UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Preservation and Improvement of Valencian Agro-diversity University Research Institute (COMAV)

Optimization of predictive ability of genomic selection models for Fusarium Head Blight resistance in wheat cultivars through nonlinear parameterization functions

Master's Thesis

Erasmus Mundus Master Programme in Plant Breeding - emPLANT +

AUTHOR: Tita , Bleck Tabeh

Tutor: Gramazio, Pietro

External cotutor: ISIDRO SANCHEZ, JULIO

ACADEMIC YEAR: 2022/2023

Universitat Politècnica de València

Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural

Instituto Universitario de Conservación y Mejora de la Agrodiversidad Valenciana

Centro de Biotecnología y Genómica de Plantas (CBGP-UPM)

Optimization of predictive ability of genomic selection models for Fusarium Head Blight resistance in wheat cultivars through nonlinear parameterization functions

Master's thesis

Master in Plant Breeding

Academic year 2022-2023

Author:

Bleck Tita

Supervisors:

Dr. Julio Isidro y Sánchez (CBGP-UPM)

Julián García-Abadillo Velasco (CBGP-UPM)

Dr. Pietro Gramazio (UPV)

Acknowledgements

I would like to express my deep appreciation and gratitude to the following individuals and organizations who have contributed to the completion of this thesis:

First and foremost, I am immensely grateful to my supervisor and & co-supervisors, Dr Julio Isidro Sánchez, Mr. Julián García-Abadillo Velasco and Dr Pietro Gramazio for their unwavering support, guidance, and expertise. Their valuable insights, constructive feedback, and continuous encouragement have been instrumental in shaping the direction of this research project and thesis.

Then, I would like to sincerely thank both past and present members of the Rocinante Lab for their invaluable contributions.

I am thankful to the Polytechnic University of Madrid (UPM), Polytechnic University of Valencia (UPV) and the Polytechnic Institute of UniLasalle, for providing the necessary resources, facilities, and research infrastructure that facilitated the smooth progress of my study. Also to BOKU University and SAATZUCHT DONAUfor providing me with my research data.

I am indebted to the funding organizations (The Rocinante Lab and the Centro de Biotecnología y Genómica de Plantas UPM) for their financial support through Beatriz Galindo Program (BEAGAL18/ 00115) from the Ministerio de Educación y Formación Profesional of Spain and the Erasmus emPLANT + Program. Their investment in this project has enabled me to pursue my academic aspirations and make meaningful contributions to the field of Genomic Selection.

Also, I want to express my deepest gratitude to my family and friends for their unwavering support, love, and understanding throughout this challenging journey. Their encouragement, patience, and belief in me have been the driving force behind my perseverance and accomplishments. And to my Wife, Thank you for all the support and encouragements and for standing by me when I had my episodes relating to this thesis. And to myself, as I reach the culmination of this research journey, I would like to take a moment to acknowledge my own personal growth, perseverance, and dedication throughout the process of completing this thesis.

Lastly, I am immensely grateful to all those who have contributed to this thesis, directly or indirectly. Their support, guidance, and encouragement have been invaluable, and I am truly honored to have had the opportunity to work with such remarkable individuals and institutions.

Resumen

La Fusariosis de la Espiga (FHB) representa una amenaza significativa para la producción global de trigo, causando pérdidas de rendimiento, contaminación por micotoxinas y poniendo en peligro la seguridad alimentaria. El desarrollo de variedades de trigo resistentes a FHB mediante métodos convencionales de mejora es un desafío debido a la compleja arquitectura genética de los rasgos de resistencia. La selección genómica (GS) ofrece una solución prometedora para acelerar la mejora del trigo en cuanto a la resistencia a FHB.

En este estudio, evaluamos la precisión de los valores de mejora genómica estimados (GEBVs) para la resistencia a FHB en una diversa colección de 865 variedades de trigo hexaploide (*Triticum aestivum*). Para modelar la curva de progresión de FHB, utilizamos funciones de regresión no lineal que incluyen logística, Gompertz, Mono-molecular, Gaussiana y Gamma. Además, se utilizó el área convencional bajo la curva de progresión de la enfermedad (AUDPC) como referencia.

Nuestros resultados muestran que la función Logística con dos parámetros mostró la correlación más alta de 0,89 con AUDPC. Para predecir los GEBV, aplicamos el método de predicción lineal mejorada genómica (GBLUP), XgBoost y varios métodos de regresión bayesiana (BayesA, BayesB, BayesC, BayesRR). Mediante validación cruzada de cinco pliegues, la precisión media a lo largo de los años fue de 0,58, 0,60, 0,55 y 0,53 para BayesA y BayesB, GBLUP y XgBoost respectivamente, cuando se utilizó el parámetro B de la función logística. Cuando se utiliza AUDPC, GBLUP, XgBoost, BayesA y BayesB consiguen precisiones de predicción de 0,54, 0,49, 0,57 y 0,59 respectivamente.

Nuestros resultados destacan la eficacia de los modelos de regresión no lineal en la selección genómica para mejorar la resistencia a FHB en trigo. El uso de la parametrización no lineal superó a la AUDPC convencional, con BayesB, BayesA, GBLUP y Xgboost en otros de fuerza mostrando una mayor precisión en la predicción de variedades resistentes a FHB. Estos resultados proporcionan información valiosa para los mejoradores que buscan desarrollar variedades de trigo resistentes a FHB utilizando GS.

Abstract

Fusarium ear blight (FHB) represents a significant threat to global wheat production, causing yield losses, mycotoxin contamination and jeopardizing food security. The development of FHB-resistant wheat varieties through conventional breeding methods is challenging due to the complex genetic architecture of the resistance traits. Genomic selection (GS) offers a promising solution to accelerate wheat breeding for FHB resistance.

In this study, we evaluated the accuracy of estimated genomic breeding values (GEBVs) for FHB resistance in a diverse collection of 865 hexaploid wheat (*Triticum aestivum*) varieties. To model the FHB progression curve, we used nonlinear regression functions including Logistic, Gompertz, Mono-molecular, Gaussian, and Gamma. In addition, the conventional Area Under the Disease Progression Curve (AUDPC) was used as a reference.

Our results show that the Logistic function with two parameters showed the highest correlation of 0.89 with AUDPC. To predict GEBVs, we applied the Genomic Linear Enhanced Linear Prediction (GBLUP) method, XgBoost and several Bayesian regression methods (BayesA, BayesB, BayesC, BayesRR). By five-fold cross-validation, the average precision across years was 0.58, 0.60, 0.55 and 0.53 for BayesA and BayesB, GBLUP, and XgBoost respectively, when using parameter B of the logistic function. When using AUDPC, GBLUP, XgBoost, BayesA, and BayesB achieved prediction accuracies of 0.54, 0.49, 0.57 and 0.59 respectively.

Our results highlight the effectiveness of nonlinear regression models in genomic selection for improving FHB resistance in wheat. The use of nonlinear parameterization outperformed conventional AUDPC, with BayesB, BayesA, GBLUP and Xgboost in other of strength showing higher accuracy in predicting FHB resistant varieties. These results provide valuable insights for breeders seeking to develop FHB resistant wheat varieties using GS.

Contents

1 Introduction

In 2021, wheat *(Triticum aestivum L.)* contributed to an estimated global production of over 770 million tons (faostat 2021), serving as a primary source of sustenance for most of the world's population, particularly in developing countries, providing 18% of human daily calorie intake and 20% of protein (faostat 2021, *Wheat Nutrition* n.d.). However, increasing population growth and changing climatic conditions have exerted pressure on wheat production, making it imperative to enhance the crop's genetic potential to meet the rising food demands and disease management is paramount to this goal (Aggarwal 2008). Effective management of diseases such as Fusarium Head Blight (FHB) (Wilde & Miedaner 2006), a fungal ailment that can cause yield losses of up to 50% and reduce grain quality by producing mycotoxins, is one of the most significant impediments to wheat production (Machado et al. 2018).

The primary causative agents of FHB are pathogenic species such as *Fusarium graminearum*, *F. culmorum*, and *F. avenaceum*, which produce toxins that are harmful to both human and animal health. These pathogens also cause significant crop losses and jeopardize food security (Nelson et al. 1993, Yu et al. 2004, Goswami et al. 2006, Lemmens et al. 2016, Perochon et al. 2015, Xia et al. 2021, Edwards 2022). Additionally, kang2008cytological report that these pathogenic species have detrimental effects on human and animal health, as well as significant impacts on crop yields in most wheat producing areas around the world. Expensive fungicides are often used in conventional management of FHB in wheat production, which can have negative environmental consequences (Kage et al. 2017, Gunupuru et al. 2019, Reyna et al. 2023). To manage FHB, researchers have proposed different strategies such as crop rotation, tillage, use of resistant varieties, and fungicide application (Pirgozliev et al. 2003, Musyimi 2009, Bernhoft et al. 2022). The most sustainable and effective approach is using resistant varieties (Cromey et al. 2001, Gosman et al. 2007). However, developing FHB-
resistant wheat varieties has been challenging due to the complex genetic architecture of disease-resistance traits (Arruda et al. 2016, Velásquez et al. 2018, Buerstmayr et al. 2020).

Initially, breeders rely on traditional methods, i.e.phenotypic selection, where breeders use their expertise to choose good offspring based on observed traits. Marker-assisted selection (MAS) (Lande & Thompson 1990, Wakchaure et al. 2015, Boopathi & Boopathi 2020) which gained considerable popularity in the past, employed the use of molecular genetic techniques to identify traits controlled by a small number of major genes. However, the limitations of MAS became evident when dealing with complex traits such as disease resistance, which are often influenced by multiple genes (Lande & Thompson 1990). As genotyping costs began to fall, it became feasible to expand upon MAS by including all loci, rather than a select few, leading to the development of Genomic Selection (GS) (Meuwissen et al. 2001). This method, an advanced derivative of MAS, leverages genome-wide markers to calculate a Genomic Estimated Breeding Value (GEBV). The core advantage of GS over MAS is that it allows for the estimation of the effects of all loci, facilitating selection for complex, quantitative traits, in contrast to MAS's focus on simpler, qualitative ones. Genomic selection utilizes a training population to establish associations between markers and phenotypes (Isidro et al. 2016, Cobb et al. 2013, Spindel et al. 2015, Crossa et al. 2017, Heffner et al. 2009, Heslot et al. 2012), subsequently constructing a predictive model of genomic markers and phenotypic traits. This innovation moves beyond the confines of Quantitative Trait Loci (QTL) detection, enabling the prediction of genetic effect values of unobserved individuals using the established model. Thus, GS presents a significant upgrade in breeding strategies, offering an effective solution to the enduring challenge of developing FHB-resistant wheat varieties.

In recent years, GS has emerged as a powerful tool for enhancing FHB resistance in wheat, particularly with the advent of contemporary DNA marker technologies such as Diversity Arrays Technology (DArT) and Single Nucleotide Polymorphism (SNP) (Ashraf 2021). As an illustration, an experimental review by (Arruda et al. 2016) demonstrated superior predictive ability for FHB resistance utilizing GS, even outperforming traditional selection methodologies. In fact, the top-performing lines selected via GS exhibited significantly elevated resistance levels (Arruda et al. 2016, Larkin et al. 2019).The efficiency of GS has been tested against MAS and conventional pedigree breeding in maize, where it has resulted in yield increments up to 16% (Jannink et al. 2010). GS has also been successfully utilized in wheat to predict rust resistance, achieving prediction accuracies ranging from 0.27 to 0.44 (González-Camacho et al. 2018). Significantly, GS facilitates the early assessment and selection of individuals, reducing the time and labor expenditure per breeding cycle and hence shortening the generation interval. This early assessment possibility might lead to a paradigm shift in the role of phenotyping, where it predominantly

serves to update prediction models rather than solely being used for line selection (Jeon et al. 2023).

GS as previously discussed, uses statistical models to identify and quantify the correlation between genetic markers and the phenotype of interest (Cobb et al. 2013, Spindel et al. 2015, Crossa et al. 2017). The majority of these statistical models used in GS are linear models, based on the assumption that the phenotype is a linear function of the genetic markers (Heffner et al. 2009, Pérez et al. 2010, Ogutu et al. 2012, Howard et al. 2014, González-Camacho et al. 2018). Key to GS is the presumption of additive effects of each marker on the phenotype (Heffner et al. 2009), coupled with the notion that at least one marker is in linkage disequilibrium with the QTL of interest. A prevalent linear model used in GS is the genomic best linear unbiased prediction (GBLUP) model. This model assumes that each marker's effect on the phenotype adheres to a normal distribution, with a mean of zero and a variance proportionate to the marker's linkage disequilibrium with the QTL affecting the trait of interest (Xiao et al. 2022). Apart from GBLUP, other linear models like least absolute shrinkage and selection operator (LASSO) and Bayesian ridge regression have been deployed to predict disease resistance in wheat (Juliana et al. 2017, Momen et al. 2018, Xu et al. 2021). Notably, works such as those by Wu et al.(2020) and Xu et al.(2021) have used LASSO regression for predicting FHB resistance in Chinese wheat landraces. They found LASSO regression proficient in identifying QTLs with minor effects, which were not detected by traditional genome-wide association studies. In addition, Heslot et al.(2015) applied Bayesian ridge regression to several quantitative traits in a spring wheat panel and reported high prediction accuracy for grain yield.

