SYNTHETIC OVERSAMPLING IMPROVES SPECTRAL DETECTION OF AFLATOXIN IN SINGLE MAIZE KERNELS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Chloe Siobhan Siegel

May 2022

ABSTRACT

Current spectral models for detection of aflatoxin (AF) in maize kernels are limited by very low representation of kernels with AF levels above the models' specified contamination thresholds. With the aim of improving prediction model sensitivity to AF-contaminated kernels, two methods were evaluated for artificially enriching the representation of kernels containing ≥150 ppb AF in a large spectral dataset: humid incubation and synthetic oversampling. The addition of synthetic training samples improved prediction accuracy of kernels containing ≥150 ppb AF from 51% to 80%. Humid incubation contributed additional kernels with intermediate levels of AF contamination (primarily 5-75 ppb), but generally did not change the overall distribution of AF contamination. Feature importance distributions overlapped among models at 329-345 nm, 380-385.5 nm, 415-425 nm, 639-668 nm, and 1,013.5-1,060 nm. These spectral ranges could be applicable to the development of limited wavelength grain sorting devices for low-resource settings.

BIOGRAPHICAL SKETCH

My plant science journey began at the Chicago Botanic Garden, a short bike ride from my family home. As an intern in their plant genetics laboratory, I laid the first bricks in my foundation as a molecular biologist. That fall, I started college at the University of Illinois in Urbana-Champaign where I kept myself busy with coursework, undergraduate research, and a part time job in a campus library. I spent the summers of my undergraduate years participating in additional molecular biology internships at the Smithsonian National Museum of Natural History in Washington, D.C. and Corteva Agriscience in Indianapolis, IN. In May of 2018, I graduated from the University of Illinois with a B.S. in Crop Sciences and a concentration in plant biotechnology and molecular biology. With my eyes set on graduate school, I searched for a program where I could incorporate my passions for plant science, microbiology, and human health. I found that intersection in Rebecca Nelson's lab at Cornell University where I was introduced to the world of mycotoxins.

To the Siegel and Halm families

ACKNOWLEDGMENTS

I extend my gratitude to my special committee, including Gary Bergstrom, Katie Gold, and my committee chair, Rebecca Nelson.

Gary -- I can't thank you enough for the generous mentorship and guidance you've given me throughout my time at Cornell. From sending me relevant papers and dropping helpful resources at my desk, to taking me to the NY Corn Congress and helping me craft DON management recommendations for a local farm, you've gone above and beyond to integrate me into the corn world. Without the exposure you've given me to that world, I don't think I could've fully grasped how multi-faceted our field really is. I feel so fortunate to have gotten a chance to work with you over the last few years.

Katie -- I have felt your support on my academic journey even before you joined the faculty at Cornell. You made yourself available to discuss my projects with me when I would visit Madison, and have offered me invaluable advice, both academic and personal, over the last few years. When I saw your seminar at Cornell at the beginning of 2019 and heard the way you talked about precision agriculture, my fingers were crossed that you would come to Cornell and be on my committee. My work has undoubtedly been improved thanks to your insights on the biochemical bases behind spectral features and the models that are best suited to exploit those features. You've instilled in me an immense appreciation for the power digital tools will have on the future of agriculture, and you've empowered me to contribute to that future.

Rebecca –I never thought I would be able work at the intersection of plant health, animal health, human health, and planetary health. It wasn't until I joined your lab that I learned about mycotoxins and realized I could connect all these passions. I immediately felt a shared sense of purpose with you when we first talked about the central role of plant science research in developing practical solutions for the people who need them most. I plan to carry forth your emphatically interdisciplinary vision of a more sustainable, safe, and equitable global food system in all my future professional endeavors. Thank you for all the busy days and long nights you've put in to help me get to this point.

I would also like to thank Mike Gore for the immensely beneficial feedback he has given to me on this thesis over the years. Thank you for raising important considerations about my humid incubation experiments and for prompting me to think critically about the practical applications of different oversampling methods.

I owe a debt of gratitude to every member of the Nelson lab during my time at Cornell, including Anthony Wenndt, Danilo Moreta, Judy Kolkman, Julian Atukuri, Shan Jin, Shimeng Gao, Tyr Wiesner-Hanks, and Will Stafstrom. Your contributions to my work in and out of lab meeting cannot be overstated. I especially want to express my thanks to Ace Huot Repka and Meriem Aoun for their unwavering dedication to my development as a scientist. Thank you for teaching me new laboratory techniques, troubleshooting with me when things went wrong, and for making yourselves available to me whenever I needed guidance. I could not have done this without either of you.

Finally, I wish to thank the 2018 cohort of plant pathology students at Cornell University, including Andrea Lugo-Torres, Alex Clarke, Eric Branch, Igor Yang, Jess Choi, Juan Luis Gonzalez Giron, Maddie Flasco, Marina Mann, and Saeed Hosseinzadeh. I could not have asked for a better group to go through grad school with than you.

TABLE OF CONTENTS

LIST OF FIGURES

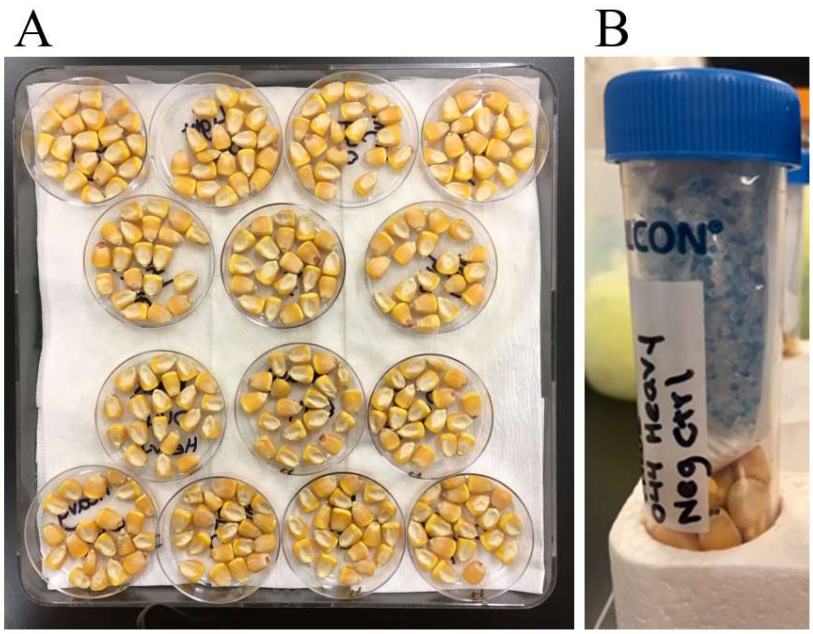

Figure 1. A) Unsealed incubation chamber. **B)** Dried kernels in 50 mL Falcon tube with indicator desiccant packet.

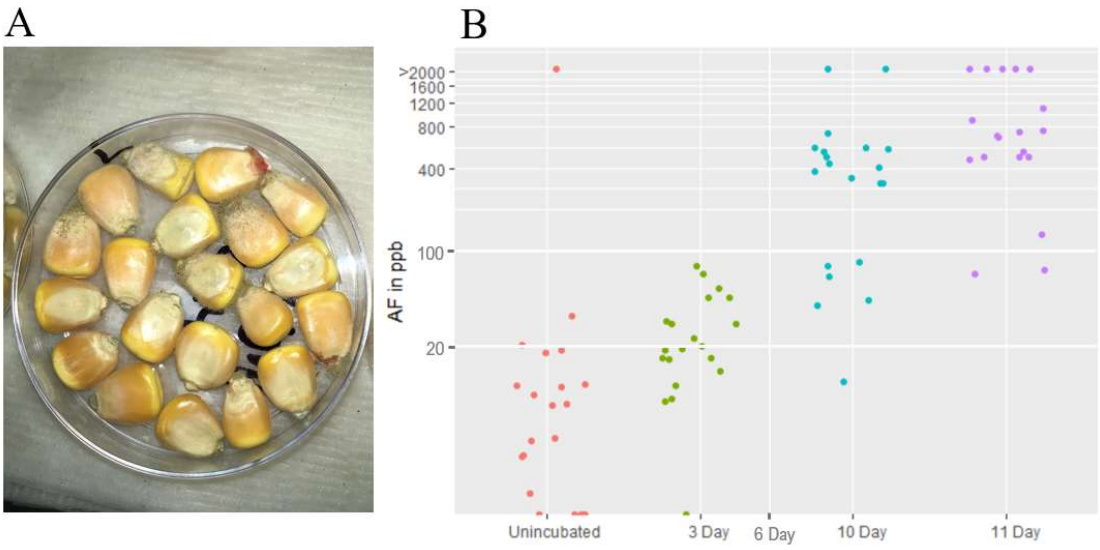

Figure 2. A) Kernels showing visible patches of *A. flavus* after 10 days of incubation. **B)** Preliminary experiments showed that incubation periods of three and six days were effective for increasing AF concentrations without inducing sporulation.

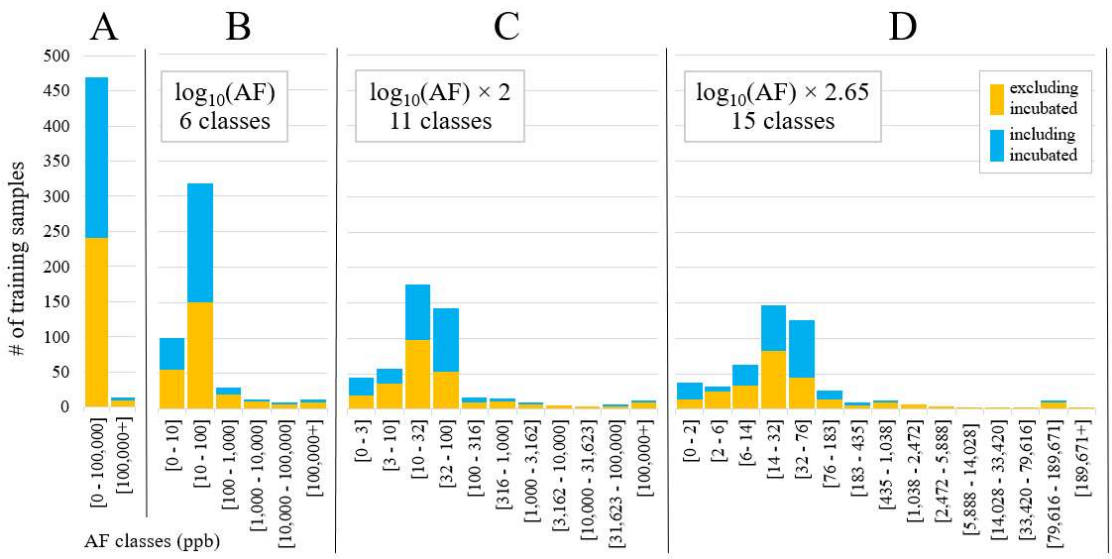

Figure 3. Partitioning AF classes along a log scale allowed for the division of more classes compared to a linear scale. Multiplying the log scale by an increasingly large number allowed for generation of a higher number of narrower classes. **A)** AF classes partitioned along a linear scale, **(B)** $\log_{10}(AF)$ scale, **C)** $\log_{10}(AF) \times 2$, and **D)** $\log_{10}(AF) \times 2.65$. AF ranges defining each class threshold are shown in ppb along the X axis. Training samples excluding and including incubated kernels are shown in yellow and blue, respectively.

	Excluding incubated kernels	Including incubated kernels
Excluding synthetic samples	-inc/-syn	+inc/-syn
Including synthetic samples	-inc/+syn	+inc/+syn

Figure 4. Contingency table of samples included in each of the four training sets.

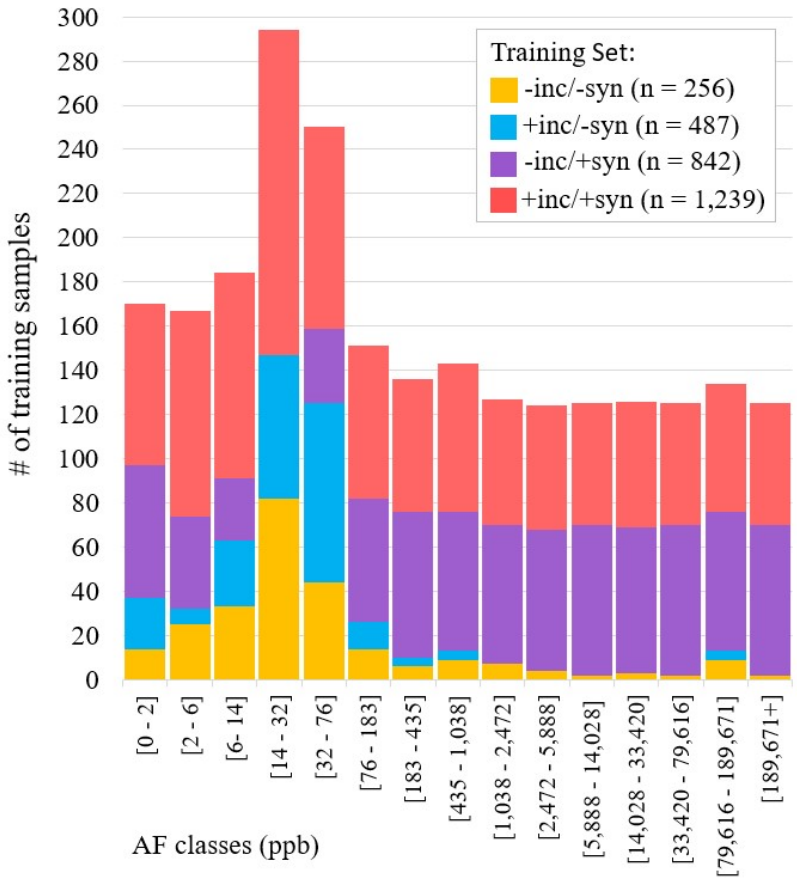

Figure 5. Distribution of training set samples along a $\log_{10}(AF) \times 2.65$ scale following incubation and synthetic oversampling. AF ranges defining each class threshold are shown in ppb along the X axis. Number of samples in each training set are shown along the Y axis.

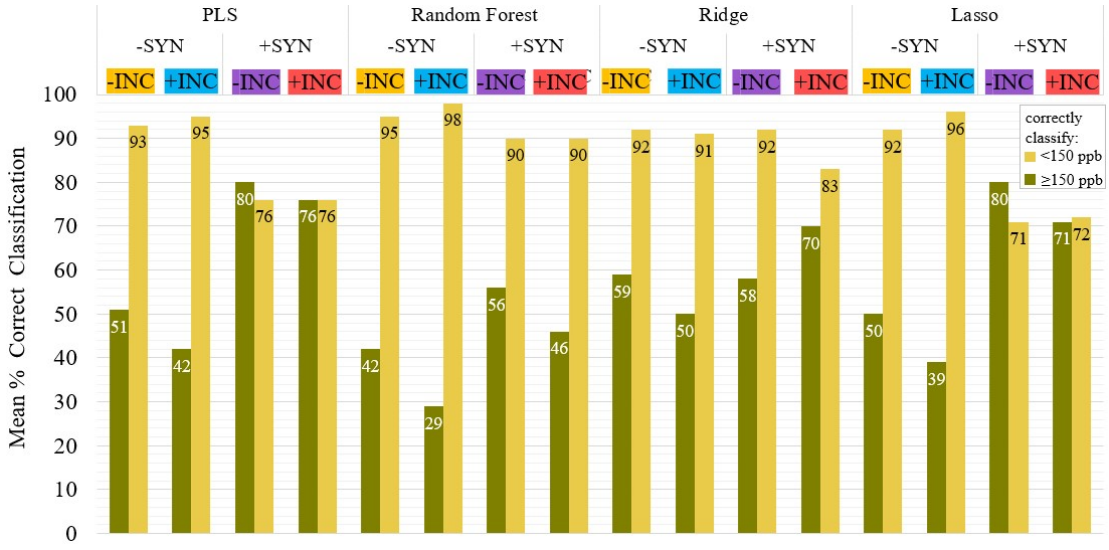

Figure 6. Mean percent sensitivity and specificity for each model and training set. Means were calculated based on 1,000 random iterations. Sensitivity and specificity are displayed in green and yellow, respectively.

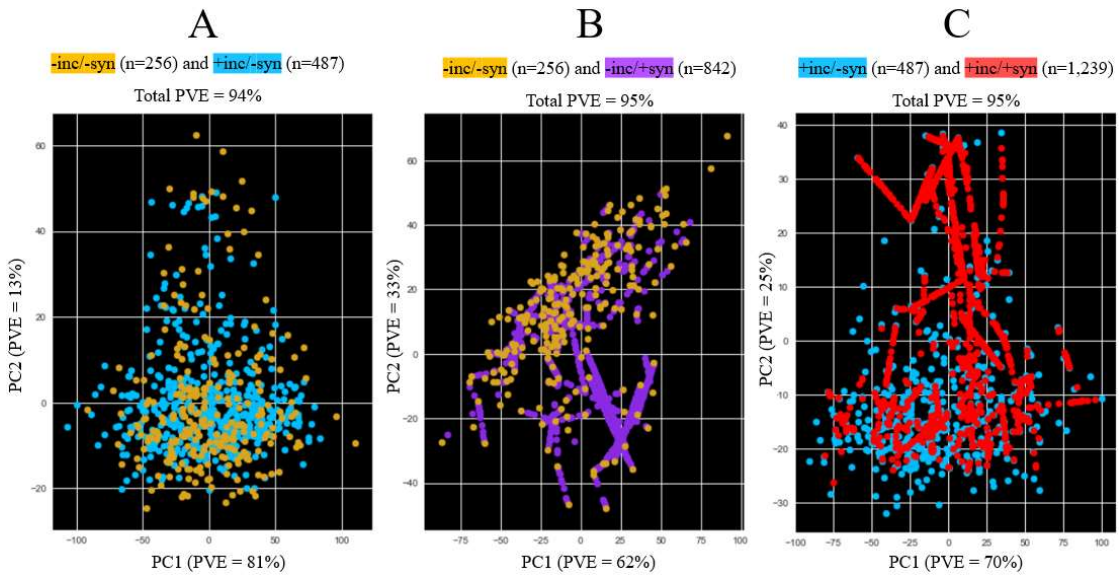

Figure 7. Principal Component Analysis of training set samples. **A)** -inc/-syn (n = 256) and +inc/-syn (n = 487). **B)** -inc/-syn (n = 256) and -inc/+syn (n = 842). **C)** +inc/-syn (n = 487) and +inc/+syn (n = 1,239) training samples. PC = principal component. PVE = percent variance explained.

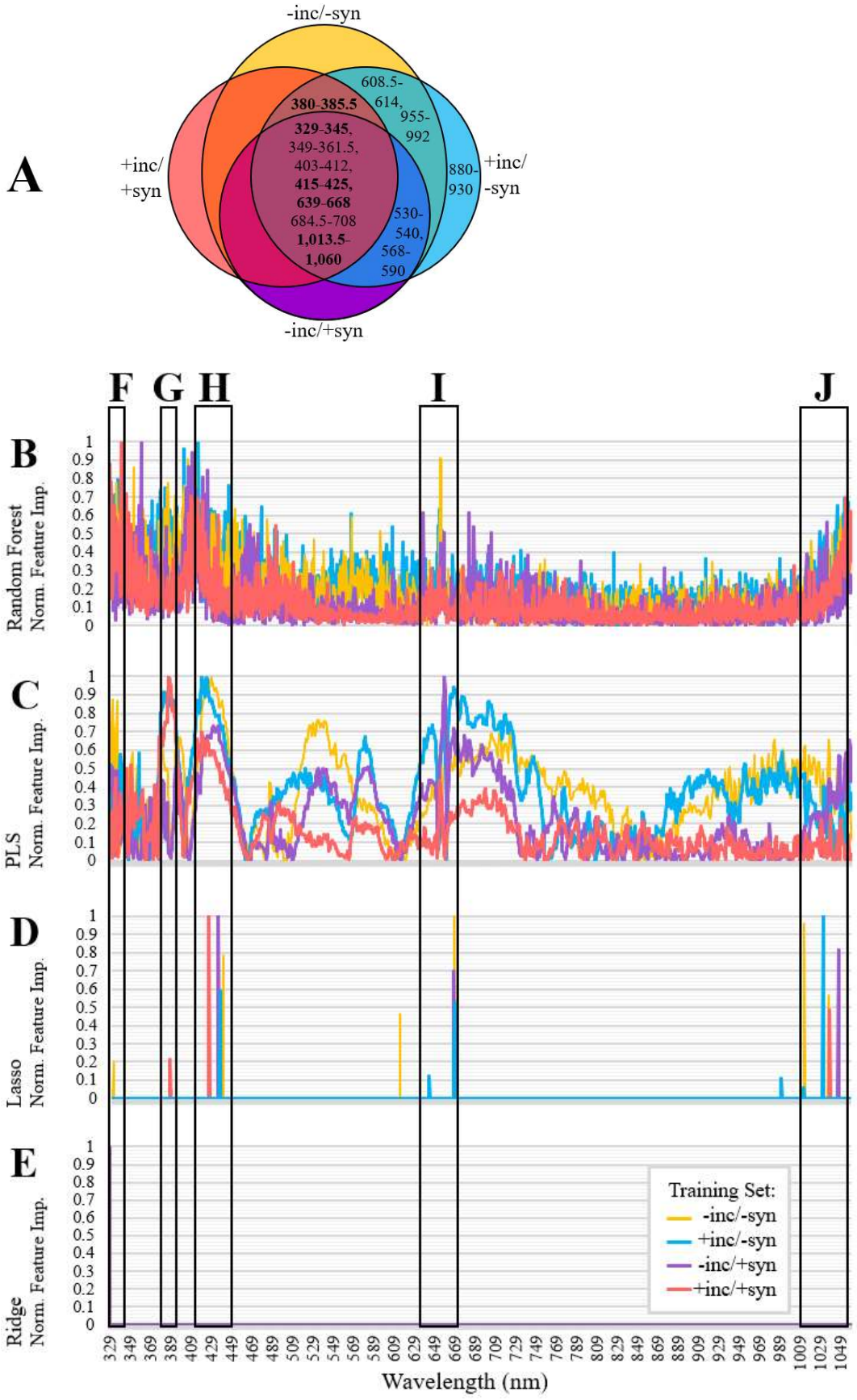
A
-inc/-syn
+inc/
+syn
+inc/
-syn
-inc/+syn
380-385.5
329-345,
349-361.5,
403-412,
415-425,
639-668
684.5-708
1,013.5-
1,060
608.5-
614,
955-
992
880-
930
530-
540,
568-
590
F
G
H
I
J
B
Random Forest
Norm. Feature Imp.
C
PLS
Norm. Feature Imp.
D
Lasso
Norm. Feature Imp.
E
Ridge
Norm. Feature Imp.
Training Set:
-inc/-syn
+inc/-syn
-inc/+syn
+inc/+syn
Wavelength (nm)

Figure 8. A) Spectral regions with high feature importance shared between the four training sets based on all model types. Spectral regions shared between at least three of the four model types are in bold. **B)** Normalized absolute value of feature importance scores for RF, **C**) PLS, **D)** Lasso, and **E)** Ridge models. Spectral regions with high feature importance are shared between at least three of the four model types at **F)** 329-345 nm, **G)** 380-385.5, **H)** 415-425 nm, **I)** 639-668 nm, and **J)** 1,013.5-1,060 nm.

LIST OF TABLES

Table 1. Specifications of the AF accumulation phenotype, maize hybrid, and incubation treatment of each of the 872 kernels included in this study.

Hybrid	Incubation Treatment			
	0 day	**3 day**	**6 day**	**All treatments:**
Moderately Resistant				
BH8900VIP3111	24	16	15	**55**
DK64-69	23	21	24	**68**
GEMS-0027 × Tx779	24	32	36	**92**
Subtotal:	**71**	**69**	**75**	**215**
Intermediately Susceptible				
Mp494 × Mp717	24	9	35	**68**
Mp494 × Mp719	24	23	22	**69**
Mp494 × PHW79	24	22	27	**73**
Subtotal:	**72**	**54**	**84**	**210**
Highly Susceptible				
Pioneer 1745R	23	29	34	**86**
Pioneer 2088R	24	25	22	**71**
SC212M × PHW79	247	0	0	**247**
TZAR106 × T173	16	21	6	**43**
Subtotal:	**310**	**75**	**62**	**447**
All hybrids:	**453**	**198**	**221**	**872**

Synthetic oversampling improves spectral detection of aflatoxin in single maize kernels*

INTRODUCTION

Aflatoxin (AF) is an invisible, highly carcinogenic metabolite produced by the fungus *Aspergillus flavus*. AF contamination in *A. flavus*-infected maize is responsible for widespread crop damage, human and livestock health issues, and economic losses. In the United States alone, annual losses due to AF can cost the maize industry an estimated $1.68 billion in drought years (Mitchell et al., 2016). An estimated 4.5 billion people are exposed to AF worldwide (Williams et al., 2004) with chronic exposure effects including liver cancer, reduced growth in children, and immunosuppression (Gong et al., 2016; Zain, 2011). In Kenya, cases of acute aflatoxin poisonings have occurred in the hundreds during major outbreak years, with many of them fatal (Lewis et al., 2005; Mutegi et al., 2018; Zain et al., 2011). Accurate AF detection methods are essential for preventing the consequences of exposure, though overly stringent grain rejection can threaten already-compromised food supplies.

