSTG_12296 | - | hypothetical protein |
| III | KNE94341-1 | PSTG_12366 | Yes | hypothetical protein |
| III | KNE94223-3 | PSTG_12451 | - | hypothetical protein |
| III | KNE94149-8 | PSTG_12479 | Yes | hypothetical protein |
| III | KNE94173-2 | PSTG_12504 | - | hypothetical protein |
| III | KNE93783-1 | PSTG_12886 | - | hypothetical protein |
| III | KNE93426-2 | PSTG_13248 | Yes | hypothetical protein |
| III | KNE93017-4 | PSTG_13591 | Yes | hypothetical protein |
| III | KNE92929-10 | PSTG_13645 | - | hypothetical protein |
| III | KNE92948-10 | PSTG_13662 | Yes | hypothetical protein |
| III | KNE92892-4 | PSTG_13740 | - | hypothetical protein |
| III | KNE92714-3 | PSTG_13922 | Yes | hypothetical protein |
| III | KNE92505-4 | PSTG_14106 | - | hypothetical protein |
| III | KNE92505-6 | PSTG_14106 | - | hypothetical protein |
| III | KNE91482-5 | PSTG_15112 | Yes | hypothetical protein |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| III | KNE91215-4 | PSTG_15379 | Yes | hypothetical protein |
| III | KNE90769-9 | PSTG_15777 | Yes | hypothetical protein |
| III | KNE90191-10 | PSTG_16373 | Yes | hypothetical protein |
| III | KNE89665-2 | PSTG_16862 | - | hypothetical protein |
| III | KNE89425-4 | PSTG_17118 | - | hypothetical protein |
| IV | KNF06510-5 | PSTG_00386 | - | hypothetical protein |
| IV | KNF02276-18 | PSTG_04484 | - | hypothetical protein |
| IV | KNF02298-4 | PSTG_04505 | Yes | hypothetical protein |
| IV | KNF00689-1 | PSTG_06104 | Yes | hypothetical protein |
| IV | KNF00413-5 | PSTG_06342 | - | hypothetical protein |
| IV | KNE99323-1 | PSTG_07440 | Yes | hypothetical protein |
| IV | KNE99148-2 | PSTG_07625 | - | hypothetical protein |
| IV | KNE98142-3 | PSTG_08606 | Yes | hypothetical protein |
| IV | KNE97642-1 | PSTG_09048 | Yes | hypothetical protein |
| IV | KNE97642-6 | PSTG_09048 | Yes | hypothetical protein |
| IV | KNE95229-3 | PSTG_11493 | Yes | hypothetical protein |
| IV | KNE94989-2 | PSTG_11687 | Yes | hypothetical protein |
| IV | KNE94444-4 | PSTG_12237 | Yes | hypothetical protein |
| IV | KNE92718-5 | PSTG_13849 | - | hypothetical protein |
| IV | KNE92366-1 | PSTG_14202 | Yes | hypothetical protein |
| IV | KNE92373-1 | PSTG_14208 | - | hypothetical protein |
| IV | KNE92149-1 | PSTG_14441 | - | hypothetical protein |
| IV | KNE92149-4 | PSTG_14441 | - | hypothetical protein |
| IV | KNE91633-11 | PSTG_14941 | Yes | hypothetical protein |
| IV | KNE91490-3 | PSTG_15087 | - | hypothetical protein |
| IV | KNE90739-4 | PSTG_15845 | - | hypothetical protein |
| IV | KNE90575-5 | PSTG_15994 | - | hydroxymethylglutaryl-CoA synthase |
| V | ENSRNAG049979156-1 | ENSRNAG049979156 | - | Bacterial large subunit |
| V | ENSRNAG049979160-1 | ENSRNAG049979160 | - | Bacterial small subunit |
| V | KNF06910-1 | PSTG_40002 | - | cytochrome c oxidase subunit 3 |
| V | KNF06911-1 | PSTG_40005 | - | NADH-ubiquinone oxidoreductase chain 4 |
| V | KNF06911-2 | PSTG_40005 | - | NADH-ubiquinone oxidoreductase chain 4 |
| V | KNF06912-1 | PSTG_40006 | - | intron_maturase 2 |
| V | KNF06915-1 | PSTG_40013 | - | cytochrome c oxidase subunit 1 |
| V | KNF06915-4 | PSTG_40013 | - | cytochrome c oxidase subunit 1 |
| V | KNF06916-1 | PSTG_40014 | - | LAGLIDADG 1 |
| V | KNF06923-3 | PSTG_40028 | - | NADH-ubiquinone oxidoreductase chain 5 |
| V | KNF06924-1 | PSTG_40033 | - | ATP synthase subunit 9 |
| V | KNF06925-1 | PSTG_40038 | - | NADH dehydrogenase subunit 2 |

| Pst ECN | Exon | Gene | Secreted? | Description |
|---|---|---|---|---|
| V | KNF06928-1 | PSTG_40046 | - | ATP synthase subunit A |
| V | KNF06929-1 | PSTG_40048 | - | cytochrome c oxidase subunit 2 |
| V | KNF06929-2 | PSTG_40048 | - | cytochrome c oxidase subunit 2 |
| VI | KNF03487-2 | PSTG_03423 | - | heat shock protein HSS1 |
| VI | KNF03487-4 | PSTG_03423 | - | heat shock protein HSS1 |
| VI | KNF03487-5 | PSTG_03423 | - | heat shock protein HSS1 |
| VI | KNF03487-6 | PSTG_03423 | - | heat shock protein HSS1 |
| VI | KNF03487-7 | PSTG_03423 | - | heat shock protein HSS1 |
| VI | KNF02405-10 | PSTG_04312 | - | heat shock protein 90 |
| VI | KNF02405-11 | PSTG_04312 | - | heat shock protein 90 |
| VI | KNF02405-6 | PSTG_04312 | - | heat shock protein 90 |
| VI | KNF02405-7 | PSTG_04312 | - | heat shock protein 90 |
| VI | KNF02405-8 | PSTG_04312 | - | heat shock protein 90 |

Legend:
- Not secreted.