Several strategies have been proposed to leverage GS in breeding for FHB. For instance, Tessema et al. (2020) proposed using optimized genotyping strategies and statistical models that can account for complex trait architectu
-res to increase the accuracy of predictions. Ashraf (2021) suggested using wild relatives to improve wheat breeding for rust resistance. These relatives can be incorporated into GS programs to increase the allele diversity used for prediction. Current research on GS for FHB resistance in wheat primarily relies on linear regression models (Arruda et al. 2016, Jiang et al. 2017, Zhang et al. 2022). However, recent studies have demonstrated that non-linear regression models can be advantageous for disease progression modeling and improving prediction accuracy (Bari et al. 2012, Heslot et al. 2014, Garcia-Abadillo et al. 2022). Non-linear models are more flexible than linear models and can capture complex relationships between genomic markers and disease progression (Blum & François 2010, Pérez-Rodríguez et al. 2012). Machine learning algorithms, such as Support Vector Machines, Random Forests, and Neural Networks, have shown potential in GS for various crop traits (González-Camacho et al. 2012). These methods can model non-linear and interaction effects among markers and improve prediction accuracy (Zingaretti et al. 2020). Some studies have explored the use of non-linear regression models to predict FHB incidence in wheat, using methods such as the generalized additive model and Support Vector Machine models (Dyba et al. 1997, Paul et al. 2005).

These findings suggest that non-linear models can provide more accurate predictions of FHB incidence in wheat, outperforming traditional linear regression models. However, the predictive ability of most models linear or non linear depends on the precise evaluation of the progression of disease and the resistance of wheat varieties, therefore, disease progress curves (DPCs), which describe the increase in disease severity over time, are commonly used to evaluate the efficacy of resistance acting as estimators and/or predictor parameters (Savary et al. 2000, 2006, Mundt 2014, Gazal et al. 2016, Saeed et al. 2022). Previous studies suggested that it is demanding to accurately model the disease progression curves using sigmoidal functions through nonlinear regression (Cao et al. 2019, Carrillo & González 2002, Berger 1981) and as a result most researchers turn to use mostly linear models which has its limitations (Berger 1981) but the applicability of traditional linear statistical models for predicting FHB incidence are usually limited, particularly in capturing the complex and nonlinear relationships between environmental factors and disease occurrence (Huang et al. 2019, Almoujahed et al. 2022, Shah et al. 2023). Therefore, non-linear regression techniques provide a better potential solution to improve FHB prediction accuracy despite its complexity (Shah et al. 2023).

As to any model based campaign for estimation and or predictions, the selection of an appropriate mathematical function to model any response variable (in the case of this study, *DCPs*) is a crucial step towards building a robust model with a good predictive ability (Mesterhazy et al. 2000, Rossi et al. 2009, Paul et al. 2005). So, this study employed the use of several non linear mathematical functions, such as *Logistic, Gompertz, Monomolecular, Gaussian, and Gamma* to model the disease progression curves of FHB in wheat and also evaluates its fitness criteria as suggested by (Mesterhazy 1995, Li et al. 2022, Xiao et al. 2022) in prior studies. The selection of these nonlinear functions is generally based on widely

used growth models for describing temporal disease epidemics, as discussed by Xu et al.(2006).

Researchers have employed some of these functions to model the pattern of some major plant disease severity in different crops, with some major breakthroughs including: using the logistic function, which usually depicts an S-shaped curve or a sigmoid with an initial exponential growth phase followed by a slowdown phase and a plateau (Arruda et al. 2016), to model the progression of certain plant diseases. The Gompertz function has been used to model the spread of plant diseases by assuming that the disease progress rate in general decreases exponentially with time (Cheng et al. 2020). The Gaussian function, which is principally used to estimate the mean and variance of a normal distribution (Nutter 2007), has been used to model the distribution of most plant diseases as well. Finally, the Gamma function has been employed to analyze the distribution of disease severity scores assuming that the severity scores follow a gamma-like distribution (Andrews & Mallows 1974). These mathematical functions have proven to be valuable tools in predicting and understanding the patterns and spread of plant diseases, aiding in the development of effective management strategies.

2 Objectives

Our study aims to investigate the most appropriate mathematical function that characterizes the progress curves of FHB disease in wheat and define the parameters that correspond to the most optimal curve and can be used for its estimate. To achieve this objective, we will use data collected over eight years of an empirical wheat breeding program and apply genomics-assisted selection tools to predict the disease's associated parameters. Our study will

- Develop a range of mathematical functions using non-linear regression analysis to fit the disease progress curves.
- Define the parameters associated with the most optimal mathematical function that characterizes the disease progress curve of FHB in wheat.

Our goal is that our findings will provide valuable insights into the dynamics of FHB disease in wheat and facilitate the development of effective breeding and selection measures for this economically important disease.

3 Materials and Methods

3.1 Materials

3.1.1 Data Description

I. **FHB Phenotypic Data:**

For our study, we utilized a dataset generated from a wheat breeding program conducted by BOKU between 2015 to 2022 (Table 1). The dataset contains information on 865 genotypes that underwent testing for eight years, with FHB scoring information at four different time points per year (Figure 1). In addition, the dataset includes other relevant phenotypic information such as plant height, anther retention, and duration of anthesis which serve as good co-variables. The field trials were conducted in Tulln (Austria) using a randomized design with two to six replications per genotype within a year (Table 2). For some genotypes, non-replicated trials were conducted across multiple years.

Table 1: Number of Genotypes per year and Repetition utilized in this study.

	Number of Genotypes in Each Reps	
Year	**Rep 1**	**Rep 2**
2015	166	166
2016	199	199
2017	193	193
2018	97	97
2019	139	134
2020	120	120
2021	127	127
2022	266	266

Table 2: Genotype Appearance Accross Years

Years	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
2015	138	34	9	2	4	1	0	0
2016	34	187	46	2	3	1	0	0
2017	9	46	126	24	9	2	5	4
2018	2	2	24	90	23	2	0	0
2019	4	3	9	23	90	19	9	9
2020	1	1	2	2	19	111	27	15
2021	0	0	5	0	9	27	126	39
2022	0	0	4	0	9	15	39	233

II. **FHB Genotypic Data:**

To evaluate the extent of FHB, the field staff at BOKU inoculated *F. graminearum* or *F. culmorum* during the anthesis stage of the field trials. The incidence and severity of the disease was calculated by determining the percentage of infected spikelets per plot (Figure 1). The genetic information used in this study was obtained from 1-week-old seedlings of the 865 different varieties which was extracted and subjected to sequencing using the TraitGenetic 25K single nucleotide polymorphism (SNP) chip and to ensure accuracy, markers that exhibited more than 5% heterozygous or missing calls and had a minor allele frequency below 5% were eliminated. After applying these filters, a total of 6677 informative markers were preserved for further analysis.

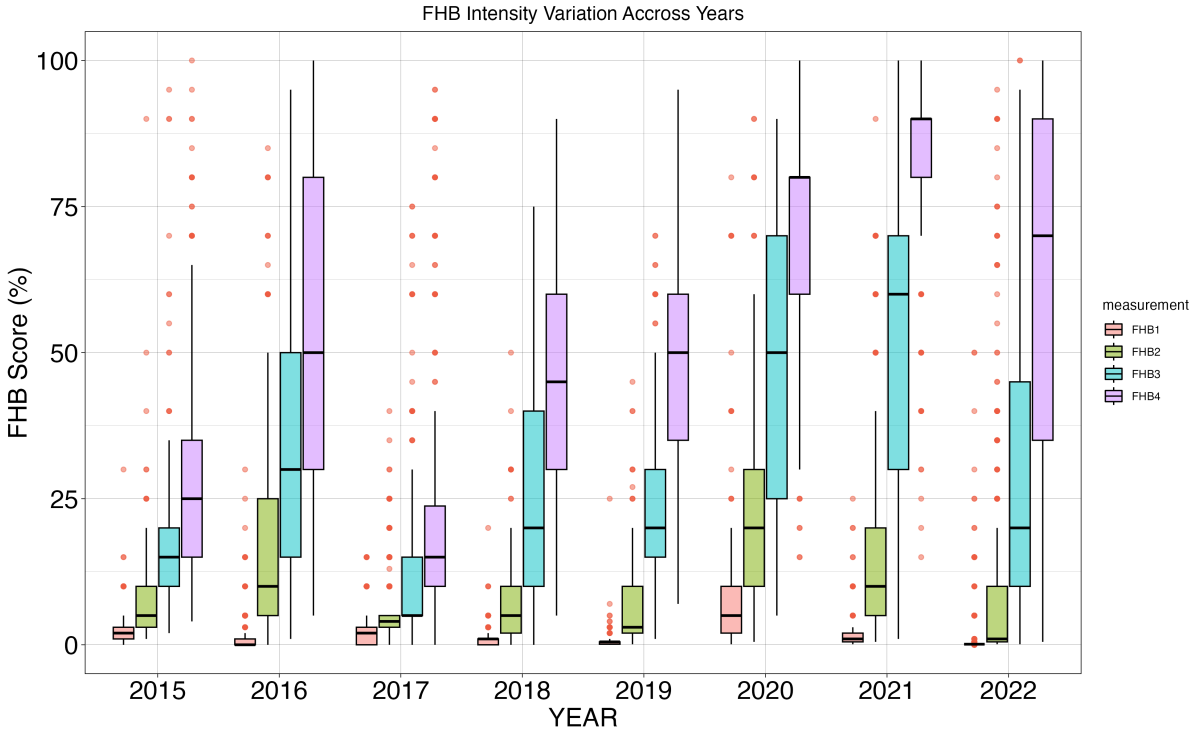

Figure 1: FHB severity variation across years: This figure represents the variation of the FHB severity measured during 8 years. The severity intensity increased relatively from 2015 to 2022 as represented by the boxplots

3.2 Methods

3.2.1 Scoring Metrics

I. **Area Under the Disease Progress Curve (AUDPC):**

To determine the overall progression of a given pathogen, it is essential to calculate the area under the disease progress curve (AUDPC). This calculation involves employing Riemann's sum with a discrete set of T assessments to approximate the definite integral of a function over the interval between x_1 and x_T. The default method for obtaining AUDPC is the Trapezoidal or mid-point rule, as suggested by (Wilcoxson et al. 1975). However, depending on experimental and statistical requirements, other algorithms may be used, as recommended by (Simko & Piepho 2012).

In this study, we propose an alternative scoring metric that utilizes the corresponding T FHB assessment frequency, represented by $X = x_1, x_2, ..., x_T$, and the set of T FHB severity evaluations denoted as $Y = y_1, y_2, ..., y_T$. To achieve this, we define a scoring metric as a function that maps X and Y to a score, which is then utilized as the response variable in further analysis.

$$X = x_1, x_2, ..., x_T; \;\; \Omega = \omega_1, \omega_2, ..., \omega_\Omega; \;\; Y = y_1, y_2, ..., y_T$$

A scoring metric is defined as:

$$S(\Omega, Y) = f(X, Y)$$

where S is the scoring metric, f is the function that maps X and Y to a score.

II. **Time needed to reach 50% of the total FHB severity (T_{50}):**

To measure partial resistance in plants and the progression of the disease, we calculated T_{50}, which is the accumulated time necessary for a plot to reach 50% of the total FHB severity. This measure is also known as the latent period in disease progression. To determine T_{50}, we used a linear interpolation algorithm that identified the data points nearest to the target disease level of 50%. The calculation for T_{50} is as follows:

$$T_{50}(X, Y) = \frac{50 - y_L}{y_R - y_L}(x_R - x_L) + x_L \tag{1}$$

where $L = \arg\max_x (X|y \leq 50)$, and $R = \arg\min_x (X|y \geq 50)$.

We can identify the indexes for the closest points to the target disease level (50%) as L and R, corresponding to the points located to the left and right of the target, respectively. Thus, the optimal points for performing a linear interpolation of T_{50} can be represented by the coordinates (x_L, y_L) and (x_R, y_R).

3.2.2 Mathematical Model Selection

We evaluated different mathematical functions, including logistic, Gompertz, Gaussian, and gamma, to determine the function that best fits the FHB phenoytpic data, assuming a sigmoid pattern for the FHB disease. We fit these models to the data using nonlinear regression analysis and selected the model with the best results.

I. **Logistic Model:**

We considered a logistic model with two parameters and three parameters to evaluate which one provided the best fit for the data when fitting the data parameters.

(a) Logistic model with only two parameters:

$$y = \frac{100}{1 + \exp(-r(x - B))} \tag{2}$$

(b) Logistic model with three parameters:

$$y = \frac{K}{1 + \exp(-r(x - B))} \tag{3}$$

Equations three and four record the severity of FHB for a given value of accumulated FHB assessment time (T), represented by the variable **y**, while the parameter **B** corresponds to the T values at any given time, represented by the variable **x**. The value of **x** when the severity of FHB reaches 50% corresponds to the ***inflection point***. The parameter **r** provides information on the rate at which the disease progresses, corresponding to the shape of the sigmoid curve, while the parameter **K** represents the threshold limits of the sigmoid curve growth. In the case of the logistic model with two parameters, this threshold limit was preset to 100, assuming that the maximum severity of FHB is attained at that point. Meanwhile, in the case of the logistic with three parameters, this threshold was set to have a variable threshold limit.

Besides the logistic model with an extra K parameter, all remaining models discussed below have similar parameters Where y corresponds to FHB severity for any given value of T (B), r represents the growth rate of the disease, and x represents the inflection point at 50% severity rate.

II. **Gompertz Model:**

Assuming severity measurements are available at four time points, namely t_1, t_2, t_3, and t_4, we can represent the severity measurements as y_1, y_2, y_3, and y_4, respectively.

The Gompertz function is used to describe the growth of FHB severity over time (t) and therefore, the Gompertz model used to fit the data is given by the following equation:

$$y(t) = K \times \exp\left(-B \times \exp\left(-r \times t\right)\right) \tag{4}$$

In this equation:

$y(t)$ represents the severity of FHB at time t.

K is the carrying capacity or the maximum severity that FHB can reach (K=100).

B is a parameter that influences the shape of the severity curve.

r controls the rate of change of severity over time.