Grain sorting protocols for AF detection are sometimes performed on a single kernel basis due to the highly heterogenous distribution of infected kernels on a single ear (Smart et al., 1990). The distribution of AF-contaminated kernels is also often highly heterogenous in bulk samples, such that a small minority of extremely toxic kernels

*In the event of publication, the authors on this manuscript will include myself, Meriem Aoun, Gary L. Windham, W. Paul Williams, and Rebecca J. Nelson (subject to additional authors)

are intermixed with kernels without detectable levels of AF (Chavez et al., 2020; Johansson & Whitaker, 2000; Mitchell et al., 2016; Stasiewicz et al., 2017; Whitaker & Johannson, 2005). However, higher proportions of kernels with dangerous levels of AF contamination have been observed in Tanzania (Aoun et al., 2020) and Kenya (Mutiga et al., 2014) where access to AF management resources is often limited. Complex interactions between management, environmental conditions, aggressiveness of the *A. flavus* fungus, and susceptibility of the maize plant can give rise to a variety of AF distribution scenarios.

Inconsistencies in the appearance of kernels with dangerous levels of AF can preclude their detection. Fungal colonization by the *A. flavus* fungus, as measured by quantitative PCR, is correlated with AF accumulation in kernels (Mideros et al., 2009), but growth can occur on internal structures of the kernel hidden from view (Marsh & Payne, 1984; Shinha & Bhatnagar, 1998; Zuber et al., 1986). Thus, AF can accumulate in kernels where no fungal colonization is visually observed (Aoun et al., 2020; Smart et al., 1990). While kernels with surface growth of *A. flavus* can be discarded based on inspection, kernels with internal fungal growth require detection methods not reliant on visible differences.

Ultraviolet (UV) fluorescence has been used as an indirect means of AF detection for several decades (Bothast & Hesseltine, 1975; Hesseltine & Shotwell, 1973; Shotwell & Hesseltine, 1981), but faces similar pitfalls as detection by visible fungal growth. The UV fluorescence is generally not produced by the AF itself, but rather the commonly co-occurring metabolite kojic acid. While AF and kojic acid are correlated,

not all UV-fluorescent kernels contain AF (Schmitt & Hurburgh, 1989; Thompson & Henke, 2000; Wicklow, 1999; Yao et al., 2010) and not all AF-contaminated kernels exhibit UV fluorescence (Brekke et al., 1975; Gloria et al., 1998; Hruska et al., 2016; Pearson et al., 2001; Yao et al., 2010).

Spectral features associated with AF contamination have demonstrated superior consistency for detection in single kernels compared to other kernel properties including appearance (Aoun et al., 2020; Mutiga et al., 2014), size (Aoun et al., 2020), or density (Aoun et al., 2020; Brekke et al., 1975; Shi et al., 2014). Spectral methods leverage the unique patterns by which kernels absorb and reflect light across the UV, visible, and near infrared (NIR) spectrum. By comparing the spectral features of many kernels containing levels of AF contamination above and below a specified threshold, machine learning models can be used to identify and classify spectral features of AF beyond what is visible to the human eye. However, the utility of these models has been limited by very low representation of kernels with AF levels above their specified contamination thresholds (Chavez et al., 2022; Cheng et al., 2019; Chu et al., 2017; Kimuli et al., 2017; Stasiewicz et al., 2017).

Models trained to predict rare occurrences, such as the presence of an extremely toxic kernel, commonly show misleadingly high accuracy by classifying all samples to the majority class (e.g. non-detectable levels of AF) by default. For example, if 1% of kernels in a lot are extremely toxic, a model could achieve 99% accuracy by predicting that none of the kernels have detectable levels of AF. Furthermore, strong performance based on a dataset with very few minority class samples is not applicable

to toxin prediction more generally, no matter how many iterations are performed. Consider an example dataset of 1,000 kernels, of which only 10 are extremely toxic; a model could subset a random sample of 3 kernels from the 10 toxic ones any number of times and the validity of the model would still be limited by the representativeness of those 10 kernels. These scenarios underscore the challenges of prediction in imbalanced datasets, which are defined by an unequal distribution of data points across different classes.

With the aim of improving prediction of AF-contaminated kernels, we evaluated the utility of two methods for artificially enriching the representation of toxic kernels in a large spectral dataset. In our models, kernels with ≥ 150 ppb AF contamination were considered "toxic kernels." For the first method, we simulated poor storage conditions associated with AF accumulation. It has been observed that hot, humid, and poorly ventilated conditions during storage facilitate rapid accumulation of AF (Pitt et al., 2013; Muga et al., 2019). Re-moistening of previously dried harvested kernels due to leaks in storage containers has also been associated with extensive AF accumulation (Bankhole et al., 2006; Trenk & Hartman, 1970). Therefore, we hypothesized that incubating kernels at 30°C in an enclosed humid chamber would increase the incidence of high AF kernels and improve model sensitivity of unincubated toxic kernels.

For the second method, we used machine learning tools to enrich training sets with synthetically generated examples of toxic kernels. Several machine learning algorithms have been developed to deal with imbalanced datasets by generating

synthetic training samples, most notably Synthetic Minority Over-sampling Technique or SMOTE (Chawla et al., 2002). Adaptive Synthetic (ADASYN) sampling technique (He et al., 2008), a variation of SMOTE, allows for the generation of a higher ratio of synthetic training samples for data in classes that are hardest to classify (e.g. toxic kernels), and fewer synthetic training samples for data in the majority class that are easier to classify (e.g. non-toxic kernels). We hypothesized that the generation of synthetic training data with the ADASYN algorithm would increase classification accuracy of non-synthetic toxic kernels.

The main objective of this study was to improve spectral detection of AF contamination in single maize kernels. Specifically, I sought to (1) increase the proportion of toxic kernels in training sets using multiple artificial enrichment methods, (2) assess if spectral artifacts were associated with artificial enrichment methods, and (3) identify key wavelengths shared between AF classification models.

METHODS

Biological Materials

Ten maize hybrids were included in this study, representing a range of phenotypes relating to AF accumulation. Among them, four hybrids were highly susceptible, three were intermediately susceptible, and three were moderately resistant to AF accumulation, as previously characterized by Aoun et al. (2020) (Table 1). Nine of the hybrids were grown in Starkville, Mississippi in 2017 as single row plots in a

complete randomized block design with three replicates; the tenth hybrid, BH8900VIP3111, was grown in 2018 at the same location.

Primary ears were injected with a 3×10^8 conidia suspension (Wilcox et al., 2013) of *A. flavus* strain NRRL 33574.26 using the side needle technique (Zummo & Scott, 1989) one week after mid-silk. When plants reached physiological maturity, primary ears were harvested, dried, and shelled. A total of 872 shelled kernels from the highly susceptible (n = 447), intermediately susceptible (n = 210), and moderately resistant (n = 215) hybrids were used for subsequent analysis (Table 1).

Humid Incubation

Approximately 50 kernels from each hybrid were incubated at 30°C in a humid chamber. The humid chamber consisted of a 230 × 230 mm polystyrene square petri dish lined with three layers of paper towels. The paper towels were dampened with deionized water, with care taken to not completely saturate the towels. To prevent the kernels from coming into direct contact with the damp paper towels, groups of kernels ($n \approx 20$) in smaller 100 × 15 mm open petri dishes were placed on the paper towels. The lid of the larger square petri dish was then replaced and sealed with parafilm (Fig. 1).

Kernels exhibiting visible mold were not excluded from the dataset, but their inclusion was limited by only incubating kernels long enough to increase AF levels yet not so long as to enable sporulation. Limiting the prevalence of visibly moldy kernels in our dataset reduced the risk of inadvertently training our models to predict external *A.*

flavus moldiness. In preliminary experiments, we observed that visible patches of *A. flavus* began to appear on the surface of kernels after approximately 10 days of incubation (Fig. 2A). Incubation periods of three and six days were effective for increasing AF concentrations without inducing sporulation (Fig. 2B). Therefore, only kernels that were incubated for zero, three, and six days were included in the final dataset. Among the incubated kernels, 198 were removed from the humid chamber after three days and 221 were removed after six days. The remaining 453 kernels in the dataset were left unincubated (Table 1).

After kernels were removed from the incubator, they were transferred to 50 mL Falcon tubes in groups of ~20. Desiccant packets containing silica beads that change color from blue to pink when saturated with water were used to indicate kernel dryness. The indicator desiccant packets were placed inside each Falcon tube with the kernels and replaced every few days until the beads remained stably blue.

Acquisition of Spectral Reflectance

Spectral measurements were acquired using an UV-Visible-NIR spectrometer (2DCCD-512x64, Control Development, Inc., South Bend, IN), UV-Visible-NIR light source (Ocean Optics, Inc., Dunedin, FL), and a light-emitting diode fiber optic probe (Ocean Optics, Inc., Dunedin, FL). The light source produced wavelengths of light every 0.5 nm from 304-1,085 nm (n = 1,563 wavelengths), which were emitted through the end of the fiber optic probe. The end of the probe was held against the surface of the embryo side of each kernel and the kernel's reflectance at each wavelength was detected by the spectrometer. The intensity of light reflected by each

sample (percent reflectance relative to a calibration reference) at each of the 1,563 wavelengths will hereafter be referred to as the kernel's spectral signature.

AF Quantification

AF concentrations of single kernels were measured using an enzyme-linked immunosorbent assay (Total AF ELISA Assay, Cat. No. 941AFL01M-96; Helica Biosystems, Inc., Santa Ana, CA), following the manufacturer's protocol. Six standards of known AF concentrations were run in tandem with the assayed samples. The optical densities of the standards and assayed samples were measured at 450 nm using a Synergy HTX multi-mode reader (BioTek Instruments, Inc., Winooski, VT). The optical densities of the standards were used to form a standard curve from which the AF concentrations of the kernels were extrapolated using Gen5 Microplate Reader and Imager Software (BioTek Instruments, Inc., Winooski, VT).

Spectral Preprocessing

Minimizing Stray Light Effects. Light detected by a spectrometer from sources other than the sample (e.g. ambient light in the room), referred to as "stray light," can introduce unwanted noise to spectral prediction models. The proportion of reflectance signal due to stray light tends to be highest at the ends of a spectrometer's measured range (Reule, 1976) and at ranges where the sample's reflectance is lowest (Salim et al., 2011). In order to minimize spectral effects from stray light, reflectance values at the first 50 and last 50 wavelengths (304-328.5 nm and 1,060-1,085 nm) were excluded from each sample's spectral signature. The resulting spectral signatures

consisted of spectral reflectance values every 0.5 nm from 329-1,060 nm (n = 1,463 wavelengths).

Minimizing Light Scattering and Path Length Effects. Preprocessing techniques such as multiplicative scatter correction and standard normal variate (SNV) can reduce some of the spectral variation that arises from differences in light scattering and path length among samples. In a subset of AF-inoculated SC212M × PHW79 kernels used in the present study, Aoun et al. (2022) found that SNV transformation consistently outperformed several preprocessing functions at AF classification tasks. In order to reduce variation from factors not related to AF in the present study, SNV transformation was implemented using the "sklearn.preprocessing.StandardScaler" function in the Scikit Learn library (Pedregosa et al., 2011). Training sets and test sets were transformed separately to prevent "leakage" of test data into the training process (Brownlee, 2020).

Defining Training and Test Sets

Samples were divided into AF classes to provide a basis for ADASYN oversampling. To allow for subsequent generation of synthetic training samples, at least three samples were required for each AF class, including two that could be used for model training (one sample and one "nearest neighbor") and one for model testing. On a linear scale (all classes of equal size), this constraint only allowed for the division of samples into two classes: [0 - 100,000 ppb] and [100,000+ ppb] (Fig. 3A). To represent the range of AF concentrations present in our samples, AF classes needed to

be divided along a logarithmic scale (i.e. lower AF kernels in narrower classes and higher AF samples in broader classes).

Multiplying a log10 scale by increasingly large factors "exaggerates" the scale, such that small values are closer together and larger values are farther apart (Fig. 3B-D). Based on this principle, we found that a log10(AF) × 2.65 scale enabled division of our samples into the maximum possible number of classes (n = 15 classes) while still maintaining at least three samples per class (Fig. 3D). Maximizing the number of AF classes in this way allowed for subsequent generation of the maximum number of synthetic training samples.

For model training, 55% of samples from each AF class (rounded to the nearest integer) were randomly selected, leaving the remaining 45% to form the test set. While train/test splits of 70/30 are most common, our 55/45 split was chosen to increase the validity of model performance on our test data given the relatively small number of kernels representing higher AF concentrations.

Adaptive Synthetic Oversampling

The "imblearn.over_sampling.ADASYN" function from the Imbalanced-learn Python toolbox (Lemaitre et al., 2017) was used to generate synthetic training samples based on spectral signatures of kernels from each AF class. Spectral signatures from toxin classes 6 through 8 (79 - 1,038 ppb) were not oversampled. This oversampling scheme proved optimal for subsequent binary classification about a 150 ppb AF threshold. The "nearest neighbor" parameter in the ADASYN algorithm was set to one, meaning that

each synthetic reflectance value was generated at a random point along the distance vector between an existing non-synthetic sample and its one nearest neighbor in Euclidean space. It is important to note that synthetically generated spectra and spectra from incubated kernels were used as training data only and never included in any test data.

AF Classification

AF concentrations were binarized as either one or zero for samples above and below a threshold of 150 ppb, respectively. SVN-transformed spectral signatures served as the independent variables (predictors) and the binary vector of AF concentrations served as the dependent variable (to be predicted) for binary classification tasks. Four training sets were partitioned based on the presence or absence of data from incubated kernels and synthetically generated samples, hereafter referred to as (1) -inc/-syn, (2) +inc/-syn, (3) -inc/+syn, and (4) +inc/+syn (Fig. 4). Binary classifiers were fit to partial least squares (PLS), random forest (RF), ridge, and lasso models using the four training sets for a total of four binary classifiers × four training sets = 16 binary classification models.

Model development was carried out using the following functions and parameters from the Scikit Learn library in Python (Pedregosa et al., 2011): "sklearn.cross_decomposition.PLSRegression" (n_components = 6), "sklearn.ensemble.RandomForestClassifier" (max_depth = 10, random_state = 0), "sklearn.linear_model.Ridge" (alpha = 10.0), and "sklearn.linear_model.Lasso" (alpha = 1.1, max_iter = 5000). Average classification performance for each model (n =

1,000 iterations) was evaluated in terms of model sensitivity (correctly classify samples ≥ 150 ppb, i.e. 1 – Type 2 error) and specificity (correctly classify samples < 150 ppb, i.e. 1 – Type 1 error).

Principal Component Analysis

Principal component analyses (PCA) were performed to assess if any spectral artifacts differentiated incubated and/or synthetically generated samples from their unincubated and/or non-synthetic counterparts. The PCAs were carried out using the "sklearn.decomposition.PCA" function (n_components = 2) from the Scikit Learn library in Python (Pedregosa et al., 2011).

Feature Importance

Regression coefficients associated with each wavelength were used to estimate feature importance for the linear models (PLS, ridge, and lasso). Coefficients were calculated using the "sklearn.cross_decomposition.PLSRegression.coef_", "sklearn.linear_model.Ridge.coef_", and "sklearn.linear_model.Lasso.coef_" functions from the Scikit Learn library in Python (Pedregosa et al., 2011). For the RF model, mean decreases in impurity (MDI) for each feature were estimated as $1 - [(P_{Over}) + (P_{Under})]^2$ where P_{over} and P_{under} is the proportion of nodes across all decision trees in the RF that classified the sample as above and below 150 ppb, respectively. This calculation was carried out using the "sklearn.ensemble.RandomForestClassifier.feature_importances_" function from the Scikit Learn library in Python (Pedregosa et al., 2011). The absolute values of feature

importance scores for all models were normalized to 1 (divided by the maximum feature importance score) to more easily identify key wavelengths shared between different models.

RESULTS

Training Set Enrichment

The addition of synthetic samples in the -inc/+syn training set strongly increased the proportion of samples ≥150 ppb compared to the -inc/-syn training set. Synthetic oversampling generated many more samples at the higher end of the AF range. The same was true when incubated kernels were included along with synthetic samples generated from incubated kernels. Humid incubation periods of three and six days (before visible moldiness became predominant) were effective at generating additional kernels between 5-75 ppb. However, these incubation treatments did not produce a significant additional number of kernels from higher toxin classes. While the number of kernels ≥150 ppb was higher in the +inc/-syn training set than the -inc/-syn training set, the proportion of kernels ≥150 ppb surprisingly decreased significantly ($p = 0.037$) from 18% to 11% (Fig. 5).

AF Classification

The addition of synthetically generated training samples greatly improved model sensitivity in all models tested. Classification errors were no longer skewed in the PLS (-inc/+syn and +inc/+syn), lasso (-inc/+syn and +inc/+syn), or ridge (+inc/+syn) models. Each of these models achieved over 70% accuracy for both sensitivity and

specificity. When enriched with synthetic training data, the PLS model performed best in the presence and absence of incubated training samples. The -inc/+syn PLS model achieved a mean specificity and sensitivity of 76% and 80%, respectively. The distribution of errors in the +inc/+syn model was remarkably balanced with a specificity and sensitivity each of 76% (Fig. 6).

The addition of humid incubated training samples (+inc/-syn) slightly increased classification accuracy of <150 ppb AF but did not increase sensitivity in any of the models compared to -inc/-syn (Fig. 6B). Mean specificities of the models remained much higher than mean sensitivities, ranging from 92-98% and 29%-50%, respectively. The addition of synthetic training samples generated from unincubated kernels (-inc/+syn) decreased specificity in all models compared to -inc/-syn, but none dipped below 70% (Fig. 6).

Principal Component Analysis

The pattern of data scattering within the data cloud differed between synthetic and non-synthetic data points. Synthetic data points were scattered along straight, trail-like patterns between pairs of non-synthetic data points in the two-dimensional Euclidean space. This straight-line pattern was a consequence of setting the nearest neighbor parameter in the ADASYN algorithm equal to one. Using that parameter, the ADASYN algorithm generated each synthetic reflectance value as a random point along the Euclidean distance vector between an existing non-synthetic sample and its nearest neighbor.

The first two PCs explained 94-95% of variance in our three comparisons (Fig. 7). Neither synthetically generated spectra nor spectra from incubated kernels appeared to cluster separately from non-synthetic and unincubated samples based on their first two PCs. The data clouds for -inc/-syn and +inc/-syn training set spectra exhibited similar densities of data points. Both -inc/-syn and +inc/-syn training set spectra had a higher density of samples with PC2 scores under 10 and markedly fewer samples with PC2 scores over 10. The density of the -inc/-syn data cloud and +inc/-syn data cloud were sparser for samples with PC2 scores below 0 and above 10, respectively, while these areas of the data cloud were densely populated with synthetic samples.

Feature Importance

Side-by-side comparison of each model's normalized feature importance distribution revealed five regions that overlapped in at least three of the four models (Fig. 8A): 329-345 nm and 380-385.5 nm in the UV range, 415-425 nm and 639-668 nm in the visible range, and 1,013.5-1,060 nm in the NIR range. Model-specific spectral regions of importance were observed in the PLS and RF models, but all regions of importance in the ridge and lasso models overlapped with at least one other model. This is logical given the much sparser distribution of feature importance values in the lasso and ridge models compared to the PLS and RF models. Differences in feature importance distributions between models were primarily due to differences in feature reduction performed by each model.

For non-linear models such as RF, regression coefficients cannot be used as estimators of feature importance. Instead, MDI, also known as the Gini index (Gini, 1912), of

each feature was compared. The "purity" of a given feature-based node on a decision tree in a RF refers to how unanimous (pure) its classification decision is at that node; the MDI, in turn, estimates how much more unanimous that decision becomes when a given feature is used as the root node of the decision tree compared to when it is not. No feature reduction is involved in this process, and therefore each individual wavelength has its own feature importance score (Fig 8B).

The PLS feature importance distribution appeared notably more continuous in nature compared to the RF (Fig. 8C). PLS does compute a feature importance score for each individual wavelength like RF, but these scores are based on a user-specified number of PCs. The PCs are computed in a similar way as in PCA with the key difference being that PLS also incorporates the variance of the predicted variable. The PLS classifier scans through each wavelength of the PCs looking for regions where the feature variance (e.g. spectral reflectance) is most correlated with the predicted variable (e.g. AF concentration). The regions with higher correlations are weighted more heavily by the classifier, which manifests as more gradual peaks and valleys in feature importance compared to RF.

Lasso and ridge are both considered "shrinkage" models, meaning they minimize (or "shrink") the importance of features with high residual errors. Like PLS, they represent feature variance through PCs, but unlike PLS, they impose "penalties" on features with high residual errors. Lasso does this by multiplying the correlation coefficient of each PC by its least absolute error. This process can cause some feature importance values to shrink to zero which excludes them from use in prediction. This

resulted in the sparser array of feature importance values in our lasso model compared to PLS or RF (Fig. 8D).

Ridge works in the same way as lasso, except it multiplies correlation coefficients by their residual squared error, i.e. the magnitude used by lasso squared. The squared error term causes more exaggerated shrinkage of correlation coefficients compared to lasso but prevents coefficients from reaching zero. Therefore, the distribution of feature importance values in ridge models tend to be even sparser than lasso, as was the case in our ridge model (Fig. 8E).

DISCUSSION

Improved Prediction of AF Contamination

The sensitivity rates of ~76-80% in our models trained with synthetic data greatly outperformed the 50% sensitivity rates previously reported using a non-synthetic adaptive bootstrapping algorithm (AdaBoost) for AF classification in single kernels (Chavez et al., 2022). Our models exhibited comparable sensitivity to other previously published models. Chu et al. (2017) achieved 75% sensitivity for classifying kernels >100 ppb AF, but this was only validated with eight >100 ppb AF kernels in their test set. Cheng et al. (2019) achieved 86% sensitivity for binary classification using a 20 ppb threshold (Cheng et al., 2019); while 22% of their kernels were above that threshold, only eight maize ears in total were used for the study.

Chavez et al. (2022) and Cheng et al. (2019) preferentially selected UV-fluorescent kernels in their studies, but the distribution of AF in their kernels remained highly

skewed. Pearson et al. (2001), Yao et al. (2010), and Zhu et al. (2016) employed the same UV enrichment strategy with more success, achieving sensitivities of 95%, 87% and 92%, respectively. Zhu et al. (2016) also incorporated hyperspectral reflectance images in their model, aiding to its high sensitivity. Kimuli et al. (2017) achieved even higher sensitivities about various AF thresholds on a test set containing 20 kernels with >100 ppb AF. While Pearson et al. (2001), Yao et al. (2010), and Zhu et al. (2016) utilized kernels from field-inoculated ears in their study, Kimuli et al. (2017) applied pure AF solution at standardized concentrations to the surfaces of kernels. Therefore, the spectral characteristics of these AF-contaminated kernels may not be representative of natural ear rot infections.

It is important to note that a model's classification sensitivity and specificity do not necessarily equate to toxin reduction in practice, i.e. 80% model sensitivity is unlikely to achieve 80% toxin reduction in a sorted sample. The degree of toxin reduction in a sample will be relatively high or low if the model removes the most toxic or least toxic kernels, respectively, above and below the contamination threshold. Future work should incorporate simulation modeling to estimate toxin reduction for various distributions of AF contamination based on a given model's sensitivity and specificity. These estimates would likely have more practical relevance to grain sorting applications than sensitivity and specificity estimates alone.

Predictive Wavelengths of AF Contamination

Key predictive wavelengths shared across multiple models provided robust insight on the spectral characteristics of AF contamination in single maize kernels. Five spectral

regions with high feature importance scores overlapped in at least three of our four models: 329-345 nm and 380-385.5 nm in the UV range, 415-425 nm and 639-668 nm in the visible range, and 1013.5-1060 nm in the NIR range. These spectral ranges could be applicable to the development of limited wavelength grain sorting devices for low-resource settings.

Within the 329-345 nm region, Aoun et al. (2022) implicated 332 nm as important for binary classification of maize kernels about 20 ppb and 100 ppb AF thresholds. Near the 380-385.5 nm range, the wavelengths 390 nm and 386.5 nm have been previously implicated in classification of AF in single maize kernels by Cheng et al. (2019) and Aoun et al. (2022), respectively. The importance of these regions is likely related to kojic acid, the UV-fluorescent secondary metabolite of *A. flavus* that commonly co-occurs with aflatoxin (Zeringue & Shih, 1998).