## Appendix 12: Wheat ECN Membership

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| I | TraesCS1A02G171000.1.E4 | TraesCS1A02G171000 | NADPH-protochlorophyllide oxidoreductase |
| I | TraesCS1A02G251200.1.E3 | TraesCS1A02G251200 | N/A |
| I | TraesCS1A02G298800.1.E3 | TraesCS1A02G298800 | N/A |
| I | TraesCS1A02G301400.1.E3 | TraesCS1A02G301400 | Serine/threonine-protein kinase STN8%2C chloroplastic |
| I | TraesCS1A02G313300.1.E6 | TraesCS1A02G313300 | Phosphoglycerate kinase |
| I | TraesCS1A02G392000.1.E2 | TraesCS1A02G392000 | Photosystem I reaction center subunit VI%2C chloroplast precursor (PSI- H) (Light-harvesting complex I 11 kDa protein) (GOS5 protein) |
| I | TraesCS1A02G403300.1.E1 | TraesCS1A02G403300 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS1B02G186300.1.E4 | TraesCS1B02G186300 | NADPH-protochlorophyllide oxidoreductase |
| I | TraesCS1B02G261800.1.E3 | TraesCS1B02G261800 | N/A |
| I | TraesCS1B02G325100.1.E6 | TraesCS1B02G325100 | Phosphoglycerate kinase |
| I | TraesCS1B02G420100.1.E2 | TraesCS1B02G420100 | Photosystem I reaction center subunit VI%2C chloroplast precursor (PSI- H) (Light-harvesting complex I 11 kDa protein) (GOS5 protein) |
| I | TraesCS1B02G432700.1.E1 | TraesCS1B02G432700 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS1D02G168700.2.E4 | TraesCS1D02G168700 | NADPH-protochlorophyllide oxidoreductase |
| I | TraesCS1D02G301100.1.E3 | TraesCS1D02G301100 | Serine/threonine-protein kinase STN8%2C chloroplastic |
| I | TraesCS1D02G400100.1.E2 | TraesCS1D02G400100 | Photosystem I reaction center subunit VI%2C chloroplast precursor (PSI- H) (Light-harvesting complex I 11 kDa protein) (GOS5 protein) |
| I | TraesCS1D02G411300.1.E1 | TraesCS1D02G411300 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS1D02G428200.1.E2 | TraesCS1D02G428200 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2A02G114400.1.E3 | TraesCS2A02G114400 | N/A |
| I | TraesCS2A02G133900.1.E5 | TraesCS2A02G133900 | N/A |
| I | TraesCS2A02G187200.1.E3 | TraesCS2A02G187200 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2A02G204800.1.E2 | TraesCS2A02G204800 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2A02G206200.1.E2 | TraesCS2A02G206200 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2A02G223600.1.E1 | TraesCS2A02G223600 | N/A |
| I | TraesCS2A02G250800.1.E10 | TraesCS2A02G250800 | Similar to Glycolate oxidase (EC 1.1.3.15) (Fragment) |
| I | TraesCS2A02G252600.1.E1 | TraesCS2A02G252600 | N/A |
| I | TraesCS2A02G252600.1.E2 | TraesCS2A02G252600 | N/A |
| I | TraesCS2A02G254600.3.E3 | TraesCS2A02G254600 | N/A |
| I | TraesCS2A02G342600.1.E2 | TraesCS2A02G342600 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2A02G344600.1.E4 | TraesCS2A02G344600 | Glyceraldehyde-3-phosphate dehydrogenase |
| I | TraesCS2A02G368300.1.E1 | TraesCS2A02G368300 | Peroxisomal protein%2C Response to salt and low nitrogen stresse |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| I | TraesCS2B02G079400.1.E2 | TraesCS2B02G079400 | Ribulose bisphosphate carboxylase small chain |
| I | TraesCS2B02G099500.1.E8 | TraesCS2B02G099500 | N/A |
| I | TraesCS2B02G157500.1.E5 | TraesCS2B02G157500 | N/A |
| I | TraesCS2B02G220100.1.E3 | TraesCS2B02G220100 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2B02G233400.1.E2 | TraesCS2B02G233400 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2B02G234500.1.E3 | TraesCS2B02G234500 | Cytochrome b6-f complex iron-sulfur subunit |
| I | TraesCS2B02G234500.1.E4 | TraesCS2B02G234500 | Cytochrome b6-f complex iron-sulfur subunit |
| I | TraesCS2B02G264100.2.E10 | TraesCS2B02G264100 | Similar to Glycolate oxidase (EC 1.1.3.15) (Fragment) |
| I | TraesCS2B02G267500.1.E3 | TraesCS2B02G267500 | N/A |
| I | TraesCS2B02G272300.1.E1 | TraesCS2B02G272300 | N/A |
| I | TraesCS2B02G272300.1.E2 | TraesCS2B02G272300 | N/A |
| I | TraesCS2B02G273900.1.E1 | TraesCS2B02G273900 | N/A |
| I | TraesCS2B02G278500.1.E4 | TraesCS2B02G278500 | Peroxidase |
| I | TraesCS2B02G340300.1.E2 | TraesCS2B02G340300 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2B02G342200.1.E4 | TraesCS2B02G342200 | Glyceraldehyde-3-phosphate dehydrogenase |
| I | TraesCS2D02G065600.1.E2 | TraesCS2D02G065600 | Ribulose bisphosphate carboxylase small chain |
| I | TraesCS2D02G115200.1.E3 | TraesCS2D02G115200 | N/A |
| I | TraesCS2D02G136100.1.E5 | TraesCS2D02G136100 | N/A |
| I | TraesCS2D02G209900.1.E2 | TraesCS2D02G209900 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2D02G214700.1.E3 | TraesCS2D02G214700 | Cytochrome b6-f complex iron-sulfur subunit%2C chloroplastic |
| I | TraesCS2D02G214700.1.E4 | TraesCS2D02G214700 | Cytochrome b6-f complex iron-sulfur subunit%2C chloroplastic |
| I | TraesCS2D02G229600.1.E1 | TraesCS2D02G229600 | N/A |
| I | TraesCS2D02G253200.1.E1 | TraesCS2D02G253200 | N/A |
| I | TraesCS2D02G253200.1.E2 | TraesCS2D02G253200 | N/A |
| I | TraesCS2D02G255100.1.E2 | TraesCS2D02G255100 | N/A |
| I | TraesCS2D02G255100.1.E3 | TraesCS2D02G255100 | N/A |
| I | TraesCS2D02G262400.2.E1 | TraesCS2D02G262400 | N/A |
| I | TraesCS2D02G320900.1.E2 | TraesCS2D02G320900 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS2D02G322900.1.E4 | TraesCS2D02G322900 | Glyceraldehyde-3-phosphate dehydrogenase |
| I | TraesCS2D02G346900.2.E1 | TraesCS2D02G346900 | N/A |
| I | TraesCS2D02G467900.1.E3 | TraesCS2D02G467900 | Aminomethyltransferase |
| I | TraesCS3A02G028700.1.E2 | TraesCS3A02G028700 | N/A |
| I | TraesCS3A02G052700.1.E1 | TraesCS3A02G052700 | N/A |
| I | TraesCS3A02G052700.1.E2 | TraesCS3A02G052700 | N/A |
| I | TraesCS3A02G094900.1.E1 | TraesCS3A02G094900 | N/A |
| I | TraesCS3A02G212000.1.E1 | TraesCS3A02G212000 | N/A |
| I | TraesCS3A02G367000.1.E8 | TraesCS3A02G367000 | Sedoheptulose-1%2C7-bisphosphatase%2C chloroplastic |
| I | TraesCS3A02G381000.1.E3 | TraesCS3A02G381000 | N/A |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| I | TraesCS3A02G381000.1.E4 | TraesCS3A02G381000 | N/A |
| I | TraesCS3A02G398500.1.E12 | TraesCS3A02G398500 | N/A |
| I | TraesCS3A02G418300.1.E4 | TraesCS3A02G418300 | N/A |
| I | TraesCS3B02G061700.1.E1 | TraesCS3B02G061700 | N/A |
| I | TraesCS3B02G061700.1.E2 | TraesCS3B02G061700 | N/A |
| I | TraesCS3B02G110400.1.E1 | TraesCS3B02G110400 | N/A |
| I | TraesCS3B02G240500.1.E2 | TraesCS3B02G240500 | N/A |
| I | TraesCS3B02G242400.1.E1 | TraesCS3B02G242400 | N/A |