III. **Mono-molecular:**

The monomolecular function is a mathematical model commonly used to describe various biological and chemical processes. It is characterized by a single exponential term and is often employed to analyze growth or decay phenomena. The function exhibits sigmoidal behavior, starting slowly, accelerating, and then reaching an asymptote. By adjusting the parameters, the monomolecular function can capture different growth patterns and provide insights into the initial value, maximum value, and rate of change of the process being modeled. It finds applications in areas such as population dynamics, enzyme kinetics, and ecological modeling.

To model the sigmoid curve representing FHB severity, we utilize the monomolecular function:

$$y = 100 \times (1 - B \times \exp(-r \times x)) \tag{5}$$

Where:

- y represents the severity of FHB at a given time point, scaled from 0 to 100.
- x denotes the independent variable, representing the time points (t_1, t_2, t_3, t_4).
- B is a parameter that influences the shape of the sigmoid curve, determining the rate of change of severity over time.
- r controls the rate at which the severity approaches its maximum value.

We utilize the severity measurements at the four time points (FHB1, FHB2, FHB3, FHB4) as data points to estimate the parameters B and r.

Estimating the values of B and r enables us to capture the characteristics of the sigmoid curve representing FHB severity. These estimated parameters provide insights into the growth pattern and rate of FHB severity development over time.

IV. **Gaussian:**

In the case of implementing the Gaussian model, the formula used to fit the parameters as shown in the equation below:

$$y = 100 \times \mathrm{p}_{\mathrm{norm}}(x, B, r) \tag{6}$$

This equation represents the Gaussian model used to fit the data, where y is the severity of FHB, 100 is the maximum value to which FHB reaches, x is the accumulated measurement time(T), B is the value of T at any given time, and r is the rate at which the disease progresses. The function $\mathrm{p}_{\mathrm{norm}}$ represents the cumulative distribution function of a standard normal distribution.

V. **Gamma:**

The gamma model is represented by the following equation:

$$y = 100 \times \mathrm{p}_{\mathrm{gamma}}(x, B, r) \tag{7}$$

where y is the severity of FHB recorded for a given value of accumulated measurement time(T), B corresponds to the values of the T at any given time, r provides information on the rate at which the disease progresses, and $\mathrm{p}_{\mathrm{gamma}}$ is the probability density function of the gamma distribution. The parameter values are estimated using nonlinear regression analysis to fit the gamma model to the data.

3.2.3 Parameter Estimation

We estimated the parameters associated with the most optimal curve using mathematical models. These parameters encompassed the maximum disease severity, time to reach the maximum severity, and the rate of disease progression. To accomplish this, we explored different methods for parameter selection, including maximum likelihood estimation (MLE) and least-squares regression. Maximum likelihood estimation is a statistical method that estimates the values of the model parameters that maximize the likelihood of observing the data, while least-squares regression minimizes the sum of squared differences between the observed data and the predicted values. We compared the results of both methods to determine the most suitable approach for parameter selection.

3.2.4 Goodness Of Fit Of Model

We evaluated the goodness of fit of each model by calculating their mean squared error, adjusted R-squared, and Akaike Information Criterion (AIC). The model that exhibited the lowest mean squared error, highest adjusted R-squared, and lowest AIC was selected as the most suitable fit for the FHB data.

3.2.5 Genomic Selection Model Development

We evaluated the genomic prediction accuracy using GBLUP method. Additionally, we employed four Bayesian methods: BayesA, BayesB, BayesC, and BRR. In our study, we considered a population of n individuals with available phenotypic records. Each individual was genotyped with m markers. We represented the genotypic information as the matrix Z, with dimensions n × m. In this matrix, z_{ij} denotes the number of selected alleles at the *jth* locus (j = 1, 2, ..., m) for the *ith* genotyped individual (i = 1, 2, ..., n).

I. **Genomic Best Linear Unbiased Prediction Model (GBLUP):**

The effectiveness of scoring metrics was assessed by employing the Genomic Best Linear Unbiased Predictors (GBLUP) model established by (VanRaden 2008).

The GBLUP model, represented by the following equations, was employed:

$$\begin{cases} y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \epsilon \\ \begin{bmatrix} u \\ \epsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{G}\sigma_a^2 & 0 \\ 0 & \mathbf{I}\sigma_\epsilon^2 \end{bmatrix} \right) \end{cases} \tag{8}$$

In this equation, y is the vector of phenotypic records with n dimensions,˘ represents the overall mean, and Z and ϵ are random vectors representing additive genetic values and errors, respectively. To further elaborate, the design matrices X and Z represent the known fixed effects (such as environmental factors (*years*), blocks (*Repetitions*), and covariates (*plant height, anthesis duration, and anthesis retention*)) and random effects (such as genotypes), respectively. The vector $\boldsymbol{\beta}$ indicates the regression coefficients for the fixed effects estimated using the least squares method. The part $\boldsymbol{u}$ comprises the random genetic additive effects or Genomic Estimated Breeding Values (GEBVs).

The random term and residual variances are denoted as $\boldsymbol{\sigma_a^2}$ and $\boldsymbol{\sigma_\epsilon^2}$, respectively. The matrices G and I define the covariance structure, with G representing the additive genomic relationship matrix based on marker information. This model is assumed to be homocedastic i.e. residuals are independent and their variance-covariance is the identity matrix (I)

II. **Bayesian Models (BAYES):**

Bayesian methods assume different marker effect distribution and allow to model complex genetic architectures.The prior densities and hyperparameters for different Bayesian methods in this model are as follows:

a. **BayesA:**
The prior density of the marker effects, β_j, follows a scaled-t distribution. To avoid computational complexity, the scaled-t distribution is implemented as a finite mixture of scaled-normal densities. Specifically, β_j follows a normal distribution with mean 0 and variance $\sigma^2_{\beta_j}$.

b. **BayesB**:
The prior distribution of the effect of each marker is a mixture of a scaled-t distribution and a distribution with a point mass at zero. The effect β_j follows a normal distribution with mean 0 and variance $\sigma^2_{\beta_j}$ with probability π, and it is equal to zero with probability $(1-\pi)$.

c. **BayesC**:
The prior densities of the marker effects are assumed to be a Gaussian mixture. β_j follows a normal distribution with mean 0 and variance $\sigma^2_{\beta_j}$ with probability π, and it is equal to zero with probability $(1-\pi)$. The parameter π follows a Beta distribution, $\pi \sim \text{Beta}(p_0, \pi_0)$, where $p_0 > 0$ and $\pi_0 \in [0,1]$ with $\text{E}(\pi) = \pi_0$ and $\text{Var}(\pi) = \frac{\pi_0(1-\pi_0)}{1+p_0}$.

d. **BayesBRR**:
The marker effects are assigned independent identically distributed (i.i.d.) Gaussian priors with the same variance, $\beta_j \sim \mathcal{N}(0, \sigma^2_\beta)$. The overall mean μ is assigned a flat prior. The variance of the marker effects and the error variance σ^2_e for all Bayesian models follow a chi-squared distribution with degrees of freedom v and shape parameter S. The shape parameter S has a prior distribution following a gamma distribution with rate parameter r and shape parameter s.

III. **XgBoost Model:** XgBoost, a powerful machine learning ensemble method, is suitable for both classification and regression tasks and is particularly effective when dealing with large datasets that exhibit complex patterns. This method utilizes an ensemble approach, where the final decision or prediction in a regression problem is based on the average outputs of multiple regression trees.

One key distinction between XgBoost and other ensemble methods like random forests lies in how the individual trees contribute to the final decision. XgBoost employs a gradient descent optimization heuristic, assigning different weights or votes to each tree based on their performance, whereas random forests assign uniform weights to all trees. This allows XgBoost to adaptively learn from the data and make more accurate predictions.

In contrast to parametric methods, XgBoost is considered non-parametric. This means that it doesn't compute specific effects for each marker. Instead, it makes predictions by grouping

instances based on their features, particularly SNP markers, and leveraging the patterns and relationships observed within these feature groups.

3.2.6 Model Validation

To evaluate the accuracy of genomic prediction, we adopted the repeated five-fold cross-validation (CV) approach (González-Camacho et al. 2012, Gianola et al. 2014, Pérez-Rodríguez et al. 2012). In this approach, the experiment was conducted 30 times, with each iteration employing a 5-fold CV procedure (Makowsky et al. 2011, Pérez-Cabal et al. 2012, Kramer et al. 2014). Specifically, the entire dataset was randomly divided into five disjoint subsets, with each subset containing approximately the same number of individuals. Four subsets were used as the training population, while the remaining subset served as the test population. This process was repeated five times, rotating the test subset in each iteration to ensure that each subset was used as a test population once (González-Camacho et al. 2012, Gianola et al. 2014, Pérez-Rodríguez et al. 2012, Makowsky et al. 2011, Pérez-Cabal et al. 2012, Kramer et al. 2014).

3.2.7 Statistical Analysis

The predictive ability of the genomic prediction methods was assessed by calculating the Pearson's correlation coefficients between the Genomic-Enabled Breeding Values (GEBV) and the observed phenotypic traits (Zambrano & Echeverri 2014). Within each fold of the five-fold cross-validation (CV), the Pearson's correlation between the predicted GEBV and phenotypic values was computed. The final accuracy was obtained by averaging the correlation coefficients over 30 experiments. To measure bias in the GEBV, the regression coefficient was computed by regressing the actual phenotype on the GEBV (Zambrano & Echeverri 2014). A regression coefficient close to 1 indicates no bias, while values less than 1 or greater than 1 indicate underestimation or overestimation of GEBV, respectively (Makowsky et al. 2011, Pérez-Cabal et al. 2012, Kramer et al. 2014).

The effectiveness of scoring metrics in predicting outcomes was assessed using the GBLUP (Genomic Best Linear Unbiased Predictors) model developed by Van Raden (VanRaden 2008). The sommer package in R was employed for implementing the model covarrubias2016genome. All the Bayesian methods were implemented using the BGLR R-package de2014bayesian.

For the BLUP methods, the heritability of the trait can be calculated as follows:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} \tag{9}$$

where σ_g^2 represents the additive genetic variance and σ_e^2 represents the residual variance.

$$h^2 = \frac{V_A}{V_A + \sigma_e^2} \tag{10}$$

where V_A is the total additive genetic variance, given by:

$$V_A = \pi \times 2\hat{\sigma}_S^2 \times \frac{N}{P} \sum_{j=1}^{m} p_j q_j \tag{11}$$

where p_j and q_j are the allelic frequencies of the jth locus, π denotes the proportion of markers with non-zero effects, N is the number of individuals, P is the number of markers, and $\hat{\sigma}_S^2$ is the estimated marker effect variance.

4 Results

4.1 Disease Severity Variation Across Trial Years

Figure 2 illustrates the results of scoring FHB severity over eight years within the BOKU wheat breeding program. BOKU performed the scoring at four time points (FHB1 to FHB4), which are plotted along the x-axis. The severity score of the disease, ranging from 0 to 100 percent and reflecting the rate of infection and spread, is represented on the y-axis. The lines within each plot show the rate of change of FHB severity between consecutive measurement points and provide information on AUDPC. Our results indicate that FHB severity was more pronounced in the years 2016, 2020, 2021, and 2022 compared to the other years (Figure 2). This observation is supported by the presence of a greater number of genotypes exhibiting higher AUDPC curves (blue lines).

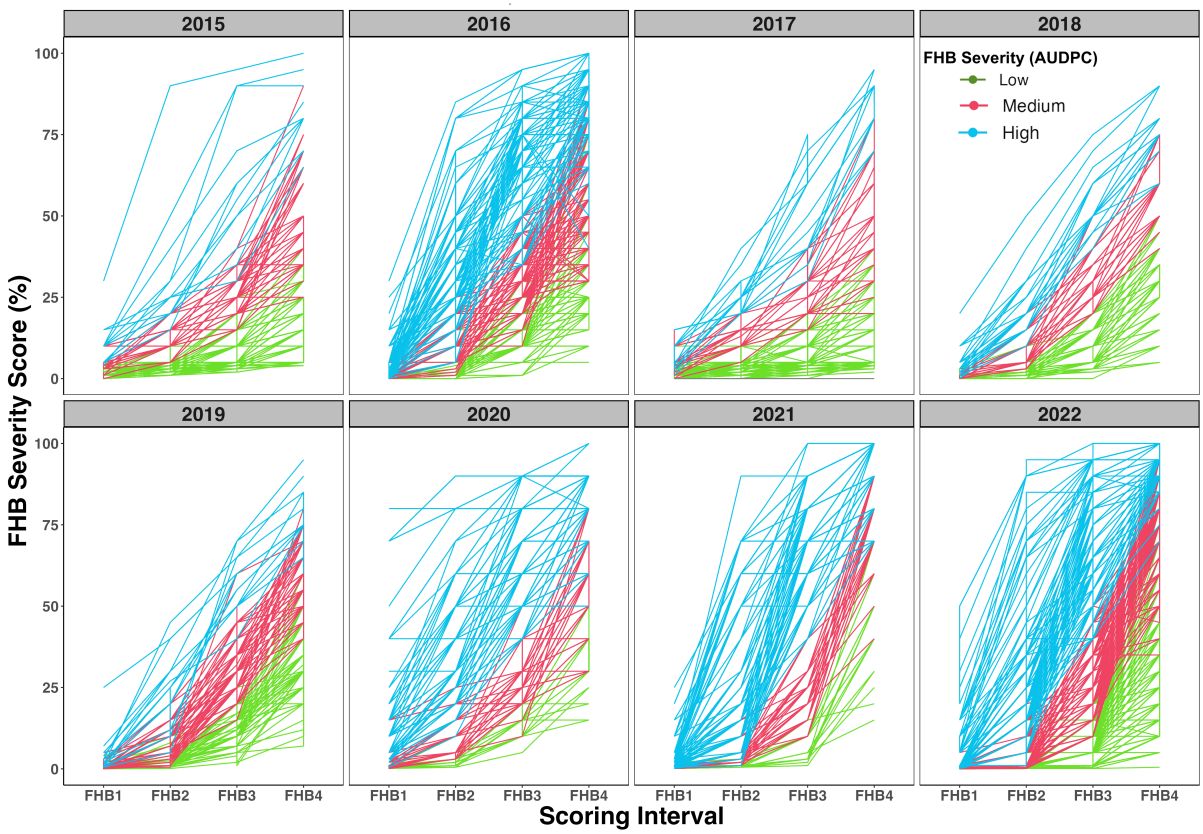

Figure 2: The figure displays eight years of FHB severity scores in the BOKU wheat breeding program. It shows four time points (FHB1 to FHB4) on the x-axis and the severity score (0-100%) on the y-axis. The lines represent the rate of change in FHB severity (AUDPC), with blue lines indicating higher severity genotypes.