Carotenoids in the kernel seed coat may underpin the importance of the 415-425 nm nm region. Within this range, 420 nm and 425 nm have been noted as key absorption spectra of various carotenoids (Fedenko & Struchko, 1991; Kumar et al., 2002). Del Foire et al. (2010) found 410 nm to be important for discriminating between kernels recently inoculated with *A. flavus* and uninoculated controls. Carotenoids in the maize seed coat can have an inhibitory effect on mycotoxin production due to their antioxidant properties (Atanasova-Penichon et al., 2016; Suwarno et al., 2019).

It is also possible that AF itself could influence the kernels' spectral properties in this region, as pure AF standards diluted in methanol have peak fluorescence at 425 nm (Carnaghan et al., 1963; Hruska et al., 2014). Previous studies have demonstrated that

kernels treated with different concentrations of pure AF can be distinguished spectrally with high accuracy (Kimuli et al., 2018; Wang et al., 2015). However, Yao et al. (2010) found that UV fluorescence was negatively correlated (r = -0.67) with AF concentration in single maize kernels at ~415-425 nm. These contrasting results in the literature are consistent with real world observations of UV-fluorescent kernels without AF (Schmitt & Hurburgh, 1989; Thompson & Henke, 2000; Wicklow, 1999; Yao et al., 2010) and non-fluorescent kernels that harbor dangerously high levels of AF (Brekke et al., 1975; Gloria et al., 1998; Hruska et al., 2016; Pearson et al., 2001; Yao et al., 2010).

Within and near the 639-668 nm range, wavelengths around 655.5-659 nm (Aoun et al., 2022), 670.2 nm, and 671.6 nm (Wang et al., 2015) were found to be important for binary and multiclass classification of single kernels about various AF thresholds. While this range is most commonly associated with chlorophyll a and b (Curran, 1989; Kumar et al., 2002; Siemanroth et al., 1980), it may also be associated with kernel hardness (Robutti, 1995). Kernel hardness, a metric often used to estimate endosperm texture, could affect AF accumulation due to the preferential colonization of starchy endosperm by the *A. flavus* fungus (Windham et al., 2018). Kernel hardness may also be related to seed coat fiber content, which has been correlated with AF accumulation in maize (Mideros et al., 2012).

Predictive wavelengths for AF classification in maize kernels have been previously identified within the 1,013.5-1060 nm range at 1,038-1051 nm (Aoun et al., 2022) and 1,050 nm (Cheng et al., 2019). Several kernel composition traits including starch, oil,

and protein exhibit characteristic spectral reflectance in the NIR range, but their predictive wavelengths are generally not reported in this specific region. However, spectral reflectance at 1,018 nm has been correlated with ergosterol, a key component in fungal cell membranes (Berardo et al., 2005). Future studies including kernels inoculated with toxigenic and atoxigenic *A. flavus* strains could help identify spectral variance due to AF contamination versus fungal biomass.

Data-dependent Considerations for Synthetic Oversampling

Our results suggest that synthetic oversampling, both with and without the addition of incubated kernels, can effectively balance Type 1 and Type 2 error rates in single maize kernels. However, precautions should be taken when applying these methods to highly imbalanced datasets to prevent the introduction of confounding spectral variables.

While "checks" such as PCA should be performed to assess the presence of spectral artifacts between sample groups, these sorts of dimensionality reduction methods inherently do not capture the entirety of spectral variation among or between sample groups. For example, if a PCA captures 99% of spectral variation among two sample groups, it is possible that critical spectral variation could be present in the other 1% and go undetected. These spectral artifacts could decrease AF model prediction power by decreasing the proportion of variance in spectral signatures explained by AF.

Each incubated kernel's spectral signature had an associated AF concentration that was measured using ELISA; this was not the case for synthetically generated spectral

signatures. Synthetic AF concentrations were generated in the same way synthetic reflectance values were generated, that is, as a random point along the Euclidean distance vector between an existing sample and its nearest neighbor. Consequently, AF concentrations of synthetic samples were subject to any bias present in the non-synthetic sample and the nearest neighbor sample(s) from which they were generated. If nearest neighbors were not characteristic of the unincubated kernels of similar AF concentration, synthetic samples generated from these kernels would not be either. This bias could be compounded by the representativeness of the nearest neighbors, the number of nearest neighbors used to generate the synthetic spectra, and the number of synthetic samples generated.

In a scenario where a sample's nearest neighbors happen to have high variance at some spectral feature λ, the generated synthetic sample may exhibit a novel spectral component not similar to either neighbor. For example, consider a sample and its nearest neighboring sample. The two samples have highly similar percent reflectance at most spectral features, but at spectral feature λ, the samples' percent reflectance values are 0.9 and 0.1. Using the ADASYN algorithm with 1 nearest neighbor, a synthetic sample generated from these two neighboring samples would have a reflectance value at a random point along the Euclidean distance vector between 0.1 and 0.9. If this point happened to fall somewhere around the middle on this vector (e.g. 0.5), the resulting reflectance value may look like neither sample. Preprocessing transformation functions should be utilized on a dataset-by-dataset basis prior to synthetic oversampling to minimize the prevalence and impact of atypical synthetic spectral components on prediction model performance.

Humid Incubation

The relatively small number of high AF kernels generated from our humid incubation was likely a consequence of incubation duration. Longer incubation periods would have increased our likelihood of generating toxic kernels, but also would have increased the likelihood of visible fungal growth on kernels. For this study, we were most interested in analyzing toxic and non-toxic kernels that could not otherwise be distinguished visually. Additional replication is required to understand the tradeoffs between incubation treatments and confounding fungal growth. Additional experiments involving kernels of various pre-incubation AF concentrations is also necessary for assessing what distributions of AF contamination can be achieved through humid incubation. The imbalanced pattern of AF contamination in maize provides a promising vehicle for improved prediction using training set enrichment techniques.

References:

Aoun, M., Siegel, C., Windham, G., Williams, P., Nelson, R. (2022). Application of reflectance spectroscopy to identify maize genotypes and aflatoxin levels in single kernels. *World Mycotoxin Journal*, in prep. accepted for publication.

Aoun, M., Stafstrom, W., Priest, P., Fuchs, J., Windham, G. L., Williams, W. P., & Nelson, R. J. (2020). Low-cost grain sorting technologies to reduce mycotoxin contamination in maize and groundnut. *Food Control*, *118*, 107363. https://doi.org/10.1016/j.foodcont.2020.107363

Atanasova-Penichon, V., Barreau, C., & Richard-Forget, F. (2016). Antioxidant secondary metabolites in cereals: Potential involvement in resistance to *Fusarium* and mycotoxin accumulation. *Frontiers in Microbiology*, *7*. https://www.frontiersin.org/article/10.3389/fmicb.2016.00566

Bankole, S., Schollenberger, M., & Drochner, W. (2006). Mycotoxins in food systems in Sub Saharan Africa: A review. *Mycotoxin Research*, *22*(3), 163–169. https://doi.org/10.1007/BF02959270

Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A., & Marocco, A. (2005). Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. *Journal of Agricultural and Food Chemistry*, *53*(21), 8128–8134. https://doi.org/10.1021/jf0512297

Bothast, R. J., & Hesseltine, C. W. (1975). Bright greenish-yellow fluorescence and aflatoxin in agricultural commodities. *Applied Microbiology*, *30*(2), 337–338.

Brekke, O. L., Peplinski, A. J., & Griffin Jr., E. L. (1975). Cleaning trials for corn containing aflatoxin. *Cereal Chemistry*, *52*(2), 198–204.

Brownlee, J. (2020). *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python* (v1.2).

Carnaghan, R. B. A., Hartley, R. D., & O'Kelly, J. (1963). Toxicity and fluorescence properties of the aflatoxins. *Nature*, *200*(4911), 1101–1101. https://doi.org/10.1038/2001101a0

Chavez, R. A., Cheng, X., Herrman, T. J., & Stasiewicz, M. J. (2022). Single kernel aflatoxin and fumonisin contamination distribution and spectral classification in commercial corn. *Food Control*, *131*, 108393. https://doi.org/10.1016/j.foodcont.2021.108393

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Cheng, X., Vella, A., & Stasiewicz, M. J. (2019). Classification of aflatoxin contaminated single corn kernels by ultraviolet to near infrared spectroscopy. *Food Control*, *98*, 253–261. https://doi.org/10.1016/j.foodcont.2018.11.037

Chu, X., Wang, W., Yoon, S.-C., Ni, X., & Heitschmidt, G. W. (2017). Detection of aflatoxin B1 (AFB1) in individual maize kernels using short wave infrared (SWIR) hyperspectral imaging. *Biosystems Engineering*, *157*, 13–23. https://doi.org/10.1016/j.biosystemseng.2017.02.005

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, *30*(3), 271–278. https://doi.org/10.1016/0034-4257(89)90069-2

Del Fiore, A., Reverberi, M., Ricelli, A., Pinzari, F., Serranti, S., Fabbri, A. A., Bonifazi, G., & Fanelli, C. (2010). Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International Journal of Food Microbiology*, *144*(1), 64–71. https://doi.org/10.1016/j.ijfoodmicro.2010.08.001

Fedenko, V. S., & Struchko, V. S. (1991). Absorption spectra of the grain of maize with different phenotypes. *Chemistry of Natural Compounds*, *27*(6), 758–759. https://doi.org/10.1007/BF00629950

Gini, C. (1912). Variabilità e Mutabilità: Contributo allo studio delle distribuzioni e delle relazioni statistiche (Variability and Mutability: Contribution to the study of distributions and statistical relations).

Gloria, E. M., Fonseca, H., Calori-Domingues, M. A., & Souza, I. M. (1998). Evaluation of the black light test for screening aflatoxin-contaminated maize in the Brazilian food industry. *Food Additives and Contaminants*, *15*(2), 181–184. https://doi.org/10.1080/02652039809374628

Gong, Y. Y., Watson, S., & Routledge, M. N. (2016). Aflatoxin exposure and associated human health effects, a review of epidemiological studies. *Food Safety*, *4*(1), 14–27. https://doi.org/10.14252/foodsafetyfscj.2015026

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic

sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969

Hesseltine, C. W., & Shotwell, O. (1973). New methods for rapid detection of aflatoxin. In P. Krogh (Ed.), *Control of Mycotoxins* (pp. 259–266). Butterworth-Heinemann. https://doi.org/10.1016/B978-0-408-70536-3.50009-0

Hruska, Z., Yao, H., Kincaid, R., Brown, R., Cleveland, T., & Bhatnagar, D. (2014). Fluorescence excitation–emission features of aflatoxin and related secondary metabolites and their application for rapid detection of mycotoxins. *Food and Bioprocess Technology*, *7*(4), 1195–1201. https://doi.org/10.1007/s11947-014-1265-2

Hruska, Z., Yao, H., Kincaid, R., Zhu, F., Brown, R. L., & Bhatnagar, D. (2016). *Evaluating the BGY fluorescence signal in maize kernels inoculated with various aflatoxin-producing fungi using fluorescence hyperspectral imaging*. 9th Conference of the World Mycotoxin Forum and the XIVth IUPAC International Symposium on Mycotoxins, Winnipeg, Canada.

Johansson, A. S., Whitaker, T. B., Giesbrecht, F. G., Hagler, W. M., & Young, J. H. (2000). Testing shelled corn for aflatoxin, part II: Modeling the observed distribution of aflatoxin test results. *Journal of AOAC International*, *83*(5), 1270–1278.

Kimuli, D., Wang, W., Lawrence, K. C., Yoon, S.-C., Ni, X., & Heitschmidt, G. W. (2018). Utilisation of visible/near-infrared hyperspectral images to classify aflatoxin B1 contaminated maize kernels. *Biosystems Engineering*, *166*, 150–160. https://doi.org/10.1016/j.biosystemseng.2017.11.018

Kumar, L., Schmidt, K., Dury, S., & Skidmore, A. (2001). *Imaging Spectrometry and Vegetation Science*. Kluwer Academic Publishers. https://doi.org/10.1007/978-0-306-47578-8_5

Lemaıtre, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, *18*(1), 559–563.

Lewis, L., Onsongo, M., Njapau, H., Schurz-Rogers, H., Luber, G., Kieszak, S., Nyamongo, J., Backer, L., Dahiye, A. M., Misore, A., DeCock, K., Rubin, C., & Group, T. K. A. I. (2005). Aflatoxin contamination of commercial maize products during an outbreak of acute aflatoxicosis in eastern and central Kenya. *Environmental Health Perspectives*, *113*(12), 1763–1767.

Marsh, S., & Payne, G. (1984). Preharvest infection of corn silks and kernels by *Aspergillus flavus*. *Phytopathology*, *74*. https://doi.org/10.1094/Phyto-74-1284

Mideros, S. X., Windham, G. L., Williams, W. P., & Nelson, R. J. (2009). *Aspergillus flavus* biomass in maize estimated by quantitative real-time polymerase chain reaction is strongly correlated with aflatoxin concentration. *Plant Disease*, *93*(11), 1163–1170. https://doi.org/10.1094/PDIS-93-11-1163

Mideros, S. X., Windham, G. L., Williams, W. P., & Nelson, R. J. (2012). Tissue-specific components of resistance to Aspergillus ear rot of maize. *Phytopathology*, *102*(8), 787–793. https://doi.org/10.1094/PHYTO-12-11-0355

Mitchell, N. J., Bowers, E., Hurburgh, C., & Wu, F. (2016). Potential economic losses to the US corn industry from aflatoxin contamination. *Food Additives & Contaminants: Part A*, *33*(3), 540–550. https://doi.org/10.1080/19440049.2016.1138545

Muga, F. C., Marenya, M. O., & Workneh, T. S. (2019). Effect of temperature, relative humidity and moisture on aflatoxin contamination of stored maize kernels. *Bulgarian Journal of Agricultural Science*, *25*(2), 271–277.

Mutegi, C. K., Cotty, P. J., & Bandyopadhyay, R. (2018). Prevalence and mitigation of aflatoxins in Kenya (1960-to date). *World Mycotoxin Journal*, *11*(3), 341–357. https://doi.org/10.3920/WMJ2018.2362

Mutiga, S. K., Were, V., Hoffmann, V., Harvey, J. W., Milgroom, M. G., & Nelson, R. J. (2014). Extent and drivers of mycotoxin contamination: Inferences from a survey of Kenyan maize mills. *Phytopathology*, *104*(11), 1221–1231. https://doi.org/10.1094/PHYTO-01-14-0006-R

Pearson, T. C., Wicklow, D. T., Maghirang, E. B., Xie, F., & F. E. (2001). Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. *Transactions of the ASAE*, *44*(5). https://doi.org/10.13031/2013.6418

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,

Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Pitt, J. I., Taniwaki, M. H., & Cole, M. B. (2013). Mycotoxin production in major crops as influenced by growing, harvesting, storage and processing, with emphasis on the achievement of Food Safety Objectives. *Food Control*, *32*(1), 205–215. https://doi.org/10.1016/j.foodcont.2012.11.023

Reule, A. G. (1976). Errors in spectrophotometry and calibration procedures to avoid them. *Journal of Research of the National Bureau of Standards. Section A, Physics and Chemistry*, *80A*(4), 609–624. https://doi.org/10.6028/jres.080A.060

Robutti, J. (1995). Maize kernel hardness estimation in beeding by near-infrared transmission analysis. *Cereal Chemistry*, *72*(6), 632–636.

Salim, S. G. R., Fox, N. P., Theocharous, E., Sun, T., & Grattan, K. T. V. (2011). Temperature and nonlinearity corrections for a photodiode array spectrometer used in the field. *Applied Optics*, *50*(6), 866–875. https://doi.org/10.1364/AO.50.000866

Schmitt, S. G., & Hurburgh, C. R. (1989). Distribution and measurement of aflatoxin in 1983 Iowa corn. *Cereal Chemistry*, *66*(3), 165–168.

Shi, H., Stroshine, R. L., & Ileleji, K. (2014). Aflatoxin reduction in corn by cleaning and sorting. ASABE and CSBE/SCGAB Annual International Meeting,

Montreal, Quebec.

Shinha, K. K., & Bhatnagar, D. (1998). *Mycotoxins in Agriculture and Food Safety*. CRC Press.

Shotwell, O. L., & Hesseltine, C. W. (1981). Use of bright greenish yellow fluorescence as a presumptive test for aflatoxin in corn. *Cereal Chemistry, 58*(2), 124–127.

Siemenroth, A., Börner, T., & Metzger, U. (1980). Biochemical studies on the *Iojap* mutant of maize. *Plant Physiology*, *65*(6), 1108–1110. https://doi.org/10.1104/pp.65.6.1108

Smart, M. G. (1990). Pathogenesis in Aspergillus ear rot of maize: Aflatoxin B_1 levels in grains around wound-inoculation sites. *Phytopathology*, *80*(12), 1283. https://doi.org/10.1094/Phyto-80-1283

Stasiewicz, M. J., Falade, T. D. O., Mutuma, M., Mutiga, S. K., Harvey, J. J. W., Fox, G., Pearson, T. C., Muthomi, J. W., & Nelson, R. J. (2017). Multi-spectral kernel sorting to reduce aflatoxins and fumonisins in Kenyan maize. *Food Control*, *78*, 203–214. https://doi.org/10.1016/j.foodcont.2017.02.038

Suwarno, W. B., Hannok, P., Palacios-Rojas, N., Windham, G., Crossa, J., & Pixley, K. V. (2019). Provitamin A aarotenoids in grain geduce aflatoxin contamination of maize while combating vitamin A deficiency. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00030

Thompson, C., & Henke, S. E. (2000). Effect of climate and type of storage container on aflatoxin production in corn and its associated risks to wildlife species. *Journal of Wildlife Diseases*, *36*(1), 172–179. https://doi.org/10.7589/0090-3558-36.1.172

Trenk, H. L., & Hartman, P. A. (1970). Effects of moisture content and temperature on aflatoxin production in corn 1. *Applied Microbiology*, *19*(5), 781–784.

Wang, W., Heitschmidt, G. W., Windham, W. R., Feldner, P., Ni, X., & Chu, X. (2015). Feasibility of detecting aflatoxin B1 on inoculated maize kernels surface using vis/NIR hyperspectral imaging. *Journal of Food Science*, *80*(1), M116–M122. https://doi.org/10.1111/1750-3841.12728

Whitaker, T. B., & Johansson, A. S. (2005). Sampling uncertainties for the detection of chemical agents in complex food matrices. *Journal of Food Protection*, *68*(6), 1306–1313. https://doi.org/10.4315/0362-028X-68.6.1306

Wicklow, D. T. (1999). Influence of *Aspergillus flavus* strains on aflatoxin and bright greenish yellow fluorescence of corn kernels. *Plant Disease*, *83*(12), 1146–1148. https://doi.org/10.1094/PDIS.1999.83.12.1146

Willcox, M. C., Davis, G. L., Warburton, M. L., Windham, G. L., Abbas, H. K., Betrán, J., Holland, J. B., & Williams, W. P. (2013). Confirming quantitative trait loci for aflatoxin resistance from Mp313E in different genetic backgrounds. *Molecular Breeding*, *32*(1), 15–26. https://doi.org/10.1007/s11032-012-9821-9

Williams, J. H., Phillips, T. D., Jolly, P. E., Stiles, J. K., Jolly, C. M., & Aggarwal, D. (2004). Human aflatoxicosis in developing countries: A review of toxicology, exposure, potential health consequences, and interventions. *The American Journal of Clinical Nutrition*, *80*(5), 1106–1122. https://doi.org/10.1093/ajcn/80.5.1106

Windham, G. L., Williams, W. P., Mylroie, J. E., Reid, C. X., & Womack, E. D. (2018). A Histological study of *Aspergillus flavus* colonization of wound inoculated maize kernels of resistant and susceptible maize hybrids in the field. *Frontiers in Microbiology*, *9*, 799. https://doi.org/10.3389/fmicb.2018.00799

Yao, H., Hruska, Z., Kincaid, R., Brown, R., Cleveland, T., & Bhatnagar, D. (2010). Correlation and classification of single kernel fluorescence hyperspectral data with aflatoxin concentration in corn kernels inoculated with *Aspergillus flavus* spores. *Food Additives & Contaminants: Part A*, *27*(5), 701–709. https://doi.org/10.1080/19440040903527368

Zain, M. E. (2011). Impact of mycotoxins on humans and animals. *Journal of Saudi Chemical Society*, *15*(2), 129–144. https://doi.org/10.1016/j.jscs.2010.06.006

Zeringue, H. J., & Shih, B. Y. (1998). Extraction and separation of the bright-greenish-yellow fluorescent material from aflatoxigenic *Aspergillus spp.* infected cotton lint by HPLC−UV/FL. *Journal of Agricultural and Food Chemistry*, *46*(3), 1071–1075. https://doi.org/10.1021/jf9707391

Zhu, F., Yao, H., Hruska, Z., Kincaid, R., Brown, R., Bhatnagar, D., & Cleveland, T.

(2016). Integration of fluorescence and reflectance visible near-infrared (VNIR) hyperspectral images for detection of aflatoxins in corn kernels. *Transactions of the ASABE*, *59*(3), 785–794. https://doi.org/10.13031/trans.59.11365

Zuber, M. S., Lillehoj, E. B., & Renfro, B. L. (1986, April 7). A proceedings of the workshop. *Aflatoxin in Maize*, El Batan, Mexico.

Zummo, N., & Scott, G. E. (1989). Evaluation of field inoculation techniques for screening maize genotypes against kernel infection by *Aspergillus flavus* in Mississippi. *Plant Disease*, *73*(4), 313–316.

APPENDIX

SECTION S1.

Deoxynivalenol Diagnostic Report & Recommendations

During the summer of 2019, I collaborated with growers from a farm in New York State who had concerns about deoxynivalenol in their corn. This is a report written for them that summarizes my diagnostic findings based on a collection of samples from their farm. The report also provides background information on Gibberella ear rot and management recommendations for preventing future DON contamination in their corn.

Deoxynivalenol Diagnostic Report & Recommendations

Summary

Ten 1.1-lb bags of white dent corn and four additional white corn samples from commercial products were submitted for diagnostic testing due to concerns about kernel appearance. Visual indicators of poor seed health were apparent to us as well, such as discoloration and kernel breakage. Based on the symptoms, we decided to test the submitted samples for one of the most common corn ear rot diseases in Central New York, and the Northeast U.S. as whole: gibberella ear rot. Gibberella ear rot is caused by the fungus *Fusarium graminearum*. This locally rampant fungus produce a toxic chemical called deoxynivalenol (DON), or vomitoxin. When livestock consume grain that has been contaminated with DON, the animal often suffers from vomitting, feed refusal, and weight loss. DON-contaminated food products pose a human health risk as well; nausea, vomitting, diarrhea, and fever are common adverse effects. As such, the U.S. Food & Drug Administration has put regulations in place for the level of DON that can be present in various grain products.

1 ppm: Finished (post-milling) grain products for human consumption

5 ppm: Grain and byproducts for swine consumption, not to exceed 20% of diet
Grain and byproducts for non-swine livestock, dog, and cat consumption, not to exceed 40% of diet
Grain and byproducts for consumption by ruminating dairy cattle older than 4 months

10 ppm: Grain and byproducts for chicken consumption, not to exceed 50% of diet
Grain and byproducts (at 88% dry matter basis) for consumption by ruminating beef and feedlot cattle older than 4 months

30 ppm: Distillers grains, brewers grains, and gluten feeds/meals derived from grains (at 88% dry matter basis) for consumption by ruminating beef cattle, feedlot cattle, and ruminating dairy cattle older than 4 months

It is important to note that many grain buyers hold significantly more stringent standards for DON levels than the FDA advisory. For example, the FDA advisory for DON in distillers grains is 30 ppm, but many buyers will only purchase grain well below this threshold because DON can become more concentrated during the distilling process.

An enzyme-linked immunosorbent assay (ELISA) is the most common test used for quantifying DON concentration in a sample. The samples were diluted to three different levels in order to ensure that their DON levels would fall within the bounds of a standard curve of assay standards. Two replicates of each dilution were assayed for a total of six replicates from each bag of grain. ELISA results showed that DON was detected in grain from each of the 10 bags, most of which exceed FDA advisory levels for human and livestock consumption.

Deoxynivalenol Diagnostic Report & Recommendations

Results

DON concentrations from the 10 bags of corn (A-J) and additional samples from commercial products are shown below.