| I | TraesCS3B02G344200.1.E2 | TraesCS3B02G344200 | Similar to Photosystem II reaction center W protein (PSII 6.1 kDa protein) (Fragment) |
| I | TraesCS3B02G398300.1.E5 | TraesCS3B02G398300 | Sedoheptulose-1%2C7-bisphosphatase%2C chloroplast%2C expressed |
| I | TraesCS3B02G398300.1.E8 | TraesCS3B02G398300 | Sedoheptulose-1%2C7-bisphosphatase%2C chloroplast%2C expressed |
| I | TraesCS3B02G410400.1.E1 | TraesCS3B02G410400 | Fructose-1%2C6-bisphosphatase%2C cytosolic%2Cputative%2Cexpressed |
| I | TraesCS3B02G410400.1.E11 | TraesCS3B02G410400 | Fructose-1%2C6-bisphosphatase%2C cytosolic%2Cputative%2Cexpressed |
| I | TraesCS3B02G413700.1.E3 | TraesCS3B02G413700 | N/A |
| I | TraesCS3B02G413700.1.E4 | TraesCS3B02G413700 | N/A |
| I | TraesCS3B02G430900.1.E12 | TraesCS3B02G430900 | N/A |
| I | TraesCS3B02G453500.1.E4 | TraesCS3B02G453500 | N/A |
| I | TraesCS3D02G051800.1.E1 | TraesCS3D02G051800 | N/A |
| I | TraesCS3D02G051800.1.E2 | TraesCS3D02G051800 | N/A |
| I | TraesCS3D02G160600.1.E4 | TraesCS3D02G160600 | N/A |
| I | TraesCS3D02G214600.1.E1 | TraesCS3D02G214600 | N/A |
| I | TraesCS3D02G309400.1.E2 | TraesCS3D02G309400 | Similar to Photosystem II reaction center W protein (PSII 6.1 kDa protein) (Fragment) |
| I | TraesCS3D02G359900.1.E5 | TraesCS3D02G359900 | N/A |
| I | TraesCS3D02G359900.1.E8 | TraesCS3D02G359900 | N/A |
| I | TraesCS3D02G370700.1.E11 | TraesCS3D02G370700 | N/A |
| I | TraesCS3D02G374200.1.E3 | TraesCS3D02G374200 | N/A |
| I | TraesCS3D02G374200.1.E4 | TraesCS3D02G374200 | N/A |
| I | TraesCS3D02G392600.1.E12 | TraesCS3D02G392600 | N/A |
| I | TraesCS4A02G093100.1.E2 | TraesCS4A02G093100 | Fructose-1%2C6-bisphosphatase%2C chloroplastic |
| I | TraesCS4A02G093100.1.E3 | TraesCS4A02G093100 | Fructose-1%2C6-bisphosphatase%2C chloroplastic |
| I | TraesCS4A02G099800.1.E1 | TraesCS4A02G099800 | N/A |
| I | TraesCS4A02G110700.1.E2 | TraesCS4A02G110700 | N/A |
| I | TraesCS4A02G177500.1.E7 | TraesCS4A02G177500 | N/A |
| I | TraesCS4A02G177500.3.E3 | TraesCS4A02G177500 | N/A |
| I | TraesCS4A02G190400.2.E1 | TraesCS4A02G190400 | N/A |
| I | TraesCS4A02G206400.1.E1 | TraesCS4A02G206400 | Fructose-bisphosphate aldolase |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| I | TraesCS4A02G206400.1.E2 | TraesCS4A02G206400 | Fructose-bisphosphate aldolase |
| I | TraesCS4A02G206400.1.E5 | TraesCS4A02G206400 | Fructose-bisphosphate aldolase |
| I | TraesCS4A02G206400.1.E6 | TraesCS4A02G206400 | Fructose-bisphosphate aldolase |
| I | TraesCS4A02G226900.3.E4 | TraesCS4A02G226900 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS4A02G246100.1.E9 | TraesCS4A02G246100 | Serine hydroxymethyltransferase |
| I | TraesCS4B02G069300.1.E9 | TraesCS4B02G069300 | Serine hydroxymethyltransferase |
| I | TraesCS4B02G089000.1.E1 | TraesCS4B02G089000 | N/A |
| I | TraesCS4B02G089500.1.E1 | TraesCS4B02G089500 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS4B02G089500.1.E6 | TraesCS4B02G089500 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS4B02G109900.1.E1 | TraesCS4B02G109900 | Fructose-bisphosphate aldolase |
| I | TraesCS4B02G109900.1.E2 | TraesCS4B02G109900 | Fructose-bisphosphate aldolase |
| I | TraesCS4B02G109900.1.E5 | TraesCS4B02G109900 | Fructose-bisphosphate aldolase |
| I | TraesCS4B02G109900.1.E6 | TraesCS4B02G109900 | Fructose-bisphosphate aldolase |
| I | TraesCS4B02G140300.1.E3 | TraesCS4B02G140300 | N/A |
| I | TraesCS4B02G140300.1.E7 | TraesCS4B02G140300 | N/A |
| I | TraesCS4B02G193400.1.E2 | TraesCS4B02G193400 | N/A |
| I | TraesCS4B02G204500.1.E1 | TraesCS4B02G204500 | N/A |
| I | TraesCS4B02G211400.1.E3 | TraesCS4B02G211400 | N/A |
| I | TraesCS4B02G323000.1.E7 | TraesCS4B02G323000 | Glyceraldehyde-3-phosphate dehydrogenase |
| I | TraesCS4D02G041200.1.E2 | TraesCS4D02G041200 | Peptidylprolyl isomerase |
| I | TraesCS4D02G086400.1.E1 | TraesCS4D02G086400 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS4D02G086400.1.E6 | TraesCS4D02G086400 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS4D02G107400.1.E1 | TraesCS4D02G107400 | Fructose-bisphosphate aldolase |
| I | TraesCS4D02G107400.1.E2 | TraesCS4D02G107400 | Fructose-bisphosphate aldolase |
| I | TraesCS4D02G107400.1.E5 | TraesCS4D02G107400 | Fructose-bisphosphate aldolase |
| I | TraesCS4D02G107400.1.E6 | TraesCS4D02G107400 | Fructose-bisphosphate aldolase |
| I | TraesCS4D02G135000.9.E7 | TraesCS4D02G135000 | N/A |
| I | TraesCS4D02G194400.1.E2 | TraesCS4D02G194400 | N/A |
| I | TraesCS4D02G205400.1.E1 | TraesCS4D02G205400 | N/A |
| I | TraesCS4D02G212000.1.E3 | TraesCS4D02G212000 | N/A |
| I | TraesCS4D02G319400.1.E7 | TraesCS4D02G319400 | Glyceraldehyde-3-phosphate dehydrogenase |
| I | TraesCS4D02G348600.1.E8 | TraesCS4D02G348600 | Peroxisomal protein%2C Response to salt and low nitrogen stresse |
| I | TraesCS5A02G036200.1.E10 | TraesCS5A02G036200 | N/A |
| I | TraesCS5A02G036200.1.E7 | TraesCS5A02G036200 | N/A |
| I | TraesCS5A02G108000.1.E1 | TraesCS5A02G108000 | Fructose-bisphosphate aldolase |
| I | TraesCS5A02G108000.1.E2 | TraesCS5A02G108000 | Fructose-bisphosphate aldolase |
| I | TraesCS5A02G110300.1.E1 | TraesCS5A02G110300 | N/A |
| I | TraesCS5A02G110300.1.E2 | TraesCS5A02G110300 | N/A |
| I | TraesCS5A02G165700.1.E1 | TraesCS5A02G165700 | Ribulose bisphosphate carboxylase small chain |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| I | TraesCS5A02G165700.1.E2 | TraesCS5A02G165700 | Ribulose bisphosphate carboxylase small chain |
| I | TraesCS5A02G229300.1.E2 | TraesCS5A02G229300 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS5A02G256900.1.E1 | TraesCS5A02G256900 | N/A |
| I | TraesCS5A02G322500.1.E2 | TraesCS5A02G322500 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS5A02G350600.1.E1 | TraesCS5A02G350600 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS5A02G402600.1.E2 | TraesCS5A02G402600 | N/A |
| I | TraesCS5A02G424400.1.E2 | TraesCS5A02G424400 | N/A |
| I | TraesCS5A02G424400.1.E9 | TraesCS5A02G424400 | N/A |
| I | TraesCS5A02G457500.1.E1 | TraesCS5A02G457500 | N/A |
| I | TraesCS5A02G461500.1.E3 | TraesCS5A02G461500 | N/A |
| I | TraesCS5A02G482800.1.E3 | TraesCS5A02G482800 | N/A |
| I | TraesCS5A02G495100.1.E7 | TraesCS5A02G495100 | Glyceraldehyde-3-phosphate dehydrogenase |
| I | TraesCS5B02G036800.1.E11 | TraesCS5B02G036800 | N/A |
| I | TraesCS5B02G036800.1.E8 | TraesCS5B02G036800 | N/A |
| I | TraesCS5B02G111800.1.E1 | TraesCS5B02G111800 | Thioredoxin M-type%2C chloroplastic |