4.2 Distribution of Plant Height, Anthesis and Anthesis Retention Across Years

In this study, we considered PH, AT, and AR as crucial traits acting as covariates within a two-step genomic prediction model. Concerning PH, we observed no substantial variation in mean values across the eight years of experimentation (Figure 3). However, the year 2020 stood out as it exhibited the shortest wheat plants overall. It is important to note that we did not have PH data available for the year 2016. In the remaining years, we observed a relatively normal distribution, with the majority of mean values falling within the range of 60 to 100 cm (Figure 3).

Similar to PH, AT and AR rate displayed comparable patterns. The years 2020 and 2021 showed the shortest duration from anthesis initiation to completion and the lowest rates of anthesis retention, respectively (Figure 3).

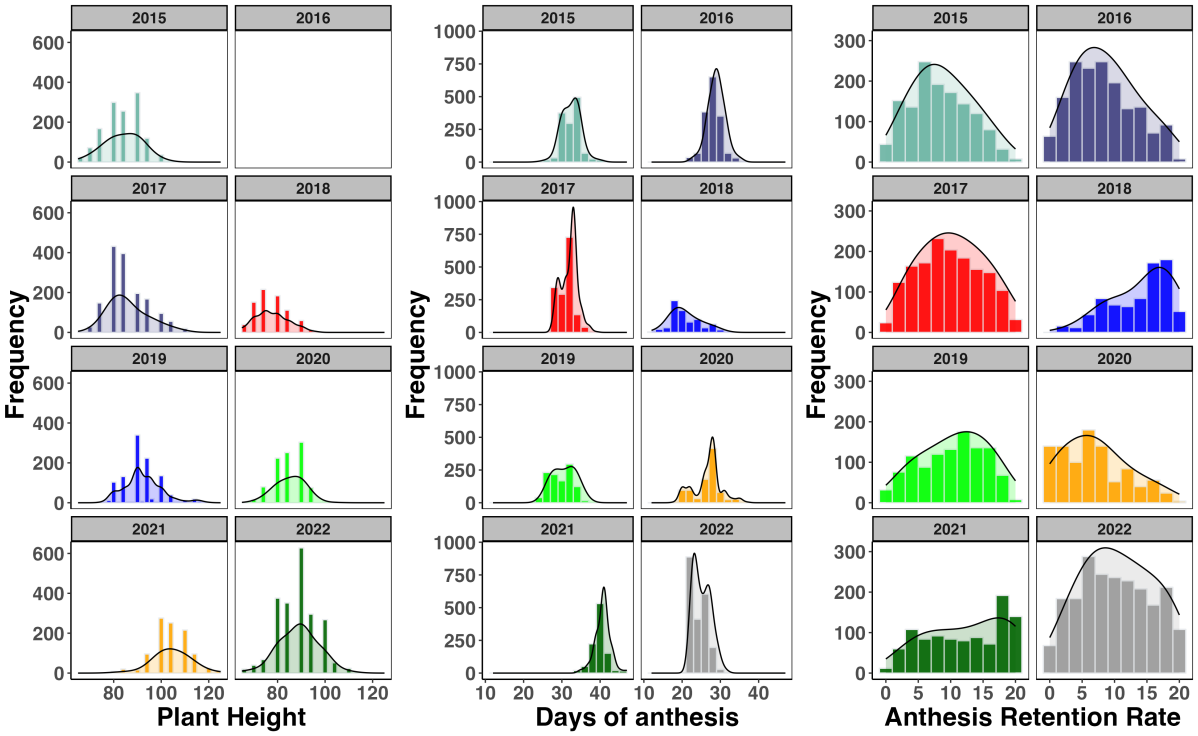

Figure 3: Variation in Plant Height, Anthesis Time, and Anthesis Retention Time across Experimental Years. (a) Plant Height: Mean values of plant height (cm) were analyzed over eight years of experimentation. (Note: Data for 2016 was unavailable) (b) Anthesis Time: The duration from anthesis initiation to completion (in days) was examined across the experimental years. (c) Anthesis Retention Time: The rate of anthesis retention was assessed over the experimental years

4.3 Correlation Of Disease Score And Co-variates

We conducted an analysis to explore the relationship between disease progression of FHB (FHB1, FHB2, FHB3, & FHB4) and three covariates: PH, AT, and AR. We utilized Spearman correlation to examine the associations among these variables. We found a negative correlation between PH and each FHB scoring, with the strength of this negative relationship slightly decreasing from FHB1 to FHB4 (Figure 4). Among the FHB scores, FHB1 showed a strong correlation (0.67) with FHB2, while we observed weaker correlations between FHB1 and FHB3 (0.41) as well as FHB4 (0.23). Similar patterns were observed when comparing FHB2 to the other traits, as well as FHB3 to FHB4, demonstrating a correlation of 0.87. Regarding the AT, we observed a weak positive correlation with FHB1 (0.35), while we found a negative correlation with FHB4 (-0.01). Notably, AR exhibited the highest positive correlation (0.42) compared to PH. AR displayed a weak positive relationship with PH (0.02) and AR (0.05). Interestingly, when assessing the relationship between AR and the FHB scores, we observed a general positive trend, with the correlation increasing from FHB1 (0.11) to FHB4 (0.3).

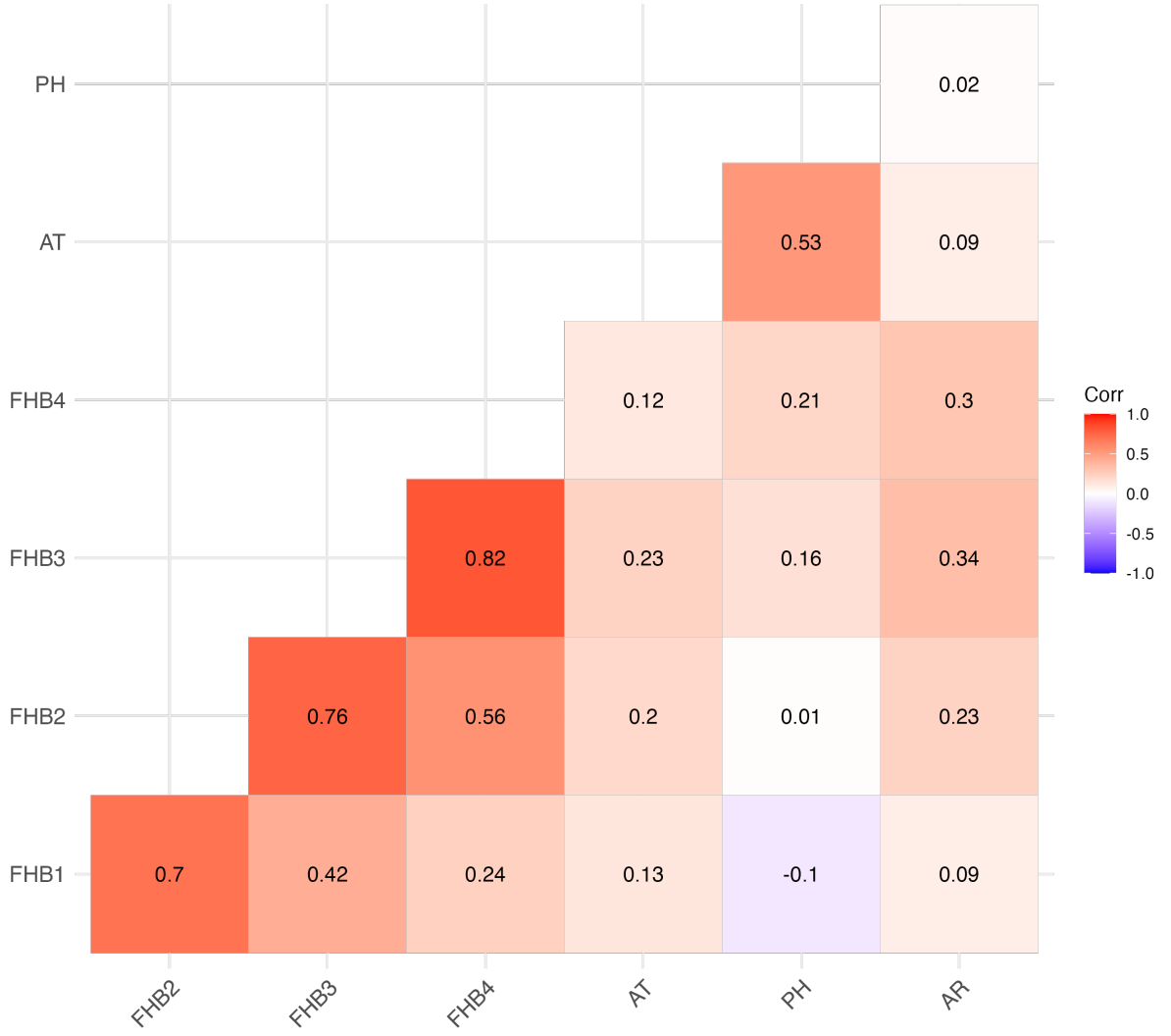

Figure 4: This figure showcases the correlation analysis between disease progression of FHB at different scoring levels and the covariates: plant height (PH), anthesis time (AT), and anthesis retention (AR). The graph illustrates the relationships observed among these variables, highlighting the trends and associations between FHB scores and the covariates

4.4 Fitting Non Linear Functions With Covariates To Generate Prediction Parameters

We performed a correlation analysis on the estimated or secondary traits derived from various the mathematical models used to simulate the progression of original FHB disease across different wheat genotypes over multiple years as shown in Figure 5. The results revealed significant correlations among the traits. The AUDPC trait, which quantifies the area under the disease progress curve, exhibited a strong negative correlation with the B parametric traits of all other models. Among them, Logistic 2B displayed the strongest negative correlation (-0.89), followed by Logistic 3B (-0.84). Conversely, AUDPC demonstrated positive correlations with other traits. Notably, strong positive associations were observed with Gompertz r (0.78), Logistic 3K (0.55), and Logistic 3r (0.34).

Another important trait, the point of inflection (inf_pnt) for each genotype, displayed noteworthy correlations. We observed a relatively strong negative correlation between inf_pnt and AUDPC (-0.61), Gamma r (-0.65), Gompertz r (-0.73), and Logistic 3K (-0.65). Additionally, inf_pnt exhibited positive relationships with Logistic 3B (0.47) and Logistic 2B (0.62).

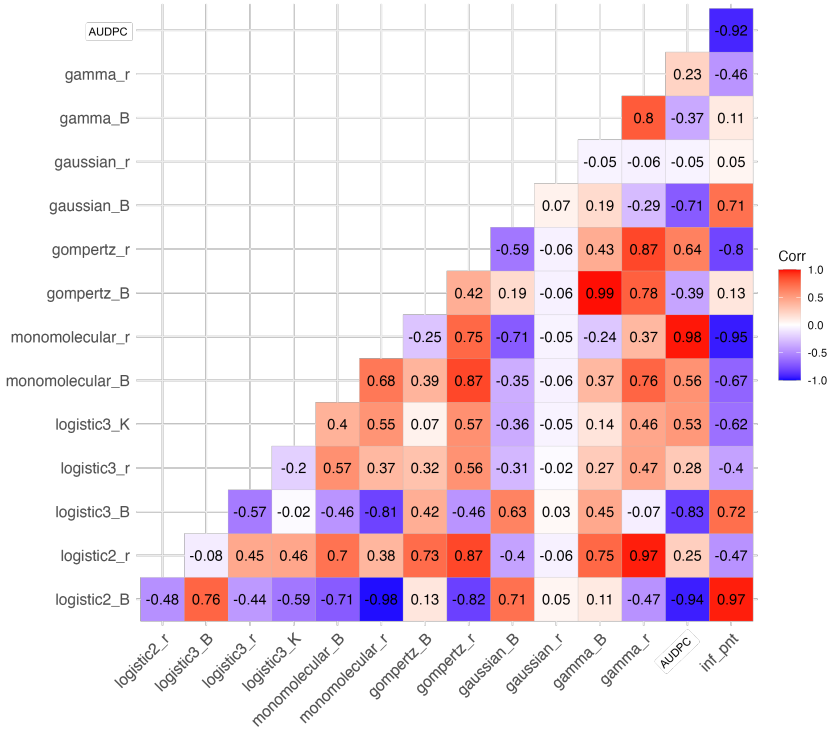

Figure 5: Correlation analysis of parametric traits derived from mathematical models simulating FHB disease progression in different wheat genotypes. The AUDPC trait shows negative correlations with B parametric traits, while positive correlations are observed with other traits. The point of inflection (inf_pnt) exhibits notable correlations, including both negative and positive associations

4.5 Goodness Of Model Fit Using Akaike Information Criterion

To assess the adequacy of non-linear functions in fitting the covariates and generating the BLUEs of FHB scores for various genotypes, we employed the Akaike Information Criterion (AIC). The AIC allowed us to identify the optimal combination of covariates that, when combined with the parameters of each function, produced the most accurate estimates. Table 3 presents the AIC values for each combination of covariates and their respective non-linear functions. By comparing the AIC values, we can determine the combination of covariates that yielded the best fit and the most reliable estimates for FHB scores. Based on the results obtained, it was found that utilizing all combinations of covariates resulted in the lowest AIC values. Therefore, for further statistical analysis, we selected AT, PH, and AR as the covariates to be incorporated. By including them in the analysis, we can better account for their effects and improve the accuracy of our statistical models.