	Bio Rep Average (ppm)	Sample Average (ppm)		Bio Rep Average (ppm)	Sample Average (ppm)
Sample A			**Sample F**		
A Undiluted	0.30	0.34	F Undiluted	1.04	1.14
A 1:5	0.30		F 1:5	1.73	
A 1:10	0.42		F 1:10	0.66	
Sample B			**Sample G**		
B Undiluted	0.55	0.41	G Undiluted	1.05	1.09
B 1:5	0.69		G 1:5	1.41	
B 1:10	0.01		G 1:10	0.82	
Sample C			**Sample H**		
C Undiluted	0.95	0.47	H Undiluted	0.50	0.98
C 1:5	0.45		H 1:5	1.06	
C 1:10	0.00		H 1:10	1.37	
Sample D			**Sample I**		
D Undiluted	0.50	0.63	I Undiluted	1.13	1.04
D 1:5	1.09		I 1:5	2.00	
D 1:10	0.30		I 1:10	0.00	
Sample E			**Sample J**		
E Undiluted	0.83	0.97	J Undiluted	0.55	0.84
E 1:5	0.91		J 1:5	1.06	
E 1:10	1.18		J 1:10	0.91	

	Bio Rep Average (ppm)	Sample Average (ppm)
Roasted White Corn Flour Lot 39		
Replicate 1	1.94	1.82
Replicate 2	1.70	
Roasted White Corn Flour Lot 36		
Replicate 1	2.22	2.16
Replicate 2	2.10	
Hulled White Corn Lot 36		
Replicate 1	2.13	2.17
Replicate 2	2.20	
White Corn Flour Lot 39		
Replicate 1	1.02	1.03
Replicate 2	1.04	

Deoxynivalenol Diagnostic Report & Recommendations

Recommendations

Based on the ELISA results, samples F, G, and I and the 4 samples from commercial products exceed the DON regulatory limit for human consumption. Samples E and H were right on the border. Since these samples were all below 5 ppm, they are consdiered suitable for all other purposes besides human consumption. The remaining samples are suitable for human consumption. Strategies can be implemented to prevent DON from causing problems in the future.

PRE-HARVEST:

Resistant Varieties - We understand that the maize varieties currently grown on your land hold great cultural and culinary significance. But it is possible that these varieties may be more susceptible to diseases than commerically-available resistant varities. If you choose to explore different varieties in the future, the following are examples of commerical hybrids that exhibit some resistance to gibberella ear rot and other diseases that you may encounter. Most large seed companies' websites provide ratings for their varieties' resistance to an array of common diseases.

- Pioneer P1197 (https://www.pioneer.com/home/site/ca/products/profile-perf?smo=CPB&productLine=010&language=01&productCode=P1197&ts=?reload)
- Pioneer P0622Q (https://www.pioneer.com/home/site/ca/products/profile-perf?smo=CPB&productLine=010&language=01&productCode=P0622Q&ts=Q,LL,RR2?reload)

Fungicides - If chemical fungicides are an option for your operation, you may wish to consider Proline® from Bayer CropScience (Label: https://s3-us-west-1.amazonaws.com/agrian-cg-fs1-production/pdfs/Proline_480_SC1h_Fungicide_Label.pdf). This NY State-approved fungicide is labeled for silk protection against ear rots, and should be applied at a rate of 5.7 fl oz/acre.

Minimizing soil compaction – Fungi tend to love moisture. Reducing soil compaction by tilling improves drainage in the field, which can reduce growth of the fungus that produces DON.

Crop rotation – If different crops are being grown on your land, rotating which crops are grown in the field from year-to-year can help degrade disease-causing fungi populations in the soil. However, the fungus that causes gibberella ear rot can also cause other diseases in wheat, so a maize/wheat rotation is not advised. Rotation with a non-cereal crop such as soybean would be a better choice.

Recommendations

PRE-HARVEST CONT'D:

Intercropping – Though it is more labor-intensive at planting and harvest, growing different crops alongside each other in a single field can reduce the buildup of host-specific pathogens.

Scouting - Scouting prior to the plants' physiological maturity can be very useful in identifying areas of the field where ear rots may be present. Plants from these areas should be harvested early and dried to 15% moisture to prevent further fungal growth.

AT HARVEST AND POST-HARVEST:

Harvest- Kernels that have been wounded or otherwise physically damaged are more susceptible to infection. Once they are colonized by a fungus, kernels tend to become physically lighter as well. If possible, adjust harvesting combine settings to reduce kernel damage and sort out the lightest kernels. These kernels are the most likely to harbor DON contamination.

Storage – Maintaining dry storage conditions where grain is protected from insect damage helps prevent the growth and spread of fungi. Always store grain with rot symptoms separately from healthy grain.

More information on gibberella ear rot can be found here:
https://www.extension.purdue.edu/extmedia/BP/BP-77-W.pdf
https://cropprotectionnetwork.org/resources/articles/diseases/gibberella-ear-rot-of-corn

SECTION S2.

Spectral detection of Gibberella ear rot in corn feed products

During the summer of 2021, I was awarded funding from the Cornell Institute for Digital Agriculture to conduct a research project on DON in corn silage. I collaborated with an industry mentor, John Winchell of Alltech, and a cow veterinarian, Daryl Nydam of Cornell University College of Veterinary Medicine. Using naturally inoculated ears harvested from Musgrave Research Farm and feed samples collected from 15 local dairies, I developed spectral prediction models for DON in multiple maize tissues.

LIST OF FIGURES

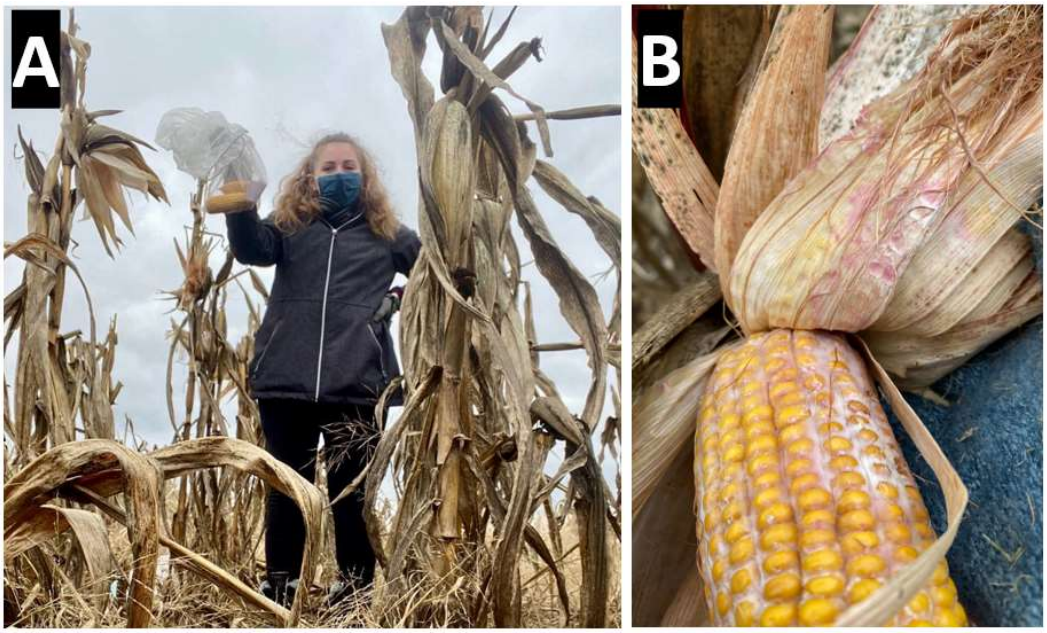

Figure 1. A) Collection of naturally infected ears in Aurora, NY. **B)** Example of a naturally infected corn plant exhibiting severe symptoms of Gibberella ear rot. Photos taken at Musgrave Research Farm on October 31, 2020.

Figure 2. A) Sample collection from a local corn silage farmer in Harford, NY. **B)** 4-6 corn plants per plot were pulled from the field and run through a tissue chopper while chopped samples are collected in plastic bags.

Figure 3. A) Collection of TMR samples from a dairy cow feeding stall. Examples of TMR for **(B)** lactating and **(C)** non-lactating cows were collected for spectral data acquisition and DON quantification.

Sample Type:

A) = Naturally infected ears

B) = Corn silage

C) = TMR

Farm #	Location	# of Samples
1	Eerie County, NY	1
2	Washington County, VT	6
3	Canaan County, CT	1
4	Marcellus, NY	~5
5	Marietta, NY	~5
6	Aurora, NY	127
7	Scipio Center, NY	60
8	Apulia, NY	~1
9	Preble, NY	~15
10	Truxton, NY	~1
11	Lansing, NY	10
12	Tompkins County, NY	1
13	Harford, NY	~30
14	Whitney Point, NY	2
15	Greene, NY	10
16	Spencer, NY	11

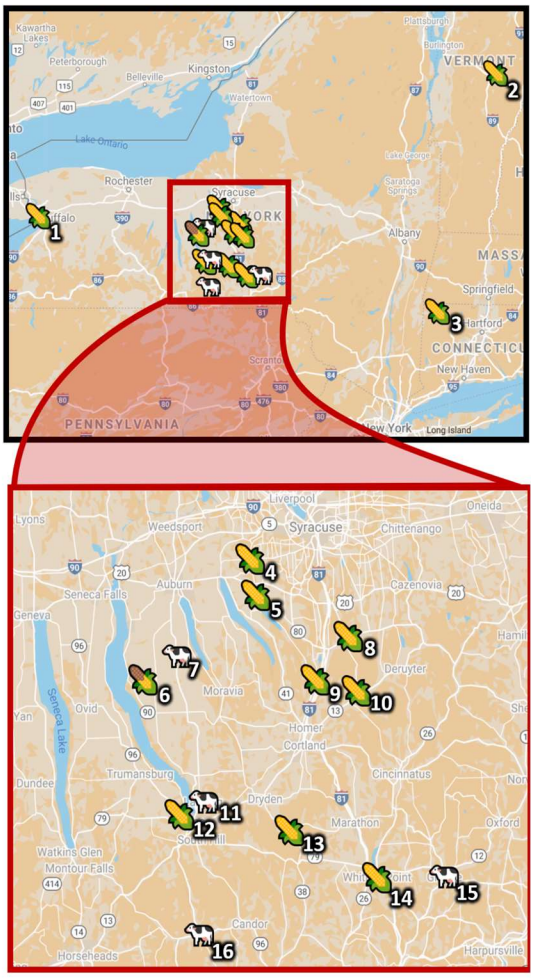

Figure 4. A) Naturally infected ears were harvested at Musgrave Research Farm in Aurora, NY in Fall 2018 and 2020. **B)** Corn silage samples were collected from 11 corn fields in the Northeast U.S. in September 2020 and 2021 with most farms located in central New York. **C)** TMR samples were collected from 4 local dairies in July and August 2021. Among the 60 TMR samples collected from Scipio Center, NY, 20 samples were collected on the morning of July 28, the evening of August 4, and the afternoon of August 12, respectively.

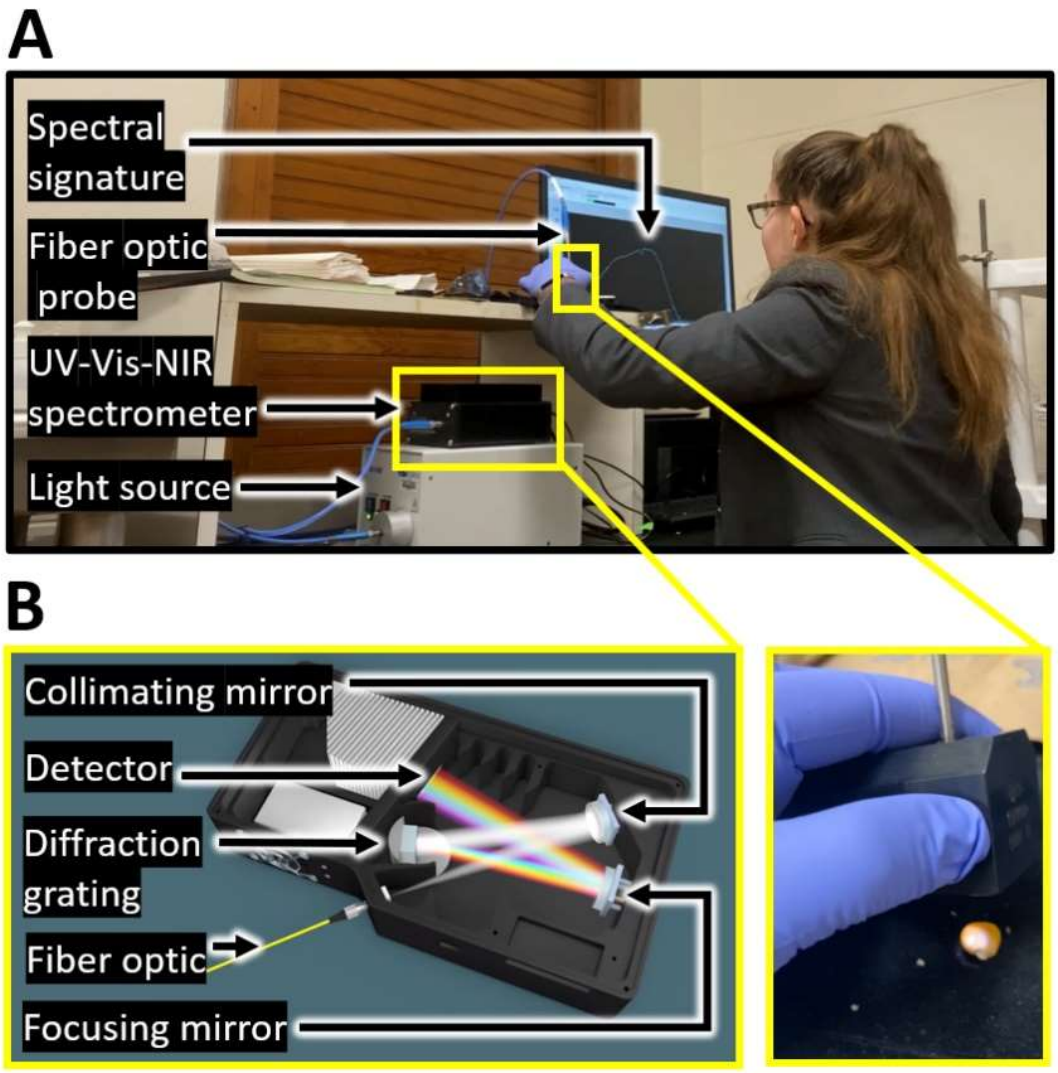

Figure 5. A) Spectroscopy system used for acquisition of UV-Vis-NIR spectral reflectance. **B)** Mechanism by which each sample's spectral reflectance is acquired. Spectrometer diagram adapted from Ocean Insight (2017).

Figure 6. A) Examples of ears exhibiting a directional infection gradient of GER symptoms from tip to base. **B)** Spectral reads were acquired from 5 excised kernels and from the rachis tissue beneath each excised kernel (n = 10 spectral reads per ear).

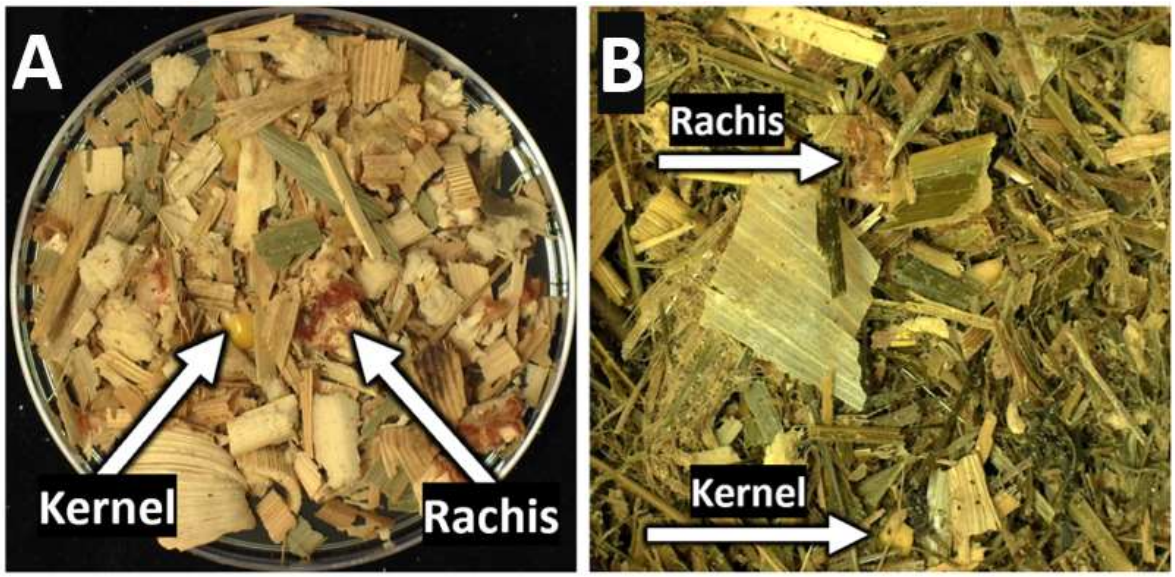

Figure 7. Spectral measurements were acquired from the center of the surface of each corn silage and TMR sample three times with sample mixing between each spectral scan. In addition, three kernel fragments and three rachis fragments from each corn silage TMR sample were removed and scanned. Arrows point to examples of kernel and rachis fragments in a **(A)** corn silage sample and **(B)** TMR sample.

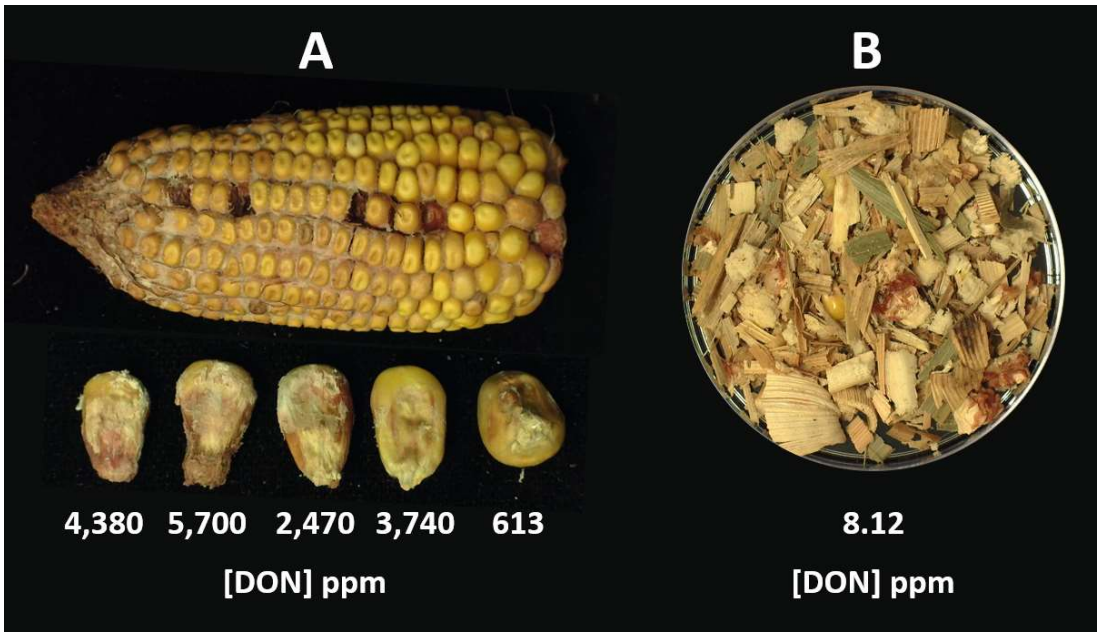

Figure 8. A) DON concentrations of kernels from a highly toxic ear harvested in Aurora, NY in November 2018. **B)** Severely contaminated corn silage sample collected from Canaan County, CT in September 2020. These examples underpin the phenomenon that DON concentration (a ratio of mass/volume) in toxic single kernels can be many orders of magnitude higher than toxic corn silage due their different scales of volume.

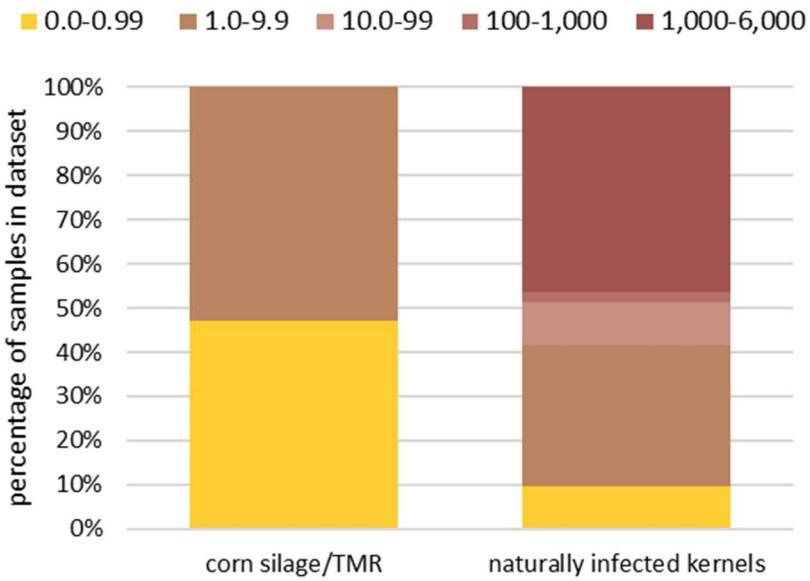

Figure 9. Distributions of DON concentration (ppm) in corn silage and TMR samples and in naturally infected kernels from ears with infection gradient of GER symptoms. Corn silage/TMR samples kernel samples were approximately evenly distributed about the 1 ppm threshold and 1,000 ppm thresholds.

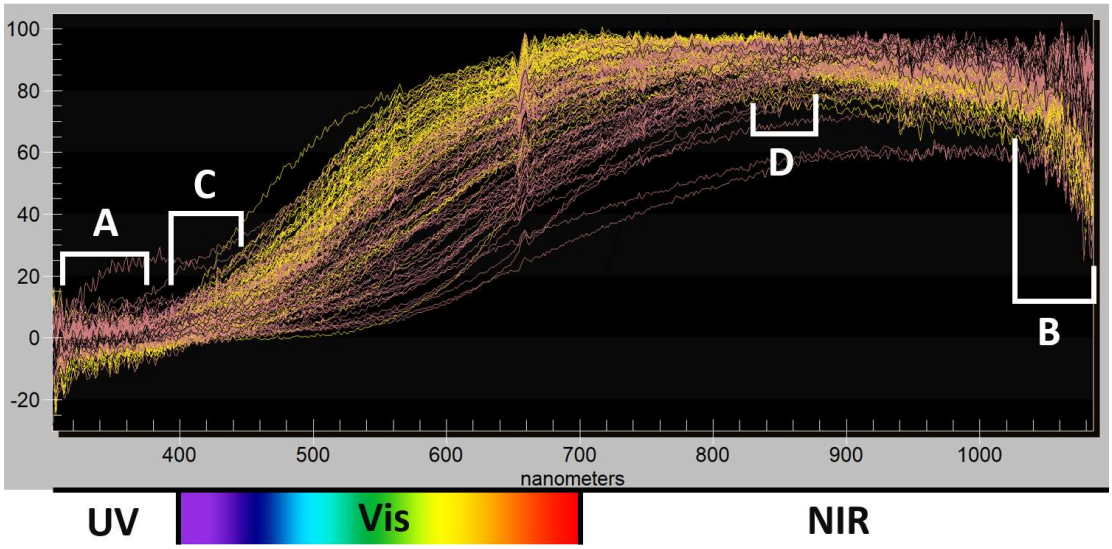

Figure 10. Percent reflectance of kernels from naturally infected ears from the 2018 field season. Each yellow line represents the spectral signature of an asymptomatic kernel and each pink line represents the spectral signature of a symptomatic kernel. **A** and **B** highlight ranges at which symptomatic and asymptomatic exhibit apparent distinction in spectral reflectance, while **C** and **D** highlight ranges of where light is reflected with more similar intensities.

