Project FY24-PB-004: Genome sequences and new annotated references for mapping Fusarium head blight traits

---

1. What are the major goals and objectives of the research project?

1. Perform additional years of greenhouse head inoculation experiments on susceptible and resistant wheat backgrounds, continuing the phenotyping of 150 *Fg* isolates from the NA1 population that we have used in our current USWBSI project for aggressiveness and amount of DON produced, for the purpose of performing genome-wide association studies (GWAS) of these traits.
2. Vastly expand the set of genetic markers segregating in this sample of 150 *Fg* isolates and available for GWAS analysis by performing inexpensive, high-throughput whole genome sequencing of this sample using an inexpensive DNA library preparation protocol to be offered by K-State's Integrated Genomics Facility.

2. What was accomplished under these goals or objectives? *(For each major goal/objective, address these three items below.)*

What were the major activities?

Goal 1): Before applying for the FY24 funding, our lab had completed two replications of inoculation experiments for most of our 150 NA1 isolates onto both wheat cultivars. After submitting the FY24 application (but before the period of performance reported on here) we attempted two more replications, though one replication on our FHB susceptible variety did not produce useable data due extensive damage from thrips that were resistant to common insecticides. Due to limitations in available greenhouse space and the need of our group to perform similar inoculation studies with a different *Fusarium* population, we did not carry out more inoculations with our NA1 isolates within the reporting period.

Goal 2): We extracted high molecular weight genomic DNA from the same 150 *F. graminearum* samples that we have been using for wheat inoculations, performed quality control measurements on a Nanodrop, and quantified the sample yield both with Nanodrop and PicoGreen plate assays. We submitted these samples, along with hundreds of others from another funded project, to K-State's Integrated Genomics Facility (IGF) for the production of Illumina DNA whole-genome sequencing (WGS) libraries through an inexpensive low-volume Nextera library prep method. The libraries were prepared with sample-specific molecular barcodes, then pooled. We sent the pooled libraries for sequencing at KUMed's Genomics Core for sequencing on an Illumina Novaseq 6000 S1 300 cycle run. Results from a small subset of isolates proved an insufficient number of reads for high-confidence WGS genotyping, and so we re-extracted DNA and re-sequenced these isolates.

What were the significant results?

Goal 1): Analysis of the first two replications has found no significant isolate by cultivar interaction. Thus, GWAS can be performed on data combined across the two cultivars, making more inoculations available for the estimation of aggressiveness for each isolate, and indicating that four to five replications for each cultivar will suffice for GWAS.

Goal 2): WGS results were excellent and are providing exactly the type of nearly complete genotype data we will use with our samples for later GWAS. However, not all samples produced sufficient read output from the initial WGS attempt (**Fig. 1**).

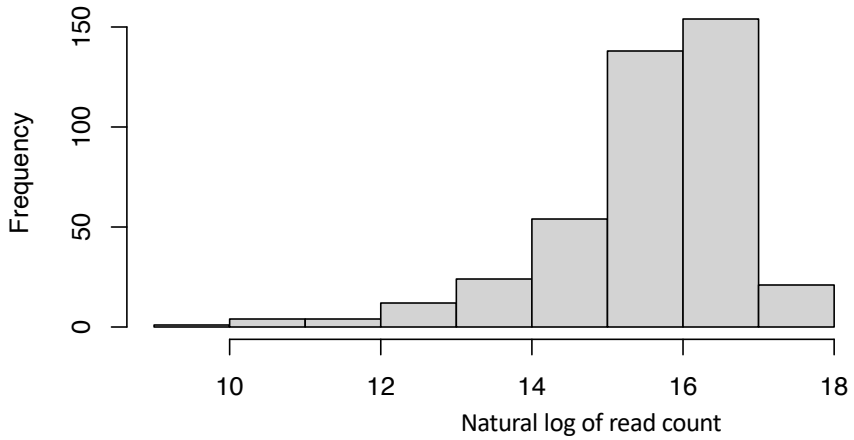

Fig. 1. Distribution of natural log of read count across all samples pooled together after parallel library preparation.