| I | TraesCS5B02G111800.1.E2 | TraesCS5B02G111800 | Thioredoxin M-type%2C chloroplastic |
| I | TraesCS5B02G115300.1.E2 | TraesCS5B02G115300 | Fructose-bisphosphate aldolase |
| I | TraesCS5B02G162800.1.E2 | TraesCS5B02G162800 | Ribulose bisphosphate carboxylase small chain |
| I | TraesCS5B02G256400.1.E1 | TraesCS5B02G256400 | Photosystem 1 subunit 5 |
| I | TraesCS5B02G322900.1.E2 | TraesCS5B02G322900 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS5B02G407400.1.E2 | TraesCS5B02G407400 | N/A |
| I | TraesCS5B02G426500.1.E9 | TraesCS5B02G426500 | N/A |
| I | TraesCS5B02G471700.1.E3 | TraesCS5B02G471700 | N/A |
| I | TraesCS5B02G496000.1.E3 | TraesCS5B02G496000 | N/A |
| I | TraesCS5B02G534400.2.E10 | TraesCS5B02G534400 | N/A |
| I | TraesCS5D02G044400.1.E10 | TraesCS5D02G044400 | N/A |
| I | TraesCS5D02G044400.1.E7 | TraesCS5D02G044400 | N/A |
| I | TraesCS5D02G090700.1.E3 | TraesCS5D02G090700 | N/A |
| I | TraesCS5D02G118500.1.E1 | TraesCS5D02G118500 | N/A |
| I | TraesCS5D02G118500.1.E2 | TraesCS5D02G118500 | N/A |
| I | TraesCS5D02G238300.1.E2 | TraesCS5D02G238300 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS5D02G265500.1.E1 | TraesCS5D02G265500 | N/A |
| I | TraesCS5D02G329200.1.E2 | TraesCS5D02G329200 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS5D02G357600.1.E1 | TraesCS5D02G357600 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS5D02G412600.1.E1 | TraesCS5D02G412600 | N/A |
| I | TraesCS5D02G422800.1.E1 | TraesCS5D02G422800 | N/A |
| I | TraesCS5D02G432900.1.E10 | TraesCS5D02G432900 | N/A |
| I | TraesCS5D02G474200.1.E2 | TraesCS5D02G474200 | N/A |
| I | TraesCS5D02G496400.2.E3 | TraesCS5D02G496400 | N/A |
| I | TraesCS5D02G502900.1.E1 | TraesCS5D02G502900 | N/A |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| I | TraesCS5D02G532100.1.E10 | TraesCS5D02G532100 | N/A |
| I | TraesCS6A02G086300.1.E1 | TraesCS6A02G086300 | N/A |
| I | TraesCS6A02G159800.1.E1 | TraesCS6A02G159800 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS6A02G159800.1.E3 | TraesCS6A02G159800 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS6A02G209800.1.E1 | TraesCS6A02G209800 | Elongation factor Tu |
| I | TraesCS6A02G267300.1.E1 | TraesCS6A02G267300 | Phosphoribulokinase |
| I | TraesCS6A02G303800.1.E1 | TraesCS6A02G303800 | PDE332 |
| I | TraesCS6A02G307700.1.E3 | TraesCS6A02G307700 | Geranylgeranyl diphosphate reductase%2C chloroplastic |
| I | TraesCS6A02G358100.1.E4 | TraesCS6A02G358100 | N/A |
| I | TraesCS6B02G015000.1.E5 | TraesCS6B02G015000 | Ferredoxin--NADP reductase%2C chloroplastic |
| I | TraesCS6B02G115900.1.E1 | TraesCS6B02G115900 | N/A |
| I | TraesCS6B02G191500.1.E1 | TraesCS6B02G191500 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS6B02G191500.1.E3 | TraesCS6B02G191500 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS6B02G197100.1.E1 | TraesCS6B02G197100 | N/A |
| I | TraesCS6B02G238600.1.E1 | TraesCS6B02G238600 | Elongation factor Tu |
| I | TraesCS6B02G248700.1.E3 | TraesCS6B02G248700 | N/A |
| I | TraesCS6B02G294700.1.E1 | TraesCS6B02G294700 | Phosphoribulokinase |
| I | TraesCS6B02G336300.1.E3 | TraesCS6B02G336300 | Geranylgeranyl diphosphate reductase%2C chloroplastic |
| I | TraesCS6B02G390600.1.E4 | TraesCS6B02G390600 | N/A |
| I | TraesCS6D02G001700.1.E8 | TraesCS6D02G001700 | hydroxypyruvate reductase |
| I | TraesCS6D02G012200.1.E5 | TraesCS6D02G012200 | Ferredoxin--NADP reductase%2C chloroplastic |
| I | TraesCS6D02G152700.1.E2 | TraesCS6D02G152700 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS6D02G152700.1.E4 | TraesCS6D02G152700 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS6D02G158500.1.E1 | TraesCS6D02G158500 | N/A |
| I | TraesCS6D02G193000.1.E1 | TraesCS6D02G193000 | Elongation factor Tu |
| I | TraesCS6D02G247400.1.E1 | TraesCS6D02G247400 | Phosphoribulokinase%2C chloroplastic |
| I | TraesCS6D02G283300.1.E1 | TraesCS6D02G283300 | PDE332 |
| I | TraesCS6D02G286900.1.E3 | TraesCS6D02G286900 | Geranylgeranyl diphosphate reductase%2C chloroplastic |
| I | TraesCS6D02G340700.1.E4 | TraesCS6D02G340700 | N/A |
| I | TraesCS6D02G370200.1.E3 | TraesCS6D02G370200 | N/A |
| I | TraesCS7A02G000400.1.E1 | TraesCS7A02G000400 | Plastocyanin |
| I | TraesCS7A02G227100.1.E5 | TraesCS7A02G227100 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS7A02G325400.1.E1 | TraesCS7A02G325400 | Ferredoxin%2C chloroplastic |
| I | TraesCS7A02G325500.1.E1 | TraesCS7A02G325500 | Ferredoxin |
| I | TraesCS7A02G381100.2.E2 | TraesCS7A02G381100 | N/A |
| I | TraesCS7B02G065200.1.E4 | TraesCS7B02G065200 | N/A |
| I | TraesCS7B02G167500.1.E4 | TraesCS7B02G167500 | N/A |
| I | TraesCS7B02G167600.1.E7 | TraesCS7B02G167600 | N/A |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| I | TraesCS7B02G192500.1.E4 | TraesCS7B02G192500 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS7B02G226100.1.E1 | TraesCS7B02G226100 | Ferredoxin%2C chloroplastic |
| I | TraesCS7B02G226200.1.E1 | TraesCS7B02G226200 | Ferredoxin |
| I | TraesCS7B02G283000.1.E2 | TraesCS7B02G283000 | Fructose-bisphosphate aldolase |
| I | TraesCS7B02G283000.1.E3 | TraesCS7B02G283000 | Fructose-bisphosphate aldolase |
| I | TraesCS7D02G000300.1.E1 | TraesCS7D02G000300 | Plastocyanin |
| I | TraesCS7D02G161000.1.E3 | TraesCS7D02G161000 | N/A |
| I | TraesCS7D02G227300.1.E4 | TraesCS7D02G227300 | Chlorophyll a-b binding protein%2C chloroplastic |
| I | TraesCS7D02G322000.1.E1 | TraesCS7D02G322000 | Ferredoxin |
| I | TraesCS7D02G377400.1.E2 | TraesCS7D02G377400 | Fructose-bisphosphate aldolase |
| I | TraesCS7D02G377400.1.E3 | TraesCS7D02G377400 | Fructose-bisphosphate aldolase |
| I | TraesCS7D02G382300.1.E1 | TraesCS7D02G382300 | N/A |
| I | TraesCSU02G077300.1.E1 | TraesCSU02G077300 | N/A |
| II | TraesCS1A02G097000.1.E1 | TraesCS1A02G097000 | N/A |
| II | TraesCS1A02G251200.1.E1 | TraesCS1A02G251200 | N/A |
| II | TraesCS1A02G392000.1.E1 | TraesCS1A02G392000 | Photosystem I reaction center subunit VI%2C chloroplast precursor (PSI- H) (Light-harvesting complex I 11 kDa protein) (GOS5 protein) |
| II | TraesCS1B02G186300.1.E3 | TraesCS1B02G186300 | NADPH-protochlorophyllide oxidoreductase |
| II | TraesCS1B02G261800.1.E1 | TraesCS1B02G261800 | N/A |
| II | TraesCS1D02G400100.1.E1 | TraesCS1D02G400100 | Photosystem I reaction center subunit VI%2C chloroplast precursor (PSI- H) (Light-harvesting complex I 11 kDa protein) (GOS5 protein) |
| II | TraesCS2A02G114400.1.E1 | TraesCS2A02G114400 | N/A |