A

		Naturally Infected Ears	
Model	**Metric**	**Kernel**	**Rachis**
PLS	Accuracy	91.6667%	85.0000%
	Sensitivity	90.0000%	83.3333%
	Specificity	93.3333%	86.6667%
RF	Accuracy	90.0000%	78.3333%
	Sensitivity	93.3333%	86.6667%
	Specificity	86.6667%	70.0000%
SVM	Accuracy	86.6667%	76.6667%
	Sensitivity	86.6667%	83.3333%
	Specificity	86.6667%	70.0000%

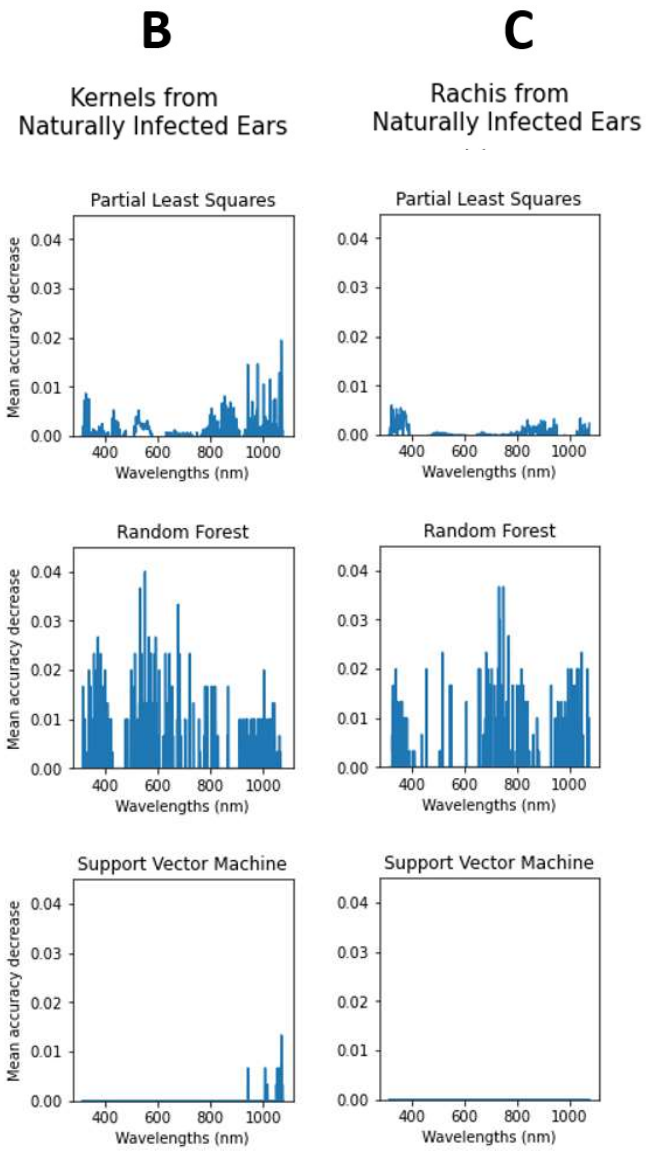

Figure 11. A) Binary classification models were PLS, RF, and SVM classifiers to predict if kernel and rachis tissues from naturally infected ears were excised from symptomatic or asymptomatic sections of the ear. Each of the three classifiers' mean accuracy decrease was calculated as a feature importance metric for spectral reflectance of **(B)** kernels and **(C)** rachis tissues at each wavelength.

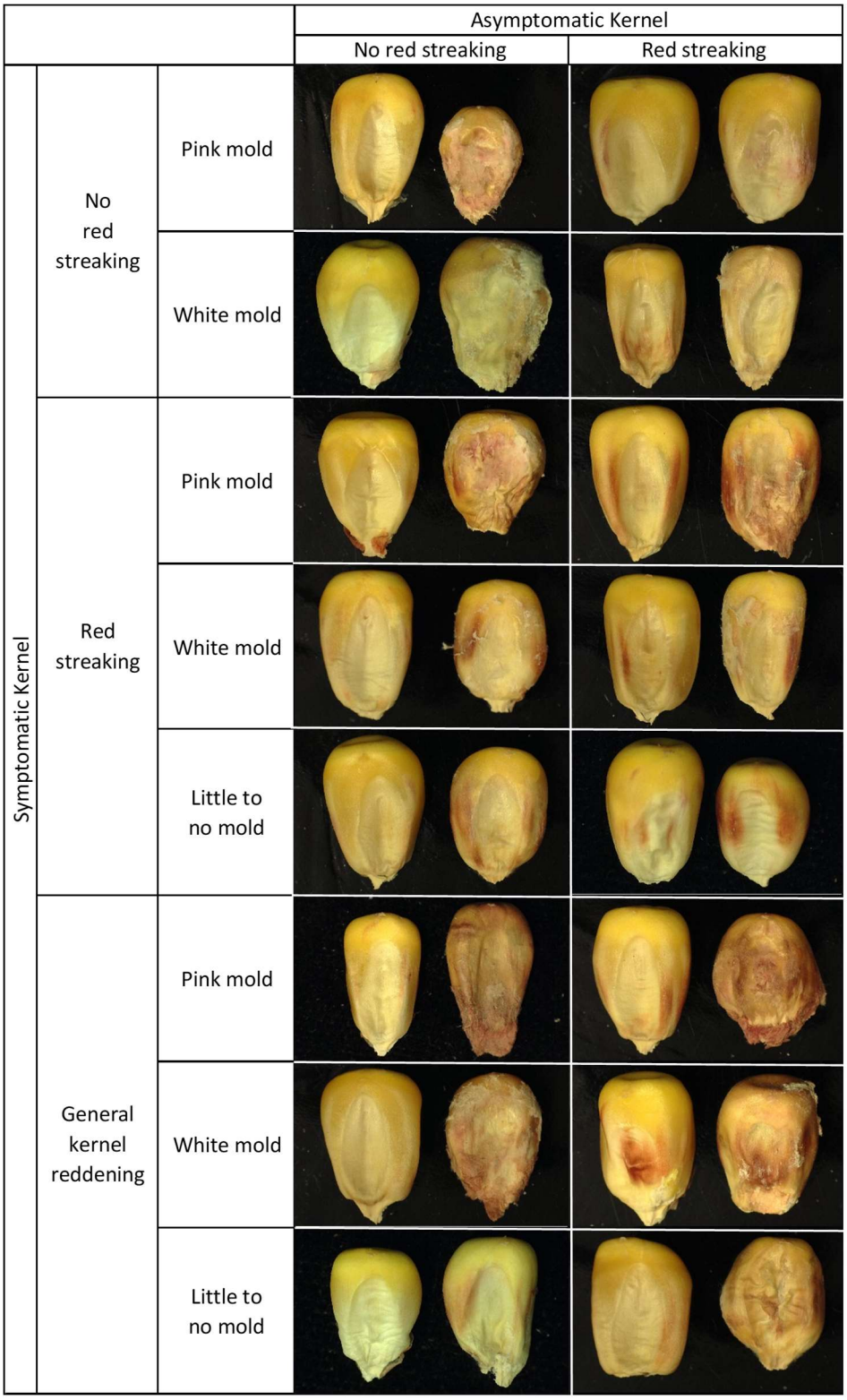

Figure 12. Examples of kernel pairs with different combinations of redness-related phenotypes. Kernels excised from asymptomatic and symptomatic sections of each ear are shown on the left and right, respectively, for each pair.

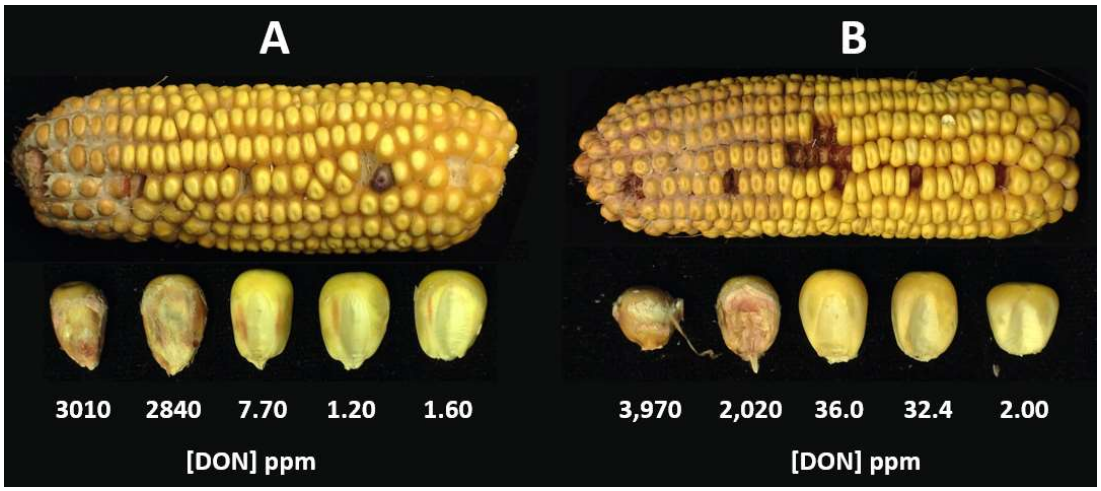

Figure 13. While red kernel streaking can sometimes be a symptom of GER, ELISA results show that **(A)** red streaking in can occur before any major accumulation of DON, and **(B)** kernels without any red streaking can have higher DON accumulation than those kernels with red streaking. This implies that DON accumulation in kernels is not always preceded by nor followed by red streaking.

A

		Corn Silage and TMR		
Model	**Metric**	**Kernel Fragments**	**Rachis Fragments**	**Scan from top of pile**
PLS	Accuracy	61.1111%	63.3333%	57.7778%
	Sensitivity	66.6667%	66.6667%	52.0833%
	Specificity	54.7619%	59.5238%	64.2857%
RF	Accuracy	67.7778%	58.8889%	61.1111%
	Sensitivity	64.5833%	56.2500%	75.0000%
	Specificity	71.4286%	61.9048%	45.2381%
SVM	Accuracy	70.0000%	55.5556%	66.6667%
	Sensitivity	70.8333%	45.8333%	72.9167%
	Specificity	69.0476%	66.6667%	59.5238%

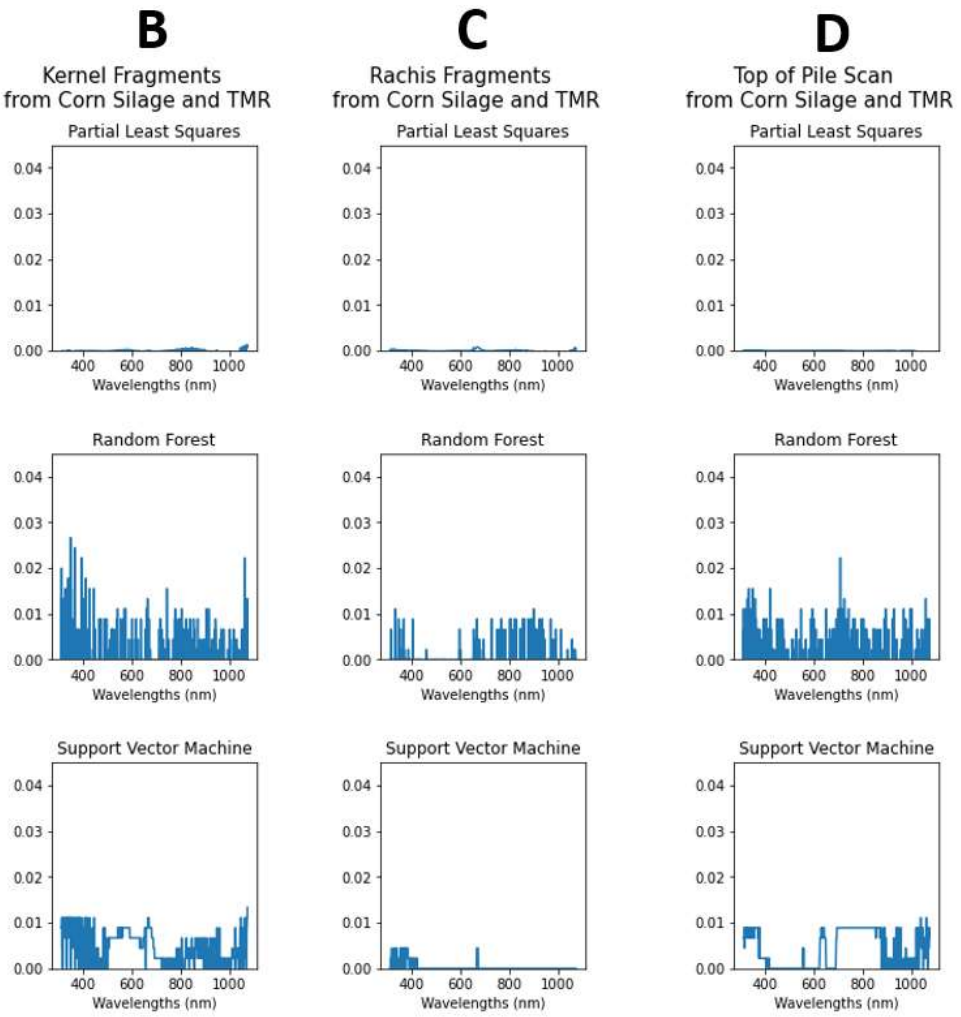

Figure 14. A) Binary classification models were trained with PLS, RF, and SVM classifiers to predict if corn silage and TMR samples had DON concentrations of over or under 1 ppm. Each of the three classifiers' mean accuracy decrease was calculated as a feature importance metric for spectral reflectance of **(B)** kernel fragments, **(C)** rachis fragments, and **(D)** scans taken from the top of a pile of silage of TMR at each wavelength.

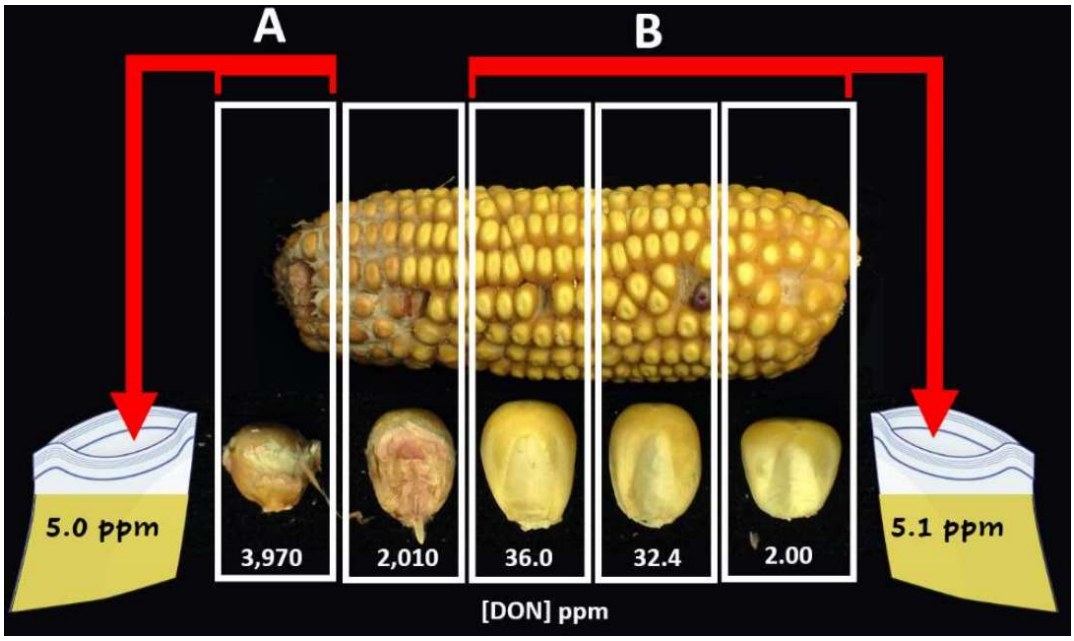

Figure 15. Example scenario highlighting that DON would exceed the legal limit of 5.0 ppm if **A)** 6 shelled kernels with DON concentrations of 3,970 ppm or **(B)** the shelled asymptomatic kernels from the lower 16 rows on the ear were present in an otherwise completely clean 0.9 kg sample of TMR.

Spectral detection of Gibberella ear rot in corn feed products*

INTRODUCTION

Deoxynivalenol (DON, also known as vomitoxin) contamination in livestock feed is a major threat to the corn and dairy industries in the eastern U.S. and Canada. DON is produced by *Fusarium graminearum*, a generalist fungal pathogen that causes Gibberella ear rot (GER) and stalk rot in corn. Corn plants contaminated with DON have lower yields and reduced crop value, and in turn, ingestion of DON-contaminated feed by dairy cows can cause feed refusal and reduced milk production. Annual losses due to DON in the U.S. have been estimated at $655 million (Schmale & Munkvold, 2018). Conventional detection of DON involves time and resource-intensive immunological lab tests such as enzyme-linked immunosorbent assays (ELISA) or liquid chromatography mass spectrometry (LC-MS). These tests are prohibitively expensive for many corn farmers, some of whom abandon their crops rather than submit samples for testing.

Inexpensive and non-destructive methods for detecting DON based on the spectral properties of corn kernels have shown promising success at various prediction tasks (Saccon et al., 2017). Kernel-based models, however, may not be effective for assessing corn silage and silage-based total mixed rations (TMR), which contain a mix of corn tissues. Available evidence suggests that the corn rachis (cob) can harbor much more DON than the kernels (Czembor & Waśkiewicz, 2015; Lauren & Di

*In the event of publication, the authors on this manuscript will include myself, John Winchell, Daryl Nydam, and Rebecca Nelson (subject to additional authors).

Menna, 1999; Oldenburg & Ellner, 2015), but rachis tissues have not been included in spectral models for DON prediction. Models that leverage spectral data from both kernel and rachis tissues could inform cost-effective DON surveillance in corn feed products.

The objective of this study was to develop a spectral model that predicts the likelihood that DON contamination in a corn silage or TMR sample exceeds 1 part per million (ppm). The ultimate envisioned application of this model is to provide extension agents and/or dairy nutritionists with a pre-screening tool to prioritize which corn silage or TMR samples to submit for immunological testing.

METHODS

Plant Materials

Naturally infected ears. Naturally infected ears of corn were sourced from harvested material from our research program's 2018 and 2020 field seasons at Musgrave Research Farm in Aurora, NY (Fig. 1A; Fig. 4A). All these plants were near-isogenic lines to the B73 reference genome. Ears exhibiting symptoms of GER were identified by visual inspection from material harvested in 2018 (n = 97 ears) and 2020 (n = 30 ears) (Fig. 1B). The likely presence of *F. graminearum* in these ears based on their symptomology was verified by an expert in the field (Gary Bergstrom, personal communication).

One kernel was excised from a symptomatic and an asymptomatic part of each ear. Whenever possible, care was taken to select symptomatic and asymptomatic kernels of

similar morphology in order to minimize differences not related to *F. graminearum* infection. However, there were some ears for which the symptomatic and asymptomatic sections of the ear did not contain morphologically similar kernels. For a small minority of ears where GER symptoms were most severe, every kernel was symptomatic; in these cases, an asymptomatic kernel was selected from a genetically identical asymptomatic ear from the same field plot.

Corn silage. Corn silage samples were collected from 11 corn fields in September 2020 (n = 9 silage samples from 4 farms) and September 2021 (n = ~68 silage samples from 7 farms) (Fig. 2A; Fig. 4B). The number of samples collected per farm ranged from 1 to ~30. Silage samples came from a mix of brown midrib (low lignin mutants) and conventional hybrids, depending on the source farm. For each corn silage sample, 4-6 whole corn plants from a given field plot were run through a gas-powered tissue chopper, then a few handfuls of the chopped plants were collected in plastic bags (Fig. 2B).

Total mixed rations. TMR samples were collected directly from dairy cow feeding stalls at local farms in July and August 2021 (n = 90 TMR samples from 4 farms) (Fig. 3A; Fig. 4C). One large handful of TMR was collected in a plastic bag for each sample. The corn component of TMR samples collected in summer 2021 came from plants grown during the 2020 field season. TMR samples varied in their composition depending on the source farm and on the lactation status of cows within farms, but all contained corn silage as a component. All TMR samples contained approximately 50% corn silage (Fig. 3B) with the exception of TMR for non-lactating cows at the

Lansing, NY source farm (Fig. 4C); these samples contained a smaller proportion of corn silage and higher proportion of wheat straw (Fig. 3C)

Spectral Data Collection

Spectral measurements were acquired using an ultraviolet-visible-near-infrared (UV-Vis-NIR) spectrometer (2DCCD-512x64, Control Development, Inc., South Bend, IN), a UV-Vis-NIR light source (Ocean Optics, Inc, Dunedin, FL), and a light-emitting diode fiber optic probe (Control Development, Inc., Dunedin, FL). The light source produced wavelengths of light every 0.5 nm from 304-1,085 nm (n = 1,563 wavelengths), which were emitted through the end of the fiber optic probe. The light source and fiber optic probe were connected directly to the spectrometer, which was connected to a computer via USB. The spectrometer interfaced with the computer via Spec32TM software (Control Development, Inc., Dunedin, FL) to display a 2-dimensional graph of the sample's spectral reflectance. The intensity of light reflected by each sample (percent reflectance relative to a background and reference) at each of the 1,563 wavelengths will henceforth be referred to as the sample's spectral signature.

The end of the probe was held directly above the sample, emitting a narrow beam of light onto the sample surface. The light reflected off the sample surface enters the spectrometer via the fiber optic probe. Inside the spectrometer, the light is reflected from the sample onto a collimating mirror, which angles the light into one uniform directional orientation. The collimated light is reflected onto the spectrometer's diffraction grating, which partitions out the light in order of wavelength (Ocean

Insight, 2017). The diffraction grating can be thought of as a prism, in which a mixture of different wavelengths enters through one side and an ordered partitioned beam of wavelengths comes out another side in the form of a rainbow. The rainbow beam of light is then reflected from the diffraction grating onto a curved focusing mirror, which funnels the light towards a single focal point in the center of the mirror. Finally, each of the 1,563 wavelengths in this focused beam of light activates a corresponding pixel on the detector (Ocean Insight, 2017). This spectroscopy system is shown in Figure 5A and the mechanism by which these instruments acquire each sample's spectral signature is outlined in Figure 5B.

Naturally infected ears. Spectral signatures were acquired from the surface of the embryo of the excised symptomatic and asymptomatic kernels and from the rachis tissue directly below each excised kernel (n = 4 spectral signatures per ear). A subset of ears exhibiting a directional infection gradient of GER symptoms from tip to base were selected (n = 14 ears) for additional analysis. For each ear in this subset, spectral signatures were acquired from 5 excised kernels and their associated rachis tissues down the length of the ear (n = 10 spectral reads per ear) (Fig. 6).

Corn silage. When collecting a spectral signature of a corn silage sample, the fiber optic probe could potentially encounter several different tissue types (e.g. fragments of kernels, rachis, stalk, leaves) due to its mixed-tissue nature. Different tissue types vary in biochemical composition which consequently influences their spectral reflectance. In order to reduce spectral variation in corn silage samples not related to DON,

spectral signatures of each silage sample were acquired from both specific tissues within the sample and from the top of a pile of mixed sample.

First, each silage sample was poured from its plastic collection bag into a bucket. The pile of silage was manually flattened out in the bucket and a spectral signature was acquired from the center of the pile's surface. The pile was then shaken to re-mix the sample, flattened out again, and another spectral signature was acquired from the center of the re-mixed pile's surface. This process was repeated one more time for a total of three spectral reads from the pile's surface. Then, three kernel fragments and three rachis fragments were manually removed from the sample and the spectral signatures of each tissue fragment were acquired (n = three kernel fragments + three rachis fragments + three scans from the top of the pile = nine spectral signatures per silage sample) (Fig. 7A).

Total mixed rations - Spectral signatures for TMR samples were acquired in the same fashion as corn silage samples (n = nine spectral signatures per TMR sample) (Fig. 7B).

DON Quantification

Naturally infected ears. The DON concentration of a highly toxic kernel is likely to be several orders of magnitudes higher than the larger silage sample from which it is drawn. A symptomatic kernel may have a DON concentration in the hundreds or thousands ppm, but these concentrations are virtually unimaginable in a homogenized corn silage or TMR sample. Likewise, 10 ppm DON in a single kernel is hardly cause

for concern when considered in the context of a larger volume of TMR, but a 10 ppm DON corn silage or TMR sample could mean a lost crop. As such, spectral features associated with *F. graminearum* symptoms in excised tissues may be of greater relevance to corn silage and TMR than excised tissues' actual DON concentrations. Excised kernel tissues were thus not chemically quantified for most ears, and for modelling purposes, were labeled categorically as symptomatic or asymptomatic.

Precise DON concentrations were of interest, however, in excised kernels from ears exhibiting directional infection gradients of GER symptoms (Fig. 6A). DON concentrations were quantified in these kernels in order to (1) assess the extent of DON accumulation in pre-symptomatic parts of the ear and (2) determine if DON concentration could be predicted spectrally before the onset of visible symptoms.

Excised kernels from this subset of ears (n = five kernels × 14 ears = 70 kernels) were each transferred to a 5 mL polycarbonate vial (SPEX SamplePrep LLC, Metuchen, NJ) with a 9.5 mm steel ball. Kernels were shaken in a 2010 Geno/Grinder® (SPEX SamplePrep LLC, Metuchen, NJ) in 1-2 minute increments until all kernels reached a powdered consistency. Unlike kernels, the rachis tissues were relatively non-brittle, which prohibited the level of grinding required for DON extraction. While larger-scale tissue grinders were effective at grinding whole plants (including rachis tissues) for corn silage samples (Fig. 2B), they were not suitable for individual pieces of excised rachis. Therefore, DON quantification of excised rachis tissues was not feasible with available instrumentation, and for modelling purposes, spectral reads of rachis tissues were labeled with the same DON concentration as their associated excised kernel.