Table 3: The figure illustrates the Akaike Information Criterion (AIC) values for different combinations of covariates, indicating the goodness of fit for each combination. The lowest AIC value represents the optimal combination of covariates that produced the most accurate estimates for FHB scores.

Covariates	Logistic2		Logistic3			Monomolecular		Gompertz		Gaussian		Gamma		AUDPC	Inflection point
	B	r	B	r	K	B	r	B	r	B	r	B	r		
None	2927	271	3716	2239	22853	-11802	-6855	7628	-1270	-771	-4191	9644	4090	25035	16860
AR	2913	273	3700	2241	22833	-11789	-6859	7625	-1268	-766	-4184	9633	4089	24992	16848
PH	2307	82	2933	1669	19389	-10084	-6069	6097	-1452	-979	-3557	7967	3066	21076	14395
PH+AR	2289	84	2918	1672	19376	-10077	-6077	6094	-1454	-974	-3552	7962	3064	21042	14387
AT	2616	248	3494	2231	22786	-11800	-7174	7612	-1340	-945	-4197	9535	4092	24533	16846
AT+AR	2605	250	3481	2233	22766	-11787	-7177	7609	-1338	-939	-4190	9525	4091	24491	16834
AT+PH	1986	35	2700	1651	19337	-10082	-6406	6074	-1521	-1162	-3562	7823	3067	20562	14376
AT+PH+AR	**1969**	**38**	**2687**	**1654**	**19325**	**-10075**	**-6413**	**6071**	**-1522**	**-1157**	**-3558**	**7819**	**3065**	**20527**	**14368**

4.6 Fitting Model Estimated Parameter

In our analysis of the estimated AUDPC for each year and FHB measurement, we utilized data from different functions, including Logistic, Gamma, Gompertz, Gaussian, and Monomolecular. The examination of the results revealed that certain functions generated a higher number of estimated fit points compared to others. This disparity can be observed by the density of lines in each function-year plot, as shown in Figure 6. Specifically, when comparing the estimated fit points generated by the Gaussian function to those produced by the Gamma and Monomolecular functions, it became apparent that the Gaussian function yielded fewer genotype AUDPC fits.

In addition to the analysis of the number of estimated fit points, we also examined the shape of the fitted curves. Our objective was to generate an S-shaped or sigmoid curve that accurately emulates the natural life-cycle of the FHB-causing agent's effect on wheat plants (Figure 6). The sigmoid curve reflects the disease progression, starting with a period of low incidence during the early growth stages of the crop. As the crop reaches the susceptible stage, the disease intensity rapidly increases, resulting in an exponential growth in severity. This phase is characterized by the pathogen's rapid spread and a higher rate of infection. The disease reaches its peak severity during the middle phase of the sigmoid curve, with the majority of infections occurring at this stage. Prominent symptoms, such as bleaching of spikelets and fungal growth, become evident. Following the peak, the disease gradually diminishes, leading to a flattening of the sigmoid curve, indicating a decrease in disease severity. Finally, the disease incidence becomes minimal, signifying a low level of spread, and the severity decreases significantly. Based on these observations, it is noteworthy that almost all the fitted curves for the different functions exhibited this sigmoid behavior, except for the Monomolecular function (Figure 6). This indicates that the chosen functions effectively capture the disease progression over time. Furthermore, in Figure 6, the green color represents genotypes with smaller AUDPC values, indicating lower disease severity, while the blue color represents genotypes with larger AUDPC values, indicating higher disease severity. This color scheme allows for easy visual differentiation and comparison of genotype performance in terms of disease severity.

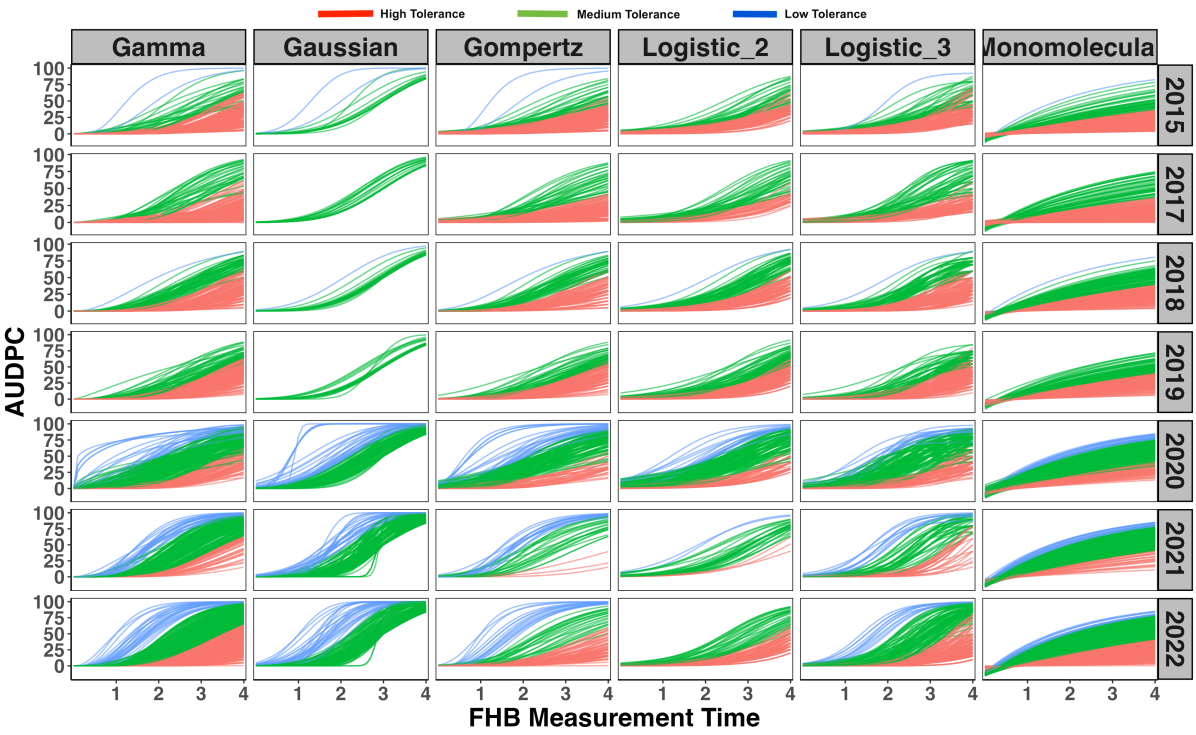

Figure 6: Comparison of estimated fit points and fitted curves for various functions used to model Fusarium head blight (FHB) disease progression. The density of lines in each function-year plot demonstrates variations in the number of estimated fit points, with the Gaussian function generating fewer genotype AUDPC fits compared to the Gamma and Monomolecular functions. The sigmoid shape of the fitted curves accurately represents the natural life-cycle of the FHB-causing agent's impact on wheat plants, except for the Monomolecular function. Genotypes with lower disease severity (smaller AUDPC values) are depicted in green, while genotypes with higher severity (larger AUDPC values) are shown in blue.

4.7 Model Predictability Accuracy

The predictability performance of various traits, including AUDPC, Logistics (2 and 3 - B, K, and r), inflection points, Gompertz (B and r), and Gamma (B and r), was assessed using multiple prediction models, including Bayesians, XgBoost, and GBLUP (Figure7). Among the traits evaluated, the best prediction accuracies were observed for AUDPC, Gompertz r, and Logistic B (two-parameter scenario) when using the BayesA and BayesB models. In terms of AUDPC, the BayesB model achieved the highest prediction accuracy of 0.58, closely followed by the BayesA model with a prediction accuracy of 0.52 (Figure7). The Bayesian Lasso model exhibited the lowest predictability for this trait, with a prediction accuracy of 0.49. It is important to note that the predictive abilities for AUDPC ranged from 0.49 to 0.58 across all models.

For the Gamma B trait, the predictability accuracies were relatively lower compared to AUDPC. The BayesB model and XgBoost demonstrated notable results with prediction accuracies of 0.35 and 0.34, respectively. Conversely, the Bayesian Lasso model had the lowest predictability for this trait, with an accuracy of 0.2. Regarding the Gamma r trait, most prediction models yielded relatively good results around 0.45, except for Bayesian Lasso and XgBoost, which exhibited lower predictability abilities of 0.40 (Figure7).

However, it is important to highlight that certain traits showed poor predictability across all models. Specifically, the prediction of Gompertz B trait resulted in low accuracy levels across all models, hovering around 0.1. XgBoost performed slightly better than the other models for this trait. Similarly, Logistic 2r, Logistic 3r, and inflection points exhibited prediction levels below 0.2 across all models (Figure7). These findings indicate variations in predictability among different traits and emphasize the importance of selecting appropriate prediction models based on the specific trait being predicted.

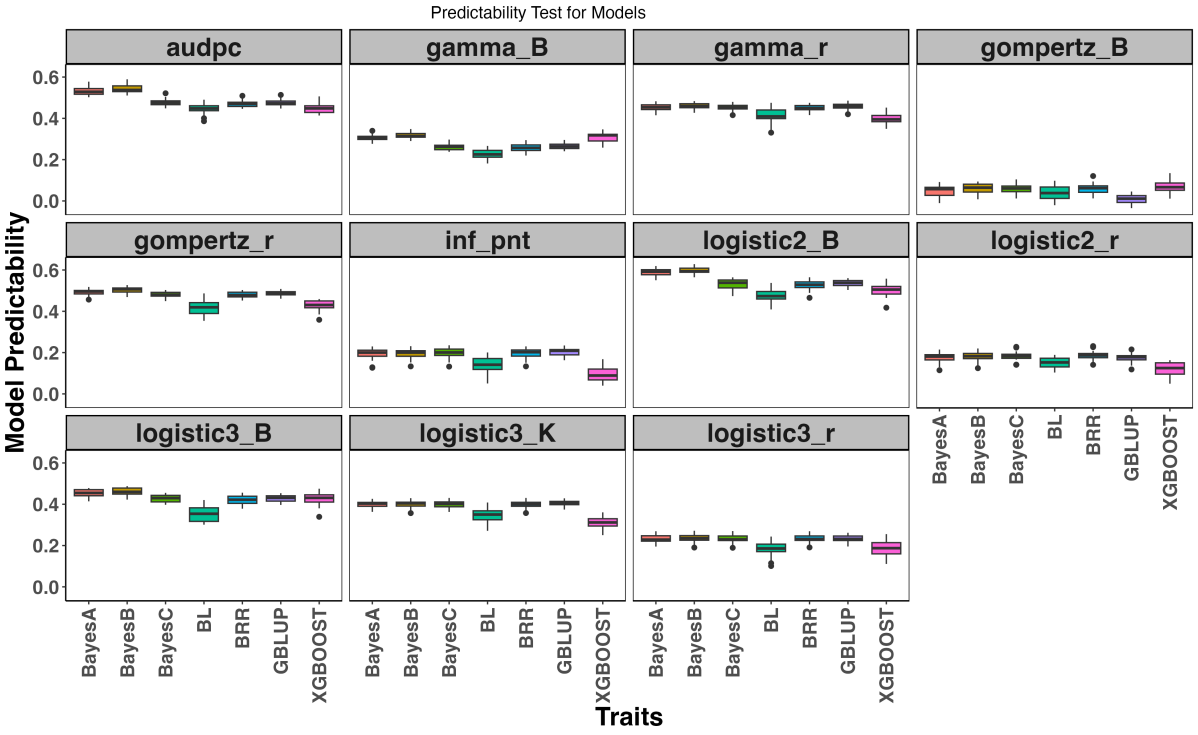

Figure 7: Model predictability of fitted parameters in the study. Each color indicates a different model. BRR: Bayesian Ridge Regression, XGBoost: Extreme Gradient Boosting.

4.8 Model Predictability Validation

We conducted cross-validation to determine top-performing genotypes for each trait relative to AUDPC. Positive correlations with AUDPC were observed for Gamma r, Logistic 2r, Logistic 3k, Logistic 3r, and Gompertz r, while the remaining traits exhibited negative correlations Figure 8. No definitive correlation pattern was discernible for Gompertz B.

Genotypes were selected based on their AUDPC correlation, drawing a cutoff to isolate the lowest 20% for positively correlated traits and the highest 80% for negatively correlated ones. This method facilitated the identification of the most performant genotypes relative to AUDPC for each trait.

The GBLUP model revealed substantial genotype overlap for traits including Gamma r, Gompertz r, Logistic 3B, Logistic 3K, Logistic 3r, inflection point, and Logistic 2B. Contrastingly, Logistic 2r and Gamma B showed minimal overlap, with Gompertz B presenting no intersecting genotypes (Figure 8)..