| II | TraesCS2A02G187200.1.E1 | TraesCS2A02G187200 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2A02G206200.1.E1 | TraesCS2A02G206200 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2A02G207900.1.E3 | TraesCS2A02G207900 | Cytochrome b6-f complex iron-sulfur subunit |
| II | TraesCS2A02G215000.1.E1 | TraesCS2A02G215000 | N/A |
| II | TraesCS2A02G225300.1.E1 | TraesCS2A02G225300 | N/A |
| II | TraesCS2A02G247300.3.E1 | TraesCS2A02G247300 | N/A |
| II | TraesCS2A02G344600.1.E3 | TraesCS2A02G344600 | Glyceraldehyde-3-phosphate dehydrogenase |
| II | TraesCS2A02G348500.1.E1 | TraesCS2A02G348500 | N/A |
| II | TraesCS2A02G371100.1.E5 | TraesCS2A02G371100 | N/A |
| II | TraesCS2A02G467900.1.E1 | TraesCS2A02G467900 | Aminomethyltransferase |
| II | TraesCS2A02G500400.2.E1 | TraesCS2A02G500400 | Glutamine synthetase |
| II | TraesCS2B02G220100.1.E1 | TraesCS2B02G220100 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2B02G233400.1.E1 | TraesCS2B02G233400 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2B02G234500.1.E2 | TraesCS2B02G234500 | Cytochrome b6-f complex iron-sulfur subunit |
| II | TraesCS2B02G240000.1.E1 | TraesCS2B02G240000 | N/A |
| II | TraesCS2B02G260200.1.E1 | TraesCS2B02G260200 | N/A |
| II | TraesCS2B02G270300.1.E1 | TraesCS2B02G270300 | Oxygen evolving enhancer protein |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| II | TraesCS2B02G340300.1.E1 | TraesCS2B02G340300 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2B02G342200.1.E3 | TraesCS2B02G342200 | Glyceraldehyde-3-phosphate dehydrogenase |
| II | TraesCS2B02G405600.1.E1 | TraesCS2B02G405600 | N/A |
| II | TraesCS2D02G115200.1.E1 | TraesCS2D02G115200 | N/A |
| II | TraesCS2D02G179700.1.E1 | TraesCS2D02G179700 | N/A |
| II | TraesCS2D02G200700.1.E1 | TraesCS2D02G200700 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2D02G200700.1.E3 | TraesCS2D02G200700 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2D02G200700.1.E4 | TraesCS2D02G200700 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2D02G209900.1.E1 | TraesCS2D02G209900 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS2D02G214700.1.E2 | TraesCS2D02G214700 | Cytochrome b6-f complex iron-sulfur subunit%2C chloroplastic |
| II | TraesCS2D02G220800.1.E1 | TraesCS2D02G220800 | N/A |
| II | TraesCS2D02G248400.2.E1 | TraesCS2D02G248400 | Oxygen evolving enhancer protein |
| II | TraesCS2D02G255100.1.E1 | TraesCS2D02G255100 | N/A |
| II | TraesCS2D02G467900.1.E1 | TraesCS2D02G467900 | Aminomethyltransferase |
| II | TraesCS3A02G250200.1.E1 | TraesCS3A02G250200 | Glycine cleavage system P protein |
| II | TraesCS3A02G304900.1.E1 | TraesCS3A02G304900 | N/A |
| II | TraesCS3A02G367000.1.E5 | TraesCS3A02G367000 | Sedoheptulose-1%2C7-bisphosphatase%2C chloroplastic |
| II | TraesCS3A02G381000.1.E1 | TraesCS3A02G381000 | N/A |
| II | TraesCS3A02G443700.1.E2 | TraesCS3A02G443700 | PnsB5 |
| II | TraesCS3B02G279700.1.E1 | TraesCS3B02G279700 | Glycine cleavage system P protein |
| II | TraesCS3B02G344200.1.E1 | TraesCS3B02G344200 | Similar to Photosystem II reaction center W protein (PSII 6.1 kDa protein) (Fragment) |
| II | TraesCS3B02G413700.1.E1 | TraesCS3B02G413700 | N/A |
| II | TraesCS3B02G477300.1.E1 | TraesCS3B02G477300 | PnsB5 |
| II | TraesCS3D02G084800.1.E1 | TraesCS3D02G084800 | N/A |
| II | TraesCS3D02G250600.1.E1 | TraesCS3D02G250600 | Glycine cleavage system P protein |
| II | TraesCS3D02G309400.1.E1 | TraesCS3D02G309400 | Similar to Photosystem II reaction center W protein (PSII 6.1 kDa protein) (Fragment) |
| II | TraesCS3D02G370700.1.E1 | TraesCS3D02G370700 | N/A |
| II | TraesCS3D02G374200.1.E1 | TraesCS3D02G374200 | N/A |
| II | TraesCS4A02G093100.1.E1 | TraesCS4A02G093100 | Fructose-1%2C6-bisphosphatase%2C chloroplastic |
| II | TraesCS4A02G120600.1.E1 | TraesCS4A02G120600 | N/A |
| II | TraesCS4A02G319000.2.E1 | TraesCS4A02G319000 | N/A |
| II | TraesCS4A02G342800.1.E1 | TraesCS4A02G342800 | N/A |
| II | TraesCS4A02G499500.1.E1 | TraesCS4A02G499500 | Plastocyanin |
| II | TraesCS4B02G043700.1.E3 | TraesCS4B02G043700 | Peptidylprolyl isomerase |
| II | TraesCS4B02G211400.1.E1 | TraesCS4B02G211400 | N/A |
| II | TraesCS4B02G274300.1.E1 | TraesCS4B02G274300 | NDH-dependent cyclic electron flow 1 |
| II | TraesCS4B02G323000.1.E6 | TraesCS4B02G323000 | Glyceraldehyde-3-phosphate dehydrogenase |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| II | TraesCS4B02G325800.1.E3 | TraesCS4B02G325800 | Catalase |
| II | TraesCS4B02G326700.1.E1 | TraesCS4B02G326700 | Signal recognition particle 43 kDa protein%2C chloroplastic |
| II | TraesCS4D02G212000.1.E1 | TraesCS4D02G212000 | N/A |
| II | TraesCS4D02G212000.1.E2 | TraesCS4D02G212000 | N/A |
| II | TraesCS4D02G319400.1.E6 | TraesCS4D02G319400 | Glyceraldehyde-3-phosphate dehydrogenase |
| II | TraesCS4D02G323400.1.E1 | TraesCS4D02G323400 | Signal recognition particle 43 kDa protein%2C chloroplastic |
| II | TraesCS5A02G414400.1.E1 | TraesCS5A02G414400 | N/A |
| II | TraesCS5A02G424400.1.E6 | TraesCS5A02G424400 | N/A |
| II | TraesCS5A02G482800.1.E1 | TraesCS5A02G482800 | N/A |
| II | TraesCS5A02G495100.1.E1 | TraesCS5A02G495100 | Glyceraldehyde-3-phosphate dehydrogenase |
| II | TraesCS5A02G495100.1.E6 | TraesCS5A02G495100 | Glyceraldehyde-3-phosphate dehydrogenase |
| II | TraesCS5A02G498000.1.E3 | TraesCS5A02G498000 | Catalase |
| II | TraesCS5A02G498700.1.E1 | TraesCS5A02G498700 | Signal recognition particle 43 kDa protein%2C chloroplastic |
| II | TraesCS5B02G417400.1.E1 | TraesCS5B02G417400 | N/A |
| II | TraesCS5B02G426500.1.E6 | TraesCS5B02G426500 | N/A |
| II | TraesCS5B02G466700.1.E1 | TraesCS5B02G466700 | N/A |
| II | TraesCS5B02G496000.1.E1 | TraesCS5B02G496000 | N/A |
| II | TraesCS5B02G560400.2.E1 | TraesCS5B02G560400 | N/A |
| II | TraesCS5D02G432900.1.E7 | TraesCS5D02G432900 | N/A |
| II | TraesCS5D02G468400.1.E1 | TraesCS5D02G468400 | N/A |
| II | TraesCS5D02G496400.2.E1 | TraesCS5D02G496400 | N/A |
| II | TraesCS5D02G569800.1.E1 | TraesCS5D02G569800 | N/A |
| II | TraesCS6A02G105400.1.E1 | TraesCS6A02G105400 | N/A |
| II | TraesCS6A02G219200.1.E1 | TraesCS6A02G219200 | N/A |
| II | TraesCS6A02G307700.1.E1 | TraesCS6A02G307700 | Geranylgeranyl diphosphate reductase%2C chloroplastic |
| II | TraesCS6A02G307700.1.E2 | TraesCS6A02G307700 | Geranylgeranyl diphosphate reductase%2C chloroplastic |
| II | TraesCS6A02G358100.1.E1 | TraesCS6A02G358100 | N/A |