Approximately 0.1 grams of ground kernel tissue was transferred to a 1.5 mL microcentrifuge tube and deionized water was added in a 1:5 w/v ratio (400 µL for 0.1 gram sample). DON was then extracted and quantified via ELISA ("Deoxynivalenol (DON) Assay, Cat. No. 941DON01M96," Helica Biosystems, Inc., Santa Ana, CA) with one modification in the toxin extraction section of the manufacturer's protocol. The manufacturer instructs that the experimenter wait for particulate matter to settle, then run the sample through Whatman filter paper and dilute the collected filtrate 1:10. As a modification after the 3 minute sample shaking and before the 1:10 sample dilution, 1.5 mL of shaken sample with particulate still in suspension was quickly transferred to a 1.5 mL microcentrifuge tube and spun down for 15 minutes at 14,000 rpm. In all other respects, the protocol was carried out according to the manufacturer's instructions. Briefly, each sample was combined with a DON-horseradish peroxidase conjugate, transferred to an antibody-coated plate, washed several times with phosphate-buffered saline, incubated with tetramethylbenzidine, and extinguished with sulphuric acid. Six standards of known concentrations were run in tandem with the assayed samples. The optical density of the standards at 450 nm was used to form a standard curve from which the DON concentrations of the samples were extrapolated.

Corn silage. Corn silage samples were dried in an industrial oven until brittle. DON was extracted and quantified via liquid chromatography and double mass spectrometry.

Total mixed rations. TMR samples were placed in paper bags and dried in an oven until brittle. Once fully dry, approximately 8 grams of tissue from each TMR sample

were transferred to a 100 mL polypropylene grinding chamber (Mfr. Model #0020012571; IKA, Inc. Staufen, Germany) and macerated with an IKA Tube Mill 100 (Ident. No.: 0020007145; IKA, Inc. Staufen, Germany). DON extraction and quantification for TMR samples were performed as described for naturally infected ears.

Spectral Prediction of DON

Defining Datasets. Spectral data from the following collections of samples were treated as 5 separate datasets: (1) kernels from naturally infected ears, (2) rachis tissues from naturally infected ears, (3) kernel fragments from corn silage and TMR, (4) rachis fragments from corn silage and TMR, and (5) top of pile scans from corn silage or TMR. Spectral reads from kernels and rachis tissues from naturally affected ears were placed into one of two file directories: {symptomatic, asymptomatic}. Spectral reads from corn silage and TMR were placed into one of two file directories depending on their DON concentration measured with ELISA or LC-MS-MS: {under 1 ppm DON, over 1 ppm DON}. Reflectance values from each file were extracted using the glob module in the Python Standard Library. All files within each directory were concatenated into a single data frame using the "concat" function in the Pandas library (McKinney, 2010).

Defining Training and Test Sets. For each of the five datasets, 70% of samples in each respective file directory were selected for model training (i.e. calibration). The remaining 30% in each directory served as test (i.e. validation) samples.

Training Set Synthetic Oversampling. The Adaptive Synthetic (ADASYN) Oversampling Technique (He et al., 2008) is an imbalanced learning tool that can be used to resample data in specific classes, e.g. samples over a specified DON threshold. The "imblearn.over_sampling.ADASYN" function from the Imbalanced-learn Python toolbox (Lemaitre et al., 2017) was used to achieve a more uniform distribution of false positives (1 ppm) and false negatives (>1 ppm samples classified as toxic).

Data Transformation. Pre-processing techniques such as multiplicative scatter correction and standard normal variate (SNV) can reduce some of the spectral variation that arises from differences in light scattering and path length between samples. Using the same spectroscopy system as the present study, Aoun et al. (2021) determined that SNV transformation was the most effective pre-processing technique when modeling aflatoxin concentration in single corn kernels. In order to reduce variation from factors not related to DON in the present study, the SNV technique was implemented using the "sklearn.preprocessing.StandardScaler" function in the Scikit Learn library (Pedregosa et al., 2011).

Binary Classification. A partial least squares (PLS) classifier, random forest (RF) classifier, and support vector machine (SVM) classifier were trained to predict if kernels and rachis tissues were symptomatic or asymptomatic. Training of these classifiers was carried out using the "sklearn.cross_decomposition.PLSRegression", "sklearn.ensemble.RandomForestClassifier", and "sklearn.svm.SVC" functions, respectively, in the Scikit Learn library (Pedregosa et al., 2011). These 3 classifiers were also used to predict if corn silage and TMR samples contained under or over 1

ppm DON. Each classifier's performance was evaluated in terms of overall model accuracy, sensitivity (ability to detect true positives), and specificity (ability to detect true negatives).

Feature Importance. In order to determine if predictive spectral ranges were shared between tissues in the 5 datasets, mean accuracy decrease was calculated for each wavelength in the 3 models (PLS, RF, and SVM). This metric represents how much a given model's prediction ability suffers when each individual wavelength is removed, therefore a higher mean decrease accuracy score indicates higher feature importance. Mean accuracy decrease was calculated for each wavelength using the "sklearn.inspection.permutation_importance" function in the Scikit Learn library (Pedregosa et al., 2011). This permutation-based feature importance method was preferred over impurity-based methods as it is less susceptible to inflating variable importance in high-dimensional data, e.g. spectral signatures.

RESULTS

Toxin distribution in corn feed products. We hypothesized that DON concentrations would be orders of magnitude higher in symptomatic kernels than in highly contaminated corn silage or TMR due to their vast differences in volume. Indeed, the DON concentrations of the most toxic symptomatic kernels were hundreds of times higher than the most toxic corn silage or TMR samples in our dataset (Fig. 8). These different toxin distributions, highlighted in Fig. 9, emphasize the importance of creating separate DON prediction models for excised tissues.

Naturally infected ears. All three classifiers had greater ease predicting if kernels were symptomatic or asymptomatic compared to DON concentration of kernel fragments in corn silage and TMR. This is unsurprising because symptomatic and asymptomatic kernels were selected based on their visual appearance. However, the spectral profiles of symptomatic and asymptomatic kernels also appeared to exhibit distinct differences beyond the visible spectrum; symptomatic kernels generally exhibited higher reflectance in the UV (Fig. 10A) and 900-1,065 nm NIR ranges (Fig. 10B). The ranges at which symptomatic and asymptomatic kernels appear least distinguishable from one another were in the visible range at 400-450 nm (Fig. 10C) and in the NIR range at approximately 840-890 nm (Fig. 10D).

While all three classifiers (PLS, RF, and SVM) had comparable performance at predicting if a kernel was symptomatic vs. asymptomatic (Fig. 11A), the classifiers' variability in mean accuracy decrease highlights key learning differences (Fig. 11B). In classifying kernels from naturally infected ears, the SVM model appeared to rely on nearly all wavelengths relatively equally with a few exceptions in the 900-1,500 nm (NIR) range. Though not the most important range for all three classifiers, this is the only range of wavelengths that exhibited better-than-random mean accuracy decrease across all 3 classifiers for kernels from naturally infected ears. The most important features in the PLS model were all in the NIR range, around 940 nm, 970 nm, and 1,060 nm. This model ascribed notably low importance to features in the 600-750 nm (red) range. The RF model had more reliance on wavelengths in the red range

compared to PLS and SVM, but the RF classifier's most important wavelengths were in the 500-600 nm (green-yellow) range.

It was initially surprising that the PLS and SVM models did not rely on the red spectral range any more than random wavelengths because the *F. graminearum* fungus itself is often pink or red in color. However, a lack of reliance on this spectral range begins to make more sense when contextualized with the diversity of redness-related phenotypes observed in the kernel pairs (Fig. 12). Both symptomatic and asymptomatic kernels had instances of red streaking, and some symptomatic kernels exhibited a general kernel reddening as opposed to a distinct streaking pattern. Instances were observed of extensive fungal colonization on red-streaked kernels, of subtle colonization on generally reddened kernels, and vise versa. The color of the mold also varied on symptomatic kernels, ranging from white to pink on both red-streaked and generally reddened phenotypes.

Beyond symptomatic vs. asymptomatic classification, redness-related phenotypes were not always present with DON accumulation in kernels. Indeed, ELISA results demonstrated that red streaking could occur before any major accumulation of DON, and kernels without any red streaking could have higher DON accumulation than those with red streaking (Fig. 13). This implies that DON accumulation in kernels is not always preceded by, nor followed by, red streaking.

While the RF rachis classifier demonstrated moderate accuracy decreases from the removal of wavelengths in the 300-400 nm and 900-1100 nm ranges, its most

important peak shifted from the ~500 nm (green) range for kernel prediction towards the ~700 nm (red) range (Fig. 11C). The RF classifier's reliance on the red spectral range in rachis tissues but not kernels could be because rachis tissues exhibit fewer potentially confounding redness-related phenotypes than kernels. For example, rachis tissues may develop gray or brownish discoloring or appear pink or white from fungal colonization (Oldenburg & Ellner, 2015), but do not exhibit anything analogous to the red streaking seen in kernels. If red kernel streaking is not a consistent predictor of GER (as seen in Fig. 13), then it would not be surprising that spectral features in the red range were better predictors of GER in rachis tissues than in kernels (as seen in Fig. 11B,C).

Corn silage and TMR. The same three binary classifiers (PLS, RF, SVM) were trained to predict if DON concentration was over or under 1 ppm in corn silage and TMR samples (Fig. 14A). PLS models for classification of kernel fragments, rachis fragments, and top of pile scans had similar overall accuracies (approximately 57-63%). SVM outperformed PLS and RF for classification of both kernel fragments (70% accuracy) and top of pile scans (~67% accuracy), but PLS outperformed other classifiers for rachis fragments (~63% accuracy). The kernel fragment SVM model had the best overall classification performance of any of the corn silage and TMR models with 70% overall accuracy, 71% sensitivity (correctly classify >1 ppm DON samples), and 69% specificity (correctly classify <1 ppm DON samples).

All binary classification models of corn silage and TMR had lower accuracy in classifying naturally infected kernel or rachis tissues as symptomatic or asymptomatic.

Kernel and rachis fragments from corn silage and TMR samples exhibited more morphological variation in shape, size, and tissue thickness compared to whole kernel and rachis tissues from naturally infected ears. Morphological variance among tissue fragments could also be underpinned by genetic variation in the corn silage and TMR samples (a combination of brown midrib and conventional hybrids) compared to the near-isogenic naturally infected ears. These sources of morphological variation likely introduced additional spectral variance in path length and scattering properties. The addition of this morphology-related variance in the corn silage and TMR samples likely decreased the proportion of variance in the models explained by DON, and in turn, decreased DON prediction accuracy.

Nearly all spectral features in corn silage and TMR models trained by PLS classifiers had mean accuracy decrease scores near 0 (Fig. 14B-D). In the top of pile scan model, this could be because the variance among spectral signatures represented by the principal components in the model was better explained by the tissue type scanned (e.g. kernel, rachis, stalk, leaf) than by DON concentration. While the same tissue type was scanned in models for kernel fragments and rachis fragments, morphological variance among kernel fragments and rachis fragments, respectively, may have explained more variance in PLS models than DON.

Spectral regions with the highest mean accuracy decrease varied between RF models. The most important regions for RF classification of kernel fragments laid at the two ends of the measured spectral range, at around 350 nm in the UV range and 1,060 nm in the NIR range (Fig. 14B). Alternatively, the most important region for top of pile

scans lied in the middle of the measured spectrum around 700 nm (red) of the visible range (Fig. 14D). Mean accuracy decreases associated spectral features were more uniformly distributed in the rachis fragment model but relied especially little on the visible spectrum (Fig. 14B). Mean accuracy decrease in the RF rachis fragment model for wavelengths in the 450 – 700 nm range all neared 0, essentially representing the whole visible spectrum except red. This implies that the visible spectrum was of less importance in the model than the UV or NIR spectra, but that wavelengths associated with red were relatively more important than other parts of the visible spectrum.

SVM models for corn silage and TMR had less distinct peaks in mean decrease accuracy compared to RF models. In the kernel fragment SVM model, however, 3 spectral regions had slightly more importance than others at ~325 (UV) – 459 nm (violet), ~675 nm (red), and ~1,060 nm (NIR) (Fig. 14B). For the top of pile SVM model, the ~1,040 nm and ~1,060 nm NIR regions had slightly more importance than the other wavelengths. This model relied especially little on the ~420 (violet) – ~550 nm (green) range with mean decrease accuracy scores near 0 (Fig. 14D). None of the mean decrease accuracy scores in the rachis fragment SVM model were as high as those in the kernel fragment or top of pile SVM models, but 2 regions were of relatively more importance than others: ~325 – 375 nm (UV) and ~675 nm (red). Mean decrease accuracies for all other wavelengths in the rachis fragment SVM model neared 0 (Fig. 14C).

DISCUSSION

The legal limit for DON in TMR for dairy cows older than 4 months of age is 5 ppm (FDA, 2010). This may sound relatively high, but a surprisingly small number of toxic single kernels can push an otherwise-clean sample over that limit. Consider that you are submitting a silage sample to a local testing facility for DON analysis. Assume that the testing facility requires a 0.9 kg sample for testing (Dairyland Laboratories, Inc. (n.d.); Dairy One, 2021). For simplicity's sake, also assume that the corn you are submitting is from reference line B73, and that the mass of every corn kernel is 0.2275 g (Pedregosa et al., 2011). If the ear of corn shown in Fig. 13B were intermixed in your corn silage sample, the inclusion of just 5 shelled kernels from the most symptomatic part of this ear would be sufficient to push the tested sample over the 5 ppm limit (Fig. 15A):

$$\frac{\frac{3{,}970\text{ mg DON}}{1\text{ kg kernel}}\times\frac{0.2275\text{ g}}{1\text{ kernel}}\times\frac{1\text{ kg kernel}}{1{,}000\text{ g kernel}}\times 5\text{ kernels}}{0.9\text{ kg sample}}=\frac{4.5\text{ mg DON}}{0.9\text{ kg sample}}=\frac{5.0\text{ mg DON}}{1\text{ kg sample}}=\mathbf{5\text{ ppm}}$$

Even if the symptomatic part of the ear were removed and never mixed into the sample, the amount of DON in the remaining asymptomatic kernels could also push the sample over the 5ppm limit (Fig. 15B):

$$\frac{\left(\frac{36.0\text{ mg DON}}{1\text{ kg kernel}}+\frac{32.0\text{ mg DON}}{1\text{ kg kernel}}+\frac{2.0\text{ mg DON}}{1\text{ kg kernel}}\right)\times\frac{0.2275\text{ g}}{1\text{ kernel}}\times\frac{1\text{ kg kernel}}{1{,}000\text{ g kernel}}\times\frac{16\text{ kernels}}{\text{row}}\times 16\text{ rows}}{0.9\text{ kg sample}}$$

$$=\frac{4.6\text{ mg DON}}{0.9\text{ kg sample}}=\frac{5.1\text{ mg DON}}{1\text{ kg sample}}=\mathbf{5.1\text{ ppm}}$$

Prediction of DON in corn feed products throughout plant development, harvest, storage, and feeding can protect dairy cows from harmful DON exposure; increasingly

accurate spectral detection of DON helps ensure that farmers' bottom lines are protected too. The scenario above illustrates that DON detection in individual toxic kernels can be of great utility for mitigating DON accumulation in corn silage and TMR, but it also highlights that symptomology prediction is not a sufficient replacement for DON prediction.

The next step in this project is to conduct ELISA on the remaining kernels from naturally infected ears (n = 42 assayed, 257 remaining). Data from these ELISAs will allow for the development of a prediction model based on DON concentration rather than symptomatic vs. asymptomatic origin of the kernel. Computer vision approaches may also be explored to assess if the combination of digital images and UV-Vis-NIR spectral signatures aids in prediction of DON in single kernels. ELISA on the remaining corn silage samples (n = 100 assayed, 59 remaining) will be completed with the aim of improving prediction accuracy in the corn silage and TMR models.

CONCLUSIONS

We developed a PLS model that successfully predicted if kernels were symptomatic or asymptomatic of Gibberella ear rot with 92% accuracy (90% sensitivity and 93% specificity). We also developed an RF model that predicted if DON concentration was over or under 1 ppm in a diverse set of corn silage and TMR samples from across the Northeastern U.S. with 70% accuracy (71% sensitivity and 69% specificity). Sensitivity and specificity errors in both models were effectively balanced using the ADASYN technique (He et al., 2008; Lemaitre et al., 2017). Spectral regions at approximately 325 (UV) – 459 nm (violet), 675 nm (red), and 1,060 nm (NIR) were

most important in the RF corn silage and TMR model. The most important spectral features in the PLS kernel fragment model were all in the NIR range at around 940 nm, 970nm, and 1,060 nm.

Acknowledgements. We would like to express our deepest appreciation to the farmers who provided us with corn silage and TMR samples. We would also like to thank the planting and harvesting teams for their assistance during the 2018 and 2020 field seasons at Musgrave Research Farm, especially lab manager Ace Repka. We are grateful to Gary Bergstrom for his expert insight on discerning GER symptoms in naturally infected ears and to Elliot McGinnity-Schneider for his valuable laboratory assistance with DON extraction of TMR samples. This work benefitted from the support of the Cornell Institute for Digital Agriculture (CIDA).

References:

Aoun, M., Siegel, C., Windham, G., Williams W., Nelson, R. (2021). Spectroscopy for genotype and aflatoxin classification in maize single kernels. *In prep.*, submitted for publication.

Czembor E, Stępień Ł, Waśkiewicz A. (2015). Effect of environmental factors on *Fusarium* species and associated mycotoxins in maize grain grown in Poland. *PLoS One* 10(7):e0133644.

Dairyland Laboratories, Inc. (n.d.) FAQS. https://www.dairylandlabs.com/resources/faqs

Dairy One. (2021, May 6). Forage FAQ. https://dairyone.com/services/forage-laboratoryservices/forage-faq/#how-much-sample.

FDA (2010, July). *Guidance for Industry and FDA: Advisory Levels for Deoxynivalenol (DON) in Finished Wheat Products for Human Consumption and Grains and Grain By-Products used for Animal Feed.* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-and-fda-advisory-levels-deoxynivalenol-don-finished-wheat-products-human

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008, June). ADASYN: Adaptive synthetic

sampling approach for imbalanced learning. IEEE international joint conference on neural networks (IEEE world congress on computational intelligence) (pp. 1322-1328).

Lemaıtre, G., & Nogueira, F. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, *18*, 1-5.

Lauren, D. R., & Di Menna, M. E. (1999). Fusaria and *Fusarium* mycotoxins in leaves and ears of maize plants: A time course study made in the Waikato region, New Zealand, in 1997. *New Zealand Journal of Crop and Horticultural Science*, *27*(3), 215-223.

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61.

Ocean Insight (2017, May 24). *How Does a Spectrometer Work?* [Video]. https://www.youtube.com/watch?v=OI3pIvLhVcc

Oldenburg, E., & Ellner, F. (2015). Distribution of disease symptoms and mycotoxins in maize ears infected by *Fusarium culmorum* and *Fusarium graminearum*. *Mycotoxin Research, 31*(3), 117-126.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Prado, S. A., López, C. G., Gambín, B. L., Abertondo, V. J., & Borrás, L. (2013). Dissecting the genetic basis of physiological processes determining maize kernel weight using the IBM (B73× Mo17) Syn4 population. *Field Crops Research*, *145*, 33-43.

Saccon, F. A., Parcey, D., Paliwal, J., & Sherif, S. S. (2017). Assessment of *Fusarium* and deoxynivalenol using optical methods. *Food and Bioprocess Technology, 10*(1), 34-50.

Schmale, D. G., & Munkvold, G. P. (2018). Mycotoxins in crops: A threat to human and domestic animal health. *The American Phytopathological Society*.

SECTION S3.

Identification of QTL conferring resistance to *Fusarium graminearum* in maize seedlings using nested near-isogenic lines

Sections S3 and S4 are project proposals focused on *Fusarium graminearum* infection in maize seedlings. I ultimately decided against pursuing these seedling-based projects as part of my M.S., and instead opted to focus my thesis on post-harvest mycotoxin detection. These proposals were shared collaborators at the University of Illinois at Urbana-Champaign for potential incorporation into an NSF-NIFA Plant-Biotic Interactions grant proposal.

LIST OF FIGURES

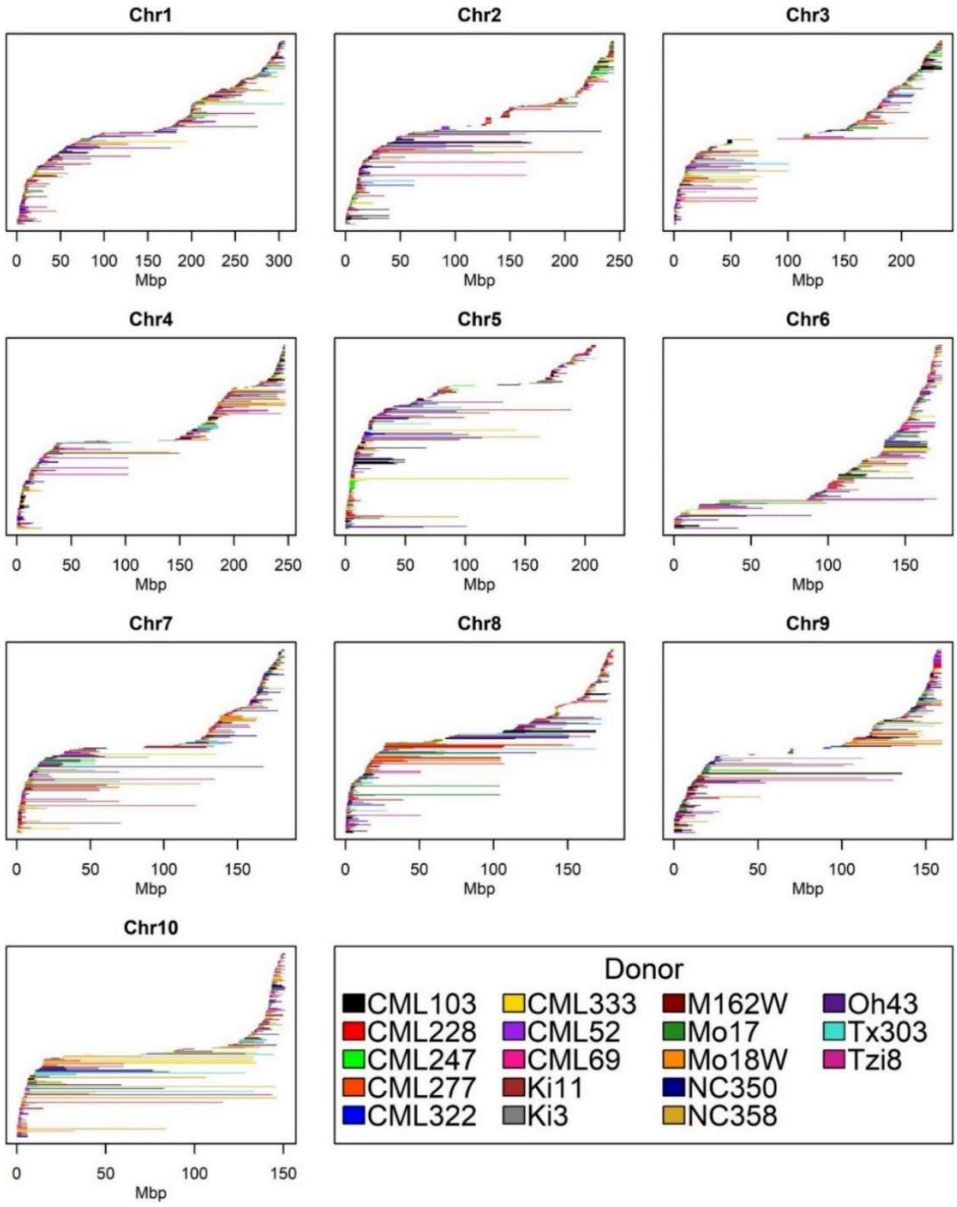

Figure 1. Introgressions from 18 diverse donor lines present in the nNILs span 100% of the maize genome. As expected, patterns of introgression size are related to chromosomal location; introgressions in high recombination regions farther from the centromere are smaller while introgressions closer to the centromere are much larger. Original figure from Morales et al. (2020).

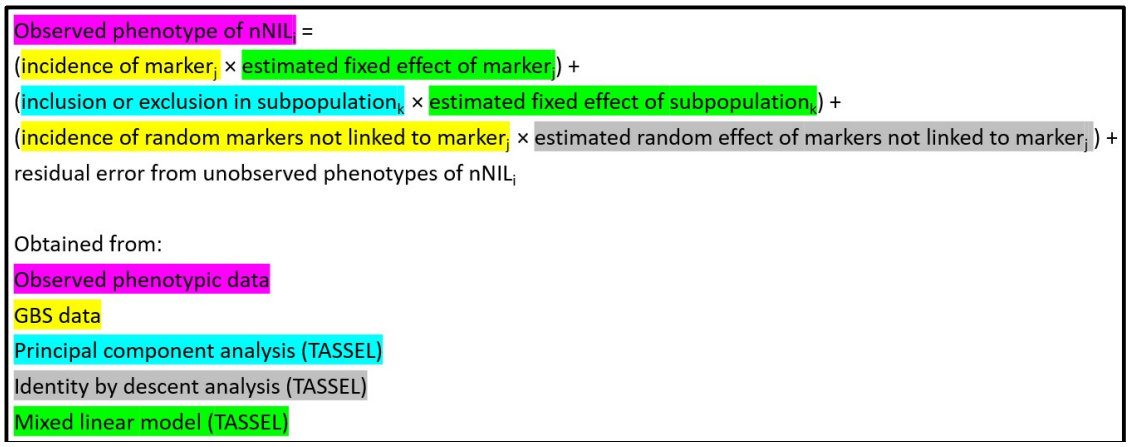

Figure 2. Conceptualization of the MLM used to estimate the effect of a genomic marker on an observed nNIL phenotype in TASSEL.