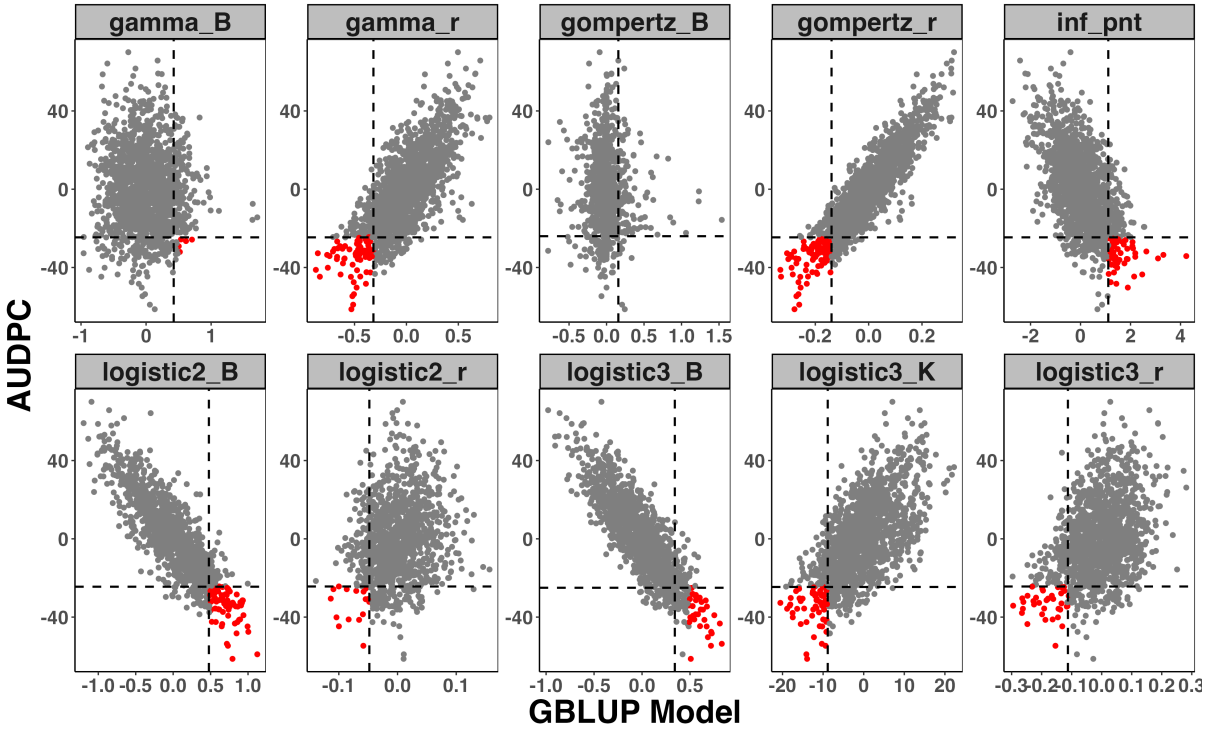

Figure 8: Genotype selection based on AUDPC correlation. The plot illustrates the chosen cutoffs that isolate the bottom 20% of genotypes for positively correlated traits (indicated in red), and the top 80% for negatively correlated ones (shown in red). This selection process enabled identification of the most performant genotypes relative to AUDPC for each trait using the GBLUP model

Conducting a parallel analysis using predictions from the XGBoost model resulted in consistent findings with the previous model Figure 9. All traits displayed intersecting individuals with the same correlation trend as previously observed. However, it is noteworthy that the total count of identified high-performing genotypes in this analysis was comparatively smaller than in the previously described model (Figure 8).

In essence, traits with fewer high-performing genotypes consistently intersected in both models. Similarly, traits with larger intersections in the prior model continued this trend in the XGBoost analysis, albeit with fewer qualifying individuals. While correlation trends and genotype intersections with AUDPC largely paralleled across both models, the XGBoost model discerned fewer high-performing genotypes per trait. This may hint at more rigorous selection criteria inherent to the XGBoost model, thereby yielding a more selective set of top performers.

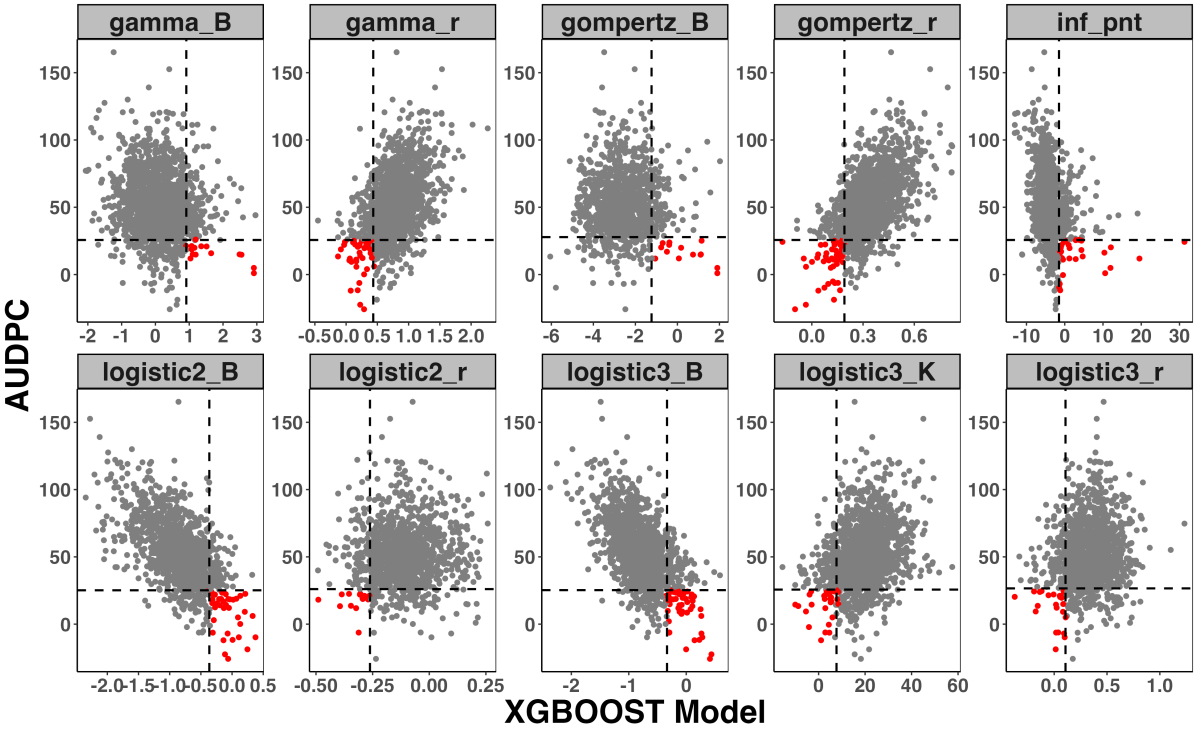

Figure 9: Selection of genotypes via AUDPC correlation employing the XGBoost model. The graphic demonstrates the applied cutoffs, isolating the lowermost 20% of genotypes for traits with positive correlations (depicted in red), and the uppermost 80% for traits showing negative correlations (also portrayed in red). This procedure facilitated the identification of superior genotypes in relation to AUDPC for each distinct trait.

5 Discussion

5.1 Implications of Trait Variability and Correlations

This study highlighted the variations in FHB severity within the BOKU wheat breeding program over eight years, pinpointing the years 2016, 2020, 2021, and 2022 as showing higher FHB severity (Figure 2). This aligns with prior research that emphasizes the critical role of environmental factors in FHB presence, especially during the wheat flowering phase (Moreno-Amores et al. 2020, Juroszek & von Tiedemann 2013, Buerstmayr et al. 2020). Plant height, anthesis time, and anthesis retention were considered covariates in our research, shaping the predictability of the genomic models used. Notably, plant height showed marginal fluctuation across the period, except for 2020, which had the shortest wheat plants, implying possible environmental or genetic influences (Hilton et al. 1999, Miedaner et al. 2023, Marino 2018, Kalih et al. 2014).

The years 2020 and 2021 showed similarities in terms of anthesis time and anthesis retention (Figure 3). In 2020, there was a noticeable decrease in plant height, anthesis duration, and anthesis retention, suggesting potential environmental impacts on these variables (Marino 2018). This reduction suggests a hastened flowering phase and potential challenges with grain maturation and maintaining yield following the anthesis stage. These observations support findings from previous studies, underscoring the role of environmental factors in disease progression and severity.

Our analysis revealed a negative correlation between plant height and each FHB severity score (Figure 4), suggesting that plant height, along with other agronomic and morphological traits, plays a significant role in influencing FHB severity. This is consistent with previous findings (Hilton et al. 1999, Kalih et al. 2014, Miedaner et al. 2023, Marino 2018, Moidu et al. 2015, Brown et al. 2010). For instance, taller plants with elongated floral structures might hinder fungal invasion into spikelets and possess superior overall health, thereby increasing their resistance to pathogens (Buerstmayr et al. 2020). Conversely, shorter plants might be more susceptible to FHB, possibly due to factors such as canopy structure and overall plant health. It is important to highlight that this is not a pleiotropy effect, where the same genes control both plant height and FHB resistance, but a strategic escape strategy where taller plants avoid FHB infection, creating a correlation between the two traits (Solovieff et al. 2013).

Correlation analysis of FHB scores at different time points helped elucidate the progression dynamics of FHB. There was a robust positive correlation between FHB1 and FHB2, implying a steady disease progression from one time point to the next (Kollers et al. 2013, Chrpová et al. 2021). On the other hand, weaker correlations between FHB1 and FHB3 and between FHB1 and FHB4 suggest potential deviations in disease progression at these later time points. Factors such as changing environmental conditions, varying pathogen activity, or the plant's defense mechanisms could contribute to these variations (Chrpová et al. 2021, Buerstmayr et al. 2003).

Lastly, we observed interesting correlations between the anthesis rate and FHB severity at different measurement points. A mild positive correlation between AT and FHB1 and a negative correlation between AT and FHB4 suggest how anthesis rate influences disease severity (Buerstmayr et al. 2003, Cowger et al. 2009, Buerstmayr et al. 2020). Furthermore, our analysis indicated that anthesis retention showed the strongest positive correlation among the traits investigated, establishing a substantial relationship with FHB severity. This implies that genotypes with better anthesis retention may experience lower FHB severity (Buerstmayr et al. 2003, Cowger et al. 2009, Buerstmayr et al. 2020).

5.2 Analysis of Non-Linear Models for Disease Progression Curve Parameters

We primarily concentrated on two-parameter models in our study, where 'B' signifies the x-axis parameter (e.g., Inflection Point, Gaussian Distribution Mean), and 'r' is the shape parameter (e.g., inflection point slope, Gaussian variance). For breeders, these parameters offer significant interpretability, particularly within certain non-linear models like logistics. The 'B' parameter, i.e. the inflection point, symbolizes a pivotal stage in the FHB progression curve, marking the peak acceleration in disease development rate (Goshu & Koya 2013). Differences in inflection points can imply varying FHB tolerance among genotypes, hence facilitating the selection of FHB-resistant lines (Zhao et al. 2004, Kuska et al. 2017, Garcia-Abadillo et al. 2022). Moreover, positive associations between inflection points and both Logistic 3B and Logistic 2B points towards a decelerated disease progression as inflection point values augment (Garcia-Abadillo et al. 2022). The 'r' parameter, representative of each model's growth rate, demonstrates a strong

positive correlation with the AUDPC. This suggests that higher disease growth rates correspond with elevated AUDPC values, meaning genotypes with faster disease growth rates are more vulnerable to FHB (Jeger & Viljanen-Rollinson 2001, Draeger et al. 2007, Yuen & Forbes 2009, Karanja et al. 2018, Amrate et al. 2023). Unique to the Logistic 3 model, the 'K' parameter interpreted as a saturation parameter for growth models where resources are constrained, 'K' represents a scenario where the disease does not reach the full scale of FHB measurements, signifies the limit of pathogen population size. Its correlation with AUDPC suggests that a higher pathogen growth limit aligns with a more severe disease manifestation in the wheat genotype, aiding in the detection of more susceptible genotypes (Lamb & Loschiavo 1981, Steffenson & Webster 1992).

By further analyzing our research findings, it becomes evident that the Gaussian function, which exhibits symmetry with respect to the inflection point, produces a reduced number of genotype AUDPC fits when compared to the Gamma and Monomolecular functions, as showed in Figure 6. This discrepancy can be attributed to the nature of the Gaussian probability function, which is symmetric with respect to the mean, while the cumulative density function, obtained through the integration of the bell-shaped curve of the data, is symmetric with respect to the inflection point representing the mean of zero in the standard normal distribution (Limpert et al. 2001, Stahl 2006) (Figure 6). For the rest of most functions, they exhibited an S-shaped sigmoidal curve, emulating the natural life-cycle effect of the FHB-causing agent, with a notable exception of the Monomolecular function due to its lack of an inflection point and outlines a drawback to using Monomolecular model to detect traits such as disease latency period or point of maximum growth (Cambaza 2018).

5.3 Predictability of Metrics: Evaluating Performance

To further explore the relationship between genotypes and traits, we undertook cross-validation, focusing on identifying top-performing genotypes relative to AUDPC. We noted positive correlations with AUDPC for our estimated or secondary derived traits such as Gamma r, Logistic 2r, Logistic 3k, Logistic 3r, and Gompertz r which are representative estimations of actual FHB field scores, whereas other traits showed negative correlations (Figure 8). We selected genotypes based on their AUDPC correlation, employing a cutoff to capture the bottom 20% for positively correlated traits, and the top 80% for negatively correlated traits. As proposed by Schrauf et al.(2021), this approach facilitated the identification of the most effective genotypes for each trait (Figure 8). Despite certain traits, such as Logistic 2B and 3B, demonstrating strong correlation with AUDPC across various prediction models, others consistently exhibit low predictability. This is largely attributed to the increased estimation error in these traits, often overlooked by the markers. For instance, poor estimations of a specific secondary trait, like Gompertz B, can lead to lower prediction accuracies, regardless of the genomic selection model's fit, corroborating the situations reported by (Mostafavi et al. 2020) (Figure 7). In addition, this study indicates that most Bayesian models outperform the standard GBLUP model, primarily due to their ability to detect significant genetic effects by presuming the presence of major genes. These results align with findings from similar research conducted by (Colombani et al. 2013).