| II | TraesCS6A02G372100.1.E1 | TraesCS6A02G372100 | Thioredoxin domain 2 containing protein |
| II | TraesCS6B02G015000.1.E1 | TraesCS6B02G015000 | Ferredoxin--NADP reductase%2C chloroplastic |
| II | TraesCS6B02G248700.1.E1 | TraesCS6B02G248700 | N/A |
| II | TraesCS6B02G336300.1.E1 | TraesCS6B02G336300 | Geranylgeranyl diphosphate reductase%2C chloroplastic |
| II | TraesCS6B02G336300.1.E2 | TraesCS6B02G336300 | Geranylgeranyl diphosphate reductase%2C chloroplastic |
| II | TraesCS6B02G390600.1.E1 | TraesCS6B02G390600 | N/A |
| II | TraesCS6D02G012200.1.E1 | TraesCS6D02G012200 | Ferredoxin--NADP reductase%2C chloroplastic |
| II | TraesCS6D02G202300.1.E1 | TraesCS6D02G202300 | N/A |
| II | TraesCS6D02G286900.1.E1 | TraesCS6D02G286900 | Geranylgeranyl diphosphate reductase%2C chloroplastic |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| II | TraesCS6D02G286900.1.E2 | TraesCS6D02G286900 | Geranylgeranyl diphosphate reductase%2C chloroplastic |
| II | TraesCS6D02G299200.1.E1 | TraesCS6D02G299200 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS6D02G340700.1.E1 | TraesCS6D02G340700 | N/A |
| II | TraesCS7A02G020500.1.E1 | TraesCS7A02G020500 | Ferredoxin--NADP reductase%2C chloroplastic |
| II | TraesCS7A02G068500.1.E1 | TraesCS7A02G068500 | N/A |
| II | TraesCS7A02G227100.1.E2 | TraesCS7A02G227100 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS7A02G227100.1.E4 | TraesCS7A02G227100 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS7A02G291000.1.E1 | TraesCS7A02G291000 | N/A |
| II | TraesCS7A02G385700.1.E1 | TraesCS7A02G385700 | N/A |
| II | TraesCS7A02G485300.1.E1 | TraesCS7A02G485300 | Peptidylprolyl isomerase |
| II | TraesCS7B02G148000.1.E3 | TraesCS7B02G148000 | N/A |
| II | TraesCS7B02G192500.1.E1 | TraesCS7B02G192500 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS7B02G192500.1.E3 | TraesCS7B02G192500 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS7B02G288600.1.E1 | TraesCS7B02G288600 | N/A |
| II | TraesCS7D02G016800.1.E1 | TraesCS7D02G016800 | Ferredoxin--NADP reductase%2C chloroplastic |
| II | TraesCS7D02G161000.1.E1 | TraesCS7D02G161000 | N/A |
| II | TraesCS7D02G227300.1.E1 | TraesCS7D02G227300 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS7D02G227300.1.E3 | TraesCS7D02G227300 | Chlorophyll a-b binding protein%2C chloroplastic |
| II | TraesCS7D02G472600.1.E1 | TraesCS7D02G472600 | Peptidylprolyl isomerase |
| III | TraesCS1D02G116400.1.E6 | TraesCS1D02G116400 | N/A |
| III | TraesCS1D02G313800.2.E6 | TraesCS1D02G313800 | Phosphoglycerate kinase |
| III | TraesCS2A02G207900.1.E5 | TraesCS2A02G207900 | Cytochrome b6-f complex iron-sulfur subunit |
| III | TraesCS2A02G261600.1.E12 | TraesCS2A02G261600 | N/A |
| III | TraesCS2A02G387000.1.E2 | TraesCS2A02G387000 | Elongation factor G%2C chloroplastic |
| III | TraesCS2A02G387000.1.E3 | TraesCS2A02G387000 | Elongation factor G%2C chloroplastic |
| III | TraesCS2A02G387000.1.E4 | TraesCS2A02G387000 | Elongation factor G%2C chloroplastic |
| III | TraesCS2A02G387800.1.E2 | TraesCS2A02G387800 | N/A |
| III | TraesCS2B02G280600.2.E12 | TraesCS2B02G280600 | N/A |
| III | TraesCS2B02G404600.1.E2 | TraesCS2B02G404600 | Elongation factor G%2C chloroplastic |
| III | TraesCS2B02G404600.1.E3 | TraesCS2B02G404600 | Elongation factor G%2C chloroplastic |
| III | TraesCS2B02G404600.1.E4 | TraesCS2B02G404600 | Elongation factor G%2C chloroplastic |
| III | TraesCS2B02G405600.1.E2 | TraesCS2B02G405600 | N/A |
| III | TraesCS2D02G057900.1.E5 | TraesCS2D02G057900 | Protein PHOTOSYSTEM I ASSEMBLY 2%2C chloroplastic |
| III | TraesCS2D02G383800.1.E2 | TraesCS2D02G383800 | Elongation factor G%2C chloroplastic |
| III | TraesCS2D02G383800.1.E3 | TraesCS2D02G383800 | Elongation factor G%2C chloroplastic |
| III | TraesCS2D02G383800.1.E4 | TraesCS2D02G383800 | Elongation factor G%2C chloroplastic |
| III | TraesCS3A02G304900.1.E6 | TraesCS3A02G304900 | N/A |
| III | TraesCS3A02G443700.1.E6 | TraesCS3A02G443700 | PnsB5 |
| III | TraesCS3B02G477300.1.E5 | TraesCS3B02G477300 | PnsB5 |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| III | TraesCS3D02G436100.1.E5 | TraesCS3D02G436100 | PnsB5 |
| III | TraesCS4A02G190400.2.E5 | TraesCS4A02G190400 | N/A |
| III | TraesCS4A02G246100.1.E15 | TraesCS4A02G246100 | Serine hydroxymethyltransferase |
| III | TraesCS4A02G246100.1.E16 | TraesCS4A02G246100 | Serine hydroxymethyltransferase |
| III | TraesCS4A02G268900.1.E2 | TraesCS4A02G268900 | N/A |
| III | TraesCS4B02G045300.1.E2 | TraesCS4B02G045300 | N/A |
| III | TraesCS4B02G069300.1.E15 | TraesCS4B02G069300 | Serine hydroxymethyltransferase |
| III | TraesCS4B02G069300.1.E17 | TraesCS4B02G069300 | Serine hydroxymethyltransferase |
| III | TraesCS4B02G134600.1.E8 | TraesCS4B02G134600 | N/A |
| III | TraesCS4B02G285000.1.E9 | TraesCS4B02G285000 | Ribulose-phosphate 3-epimerase |
| III | TraesCS4B02G355100.1.E8 | TraesCS4B02G355100 | Putative PEX11-1 protein |
| III | TraesCS4D02G045200.1.E2 | TraesCS4D02G045200 | N/A |
| III | TraesCS4D02G068100.1.E15 | TraesCS4D02G068100 | Serine hydroxymethyltransferase |
| III | TraesCS4D02G068100.1.E16 | TraesCS4D02G068100 | Serine hydroxymethyltransferase |
| III | TraesCS4D02G283800.1.E9 | TraesCS4D02G283800 | Ribulose-phosphate 3-epimerase |
| III | TraesCS5A02G030400.1.E18 | TraesCS5A02G030400 | N/A |
| III | TraesCS5A02G230400.1.E5 | TraesCS5A02G230400 | N/A |
| III | TraesCS5A02G524400.1.E8 | TraesCS5A02G524400 | Peroxisomal protein%2C Response to salt and low nitrogen stresse |
| III | TraesCS5B02G035200.1.E3 | TraesCS5B02G035200 | N/A |
| III | TraesCS5B02G412500.3.E9 | TraesCS5B02G412500 | N/A |
| III | TraesCS5D02G122700.1.E2 | TraesCS5D02G122700 | Fructose-bisphosphate aldolase |
| III | TraesCS5D02G237400.1.E5 | TraesCS5D02G237400 | N/A |
| III | TraesCS5D02G417600.1.E9 | TraesCS5D02G417600 | Malate dehydrogenase |
| III | TraesCS5D02G569800.1.E10 | TraesCS5D02G569800 | N/A |
| III | TraesCS6A02G208200.1.E3 | TraesCS6A02G208200 | Phosphate transporter |
| III | TraesCS6A02G267300.1.E4 | TraesCS6A02G267300 | Phosphoribulokinase |
| III | TraesCS6A02G267300.1.E5 | TraesCS6A02G267300 | Phosphoribulokinase |
| III | TraesCS6A02G308100.1.E4 | TraesCS6A02G308100 | N/A |
| III | TraesCS6B02G015000.1.E6 | TraesCS6B02G015000 | Ferredoxin--NADP reductase%2C chloroplastic |
| III | TraesCS6B02G294700.1.E4 | TraesCS6B02G294700 | Phosphoribulokinase |