LIST OF TABLES

Table 1. Examples of metabolic inhibitors for potential seedling treatments.

Metabolic Inhibitor	Inhibited Metabolic Process	Inhibited Enzyme	Reference
α-aminooxyacetate, α-aminooxy-β-phenylpropionic acid, and (1-amino-2-phenylethyl) phosphonic acid	1st step of PPP	PAL	Amrhein & Godeke, 1977
aminoethoxyvinylglycine	ethylene signaling	ACC synthase	Mulkey et al., 1982
piperonylic acid	2nd step of PPP	C4H	Schalk et al., 1998
1-methylcyclopropene	ethylene signaling	ACC oxidase and synthase	Serek et al., 1995
3,4-(methylenedioxy)cinnamic acid	3rd step of PPP	4CL	Grand et al., 1985
Sodium azide	Cell wall cross linking	Peroxidase	Schopfer, 1996
Brefeldin A	Extracellular vesicular transport	Golgi GTPase	Schindler et al. 1994
Sodium orthovanadate	Flavonoid shuttling	glutathione S-transferase	Marrs et al., 1995
CI-976	Intracellular vesicular transport	LPLAT	Brown & Schmidt, 2005
OHPAS, NH2PAS	Lignification	CAD	Grand et al., 1985
Bafilomycin A1	Vacuolar proton pumping	Vacuolar H(+)-ATPase	Bowman et al., 1988

Abbreviations:
4CL = 4-Coumarate-CoA ligase
ACC = 1-aminocyclopropane-1-carboxylate
C4H = cinnamate 4-monooxygenase
CAD = cinnamyl-alcohol dehydrogenase
CI-976 = 2,2-methyl-N-(2,4,6,-trimethoxyphenyl)dodecanamide
LPLAT = lysophospholipid acyltransferase
NH2PAS = N(O-aminophenyl)sulfinamoyl-tertiobutyl acetate
OHPAS = N(O-hydroxyphenyl)sulfinamoyl-tertiobutyl acetate
PAL = phenylalanine ammonia-lyase
PPP = phenylpropanoid pathway

Identification of QTL conferring resistance to *Fusarium graminearum* in maize seedlings using nested near-isogenic lines

INTRODUCTION

Fusarium graminearum is a widespread generalist pathogen of cereals and produces dangerous mycotoxins. In maize, *F. graminearum* causes Gibberella ear rot (GER), Gibberella stalk rot (GSR), and seedling blight (SB). While SB has received markedly less research investigation than FSR or GER, estimated maize yield losses due to SB in the U.S. and Canada exceeded those due to FSR and GER in 3 of the last 8 years (Mueller et al., 2015, Mueller et al., 2016; Mueller et al., 2020). Quantitative trait loci (QTL) influencing GSR (Chen et al., 2017; Ma et al., 2017; Yang et al., 2010; Zhang et al., 2012) and GER resistance (Ali et al., 2005; Giomi et al., 2016; Kebede et al., 2016; Martin et al., 2011; Martin et al., 2012; Tembo et al., 2014; Yuan et al., 2020) have been the subject of extensive research, but QTL research on *F. graminearum* SB is still in its relative infancy (Zhou et al., 2019).

SB resistance is not always correlated with resistance in mature maize (Stagnati et al., 2019) or wheat (Collard et al., 2005; Shin et al., 2014; Tamburic-Ilincic et al., 2009), suggesting the notion that different loci underpin resistance at different stages of plant development. Plants allocate most of their energetic resources to rapid growth during the seedling stage, and consequently, it has been suggested that seedlings rely heavily on induced chemical defense rather than more energetically costly constitutive defenses observed in later developmental stages (Barton & Boege, 2017; Cipollini & Heil, 2010). In maize seedlings, these induced defenses consist primarily of

specialized secondary metabolites with antimicrobial activity (Zhou et al., 2018; Zhou et al., 2019). The objectives of this study are to (i) identify and map QTL conferring *F. graminearum* SB resistance and (ii) shed light on potential metabolic defense mechanisms implicated by the associated QTL. These findings could inform breeding efforts for SB resistance and prevent future yield losses.

To screen for sources of *F. graminearum* SB resistance, I will utilize a relatively new genetic resource: the nested near-isogenic lines (nNILs; Kolkman et al., 2020; Morales et al., 2020). The nNILs consist of ~1,200 lines carrying chromosomal segments from 18 of the 25 donor lines from the nested association mapping (NAM) population (Yu et al., 2008) introgressed into the recurrent parent B73. Introgressions in the nNILs span 100% of the genome with considerable overlap in areas of higher recombination (Kolkman et al., 2020; Morales et al., 2020) (Fig. 1). Our program has previously genotyped the nNILs by sequencing (GBS; Elshire et al., 2011), which we have used to identify QTL for resistance to several other diseases including northern leaf blight, southern leaf blight, and gray leaf streak (Kolkman et al., 2020; Morales et al., 2020). We have also phenotyped these lines extensively and found variation in stand count, stalk strength, and susceptibility to fumonisin mycotoxins, among other agronomic traits. The diverse donor lines have been fully sequenced (NAM Genome Project, 2020; Hufford et al., 2021), and data on the genomic features of each genome are publicly available via JBrowse on MaizeGDB (Portwood et al., 2019). I will leverage new insights on the genetic variation present in the nNILs to associate genomic features with SB resistance and metabolic pathways in which associated genomic features are involved.

METHODS

SB Assay

The full panel of ~1,200 sequenced nNILs will be analyzed for *F. graminearum* SB severity. The nNILs will be germinated in three replications in germination pouches (PhytoAB Inc., San Jose, CA) in growth chambers. Once primary roots have reached 9 cm in length in the majority of nNILs, a spore suspension of *F. graminearum* isolate Gz014NY98 will be added to the germination pouches. This isolate is an aggressive disease agent in both maize ears and seedlings, and produces deoxynivalenol (DON) and 15-acetyldeoxynivalenol (15-ADON) mycotoxins (Kuhnem et al., 2015). An initial screen on a limited number of lines will be performed to establish the concentration of spores that provides optimal differentiation of symptoms, followed by inoculation of the full panel of lines. At V1, I will assess the germination rate of each nNIL. At V1 and V3, roots will be photographed and analyzed using RootReader 2D software (Clark et al., 2013). Also at V3, I will measure root length and biomass, shoot height and biomass, survival rate among germinated seedlings, and SB severity among surviving seedlings.

Identification of SB resistant and susceptible nNILs

I will use Dunnett's multiple comparison to identify nNILs that are significantly more resistant or susceptible to B73. Analysis will be performed in R using the DunnettTest() function from the library DescTools (Signorell et al., 2016). A nNIL will be considered significantly more resistant or susceptible than B73 if its mean

survival rate among germinated seedlings or SB severity among surviving seedlings is higher or lower, respectively, than B73 within a 95% confidence interval. nNILs determined to be statistically more resistant or susceptible than the recurrent parent B73 will be subject to another round of inoculation and phenotyping as described above. Then, Dunnett's multiple comparison will be repeated on these lines to assess if they still exhibit significantly more resistance or susceptibility compared to B73. Lines that are significantly more resistant or susceptible than B73 in both Dunnett's tests will be characterized in greater detail (see chapter 3).

Identification of SB resistant and susceptible loci

I will perform a genome-wide association study (GWAS) using a mixed linear model (MLM) in the software TASSEL (Trait Analysis by Association, Evolution and Linkage; Bradbury et al., 2007). Available GBS data of the nNILs (Kolkman et al., 2020; Morales et al., 2020) will be used to associate genomic markers with observed phenotypes for (i) survival rate among germinated seedlings and (ii) SB severity among surviving seedlings. Markers with minor allele frequencies greater than 0.001 that are represented in at least 20 GBS reads will be included in association analysis (Kolkman et al., 2020). The uniform B73 background is expected to minimize the rate of false associations due to population structure and kinship among the nNILs, but the effects of both features will still be estimated using a principal component analysis and the identity by descent method, respectively, in TASSEL. These estimates will be included in a MLM to estimate the effect of each analyzed marker on the observed disease phenotype (Fig. 2). Associations will be considered statistically significant if

the estimated effect of the marker on the observed phenotype exceeds the false discovery rate (Benjamini & Hochberg, 1995) with a probability greater than -log10(α/N), where α = 0.05 and N = the number of markers analyzed (Bonferroni, 1936). After running the GWAS, TASSEL outputs a file of MLM marker statistics and a file of MLM allele effects. A file of linkage disequilibrium (LD) estimates will also be produced in TASSEL. These 3 files will be used for subsequent pathway association analysis.

Pathway association of SB resistance and susceptible loci

I will use the Pathway Association Study Tool (PAST; Thrash et al., 2020) guided user interface (GUI) application to identify metabolic pathways associated with SB resistance and susceptible loci from GWAS. The MLM marker statistics file, MLM allele effects file, and LD file from TASSEL will be input into the PAST GUI along with a file of genomic annotations for each of the 18 diverse donor lines compiled using the JBrowse feature (Buels et al., 2016) on MaizeGDB (Portwood et al., 2019). These 4 files will be input into the PAST function assign_SNPs_to_genes. The mode parameter of this function will be specified as "increasing" for survival rate GWAS hits and "decreasing" for SB severity GWAS hits to indicate if higher or lower values for the observed phenotypes are desirable. An additional file of metabolic pathways of interest and their associated genes will be downloaded through CornCyc 9.0 (Walsh et al., 2016) on MaizeGDB (Portwood et al., 2019) and input into the PAST function find_pathway_significance. This function calculates the extent to which each metabolic pathway is associated with each GWAS hit.

Validation of metabolic defense pathways

To validate the role of associated metabolic pathways from PAST analysis on SB resistance, I will treat nNILs identified as resistant from Dunnet's multiple comparison with metabolic inhibitors. A solution of metabolic inhibitor will be added to germination pouches once root length of most lines has reached 9cm followed by inoculation with an *F. graminearum* spore suspension the following day. The metabolic inhibitor treatment applied to each resistant nNIL will depend on which pathways are significantly associated with SB resistance GWAS hits. For example, if the benzoxazinoid biosynthesis pathway is implicated in PAST analysis, germination pouches will be treated with the Golgi disruptor brefeldin A (BFA) (Schindler et al. 1994; Zhou, 2018). If a nNIL no longer exhibits SB resistance after BFA treatment, Golgi-dependent benzoxazinoids may underpin resistance conferred by that nNIL's introgression. Besides BFA, several other available pathway-specific metabolic inhibitors may be used as seedling treatments (Table 1).

References:

Ali, M. L., Taylor, J. H., Liu Jie, Genlou Sun, William, M., Kasha, K. J., Reid, L. M., Pauls, K. P., & Scoles, G. J. (2005). Molecular mapping of QTLs for resistance to Gibberella ear rot, in corn, caused by *Fusarium graminearum*. *Genome, 48*(3), 521–533. https://doi.org/10.1139/G05-014

Amrhein, N., & Gödeke, K.-H. (1977). α-aminooxy-β-phenylpropionic acid—A potent inhibitor of L-phenylalanine ammonia-lyase in vitro and in vivo. *Plant Science Letters, 8*(4), 313–317. https://doi.org/10.1016/0304-4211(77)90148-1

Baltas, M., Cazaux, L., Gorrichon, L., Maroni, P., & Tisnes, P. (1988). Sulphinamoylacetates as sulphine precursors. Mechanism of basic hydrolysis and scheme of irreversible inactivation of cinnamoyl alcohol dehydrogenase, an enzyme of the lignification process. *Journal of the Chemical Society, Perkin Transactions 2, 0*(8), 1473–1478. https://doi.org/10.1039/P29880001473

Barton, K. E., & Boege, K. (2017). Future directions in the ontogeny of plant defence: Understanding the evolutionary causes and consequences. *Ecology Letters, 20*(4), 403–411. https://doi.org/10.1111/ele.12744

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289–300.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità (8th

ed.). Istituto Superiore di Scienze Economiche e Commerciali di Firenze.

Bowman, E. J., Siebers, A., & Altendorf, K. (1988). Bafilomycins: A class of inhibitors of membrane ATPases from microorganisms, animal cells, and plant cells. *Proceedings of the National Academy of Sciences, 85*(21), 7972–7976. https://doi.org/10.1073/pnas.85.21.7972

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brown, W. J., & Schmidt, J. A. (2005). Use of acyltransferase inhibitors to block vesicular traffic between the ER and Golgi complex. In Methods in Enzymology (Vol. 404, pp. 115–125). Academic Press. https://doi.org/10.1016/S0076-6879(05)04012-7

Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., Goodstein, D. M., Elsik, C. G., Lewis, S. E., Stein, L., & Holmes, I. H. (2016). JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biology, 17*(1), 66. https://doi.org/10.1186/s13059-016-0924-1

Chen, Q., Song, J., Du, W.-P., Xu, L.-Y., Jiang, Y., Zhang, J., Xiang, X.-L., & Yu, G.-R. (2017). Identification, mapping, and molecular marker development for Rgsr8.1: A new quantitative trait locus conferring resistance to Gibberella stalk

rot in maize (*Zea mays* L.). *Frontiers in Plant Science, 8*. https://doi.org/10.3389/fpls.2017.01355

Cipollini, D., & Heil, M. (2010). Costs and benefits of induced resistance to herbivores and pathogens in plants. *CAB Reviews, 5,* No. 005.

Clark, R. T., Famoso, A. N., Zhao, K., Shaff, J. E., Craft, E. J., Bustamante, C. D., Mccouch, S. R., Aneshansley, D. J., & Kochian, L. V. (2013). High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. *Plant, Cell & Environment, 36*(2), 454–466. https://doi.org/10.1111/j.1365-3040.2012.02587.x

Collard, B. C. Y., Grams, R. A., Bovill, W. D., Percy, C. D., Jolley, R., Lehmensiek, A., Wildermuth, G., & Sutherland, M. W. (2005). Development of molecular markers for crown rot resistance in wheat: Mapping of QTLs for seedling resistance in a '2-49' × 'Janz' population. *Plant Breeding, 124*(6), 532–537. https://doi.org/10.1111/j.1439-0523.2005.01163.x

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE, 6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Giomi, G. M., Kreff, E. D., Iglesias, J., Fauguel, C. M., Fernandez, M., Oviedo, M. S.,

& Presello, D. A. (2016). Quantitative trait loci for Fusarium and Gibberella ear rot resistance in Argentinian maize germplasm. *Euphytica, 211*(3), 287–294. https://doi.org/10.1007/s10681-016-1725-z

Grand, C., Sarni, F., & Boudet, A. M. (1985). Inhibition of cinnamyl-alcohol-dehydrogenase activity and lignin synthesis in poplar (*Populus* x *euramericana* Dode) tissues by two organic compounds. *Planta, 163*(2), 232–237. https://doi.org/10.1007/BF00393512

Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., Ricci, W. A., Guo, T., Olson, A., Qiu, Y., Coletta, R. D., Tittes, S., Hudson, A. I., Marand, A. P., Wei, S., Lu, Z., Wang, B., Tello-Ruiz, M. K., Piri, R. D., … Dawe, R. K. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. BioRxiv, 2021.01.14.426684. https://doi.org/10.1101/2021.01.14.426684

Ju, M., Zhou, Z., Mu, C., Zhang, X., Gao, J., Liang, Y., Chen, J., Wu, Y., Li, X., Wang, S., Wen, J., Yang, L., & Wu, J. (2017). Dissecting the genetic architecture of *Fusarium verticillioides* seed rot resistance in maize by combining QTL mapping and genome-wide association analysis. *Scientific Reports, 7*(1), 46446. https://doi.org/10.1038/srep46446

Kebede, A. Z., Woldemariam, T., Reid, L. M., & Harris, L. J. (2016). Quantitative trait loci mapping for Gibberella ear rot resistance and associated agronomic traits using genotyping-by-sequencing in maize. TAG. Theoretical and Applied

Genetics. *Theoretische Und Angewandte Genetik, 129*(1), 17–29. https://doi.org/10.1007/s00122-015-2600-3

Kolkman, J. M., Strable, J., Harline, K., Kroon, D. E., Wiesner-Hanks, T., Bradbury, P. J., & Nelson, R. J. (2020). Maize introgression library provides evidence for the involvement of *liguleless1* in resistance to northern leaf blight. *G3 Genes|Genomes|Genetics, 10*(10), 3611–3622. https://doi.org/10.1534/g3.120.401500

Kuhnem, P. R., Del Ponte, E. M., Dong, Y., & Bergstrom, G. C. (2014). *Fusarium graminearum* isolates from wheat and maize in New York show similar range of aggressiveness and toxigenicity in cross-species pathogenicity tests. *Phytopathology, 105*(4), 441–448. https://doi.org/10.1094/PHYTO-07-14-0208-R

Ma, C., Ma, X., Yao, L., Liu, Y., Du, F., Yang, X., & Xu, M. (2017). QRfg3, a novel quantitative resistance locus against Gibberella stalk rot in maize. *Theoretical and Applied Genetics, 130*(8), 1723–1734. https://doi.org/10.1007/s00122-017-2921-5

Marrs, K. A., Alfenito, M. R., Lloyd, A. M., & Walbot, V. (1995). A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. *Nature, 375*(6530), 397–400. https://doi.org/10.1038/375397a0

Martin, M., Miedaner, T., Dhillon, B. S., Ufermann, U., Kessel, B., Ouzunova, M.,

Schipprack, W., & Melchinger, A. E. (2011). Colocalization of QTL for Gibberella ear rot resistance and low mycotoxin contamination in early European maize. *Crop Science, 51*(5), 1935–1945. https://doi.org/10.2135/cropsci2010.11.0664

Martin, M., Miedaner, T., Schwegler, D. D., Kessel, B., Ouzunova, M., Dhillon, B. S., Schipprack, W., Utz, H. F., & Melchinger, A. E. (2012). Comparative quantitative trait loci mapping for Gibberella ear rot resistance and reduced deoxynivalenol contamination across connected maize populations. *Crop Science, 52*(1), 32–43. https://doi.org/10.2135/cropsci2011.04.0214

Morales, L., Repka, A. C., Swarts, K. L., Stafstrom, W. C., He, Y., Sermons, S. M., Yang, Q., Lopez-Zuniga, L. O., Rucker, E., Thomason, W. E., Nelson, R. J., & Balint-Kurti, P. J. (2020). Genotypic and phenotypic characterization of a large, diverse population of maize near-isogenic lines. *The Plant Journal, 103*(3), 1246–1255. https://doi.org/10.1111/tpj.14787

Mueller, D. S., Wise, K. A., Sisson, A. J., Allen, T. W., Bergstrom, G. C., Bissonnette, K. M., Bradley, C. A., Byamukama, E., Chilvers, M. I., Collins, A. A., Esker, P. D., Faske, T. R., Friskop, A. J., Hagan, A. K., Heiniger, R. W., Hollier, C. A., Isakeit, T., Jackson-Ziems, T. A., Jardine, D. J., … Wiebold, W. J. (2020). Corn yield loss estimates due to diseases in the United States and Ontario, Canada, from 2016 to 2019. *Plant Health Progress, 21*(4), 238–247. https://doi.org/10.1094/PHP-05-20-0038-RS

Mueller, D. S., Wise, K. A., Sisson, A. J., Allen, T. W., Bergstrom, G. C., Bosley, D. B., Bradley, C. A., Broders, K. D., Byamukama, E., Chilvers, M. I., Collins, A., Faske, T. R., Friskop, A. J., Heiniger, R. W., Hollier, C. A., Hooker, D. C., Isakeit, T., Jackson-Ziems, T. A., Jardine, D. J., … Warner, F. (2016). Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. *Plant Health Progress, 17*(3), 211–222. https://doi.org/10.1094/PHP-RS-16-0030

Mueller, D., Wise, K., & Sisson, A. (2015) Corn disease loss estimates from the United States and Ontario, Canada - 2014. Crop Protection Network, Publication CPN-2007-14-W.

Mulkey, T. J., Kuzmanoff, K. M., & Evans, M. L. (1982). Promotion of growth and hydrogen ion efflux by auxin in roots of maize pretreated with ethylene biosynthesis inhibitors. *Plant Physiology, 70*(1), 186–188. https://doi.org/10.1104/pp.70.1.186

NAM Genomes Project. (2020). Whole-Genome Assembly of the Maize NAM Founders. https://nam-genomes.org/

Portwood, J. L., II, Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Walsh, J. R., Sen, T. Z., Cho, K. T., Schott, D. A., Braun, B. L., Dietze, M., Dunfee, B., Elsik, C. G., Manchanda, N., Coe, E., Sachs, M., Stinard, P., Tolbert, J., … Andorf, C. M. (2019). MaizeGDB 2018: The maize

multi-genome genetics and genomics database. *Nucleic Acids Research, 47(*D1), D1146–D1154. https://doi.org/10.1093/nar/gky1046

Schalk, M., Cabello-Hurtado, F., Pierrel, M.-A., Atanassova, R., Saindrenan, P., & Werck-Reichhart, D. (1998). Piperonylic acid, a selective, mechanism-based inactivator of the trans-cinnamate 4-hydroxylase: A new tool to control the flux of metabolites in the phenylpropanoid pathway. *Plant Physiology, 118*(1), 209–218. https://doi.org/10.1104/pp.118.1.209

Schindler, T., Bergfeld, R., Hohl, M., & Schopfer, P. (1994). Inhibition of Golgi-apparatus function by brefeldin A in maize coleoptiles and its consequences on auxin-mediated growth, cell-wall extensibility and secretion of cell-wall proteins. *Planta, 192*(3), 404–413. https://doi.org/10.1007/BF00198577

Schopfer, P. (1996). Hydrogen peroxide-mediated cell-wall stiffening in vitro in maize coleoptiles. *Planta, 199*(1), 43–49.

Septiani, P., Lanubile, A., Stagnati, L., Busconi, M., Nelissen, H., Pè, M. E., Dell'Acqua, M., & Marocco, A. (2019). Unravelling the genetic basis of Fusarium seedling rot resistance in the MAGIC maize population: Novel targets for breeding. *Scientific Reports, 9*(1), 5665. https://doi.org/10.1038/s41598-019-42248-0

Serek, M., Tamari, G., Sisler, E. C., & Borochov, A. (1995). Inhibition of ethylene-

induced cellular senescence symptoms by 1-methylcyclopropene, a new inhibitor of ethylene action. *Physiologia Plantarum, 94*(2), 229–232. https://doi.org/10.1111/j.1399-3054.1995.tb05305.x

Shin, S., Kim, K.-H., Kang, C.-S., Cho, K.-M., Park, C. S., Okagaki, R., & Park, J.-C. (2014). A simple method for the assessment of Fusarium head blight resistance in Korean wheat seedlings inoculated with *Fusarium graminearum*. *The Plant Pathology Journal, 30*(1), 25–32. https://doi.org/10.5423/PPJ.OA.06.2013.0059

Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., & Arppe, A. (2016). DescTools: Tools for descriptive statistics. R package version 0.99, 18.