Furthermore, this study identifies parameters that optimally align with AUDPC predictions, suggesting their potential as selection alternatives. Specifically, lines with higher values for traits such as inflection point, Logistic B, Logistic K, and Gompertz r are desirable for lower AUDPC. By incorporating these traits, we can utilize them for growth curve prediction and inform selection decisions. Notably, these selected parameters harmonize with conventional AUDPC-driven selection, providing an efficient complementary method to boost genotype performance (Madden et al. 1987).

6 Conclusion

This study highlights the significant influence of environmental conditions on FHB severity, as well as the crucial roles of plant height, anthesis time, and anthesis retention. The traits identified suggest that taller wheat genotypes with elongated floral structures show enhanced FHB resistance. These findings, combined with the identification of key non-linear model parameters like Inflection Point and growth rate, provide valuable tools for breeders. Additionally, the superior performance of Bayesian models in detecting genetic effects underscores their potential in improving wheat breeding programs. Incorporation of these models with estimated BLUES from fitting non-linear functions could enhance the detection of FHB-resistant genotypes. However, the variability in trait predictability due to prediction and estimation errors underscores the necessity for cautious interpretation and application of certain traits. The insights gained from this study offer important implications for breeding strategies, particularly in the selection of FHB-resistant wheat genotypes.

7 Bibliography

Aggarwal, P. K. (2008), 'Global climate change and indian agriculture: impacts, adaptation and mitigation', *Indian Journal of Agricultural Sciences* **78**(11), 911.

Almoujahed, M. B., Rangarajan, A. K., Whetton, R. L., Vincke, D., Eylenbosch, D., Vermeulen, P. & Mouazen, A. M. (2022), 'Detection of fusarium head blight in wheat under field conditions using a hyperspectral camera and machine learning', *Computers and Electronics in Agriculture* **203**, 107456.

Amrate, P. K., Shrivastava, M., Bhale, M., Agrawal, N., Kumawat, G., Shivakumar, M. & Nataraj, V. (2023), 'Identification and genetic diversity analysis of high-yielding charcoal rot resistant soybean genotypes', *Scientific Reports* **13**(1), 8905.

Andrews, D. W. & Mallows, C. L. (1974), 'Scale mixtures of normal distributions', *Journal of the Royal Statistical Society: Series B (Methodological)* **36**(1), 99–102.

Arruda, M. P., Brown, P., Brown-Guedira, G., Krill, A. M., Thurber, C., Merrill, K. R., Foresman, B. J. & Kolb, F. L. (2016), 'Genome-wide association mapping of fusarium head blight resistance in wheat using genotyping-by-sequencing', *The Plant Genome* **9**(1), plantgenome2015–04.

Ashraf, R. (2021), 'Utilization of wheat relatives to improve wheat breeding for rust resistance'.

Bari, A., Street, K., Mackay, M., Endresen, D. T. F., De Pauw, E. & Amri, A. (2012), 'Focused identification of germplasm strategy (figs) detects wheat stem rust resistance linked to environmental variables', *Genetic Resources and Crop Evolution* **59**, 1465–1481.

Berger, R. (1981), 'Comparison of the gompertz and logistic equations to describe plant disease progress.', *Phytopathology* **71**(7), 716–719.

Bernhoft, A., Wang, J. & Leifert, C. (2022), 'Effect of organic and conventional cereal production methods on fusarium head blight and mycotoxin contamination levels', *Agronomy* **12**(4), 797.

Blum, M. G. & François, O. (2010), 'Non-linear regression models for approximate bayesian computation', *Statistics and computing* **20**, 63–73.

Boopathi, N. M. & Boopathi, N. M. (2020), 'Marker-assisted selection (mas)', *Genetic Mapping and Marker Assisted Selection: Basics, Practice and Benefits* pp. 343–388.

Brown, N. A., Urban, M., Van de Meene, A. M. & Hammond-Kosack, K. E. (2010), 'The infection biology of fusarium graminearum: defining the pathways of spikelet to spikelet colonisation in wheat ears', *Fungal biology* **114**(7), 555–571.

Buerstmayr, H., Steiner, B., Hartl, L., Griesser, M., Angerer, N., Lengauer, D., Miedaner, T., Schneider, B. & Lemmens, M. (2003), 'Molecular mapping of qtls for fusarium head blight resistance in spring wheat. ii. resistance to fungal penetration and spread', *Theoretical and Applied Genetics* **107**, 503–508.

Buerstmayr, M., Steiner, B. & Buerstmayr, H. (2020), 'Breeding for fusarium head blight resistance in wheat—progress and challenges', *Plant Breeding* **139**(3), 429–454.

Cambaza, E. (2018), 'Which microbial growth model best fits to fusarium graminearum?'.

Cao, L., Shi, P.-J., Li, L. & Chen, G. (2019), 'A new flexible sigmoidal growth model', *Symmetry* **11**(2), 204.

Carrillo, M. & González, J. M. (2002), 'A new approach to modelling sigmoidal curves', *Technological Forecasting and Social Change* **69**(3), 233–241.

Cheng, B., Gao, X., Cao, N., Ding, Y., Gao, Y., Chen, T., Xin, Z. & Zhang, L. (2020), 'Genome-wide association analysis of stripe rust resistance loci in wheat accessions from southwestern china', *Journal of Applied Genetics* **61**, 37–50.

Chrpová, J., Grausgruber, H., Weyermann, V., Buerstmayr, M., Palicová, J., Kozová, J., Trávníčková, M., Nguyen, Q. T., Moreno Amores, J. E., Buerstmayr, H. et al. (2021), 'Resistance of winter spelt wheat [triticum aestivum subsp. spelta (l.) thell.] to fusarium head blight', *Frontiers in Plant Science* **12**, 661484.

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R. & McCouch, S. (2013), 'Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement', *Theoretical and Applied Genetics* **126**, 867–887.

Colombani, C., Legarra, A., Fritz, S., Guillaume, F., Croiseau, P., Ducrocq, V. & Robert-Granié, C. (2013), 'Application of bayesian least absolute shrinkage and selection operator (lasso) and bayescπ methods for genomic selection in french holstein and montbéliarde breeds', *Journal of Dairy Science* **96**(1), 575–591.

Cowger, C., Patton-Özkurt, J., Brown-Guedira, G. & Perugini, L. (2009), 'Post-anthesis moisture increased fusarium head blight and deoxynivalenol levels in north carolina winter wheat', *Phytopathology* **99**(4), 320–327.

Cromey, M., Lauren, D., Parkes, R., Sinclair, K., Shorter, S. & Wallace, A. (2001), 'Control of fusarium head blight of wheat with fungicides', *Australasian Plant Pathology* **30**(4), 301–308.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y. et al. (2017), 'Genomic selection in plant breeding: methods, models, and perspectives', *Trends in plant science* **22**(11), 961–975.

Draeger, R., Gosman, N., Steed, A., Chandler, E., Thomsett, M., Srinivasachary, Schondelmaier, J., Buerstmayr, H., Lemmens, M., Schmolke, M. et al. (2007), 'Identification of qtls for resistance to fusarium head blight, don accumulation and associated traits in the winter wheat variety arina', *Theoretical and applied genetics* **115**, 617–625.

Dyba, T., Hakulinen, T. & Päivärinta, L. (1997), 'A simple non-linear model in incidence prediction', *Statistics in medicine* **16**(20), 2297–2309.

Edwards, S. G. (2022), 'Pydiflumetofen co-formulated with prothioconazole: a novel fungicide for fusarium head blight and deoxynivalenol control', *Toxins* **14**(1), 34.

faostat (2021), 'FAOSTAT Database'. Accessed: April 12, 2023.
URL: *http://www.fao.org/faostat/en/data/QC/*

Garcia-Abadillo, J., Morales, L., Buerstmayr, H., Michel, S., Lillemo, M., Holzapfel, J., Hartl, L., Akdemir, D., Carvalho, H. F. & Isidro-Sánchez, J. (2022), 'Alternative scoring methods of fusarium head blight resistance for genomic assisted breeding', *Frontiers in Plant Science* **13**.

Gazal, A., Dar, Z., Wani, S., Lone, A., Shikari, A., Ali, G. & Abidi, I. (2016), 'Molecular breeding for enhancing resilience against biotic and abiotic stress in major cereals.', *SABRAO Journal of Breeding & Genetics* **48**(1).

Gianola, D., de los Campos, P., Hill, G. & Manfredi, E. (2014), 'Additive genetic variability and the bayesian alphabet', *Genetics* **198**(2), 497–508.

González-Camacho, J., de Los Campos, G., Pérez, P., Gianola, D., Cairns, J., Mahuku, G., Babu, R. & Crossa, J. (2012), 'Genome-enabled prediction of genetic values using radial basis function neural networks', *Theoretical and Applied Genetics* **125**, 759–771.

González-Camacho, J. M., Ornella, L., Pérez-Rodríguez, P., Gianola, D., Dreisigacker, S. & Crossa, J. (2018), 'Applications of machine learning methods to genomic selection in breeding wheat for rust resistance', *The plant genome* **11**(2), 170104.

Goshu, A. T. & Koya, P. R. (2013), 'Derivation of inflection points of nonlinear regression curves-implications to statistics', *Am. J. Theor. Appl. Stat* **2**(6), 268–272.

Gosman, N., Bayles, R., Jennings, P., Kirby, J. & Nicholson, P. (2007), 'Evaluation and characterization of resistance to fusarium head blight caused by fusarium culmorum in uk winter wheat cultivars', *Plant Pathology* **56**(2), 264–276.

Goswami, R. S., Xu, J.-R., Trail, F., Hilburn, K. & Kistler, H. C. (2006), 'Genomic analysis of host–pathogen interaction between fusarium graminearum and wheat during early stages of disease development', *Microbiology* **152**(6), 1877–1890.

Gunupuru, L., Patel, J., Sumarah, M., Renaud, J., Mantin, E. & Prithiviraj, B. (2019), 'A plant biostimulant made from the marine brown algae ascophyllum nodosum and chitosan reduce fusarium head blight and mycotoxin contamination in wheat', *PLoS One* **14**(9), e0220562.

Heffner, E. L., Sorrells, M. E. & Jannink, J.-L. (2009), 'Genomic selection for crop improvement', *Crop Science* **49**(1), 1–12.

Heslot, N., Akdemir, D., Sorrells, M. E. & Jannink, J.-L. (2014), 'Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions', *Theoretical and applied genetics* **127**, 463–480.

Heslot, N., Jannink, J.-L. & Sorrells, M. E. (2015), 'Perspectives for genomic selection applications and research in plants', *Crop Science* **55**(1), 1–12.

Heslot, N., Yang, H.-P., Sorrells, M. E. & Jannink, J.-L. (2012), 'Genomic selection in plant breeding: a comparison of models', *Crop science* **52**(1), 146–160.

Hilton, A., Jenkinson, P., Hollins, T., Parry, D. et al. (1999), 'Relationship between cultivar height and severity of fusarium ear blight in wheat.', *Plant Pathology* **48**(2), 202–208.

Howard, R., Carriquiry, A. L. & Beavis, W. D. (2014), 'Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures', *G3: Genes, Genomes, Genetics* **4**(6), 1027–1046.

Huang, L., Wu, Z., Huang, W., Ma, H. & Zhao, J. (2019), 'Identification of fusarium head blight in winter wheat ears based on fisher's linear discriminant analysis and a support vector machine', *Applied Sciences* **9**(18), 3894.

Isidro, J., Akdemir, D. & Burke, J. (2016), 'Genomic selection', *The world wheat book: a history of wheat breeding* **3**, 1001–1023.

Jannink, J.-L., Lorenz, A. J. & Iwata, H. (2010), 'Genomic selection in plant breeding: from theory to practice', *Briefings in functional genomics* **9**(2), 166–177.

Jeger, M. & Viljanen-Rollinson, S. (2001), 'The use of the area under the disease-progress curve (audpc) to assess quantitative disease resistance in crop cultivars', *Theoretical and Applied Genetics* **102**, 32–40.

Jeon, D., Kang, Y., Lee, S., Choi, S., Sung, Y., Lee, T.-H. & Kim, C. (2023), 'Digitalizing breeding in plants: A new trend of next-generation breeding based on genomic prediction', *Frontiers in Plant Science* **14**, 1092584.

Jiang, Y., Schulthess, A. W., Rodemann, B., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G. et al. (2017), 'Validating the prediction accuracies of marker-assisted and genomic selection of fusarium head blight resistance in wheat using an independent sample', *Theoretical and applied genetics* **130**, 471–482.

Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Rutkoski, J. E., Poland, J. A., Bergstrom, G. C. & Sorrells, M. E. (2017), 'Comparison of models and whole-genome profiling approaches for genomic-enabled prediction of septoria tritici blotch, stagonospora nodorum blotch, and tan spot resistance in wheat', *The Plant Genome* **10**(2), plantgenome2016–08.

Juroszek, P. & von Tiedemann, A. (2013), 'Climate change and potential future risks through wheat diseases: a review', *European Journal of Plant Pathology* **136**, 21–33.

Kage, U., Yogendra, K. N. & Kushalappa, A. C. (2017), 'Tawrky70 transcription factor in wheat qtl-2dl regulates downstream metabolite biosynthetic genes to resist fusarium graminearum infection spread within spike', *Scientific Reports* **7**(1), 1–14.

Kalih, R., Maurer, H. P., Hackauf, B. & Miedaner, T. (2014), 'Effect of a rye dwarfing gene on plant height, heading stage, and fusarium head blight in triticale (× triticosecale wittmack)', *Theoretical and Applied Genetics* **127**, 1527–1536.