This histogram shows the distribution of the log of read count across over 400 samples. Full genome-wide genotyping can be achieved in the haploid *F. graminearum* as we approach an average of 10x depth of coverage (a natural log of read number between 14 and 15), and the histogram reveals a long left tail of samples that performed poorly. Fortunately, the bulk of the samples exceed this threshold. We had not tried this process for fungal samples previously, and we have learned that we can make further improvements to avoid missing data (such as including an extra step of column-based genomic DNA cleanup to remove any remaining contaminating impurities). After mapping reads to a reference genome and calling variants across samples, the samples with low read counts result in proportions of genotype calls at identified SNPs below the plateau of >90% that we see for most samples (**Fig. 2**).

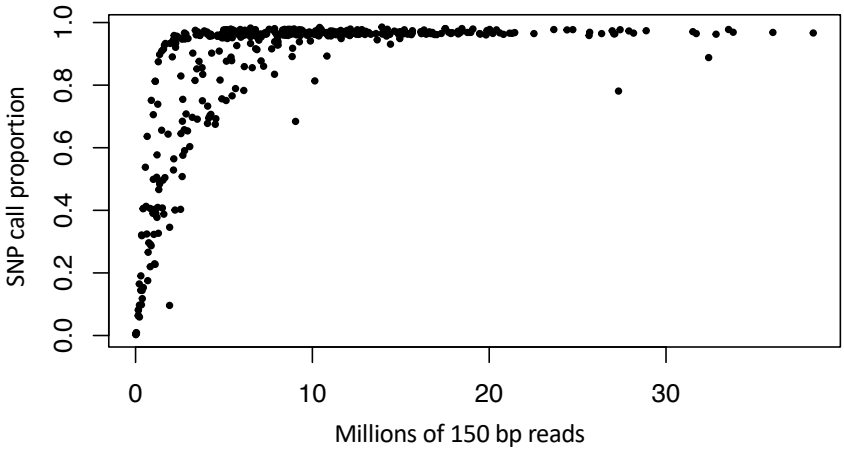

Fig. 2. Scatterplot of the proportion of identified SNPs called against millions of 150 bp reads produced across all samples included in the parallel low-volume whole-genome sequencing library preparation.

Again, the majority of samples have reached our plateau of high SNP calls, and for those that didn't, we re-extracted DNA and re-sequenced these isolates. When we focus on only our isolates that come from the **NA1 population** and correspond to this Goal, we recover over **450,000 SNPs**. Of these, > 200,000 are found at 5% or greater minor allele

frequency, meaning that they will be retained in subsequent GWAS. That number of SNPs translates to roughly **1 per 200 bp** across the *F. graminearum* genome. That density means that our SNP coverage well-exceeds what is required given the extent of linkage disequilibrium in *F. graminearum* populations, and we are unlikely to miss genome regions underlying important trait variation.

List key outcomes or other achievements.

Goal 1): We currently have data for a total of three replications of inoculations on our FHB susceptible variety, and four replications of inoculations on our variety with partial FHB resistance. This puts us on track to complete the proposed five total replications by the end of the reporting period.

Goal 2): We have completed all the WGS data collection for our NA1 samples in this project, though our latest WGS data were produced near the end of this reporting period (April 2025). Thus, we have yet to integrate the new data, which serve to supplement sequence reads from the 40 isolates that had moderate to low read levels, into the SNP genotypes previously called. Broadly, our sequencing of the pooled low-volume Nextera WGS libraries of our filamentous fungal samples has taught us much about what results to expect from different fungi or genomics applications. The K-State IGF has extensive results from this protocol for crop samples, but not fungi. The low cost of the library preps facilitates large-scale WGS across very large samples, regardless of target species. In our case, the small and very low repeat *F. graminearum* genome means that even low depth of coverage across the genome can generate full WGS genotypes for SNPs that are called against a high-quality reference genome. However, we have found that the libraries produced can have high levels of PCR duplicates (>50%), which lead to redundant sequences and inflated read depth. So, a researcher expecting 50x depth of coverage based on the count of raw reads may only get 25x after the removal of duplicate reads. This is important, as applications like *de novo* genome assembly require higher depth of coverage, and so may still require the best quality (and more expensive) library preparation protocols.