Stagnati, L., Lanubile, A., Samayoa, L. F., Bragalanti, M., Giorni, P., Busconi, M., Holland, J. B., & Marocco, A. (2019). A genome wide association study reveals markers and genes associated with resistance to *Fusarium verticillioides* infection of seedlings in a maize diversity panel. *G3 Genes|Genomes|Genetics, 9*(2), 571–579. https://doi.org/10.1534/g3.118.200916

Tamburic-Ilincic, L., Somers, D., Fedak, G., & Schaafsma, A. (2009). Different quantitative trait loci for *Fusarium* resistance in wheat seedlings and adult stage in the Wuhan/Nyubai wheat population. *Euphytica, 165*(3), 453–458. https://doi.org/10.1007/s10681-008-9747-9

Tembo, L., Asea, G., Gibson, P. T., & Okori, P. (2014). Quantitative trait loci for

resistance to *Stenocarpella maydis* and *Fusarium graminearum* cob rots in tropical maize. *Journal of Crop Improvement, 28*(2), 214–228. https://doi.org/10.1080/15427528.2013.872750

Thrash, A., Tang, J. D., DeOrnellis, M., Peterson, D. G., & Warburton, M. L. (2020). PAST: The pathway association studies tool to infer biological meaning from GWAS datasets. *Plants, 9*(1), 58. https://doi.org/10.3390/plants9010058

Walsh, J. R., Schaeffer, M. L., Zhang, P., Rhee, S. Y., Dickerson, J. A., & Sen, T. Z. (2016). The quality of metabolic pathway resources depends on initial enzymatic function assignments: A case for maize. *BMC Systems Biology, 10*(1), 129. https://doi.org/10.1186/s12918-016-0369-x

Yang, Q., Yin, G., Guo, Y., Zhang, D., Chen, S., & Xu, M. (2010). A major QTL for resistance to Gibberella stalk rot in maize. *Theoretical and Applied Genetics, 121*(4), 673–687. https://doi.org/10.1007/s00122-010-1339-0

Yu, J., Holland, J. B., McMullen, M. D., & Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics, 178*(1), 539–551. https://doi.org/10.1534/genetics.107.074245

Yuan, G., Chen, B., Peng, H., Zheng, Q., Li, Y., Xiang, K., Liu, L., Zou, C., Lin, H., Ding, H., Pan, G., & Zhang, Z. (2020). QTL mapping for resistance to ear rot caused by *Fusarium graminearum* using an IBM Syn10 DH population in maize. *Molecular Breeding, 40*(9), 91. https://doi.org/10.1007/s11032-020-01158-0

Zhang, D., Liu, Y., Guo, Y., Yang, Q., Ye, J., Chen, S., & Xu, M. (2012). Fine-mapping of qRfg2, a QTL for resistance to Gibberella stalk rot in maize. *Theoretical and Applied Genetics, 124*(3), 585–596. https://doi.org/10.1007/s00122-011-1731-4

Zhou, S. (2018). Genetic and physiological regulation of specialized metabolism and biochemical defense in *Zea mays*. http://ecommons.cornell.edu/handle/1813/59758

SECTION S4.

Characterization of defensive polyphenols in near-isogenic maize lines with resistance and susceptibility to *Fusarium graminearum* seedling blight

LIST OF FIGURES

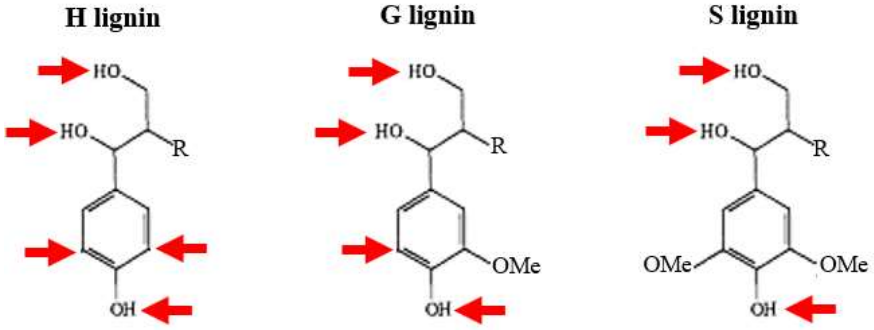

Figure 1. H, G, and S lignin contain 5, 4, and 3 potential polymerization sites, respectively. Abbreviations; H lignin: p-hydroxyphenyl subunit, G lignin: guaiacyl subunit, S lignin: syringyl subunit, R: the rest of the lignin subunit. Figure modified and adapted from Jung & Deetz (1993).

	Fiber[1]	DIMBOA-Glc[2]	HDMBOA-Glc[2]	smiglaside C[3]	smilaside A[3]	DON[3]
High check	Mo17	Mo17	CML322	B73	Mo17	B73
Low check	NC350	NC350	B73	Mo17	B73	Mo17
1.	Venkatesh et al. (2015)					
2.	Meihls et al. (2013)					
3.	Zhou et al. (2019)					

Figure 2. The inbred lines B73, Mo17, CML322, and NC350 will be included in phenotypic analysis to serve as "checks" for phenotypes of interest based on their characterization in previous studies (Meihls et al., 2013; Venkatesh et al., 2015; Zhou et al., 2019). Fiber (lignin, cellulose, and hemicellulose) is highly correlated with acid detergent lignin, in vitro dry matter digestibility, and i*n vitro* neutral detergent fiber digestibility in maize (Wang et al., 2020). Hence, lines with contrasting phenotypes for fiber content are expected to provide contrasting phenotypes for cell wall composition.

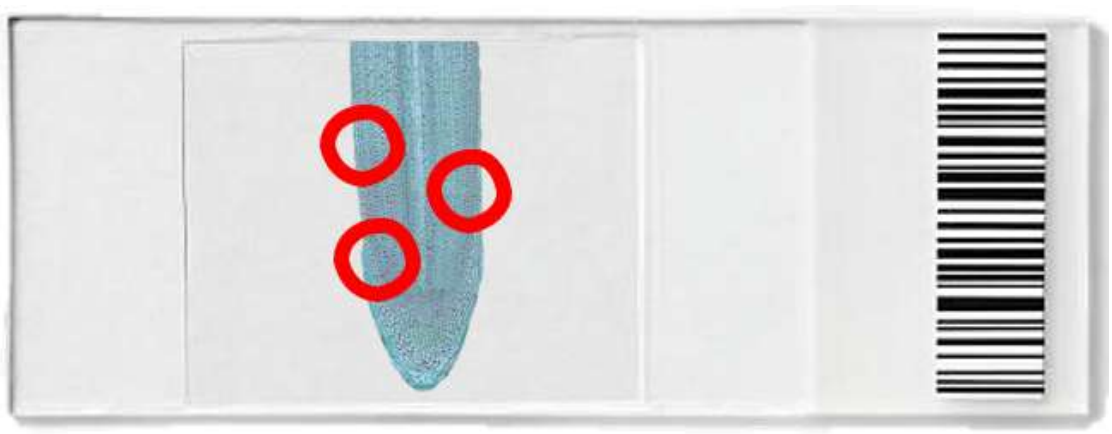

Figure 3. Example of a microscope slide containing a stained longitudinal section of a maize seedling root. Three locations will be marked on the microscope slide cover slip for each longitudinal section, then each location will be viewed under the microscope and digitally photographed. Amount of invasive hyphae will be estimated as the average "blueness" using RGB values of the 3 images.

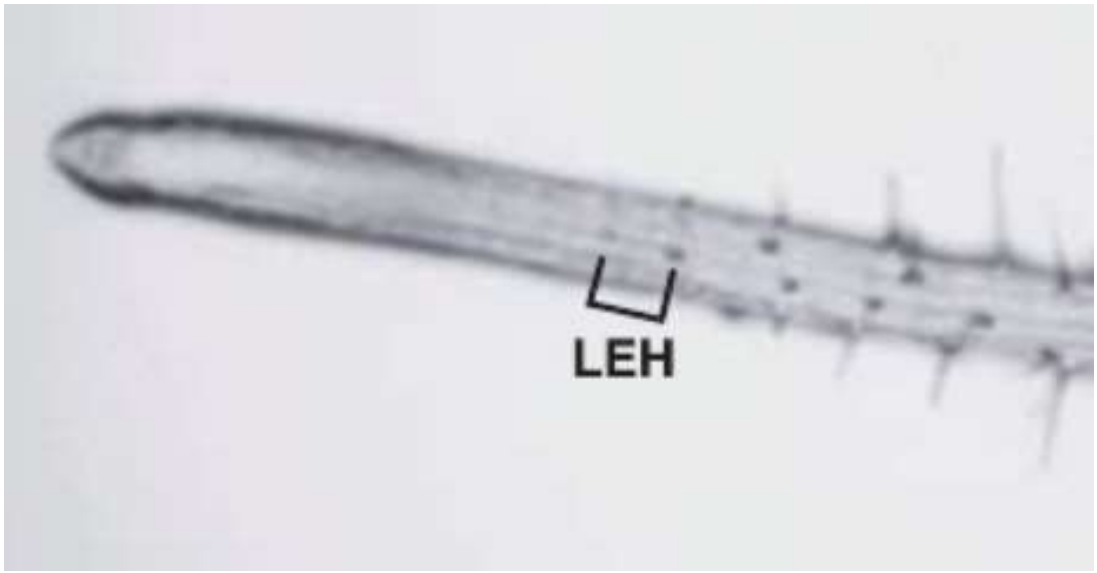

Figure 4. Example of LEH on an *Arabidopsis* root. Original figure from Cnodder et al. (2005).

Characterization of defensive polyphenols in near-isogenic maize lines with resistance and susceptibility to *Fusarium graminearum* seedling blight

INTRODUCTION

In section S3, I will have identified QTL associated with *Fusarium graminearum* SB resistance and susceptibility and gained some insight on the metabolic pathways implicated by those loci using PAST analysis. This chapter aims to further characterize the metabolic bases of SB resistance and susceptibility conferred by these QTL. While evidence from the related *F. verticillioides* pathosystem has shown that ear rot resistance is not always correlated with SB resistance, (Stagnati et al., 2019), some of the metabolic processes implicated in *F. graminearum* resistance at the ear level have also been implicated in seedlings. QTL influencing these metabolic processes are of particular interest due to their potential for pleiotropic *F. graminearum* resistance in both seedlings and mature plants.

A transcriptomics study analyzing *F. graminearum* resistance at the ear level revealed that genes involved in secondary metabolism, cell wall biosynthesis and modification, chitin catabolism, oxidation reduction processes, immune response, and nutrient reservoir activity were more upregulated in a GER-resistant inbred line compared to B73 (Kebede et al. 2018). At the seedling level, Zhou et al. (2018, 2019) demonstrated the importance of benzoxazinoids (BXs) and diferuloylsucroses, namely the ratio of smilaside A:smiglaside C, as key defense compounds in *F. graminearum* resistance.

Because BXs are only produced during the seedling (Niemeyer 2009) and early juvenile (Oikawa et al. 2004; Zheng et al. 2015) stages in maize, they are generally not thought to influence ear rot infection directly (Niemeyer 2009). However, bioactive forms of BXs exuded from root tissue during early developmental stages can persist in soil throughout the growing season (Hu et al. 2018), making them relevant to *F. graminearum* infection throughout the host life cycle. Mechanistically, it is thought that maize BXs primarily contribute to *F. graminearum* resistance by binding to fungal DNA and proteins, resulting in lethal mutagenesis of the fungal cells (Hashimoto & Shudo 1996). These antifungal properties can manifest as inhibited conidial germination and reduced germ tube elongation (Ahmad et al., 2011; Oikawa et al., 2004). Due to the fungitoxic properties of BXs, I hypothesize that inoculated root and shoot tissue from SB-resistant (SBR) lines will have a more invasive hyphae compared to SB-susceptible (SBS) lines, and that amount of invasive hyphae in inoculated tissues will be positively correlated with levels of BXs.

My investigation on the role of diferuloylsucroses in *F. graminearum* SB resistance will build of off the following key findings from Zhou et al., (2018, 2019): (i) Mo17 is more resistant to *F. graminearum* SB than B73, (ii) *F. graminearum* induced reduced root growth and elongated shoot growth in B73 but not in Mo17, (iii) B73 has a low smilaside A:smiglaside C ratio while Mo17 has a high ratio, and (iv) smilaside A:smiglaside C ratio is negatively regulated by ethylene sensitivity. As a well-studied growth regulating hormone, ethylene has been shown to reduce root elongation in maize seedlings in a concentration-dependent manner (Alarcón et al. 2009, 2012). This suggests that *F. graminearum*-induced root reduction in the SBS line B73 and

unaffected root development in the SBR line Mo17 could be due to high and low ethylene sensitivity, respectively. I hypothesize that SBS lines will exhibit reduced root cell elongation in inoculated plants compared to uninoculated plants and the SBR lines will not exhibit differences in root cell elongation between uninoculated vs. inoculated plants.

Unlike many pathogenic fungi, *Fusarium* species do not form tissue-penetrating structures like appressoria, and instead rely heavily on cell wall degrading enzymes for host penetration (Kikot et al., 2009). It has been speculated that the chemical structure of diferuloylsucroses allows for the formation of bonds with lignin monomers, and due to their stereochemistry, these bonds would be inaccessible to cell wall degrading enzymes released from pathogenic fungi (Vanholme et al., 2012). Beyond total lignin content, ratio of syringyl (S) lignin monomers is considered a key factor in host cell wall degradability (He et al., 2018). Jung & Deetz (1993) and Tai et al. (1983) suggest that S lignin is more susceptible to degradation by fungal pathogens than other lignin monomers because it contains fewer sites through which it can polymerize with other phenolic compounds (e.g. smilaside A, smiglaside C), rendering it less protected from cell wall degrading enzymes (Fig. 1). I hypothesize that root and shoot cell walls of SBR lines will have a lower proportion of S lignin than SBS lines, and that proportion of S lignin will be negatively correlated with smilaside A : smiglaside C ratio.

METHODS

Estimating amount of invasive hyphae

Root and shoot tissue from inoculated plants will be hand cut into longitudinal sections using a scalpel, cleared using KOH, stained with lactophenol/cotton blue, viewed under a light microscope, and digitally photographed. Amount of invasive hyphae will be estimated as the average "blueness" of the image using RGB values within a given field of view at 10x magnification. Blueness scores will be averaged from images of 3 locations on each root and shoot longitudinal section (Fig. 3). I will use an ANOVA to determine if SBR lines and SBS lines have statistically different means for amount of invasive hyphae within a 95% confidence interval. If the ANOVA test is significant, I will then use a one-sided t-test to determine if SBR lines have less invasive hyphae than SBS lines within a 95% confidence interval.

Estimating root elongation

Root tissue from inoculated and uninoculated plants will be hand cut into longitudinal sections, cleared with KOH, stained with toluidine blue, and viewed under a light microscope. Root elongation can be estimated as the length between the first 2 visible root hairs or root hair bulges (LEH) nearest the root tip (Cnodder et al., 2005) (Fig. 4). Degree of reduced root elongation will be estimated as the inoculated LEH / uninoculated LEH for each line. I will use an ANOVA to determine if the mean degree of reduced root elongation is significantly different between SBR and SBS lines within a 95% confidence interval. If the difference in means is significant, I will use a one-sided t-test to determine if SBR lines have a lower degree of reduced shoot elongation within a 95% confidence interval.

Estimating proportion of S lignin

Root and shoot tissue from inoculated and uninoculated plants will be hand cut into transverse sections using a scalpel. Tissues will be stained with Wiesner reagent to visualize total lignin and with Maule reagent to selectively visualize S lignin. Stained transverse sections will be viewed under a light microscope at 10X magnification and digitally photographed. Proportion of S lignin in root and shoot sections of each line will be estimated by image color intensity from Maule stain / image color intensity from Wiesner stain. I will use an ANOVA to determine if the mean estimated proportion of S lignin is significantly different between SBR and SBS lines within a 95% confidence interval. If the difference in means is significant, I will use a one-sided t-test to determine if SBR lines have a lower proportion of S lignin than SBS lines within a 95% confidence interval.

Estimating polyphenol abundance

Both BXs and diferuloyl sucroses are polyphenolic compounds. Prior to metabolite quantification of BXs and diferuloyl sucroses, I will first assess if amount of invasive hyphae, root cell elongation, and/or proportion of S lignin appear to be related to polyphenols in general. If my hypotheses about the role(s) of BXs and/or diferuloyl sucroses in SBR are supported, then I would expect SBR lines to have higher abundance of polyphenols in general. To visualize abundance of polyphenolic compounds, transverse sections of roots and shoots will be hand cut with a scalpel, cleared with KOH, stained with the polyphenol stain ferric chloride ($FeCl_3$), viewed under a light microscope, and digitally photographed. Polyphenols have a high iron affinity, so polyphenol-rich tissue will stain brown and other tissue will remain

unstained. Polyphenol abundance will be estimated as a function of the image's color intensity. I will use an ANOVA to determine if estimated polyphenol abundance is significantly different between SBR and SBS lines within a 95% confidence interval. If the difference in means is significant, I will do a one-sided t-test to determine is SBR lines have a higher estimated polyphenol abundance than SBS lines within a 95% confidence interval. A lack of SBR lines that stain brown ($FeCl_3$+) SBR lines or SBS lines that resist staining ($FeCl_3$-) would suggest that SB resistance and SB susceptibility conferred by the QTL in these nNILs may be conferred by mechanisms not related to polyphenols.

Targeted Metabolomics

$FeCl_3$+ SBR lines, $FeCl_3$- SBS lines, and the "check" lines (B73, Mo17, CML322, NC350) will be further analyzed using LC-MS according to Zhou et al. (2019). Metabolites to be quantified include the most abundant maize BX glucosides (DIMBOA-Glc, HDMBOA-Glc), their bioactive derivatives (DIMBOA, HDMBOA, and MBOA), and diferuoyl sucroses (smilaside A and smiglaside C). Quantity for each metabolite will be represented by the normalized peak area under the curve in negative electron spray ionization mode (Zhou et al., 2019). I will use a multivariate ANOVA (MANOVA) to determine if the means for any of the quantified metabolites differ between in $FeCl_3$+ SBR and $FeCl_3$- SBS lines within a 95% confidence interval. Lack of significance from the MANOVA would suggest that SB resistance and susceptibility conferred by the QTL in these nNILs may be conferred by polyphenols other than DIMBOA-Glc, HDMBOA-Glc, DIMBOA, HDMBOA, MBOA, smilaside

A or smiglaside C. If the MANOVA is significant, I will perform linear discriminant analysis to assess which metabolite or combination of metabolites provide the most distinction between $FeCl_3$+ SBR and $FeCl_3$- SBS lines. Then, I will fit a multiple linear regression model such that the amount of invasive hyphae = $b_0 + b_1$(DIMBOA-Glc) + b_2(HDMBOA-Glc) + b_3(DIMBOA) + b_4(HDMBOA) + b_5(MBOA) where the correlation coefficients of each BX (b_1-b_5) correspond to the amount of variance in invasive hyphae explained by the respective BX. The extent to which these BXs together explain variation in amount of invasive hyphae will be estimated using Pearson's correlation (R^2). Two simple linear regression models will also be fit such that proportion of S lignin = $b_0 + b_1$(smilaside A:smiglaside C) and lack of reduced root cell elongation = $b_0 + b_1$(smilaside A:smiglaside C). These models will also be interpreted using R^2.

References:

Ahmad, S., Veyrat, N., Gordon-Weeks, R., Zhang, Y., Martin, J., Smart, L., Glauser, G., Erb, M., Flors, V., Frey, M., & Ton, J. (2011). Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. *Plant Physiology, 157*(1), 317–327. https://doi.org/10.1104/pp.111.180224

Alarcón, María Victoria, Lloret, P. G., Iglesias, D. J., Talón, M., & Salguero, J. (2012). Comparison of growth responses to auxin 1-naphthaleneacetic acid and the ethylene precursor 1-aminocyclopropane-1-carboxilic acid in maize seedling root. Acta Biologica Cracoviensia: Series: *Botanica, 54*(1), 16. http://dx.doi.org.proxy.library.cornell.edu/10.2478/v10182-012-0001-3

Alarcón, Maria Victoria, Lloret-Salamanca, A., Lloret, P. G., Iglesias, D. J., Talón, M., & Salguero, J. (2009). Effects of antagonists and inhibitors of ethylene biosynthesis on maize root elongation. *Plant Signaling & Behavior, 4*(12), 1154–1156. https://doi.org/10.4161/psb.4.12.9948

Cnodder, T. D., Vissenberg, K., Straeten, D. V. D., & Verbelen, J.-P. (2005). Regulation of cell length in the *Arabidopsis thaliana* root by the ethylene precursor 1-aminocyclopropane- 1-carboxylic acid: A matter of apoplastic reactions. *New Phytologist, 168*(3), 541–550. https://doi.org/10.1111/j.1469-8137.2005.01540.x

Hashimoto, Y., & Shudo, K. (1996). Chemistry of biologically active benzoxazinoids. *Phytochemistry, 43*(3), 551–559.

https://doi.org/10.1016/0031-9422(96)00330-5

He, Y., Mouthier, T. M., Kabel, M. A., Dijkstra, J., Hendriks, W. H., Struik, P. C., & Cone, J. W. (2018). Lignin composition is more important than content for maize stem cell wall degradation. *Journal of the Science of Food and Agriculture, 98*(1), 384–390. https://doi.org/10.1002/jsfa.8630

Hu, L., Robert, C. A. M., Cadot, S., Zhang, X., Ye, M., Li, B., Manzo, D., Chervet, N., Steinger, T., van der Heijden, M. G. A., Schlaeppi, K., & Erb, M. (2018). Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nature Communications, 9*(1), 2738. https://doi.org/10.1038/s41467-018-05122-7

Jung, H. G., & Deetz, D. A. (1993). Cell Wall Lignification and Degradability. In Forage Cell Wall Structure and Digestibility (pp. 315–346). John Wiley & Sons, Ltd. https://doi.org/10.2134/1993.foragecellwall.c13

Kebede, A. Z., Johnston, A., Schneiderman, D., Bosnich, W., & Harris, L. J. (2018). Transcriptome profiling of two maize inbreds with distinct responses to Gibberella ear rot disease to identify candidate resistance genes. BMC *Genomics, 19*(1), 131. https://doi.org/10.1186/s12864-018-4513-4

Kikot, G. E., Hours, R. A., & Alconada, T. M. (2009). Contribution of cell wall degrading enzymes to pathogenesis of *Fusarium graminearum*: A review. *Journal of Basic Microbiology, 49*(3), 231–241. https://doi.org/10.1002/jobm.200800231

Meihls, L. N., Handrick, V., Glauser, G., Barbier, H., Kaur, H., Haribal, M. M., Lipka, A. E., Gershenzon, J., Buckler, E. S., Erb, M., Köllner, T. G., & Jander, G. (2013). Natural variation in maize aphid resistance is associated with 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside methyltransferase activity. *The Plant Cell, 25*(6), 2341–2355. https://doi.org/10.1105/tpc.113.112409

Niemeyer, H. M. (2009). Hydroxamic acids derived from 2-hydroxy-2 h -1,4-benzoxazin-3(4 h)-one: Key defense chemicals of cereals. *Journal of Agricultural and Food Chemistry, 57*(5), 1677–1696. https://doi.org/10.1021/jf8034034

Oikawa, A., Ishihara, A., Tanaka, C., Mori, N., Tsuda, M., & Iwamura, H. (2004). Accumulation of HDMBOA-Glc is induced by biotic stresses prior to the release of MBOA in maize leaves. *Phytochemistry, 65*(22), 2995–3001. https://doi.org/10.1016/j.phytochem.2004.09.006

Stagnati, L., Lanubile, A., Samayoa, L. F., Bragalanti, M., Giorni, P., Busconi, M., Holland, J. B., & Marocco, A. (2019). A genome wide association study reveals markers and genes associated with resistance to *Fusarium verticillioides* infection of seedlings in a maize diversity panel. *G3 Genes|Genomes|Genetics, 9*(2), 571–579. https://doi.org/10.1534/g3.118.200916

Tai, D., Terazawa, M., Chen, C. L., Chang, H. M., & Kirk, T. K. (1983).

Biodegradation of guaiacyl and guaiacyl-syringyl lignins in wood by *Phanerochaete chrysosporium*. *Recent advances in lignin biodegradation research: Fundamentals and biotechnology*, 44-63.

Vanholme, R., Morreel, K., Darrah, C., Oyarce, P., Grabber, J. H., Ralph, J., & Boerjan, W. (2012). Metabolic engineering of novel lignin in biomass crops. *New Phytologist, 196*(4), 978–1000. https://doi.org/10.1111/j.1469-8137.2012.04337.x

Wang, Q., Li, K., Hu, X., Shi, H., Liu, Z., Wu, Y., Wang, H., & Huang, C. (2020). Genetic analysis and QTL mapping of stalk cell wall components and digestibility in maize recombinant inbred lines from B73×By804. *The Crop Journal, 8*(1), 132–139. https://doi.org/10.1016/j.cj.2019.06.009

Zheng, L., McMullen, M. D., Bauer, E., Schön, C.-C., Gierl, A., & Frey, M. (2015). Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in *Zea mays*. *Journal of Experimental Botany, 66*(13), 3917–3930. https://doi.org/10.1093/jxb/erv192

Zhou, S., Richter, A., & Jander, G. (2018). Beyond defense: Multiple functions of benzoxazinoids in maize metabolism. *Plant and Cell Physiology, 59*(8), 1528–1537. https://doi.org/10.1093/pcp/pcy064

Zhou, S., Zhang, Y. K., Kremling, K. A., Ding, Y., Bennett, J. S., Bae, J. S., Kim, D.

K., Ackerman, H. H., Kolomiets, M. V., Schmelz, E. A., Schroeder, F. C., Buckler, E. S., & Jander, G. (2019). Ethylene signaling regulates natural variation in the abundance of antifungal acetylated diferuloylsucroses and *Fusarium graminearum* resistance in maize seedling roots. *New Phytologist, 221*(4), 2096–2111. https://doi.org/10.1111/nph.15520