| III | TraesCS6B02G294700.1.E5 | TraesCS6B02G294700 | Phosphoribulokinase |
| III | TraesCS6D02G012200.1.E6 | TraesCS6D02G012200 | Ferredoxin--NADP reductase%2C chloroplastic |
| III | TraesCS6D02G183500.1.E14 | TraesCS6D02G183500 | N/A |
| III | TraesCS6D02G247400.1.E4 | TraesCS6D02G247400 | Phosphoribulokinase%2C chloroplastic |
| III | TraesCS6D02G247400.1.E5 | TraesCS6D02G247400 | Phosphoribulokinase%2C chloroplastic |
| III | TraesCS6D02G283300.1.E2 | TraesCS6D02G283300 | PDE332 |
| III | TraesCS6D02G287300.1.E4 | TraesCS6D02G287300 | N/A |
| III | TraesCS6D02G299200.1.E6 | TraesCS6D02G299200 | Chlorophyll a-b binding protein%2C chloroplastic |
| III | TraesCS7A02G020500.1.E6 | TraesCS7A02G020500 | Ferredoxin--NADP reductase%2C chloroplastic |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| III | TraesCS7A02G206700.2.E5 | TraesCS7A02G206700 | N/A |
| III | TraesCS7A02G294700.1.E6 | TraesCS7A02G294700 | N/A |
| III | TraesCS7B02G114100.1.E5 | TraesCS7B02G114100 | N/A |
| III | TraesCS7B02G189500.1.E6 | TraesCS7B02G189500 | N/A |
| III | TraesCS7B02G313800.1.E4 | TraesCS7B02G313800 | N/A |
| III | TraesCS7D02G016800.1.E6 | TraesCS7D02G016800 | Ferredoxin--NADP reductase%2C chloroplastic |
| III | TraesCS7D02G209600.2.E5 | TraesCS7D02G209600 | N/A |
| III | TraesCS7D02G472600.1.E3 | TraesCS7D02G472600 | Peptidylprolyl isomerase |
| IV | ENSRNA050013832.E1 | ENSRNA050013832 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050013838.E1 | ENSRNA050013838 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050013842.E1 | ENSRNA050013842 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050013848.E1 | ENSRNA050013848 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050013851.E1 | ENSRNA050013851 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050013852.E1 | ENSRNA050013852 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050013859.E1 | ENSRNA050013859 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050013861.E1 | ENSRNA050013861 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050013864.E1 | ENSRNA050013864 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050015671.E1 | ENSRNA050015671 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050015672.E1 | ENSRNA050015672 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050015673.E1 | ENSRNA050015673 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050015982.E1 | ENSRNA050015982 | Eukaryotic large subunit ribosomal RNA |
| IV | ENSRNA050016387.E1 | ENSRNA050016387 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050016549.E1 | ENSRNA050016549 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050016783.E1 | ENSRNA050016783 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050016807.E1 | ENSRNA050016807 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050018652.E1 | ENSRNA050018652 | Eukaryotic large subunit ribosomal RNA |
| IV | ENSRNA050018659.E1 | ENSRNA050018659 | Eukaryotic large subunit ribosomal RNA |
| IV | ENSRNA050018663.E1 | ENSRNA050018663 | Eukaryotic large subunit ribosomal RNA |
| IV | ENSRNA050018909.E1 | ENSRNA050018909 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050018910.E1 | ENSRNA050018910 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050018911.E1 | ENSRNA050018911 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050018913.E1 | ENSRNA050018913 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050018926.E1 | ENSRNA050018926 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050019700.E1 | ENSRNA050019700 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050019891.E1 | ENSRNA050019891 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050019892.E1 | ENSRNA050019892 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050019893.E1 | ENSRNA050019893 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050019894.E1 | ENSRNA050019894 | Eukaryotic large subunit ribosomal RNA |
| IV | ENSRNA050020636.E1 | ENSRNA050020636 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050020637.E1 | ENSRNA050020637 | Eukaryotic small subunit ribosomal RNA |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| IV | ENSRNA050020641.E1 | ENSRNA050020641 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050020881.E1 | ENSRNA050020881 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050021225.E1 | ENSRNA050021225 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050021761.E1 | ENSRNA050021761 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024295.E1 | ENSRNA050024295 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024297.E1 | ENSRNA050024297 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024298.E1 | ENSRNA050024298 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024357.E1 | ENSRNA050024357 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024358.E1 | ENSRNA050024358 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024359.E1 | ENSRNA050024359 | Eukaryotic large subunit ribosomal RNA |
| IV | ENSRNA050024363.E1 | ENSRNA050024363 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024364.E1 | ENSRNA050024364 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024366.E1 | ENSRNA050024366 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024521.E1 | ENSRNA050024521 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024524.E1 | ENSRNA050024524 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024571.E1 | ENSRNA050024571 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024578.E1 | ENSRNA050024578 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024580.E1 | ENSRNA050024580 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024582.E1 | ENSRNA050024582 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024755.E1 | ENSRNA050024755 | Eukaryotic small subunit ribosomal RNA |
| IV | ENSRNA050024756.E1 | ENSRNA050024756 | Eukaryotic small subunit ribosomal RNA |
| IV | TraesCS1A02G001700.1.E1 | TraesCS1A02G001700 | N/A |
| IX | TraesCS3B02G268900.1.E1 | TraesCS3B02G268900 | N/A |
| IX | TraesCS3D02G194000.1.E2 | TraesCS3D02G194000 | Similar to 50S ribosomal protein L40 |
| IX | TraesCS3D02G240800.1.E1 | TraesCS3D02G240800 | N/A |
| IX | TraesCS4A02G091100.1.E1 | TraesCS4A02G091100 | RPL4 |
| IX | TraesCS4B02G213600.1.E1 | TraesCS4B02G213600 | RPL4 |
| IX | TraesCS4B02G344600.2.E3 | TraesCS4B02G344600 | 50S ribosomal protein L5%2C chloroplastic |
| IX | TraesCS4D02G339700.1.E3 | TraesCS4D02G339700 | 50S ribosomal protein L5%2C chloroplastic |