Outside of the main goals of the current project, but related to goals from past USWBSI-funded projects, we have published a large population genomics study of American *F. graminearum* populations that is open access. It sheds light on differences between populations as well as their relationships and inferred demographic histories.

3. What opportunities for training and professional development has the project provided?

The project has provided training for my PhD students Sandhya Gopisetty and Sumit Chowdhury, including presenting a poster at the APS North Central Division meeting in summer 2024. Both have been trained on culturing and performing experiments with *Fusarium* in the lab, as well as greenhouse wheat head inoculations. They have begun their training in data analysis and a range of bioinformatics and genomics analyses. The project has also provided training for two undergraduate hourly workers, Emily Gipson and Jamia Roberts. Both helped with greenhouse inoculations and grain sample preparation for DON testing, as well as learned additional *Fusarium* culturing techniques. Emily also learned PCR technique to confirm *TRI* genotype on a subset of *Fusarium* strains involved in this project. PI Toomajian used data from this project in teaching his graduate Population Genetics class to 24 students in the Spring of 2025. The project also provided professional development and training to 26 participants in the 2024 Fusarium Laboratory Workshop. Finally, the PI had professional development opportunities through participation in the 2024 Scab Forum.

4. How have the results been disseminated to communities of interest?

The results of this project have been disseminated through a presentation at the APS North Central annual meeting and through the publication of a peer-reviewed manuscript in *Fungal Genetics and Biology*. Additional manuscripts will be submitted during the second year of this project and subsequently. PI Toomajian also presented some results in June 2024 through his lecture on population genetics for the Fusarium Laboratory Workshop at Kansas State University. PI Toomajian participated in Kansas State University's annual open house for the purpose of outreach to enhance public understanding of plant pathology and increase interest in careers in this field.

5. What do you plan to do during the next reporting period to accomplish the goals and objectives?

Goal 1): Perform one more replication of inoculations of the NA1 sample onto our cultivar with partial FHB resistance, and two more replications of inoculations onto the FHB susceptible cultivar. This will bring us to a total of five replications on each wheat cultivar. Perform a comprehensive analysis of isolate aggressiveness across all replications. Complete processing of inoculated heads before sending them to the Virginia Tech USWBSI DON testing lab. Perform GWAS for both isolate aggressiveness and *in planta* DON levels, and publish these results.

Goal 2): Incorporate the most recent WGS data produced in April 2025, into our SNP calling pipeline that includes our older WGS data, to generate full WGS genotyping data for our sample. Our initial analyses performed read mapping only against the PH-1 reference genome, and focused on SNP variation. For the next period we will test read mapping against additional references to uncover accessory genome variants. We will also incorporate nonSNP variation such as structural variants (e.g., variants over 50 bp in size) that have strong support in our read data. These efforts will minimize the chance that we miss the functional variation this project seeks to capture and map. We will then repeat GWAS of previously collected trait data (*e.g.*, fungicide sensitivity, *in vitro* DON production) using the WGS SNPs. To reduce the multiple testing involved in GWAS, we will filter our WGS variants based on linkage disequilibrium, to remove redundant SNPs that would give us the same signal as others in the dataset. We will then publish these results: first, focusing on the GWAS results using the new WGS data, and then secondly looking to report on the full WGS data of the NA1 sample in the context of variation found in other NA1 isolates or in other *Fusarium* populations. These data can help further our understanding of genetic variation in North American populations of *F. graminearum*.