Karanja, J., Derera, J., Gubba, A., Mugo, S. & Wangai, A. (2018), 'Response of selected maize inbred germplasm to maize lethal necrosis disease and its causative viruses (sugarcane mosaic virus and maize chlorotic mottle virus) in kenya', *The Open Agriculture Journal* **12**(1).

Kollers, S., Rodemann, B., Ling, J., Korzun, V., Ebmeyer, E., Argillier, O., Hinze, M., Plieske, J., Kulosa, D., Ganal, M. W. et al. (2013), 'Whole genome association mapping of fusarium head blight resistance in european winter wheat (triticum aestivum l.)', *PLoS One* **8**(2), e57500.

Kramer, M., Vera-Álvarez, C., van der Linden, P., Boer, M. & Veerkamp, R. (2014), 'An extended genomic relationship matrix to increase the reliability of genomic breeding values', *Journal of Dairy Science* **97**(11), 6874–6883.

Kuska, M. T., Brugger, A., Thomas, S., Wahabzada, M., Kersting, K., Oerke, E.-C., Steiner, U. & Mahlein, A.-K. (2017), 'Spectral patterns reveal early resistance reactions of barley against blumeria graminis f. sp. hordei', *Phytopathology* **107**(11), 1388–1398.

Lamb, R. & Loschiavo, S. (1981), 'Diet, temperature, and the logistic model of developmental rate for tribolium confusum (coleoptera: Tenebrionidae) 1', *The Canadian Entomologist* **113**(9), 813–818.

Lande, R. & Thompson, R. (1990), 'Efficiency of marker-assisted selection in the improvement of quantitative traits.', *Genetics* **124**(3), 743–756.

Larkin, D. L., Lozada, D. N. & Mason, R. E. (2019), 'Genomic selection—considerations for successful implementation in wheat breeding programs', *Agronomy* **9**(9), 479.

Lemmens, M., Steiner, B., Sulyok, M., Nicholson, P., Mesterhazy, A. & Buerstmayr, H. (2016), 'Masked mycotoxins: does breeding for enhanced fusarium head blight resistance result in more deoxynivalenol-3-glucoside in new wheat varieties?', *World Mycotoxin Journal* **9**(5), 741–754.

Li, L., Dong, Y., Xiao, Y., Liu, L., Zhao, X. & Huang, W. (2022), 'Combining disease mechanism and machine learning to predict wheat fusarium head blight', *Remote Sensing* **14**(12), 2732.

Limpert, E., Stahel, W. A. & Abbt, M. (2001), 'Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question', *BioScience* **51**(5), 341–352.

Machado, A. K., Brown, N. A., Urban, M., Kanyuka, K. & Hammond-Kosack, K. E. (2018), 'Rnai as an emerging approach to control fusarium head blight disease and mycotoxin contamination in cereals', *Pest management science* **74**(4), 790–799.

Madden, L. et al. (1987), 'Potential effects of air pollutants on epidemics of plant diseases', *Agriculture, ecosystems & environment* **18**(3), 251–262.

Makowsky, R., Pajewski, L., Klimentidis, M., Vazquez, Y., Duarte, C. & Allison, M. (2011), 'Beyond missing heritability: Prediction of complex traits', *PLoS Genetics* **7**(4), e1002051.

Marino, T. P. (2018), *Genomic Prediction and QTL Validation for Resistance to Fusarium Ear Rot and Fumonisin in Maize*, North Carolina State University.

Mesterhazy, A. (1995), 'Types and components of resistance to fusarium head blight of wheat', *Plant breeding* **114**(5), 377–386.

Mesterhazy, A., Bart'ok, T., Mirocha, C. & Komor'oczy, R. (2000), 'Types and components of resistance to fusarium head blight of wheat', *Plant Breeding* **119**, 1–16.

Meuwissen, T. H., Hayes, B. J. & Goddard, M. (2001), 'Prediction of total genetic value using genome-wide dense marker maps', *genetics* **157**(4), 1819–1829.

Miedaner, T., Flamm, C. & Oberforster, M. (2023), 'The importance of fusarium head blight resistance in the cereal breeding industry: Case studies from germany and austria', *Plant Breeding* .

Moidu, H., Brownlee, J., Wang, X., Deschiffart, I., Langille, L., Voldeng, H. & Khanizadeh, S. (2015), 'Effect of plant height on fusarium head blight in spring wheat', *Journal of Plant Studies; Vol* **4**(2).

Momen, M., Mehrgardi, A. A., Sheikhi, A., Kranis, A., Tusell, L., Morota, G., Rosa, G. J. & Gianola, D. (2018), 'Predictive ability of genome-assisted statistical models under various forms of gene action', *Scientific reports* **8**(1), 12309.

Moreno-Amores, J., Michel, S., Löschenberger, F. & Buerstmayr, H. (2020), 'Dissecting the contribution of environmental influences, plant phenology, and disease resistance to improving genomic predictions for fusarium head blight resistance in wheat', *Agronomy* **10**(12), 2008.

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K. & Przeworski, M. (2020), 'Variable prediction accuracy of polygenic scores within an ancestry group', *elife* **9**, e48376.

Mundt, C. C. (2014), 'Durable resistance: a key to sustainable management of pathogens and pests', *Infection, Genetics and Evolution* **27**, 446–455.

Musyimi, L. (2009), Management of fusarium head blight of wheat through host resistance and microbial agents, PhD thesis, University of Nairobi.

Nelson, P. E., Desjardins, A. E. & Plattner, R. D. (1993), 'Fumonisins, mycotoxins produced by fusarium species: biology, chemistry, and significance', *Annual review of phytopathology* **31**(1), 233–252.

Nutter, F. F. (2007), 'The role of plant disease epidemiology in developing successful integrated disease management programs', *General concepts in integrated pest and disease management* pp. 45–79.

Ogutu, J. O., Schulz-Streeck, T. & Piepho, H.-P. (2012), Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions, *in* 'BMC proceedings', Vol. 6, Springer, pp. 1–6.

Paul, P., El-Allaf, S., Lipps, P. & Madden, L. (2005), 'Relationships between incidence and severity of fusarium head blight on winter wheat in ohio', *Phytopathology* **95**(9), 1049–1060.

Pérez-Cabal, M., Toro, M. & Fernández, C. (2012), 'Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels', *Journal of Dairy Science* **95**(7), 4114–4125.

Pérez, P., de Los Campos, G., Crossa, J. & Gianola, D. (2010), 'Genomic-enabled prediction based on molecular markers and pedigree using the bayesian linear regression package in r', *The plant genome* **3**(2).

Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y. & Dreisigacker, S. (2012), 'Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat', *G3: Genes| Genomes| Genetics* **2**(12), 1595–1605.

Perochon, A., Jianguang, J., Kahla, A., Arunachalam, C., Scofield, S. R., Bowden, S., Wallington, E. & Doohan, F. M. (2015), 'Tafrog encodes a pooideae orphan protein that interacts with snrk1 and enhances resistance to the mycotoxigenic fungus fusarium graminearum', *Plant physiology* **169**(4), 2895–2906.

Pirgozliev, S. R., Edwards, S. G., Hare, M. C. & Jenkinson, P. (2003), 'Strategies for the control of fusarium head blight in cereals', *European Journal of Plant Pathology* **109**, 731–742.

Reyna, M., Macor, E. P., Vilchez, A. C. & Villasuso, A. L. (2023), 'Response in barley roots during interaction with bacillus subtilis and fusarium graminearum', *Biological Control* **179**, 105128.

Rossi, V., Salinari, F., Bernazzani, R., Giosuè, S. & Mazzoni, E. (2009), 'Models for pest's epidemiology: review, documentation and evaluation for pest risk analysis (mopest)', *EFSA Supporting Publications* **6**(9), 28E.

Saeed, M., Ahmad, W., Ibrahim, M., Khan, M., Ullah, F., Bari, A., Ali, S., Shah, L., Ali, M., Munsif, F. et al. (2022), 'Differential responses to yellow-rust stress assist in the identification of candidate wheat (triticum aestivum l.) genotypes for resistance breeding', *Agronomy* **12**(9), 2038.

Savary, S., Teng, P. S., Willocquet, L. & Nutter Jr, F. W. (2006), 'Quantification and modeling of crop losses: a review of purposes', *Annu. Rev. Phytopathol.* **44**, 89–112.

Savary, S., Willocquet, L., Elazegui, F. A., Castilla, N. P. & Teng, P. S. (2000), 'Rice pest constraints in tropical asia: quantification of yield losses due to rice pests in a range of production situations', *Plant disease* **84**(3), 357–369.

Schrauf, M. F., de Los Campos, G. & Munilla, S. (2021), 'Comparing genomic prediction models by means of cross validation', *Frontiers in Plant Science* p. 2648.

Shah, D. A., De Wolf, E., Paul, P. A. & Madden, L. V. (2023), 'Into the trees: random forests for predicting fusarium head blight epidemics of wheat in the united states', *Phytopathology* (ja).

Simko, I. & Piepho, H.-P. (2012), 'The area under the disease progress stairs: calculation, advantage, and application', *Phytopathology* **102**(4), 381–389.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. (2013), 'Pleiotropy in complex traits: challenges and strategies', *Nature Reviews Genetics* **14**(7), 483–495.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., Atlin, G., Jannink, J.-L. & McCouch, S. R. (2015), 'Genomic selection and association mapping in rice (oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines', *PLoS genetics* **11**(2), e1004982.

Stahl, S. (2006), 'The evolution of the normal distribution', *Mathematics magazine* **79**(2), 96–113.

Steffenson, B. J. & Webster, R. (1992), 'Quantitative resistance to pyrenophora teres f. teres in barley.', *Phytopathology* **82**(4), 407–411.

Tessema, B. B., Wu, J., Rudd, J. C., Akhunov, E., Liu, S., Dillon, S. L., Muleta, K. T., Morris, G. P., Xu, S., Jin, Y. et al. (2020), 'Strategies using genomic selection to increase genetic gain in breeding programs for wheat', *Frontiers in genetics* **11**, 578123.

VanRaden, P. M. (2008), 'Efficient methods to compute genomic predictions', *Journal of dairy science* **91**(11), 4414–4423.

Velásquez, A. C., Castroverde, C. D. M. & He, S. Y. (2018), 'Plant–pathogen warfare under changing climate conditions', *Current biology* **28**(10), R619–R634.

Wakchaure, R., Ganguly, S., Praveen, P., Kumar, A., Sharma, S. & Mahajan, T. (2015), 'Marker assisted selection (mas) in animal breeding: a review', *J. Drug. Metab. Toxicol* **6**(5), e127.

Wheat Nutrition (n.d.), `https://wheat.org/nutrition/`. Accessed: April 23, 2023.

Wilcoxson, R. D., Skovmand, B. & Atif, A. (1975), 'Evaluation of wheat cultivars for ability to retard development of stem rust', *Annals of Applied Biology* **80**(3), 275–281.

Wilde, F. & Miedaner, T. (2006), 'Selection for fusarium head blight resistance in early generations reduces the deoxynivalenol (don) content in grain of winter and spring wheat', *Plant Breeding* **125**(1), 96–98.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J. et al. (2020), 'Genome composition and divergence of the novel coronavirus (2019-ncov) originating in china', *Cell host & microbe* **27**(3), 325–328.

Xia, R., Schaafsma, A., Limay-Rios, V. & Hooker, D. (2021), 'Effectiveness of a novel fungicide pydiflumetofen against fusarium head blight and mycotoxin accumulation in winter wheat', *World Mycotoxin Journal* **14**(4), 477–493.

Xiao, Y., Dong, Y., Huang, W. & Liu, L. (2022), 'Regional prediction of fusarium head blight occurrence in wheat with remote sensing based susceptible-exposed-infectious-removed model', *International Journal of Applied Earth Observation and Geoinformation* **114**, 103043.

Xu, Y., Ma, K., Zhao, Y., Wang, X., Zhou, K., Yu, G., Li, C., Li, P., Yang, Z., Xu, C. et al. (2021), 'Genomic selection: A breakthrough technology in rice breeding', *The Crop Journal* **9**(3), 669–677.

Xu, Y., Zhang, D., Jin, Z., Li, M. & Yang, J.-Y. (2006), 'A fast kernel-based nonlinear discriminant analysis for multi-class problems', *Pattern Recognition* **39**(6), 1026–1033.

Yu, J., Proctor, R. H., Brown, D. W., Abe, K., Gomi, K., Machida, M., Hasegawa, F., Nierman, W. C., Bhatnagar, D., Cleveland, T. E. et al. (2004), 'Genomics of economically significant aspergillus and fusarium species', *Applied mycology and biotechnology* **4**, 249–283.

Yuen, J. E. & Forbes, G. A. (2009), 'Estimating the level of susceptibility to phytophthora infestans in potato genotypes', *Phytopathology* **99**(6), 782–786.

Zambrano, J. C. & Echeverri, J. (2014), 'Genetic and environmental variance and covariance parameters for some reproductive traits of holstein and jersey cattle in antioquia (colombia)', *Revista Brasileira de Zootecnia* **43**, 132–139.

Zhang, J., Gill, H. S., Brar, N. K., Halder, J., Ali, S., Liu, X., Bernardo, A., Amand, P. S., Bai, G., Gill, U. S. et al. (2022), 'Genomic prediction of fusarium head blight resistance in early stages using advanced breeding lines in hard winter wheat', *The Crop Journal* **10**(6), 1695–1704.

Zhao, W., Zhu, J., Gallo-Meagher, M. & Wu, R. (2004), 'A unified statistical model for functional mapping of environment-dependent genetic expression and genotype× environment interactions for ontogenetic development', *Genetics* **168**(3), 1751–1762.

Zingaretti, L. M., Gezan, S. A., Ferrão, L. F. V., Osorio, L. F., Monfort, A., Muñoz, P. R., Whitaker, V. M. & Pérez-Enciso, M. (2020), 'Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species', *Frontiers in plant science* **11**, 25.