| IX | TraesCS5A02G368400.1.E3 | TraesCS5A02G368400 | N/A |
| IX | TraesCS5B02G370700.1.E3 | TraesCS5B02G370700 | N/A |
| IX | TraesCS5D02G414600.1.E2 | TraesCS5D02G414600 | 30S ribosomal protein S9%2C chloroplastic |
| IX | TraesCS7A02G480700.1.E3 | TraesCS7A02G480700 | Mg-protoporphyrin IX chelatase |
| IX | TraesCS7D02G467500.1.E3 | TraesCS7D02G467500 | Mg-protoporphyrin IX chelatase |
| V | TraesCS1A02G159100.1.E1 | TraesCS1A02G159100 | N/A |
| V | TraesCS1B02G174800.1.E1 | TraesCS1B02G174800 | N/A |
| V | TraesCS2A02G170800.1.E1 | TraesCS2A02G170800 | N/A |
| V | TraesCS2D02G111100.1.E1 | TraesCS2D02G111100 | N/A |
| V | TraesCS2D02G178300.1.E1 | TraesCS2D02G178300 | N/A |
| V | TraesCS2D02G263000.1.E1 | TraesCS2D02G263000 | N/A |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| V | TraesCS3A02G313500.1.E1 | TraesCS3A02G313500 | N/A |
| V | TraesCS3B02G158400.1.E1 | TraesCS3B02G158400 | N/A |
| V | TraesCS3D02G140800.1.E1 | TraesCS3D02G140800 | N/A |
| V | TraesCS4A02G190200.1.E1 | TraesCS4A02G190200 | N/A |
| V | TraesCS4D02G123000.1.E1 | TraesCS4D02G123000 | N/A |
| V | TraesCS5A02G082800.1.E1 | TraesCS5A02G082800 | N/A |
| V | TraesCS5A02G458900.1.E1 | TraesCS5A02G458900 | N/A |
| V | TraesCS5B02G088900.1.E1 | TraesCS5B02G088900 | N/A |
| V | TraesCS5B02G468500.1.E1 | TraesCS5B02G468500 | N/A |
| V | TraesCS5D02G470000.1.E1 | TraesCS5D02G470000 | N/A |
| V | TraesCS7B02G444400.1.E1 | TraesCS7B02G444400 | 60S ribosomal protein L27 |
| VI | TraesCS2B02G313700.1.E7 | TraesCS2B02G313700 | N/A |
| VI | TraesCS2D02G295500.1.E7 | TraesCS2D02G295500 | N/A |
| VI | TraesCS5A02G368400.1.E4 | TraesCS5A02G368400 | N/A |
| VI | TraesCS5A02G468600.1.E2 | TraesCS5A02G468600 | 50S ribosomal protein L17%2C chloroplastic |
| VI | TraesCS5B02G370700.1.E4 | TraesCS5B02G370700 | N/A |
| VI | TraesCS5B02G481100.1.E2 | TraesCS5B02G481100 | 50S ribosomal protein L17%2C chloroplastic |
| VI | TraesCS5D02G378000.1.E4 | TraesCS5D02G378000 | N/A |
| VI | TraesCS5D02G481500.1.E2 | TraesCS5D02G481500 | 50S ribosomal protein L17%2C chloroplastic |
| VI | TraesCS6A02G084200.1.E3 | TraesCS6A02G084200 | N/A |
| VI | TraesCS6A02G105400.1.E2 | TraesCS6A02G105400 | N/A |
| VI | TraesCS6B02G112800.1.E3 | TraesCS6B02G112800 | N/A |
| VI | TraesCS6B02G134800.1.E2 | TraesCS6B02G134800 | N/A |
| VI | TraesCS6D02G094200.1.E2 | TraesCS6D02G094200 | N/A |
| VI | TraesCS7B02G176100.1.E4 | TraesCS7B02G176100 | N/A |
| VI | TraesCS7D02G278100.1.E4 | TraesCS7D02G278100 | N/A |
| VII | ENSRNA050013833.E1 | ENSRNA050013833 | Eukaryotic small subunit ribosomal RNA |
| VII | ENSRNA050013840.E1 | ENSRNA050013840 | Eukaryotic small subunit ribosomal RNA |
| VII | ENSRNA050013857.E1 | ENSRNA050013857 | Eukaryotic small subunit ribosomal RNA |
| VII | ENSRNA050013875.E1 | ENSRNA050013875 | Eukaryotic large subunit ribosomal RNA |
| VII | ENSRNA050019260.E1 | ENSRNA050019260 | Eukaryotic small subunit ribosomal RNA |
| VII | ENSRNA050019702.E1 | ENSRNA050019702 | Eukaryotic large subunit ribosomal RNA |
| VII | ENSRNA050020224.E1 | ENSRNA050020224 | Eukaryotic large subunit ribosomal RNA |
| VII | ENSRNA050021222.E1 | ENSRNA050021222 | Eukaryotic small subunit ribosomal RNA |
| VII | ENSRNA050024312.E1 | ENSRNA050024312 | Eukaryotic small subunit ribosomal RNA |
| VII | ENSRNA050024360.E1 | ENSRNA050024360 | Eukaryotic large subunit ribosomal RNA |
| VII | ENSRNA050024528.E1 | ENSRNA050024528 | Eukaryotic small subunit ribosomal RNA |
| VII | ENSRNA050024795.E1 | ENSRNA050024795 | Eukaryotic small subunit ribosomal RNA |
| VII | TraesCS1A02G000900.1.E2 | TraesCS1A02G000900 | N/A |
| VII | TraesCS6B02G117900.1.E1 | TraesCS6B02G117900 | N/A |

| Wheat ECN | Exon | Gene | Description |
|---|---|---|---|
| VIII | TraesCS2A02G066700.1.E2 | TraesCS2A02G066700 | Ribulose bisphosphate carboxylase small chain |
| VIII | TraesCS2A02G066800.1.E2 | TraesCS2A02G066800 | Ribulose bisphosphate carboxylase small chain |
| VIII | TraesCS2B02G079200.1.E2 | TraesCS2B02G079200 | Ribulose bisphosphate carboxylase small chain |
| VIII | TraesCS2D02G065200.1.E2 | TraesCS2D02G065200 | Ribulose bisphosphate carboxylase small chain |
| VIII | TraesCS2D02G065300.1.E2 | TraesCS2D02G065300 | Ribulose bisphosphate carboxylase small chain |
| VIII | TraesCS3A02G230000.3.E9 | TraesCS3A02G230000 | Carbonic anhydrase |
| VIII | TraesCS3B02G259300.2.E9 | TraesCS3B02G259300 | Carbonic anhydrase |
| VIII | TraesCS3D02G223300.3.E9 | TraesCS3D02G223300 | Carbonic anhydrase |
| VIII | TraesCS6A02G312200.1.E4 | TraesCS6A02G312200 | Glycine cleavage system H protein |
| VIII | TraesCS6B02G342400.1.E4 | TraesCS6B02G342400 | Glycine cleavage system H protein |
| VIII | TraesCS7D02G016800.1.E7 | TraesCS7D02G016800 | Ferredoxin--NADP reductase%2C chloroplastic |
| VIII | TraesCS7D02G343300.1.E8 | TraesCS7D02G343300 | CBBY-like protein |
| X | TraesCS6A02G173100.1.E1 | TraesCS6A02G173100 | N/A |
| X | TraesCS6B02G200800.1.E1 | TraesCS6B02G200800 | N/A |
| X | TraesCS6D02G094200.1.E1 | TraesCS6D02G094200 | N/A |
| X | TraesCS6D02G162200.1.E1 | TraesCS6D02G162200 | N/A |
| X | TraesCS6D02G174500.1.E1 | TraesCS6D02G174500 | 50S ribosomal protein L21%2C chloroplastic |
| X | TraesCS6D02G224200.1.E1 | TraesCS6D02G224200 | N/A |
| X | TraesCS7A02G267900.1.E1 | TraesCS7A02G267900 | plastid-specific 50S ribosomal protein 6 |
| X | TraesCS7A02G278300.1.E1 | TraesCS7A02G278300 | N/A |
| X | TraesCS7B02G176100.1.E1 | TraesCS7B02G176100 | N/A |
| X | TraesCS7D02G267000.1.E1 | TraesCS7D02G267000 | plastid-specific 50S ribosomal protein 6 |
| X | TraesCS7D02G278100.1.E1 | TraesCS7D02G278100 | N/A |
| XI | ENSRNA050016680.E1 | ENSRNA050016680 | Antisense RNA which regulates isiA expression |
| XI | ENSRNA050021557.E1 | ENSRNA050021557 | Antisense RNA which regulates isiA expression |
| XI | ENSRNA050022801.E1 | ENSRNA050022801 | Antisense RNA which regulates isiA expression |
| XI | TraesCS1D02G179900.1.E1 | TraesCS1D02G179900 | Photosystem II protein D1 |
| XI | TraesCS2B02G487300.1.E1 | TraesCS2B02G487300 | 30S ribosomal protein S14%2C chloroplastic |
| XI | TraesCS2B02G487400.1.E1 | TraesCS2B02G487400 | N/A |
| XI | TraesCS5D02G559400.1.E1 | TraesCS5D02G559400 | N/A |
| XI | TraesCS5D02G559700.1.E1 | TraesCS5D02G559700 | N/A |
| XI | TraesCS6B02G037800.1.E1 | TraesCS6B02G037800 | N/A |
| XI | TraesCSU02G269500.1.E1 | TraesCSU02G269500 | Cytochrome b6-f complex subunit 4 |

Legend:
N/A Not applicable.